数 理 统 计

Mathematical Statistics

2019 U.S. News Best Jobs Rankings

100 Best Jobs

1. Software Developer
2. Statistician
3. Physician Assistant
4. Dentist
5. Orthodontist (tie)
5. Nurse Anesthetist (tie)

Best Business Jobs

1. Statistician
2. Mathematician
3. Accountant
4. Financial Manager
5. Medical and Health Services Manager

在终极的分析中，一切知识都是历史

在抽象的意义下，一切科学都是数学

在理性的基础上，所有的判断都是统计学

All knowledge is, in final analysis, history.

All sciences are, in the abstact, mathematics.

All judgements are, in their rationale, statistics

C.R.Rao

第1章 绪论

§1.1 什么叫数理统计学

- (数理)统计学是"收集和分析数据的科学与艺术" (出自《不列颠百科全书》);

第1章 绪论

§1.1 什么叫数理统计学

- (数理)统计学是"收集和分析数据的科学与艺术" (出自《不列颠百科全书》);
- 用有效的方法;

第1章 绪论

§1.1 什么叫数理统计学

- (数理)统计学是”收集和分析数据的科学与艺术” (出自《不列颠百科全书》);
- 用有效的方法;
- 它分析处理的对象是带有随机性的数据(random data).

第1章 绪论

§1.1 什么叫数理统计学

- (数理)统计学是"收集和分析数据的科学与艺术" (出自《不列颠百科全书》);
- 用有效的方法;
- 它分析处理的对象是带有随机性的数据(random data).
- 数理统计学是数学的一个分支，它是一门用有效的方法收集和分析带有随机影响的数据的学科，且其目的是解决特定的问题. (陈希孺院士)

- 数理统计的历史、发展和应用

  《数理统计简史》，陈希孺编著

  "大数据时代"

- 数理统计的历史、发展和应用
  《数理统计简史》，陈希孺编著

  “大数据时代”
- 无处不在的统计
  如: 利用“Google”对互联网进行搜索, 得到的结果是: 包含“统计”这一词汇的网页高达1,100,000,000项, 包含“粮食”这一词汇的网页有56,300,000项, 前者是后者的60多倍.

- 统计研究与应用包括

$$\left\{\begin{array}{l}\text{抽样调查}\\\text{试验设计}\\\text{回归分析}\\\text{多元统计分析}\\\text{时间序列分析}\\\text{数据挖掘}\\\text{生存分析}\\\text{非参数统计}\\\cdots\end{array}\right.$$

统计学 Vs. 概率论

在统计学中，研究对象的全体H称为总体(population)，随机抽取的X_1, X_2, $\cdots$, X_n称为样本(sample).

- 在概率论中，我们先是知道H的情况，再推导X_1, X_2,$\cdots$, X_n的性质. 总体是因,样本是果. **知因而后知果**.
- 在统计学中，总体H是未知的，将通过可以得到的样本X_1, X_2,$\cdots$, X_n来了解总体的情况. **由果来推断因**.
- 由于不同的因可以产生同一结果，有时同一因可产生不同的结果，由样本推断总体时不可避免会犯错误. 如何控制犯错误的概率是统计学中的一个重要内容.

统计学 Vs. 数学

- 数学是"演绎式"的，得出的结论是从一些假设、命题或已知事实出发，按一定的逻辑推理得出来的；
- 统计学方法本质是"归纳式"的，作出结论时，是根据所观察到的大量"个别"情况"归纳"起来所得.

统计推断属于归纳推理方法，归纳推理作出的推断不是100%可靠，但它的可靠程度(即结论的正确程度)可以通过概率来度量.

统计学与专门学科的关系

统计学是从事物外在数量上的表现去推断该事物可能的规律性.

统计学本身不能说明何以会有这个规律性, 这是各个专门学科的任务.

统计学为各个专门学科提供“雾里看花”的慧眼.

- R, SAS, SPSS, S-plus, EViews, Excel, Matlab, C, C++, Python,...

- R, SAS, SPSS, S-plus, EViews, Excel, Matlab, C, C++, Python,...
- 预修课程:概率论($\geq$3学分)

- R, SAS, SPSS, S-plus, EViews, Excel, Matlab, C, C++, Python,...
- 预修课程:概率论($\geq$3学分)
- 参考资料:

- R, SAS, SPSS, S-plus, EViews, Excel, Matlab, C, C++, Python,...
- 预修课程:概率论(≥3学分)
- 参考资料:
- 课件下载:学在浙大

- R, SAS, SPSS, S-plus, EViews, Excel, Matlab, C, C++, Python,...
- 预修课程:概率论(≥3学分)
- 参考资料:
- 课件下载:学在浙大
- 联系邮箱: stazlx@zju.edu.cn

- R, SAS, SPSS, S-plus, EViews, Excel, Matlab, C, C++, Python,...
- 预修课程:概率论($\geq$3学分)
- 参考资料:
- 课件下载:学在浙大
- 联系邮箱: stazlx@zju.edu.cn
- 作业: 每周一交

课程安排

- 基本概念
- 统计量及其分布
- 参数估计
  - 点估计
  - 区间估计
- 假设检验
  - 参数假设检验(包括方差分析)
  - 非参数假设检验

§1.2 数理统计的若干基本概念

(一)总体及总体分布

随机数据的来源是在统计分析问题中的研究对象的全体所构成的集合，这样的集合就称为总体(population). 总体中的每个元素称为个体(individual)，是数据的载体.

§1.2 数理统计的若干基本概念

(一)总体及总体分布

随机数据的来源是在统计分析问题中的研究对象的全体所构成的集合，这样的集合就称为总体(population). 总体中的每个元素称为个体(individual)，是数据的载体.

总体

§1.2 数理统计的若干基本概念

(一)总体及总体分布

随机数据的来源是在统计分析问题中的研究对象的全体所构成的集合，这样的集合就称为总体(population)．总体中的每个元素称为个体(individual)，是数据的载体．

$$\text{总体} \\ \updownarrow \\ \text{指标值的全体}\quad X(\omega)$$

§1.2 数理统计的若干基本概念

(一)总体及总体分布

随机数据的来源是在统计分析问题中的研究对象的全体所构成的集合，这样的集合就称为总体(population). 总体中的每个元素称为个体(individual)，是数据的载体.

$$\text{总体}$$

$$\updownarrow$$

$$\text{指标值的全体} \quad X(\omega)$$

$$\updownarrow$$

$$\text{用随机变量(向量)及其分布来表征}$$

因此,在统计学中，我们用一个随机变量(一维或多维随机变量)表示总体，或用一个分布函数表示总体.

说明:

- 当总体中个体的数目有限时，该总体称为有限总体；否则就称为无限总体.

说明:

- 当总体中个体的数目有限时，该总体称为有限总体；否则就称为无限总体.
- 具体对象可以是同一个,但具体研究总体针对不同的问题可以有很多个.

统计分析的真正目的

不在于了解总体中某个或某些个体的确切情况，

而在于了解总体分布.

一旦了解了总体分布，就可以进一步了解它的统计数学特征.

统计分析的真正目的

不在于了解总体中某个或某些个体的确切情况,

而在于了解总体分布.

一旦了解了总体分布, 就可以进一步了解它的统计数学特征.

例如, 我们可以了解它的理论均值(即数学期望), 它的理论方差, 等等.

这些概念在概率论中都有严格的定义.

统计分析的真正目的

不在于了解总体中某个或某些个体的确切情况，
而在于了解总体分布.
一旦了解了总体分布，就可以进一步了解它的统计数学特征.
例如，我们可以了解它的理论均值(即数学期望)，它的理论方差，等等.
这些概念在概率论中都有严格的定义.

根据从总体中按照一定的方法所获得的随机数据来推断总体分布或它的某些方面的统计特性，就是统计分析或统计推断的任务.

但事实上我们一般无法确切知道总体分布，除非我们对总体进行穷尽的观测.

(二)分布族

总体变量X具有一定的分布, 称为总体分布.

总体分布一般是不完全已知的, 但常常会对其做某些假定.

(二)分布族

总体变量X具有一定的分布，称为总体分布.

总体分布一般是不完全已知的，但常常会对其做某些假定.

数理统计研究总体的前提条件:

假定总体分布的范围是已知的,即,假定总体X是取自某一分布族的,或者说，假定总体X的分布具有某种形式.

Example

例1 为检验某灯泡厂生产的产品质量,检测灯泡寿命(h).

此时总体为灯泡的寿命X(以h为单位),我们可假定X取值于$[0, \infty)$.

Example

例1 为检验某灯泡厂生产的产品质量,检测灯泡寿命(h).

此时总体为灯泡的寿命X(以h为单位),我们可假定X取值于$[0,\infty)$.
在实际中，根据经验，通常假定灯泡的寿命服从一定类型的分布.
这里假定为指数分布，即

$$F(x) = 1 - e^{-\lambda x}, \qquad 0 \le x < \infty,$$

其中$\lambda > 0$为待定的未知参数，$1/\lambda$是总体分布的均值.

Example

例1 为检验某灯泡厂生产的产品质量,检测灯泡寿命(h).

此时总体为灯泡的寿命X(以h为单位),我们可假定X取值于$[0,\infty)$.
在实际中，根据经验，通常假定灯泡的寿命服从一定类型的分布.
这里假定为指数分布，即

$$F(x)=1-e^{-\lambda x},\qquad 0\le x<\infty,$$

其中$\lambda>0$为待定的未知参数，$1/\lambda$是总体分布的均值.当我们知道λ的真值时，就完全掌握了灯泡寿命的分布. 这样，对灯泡寿命分布的推断就可以转化为对参数λ的推断. 总体分布族为指数分布族:

$$\mathcal{F}=\{E(\lambda):\lambda>0\}.$$

Example

例2 在上例中对同样一批灯泡，假定按照某种规定的标准，灯泡的寿命在3000h以上为正品，否则即为次品．而考察的只是区分正品或次品．

此时考察的是一个非数量化指标，通常也称为"记名变量"，这里仅取两个"值"：正品或次品．

Example

例2 在上例中对同样一批灯泡，假定按照某种规定的标准，灯泡的寿命在3000h以上为正品，否则即为次品．而考察的只是区分正品或次品．

此时考察的是一个非数量化指标，通常也称为"记名变量"，这里仅取两个"值"：正品或次品．为能够进行数据分析，我们可以将这个指标数量化，定义总体(变量)为

$$X = \begin{cases} 1, & \text{若灯泡为次品} \\ 0, & \text{若灯泡为正品.} \end{cases}$$

则可以假定总体分布族为

$$\mathcal{F} = \{B(1,p) : 0 < p < 1\},$$

其中$p = P(X = 1)$为次品率，未知，是我们需要推断的.

Example

例3 一个实验者对一个物理量μ的值进行测定.

此时，一切可能的测量结果就是总体. 它可用随机变量X(测量值)表示.

Example

例3 一个实验者对一个物理量μ的值进行测定.
此时，一切可能的测量结果就是总体. 它可用随机变量X(测量值)表示.

大家知道,测量者所测得的结果会受到各种随机因素的影响,
经验表明,测量结果X可以认为是μ加上随机误差ϵ而得，即得

$$X = \mu + \epsilon.$$

1. 在正常情况下，随机误差$\epsilon \sim N(0, \sigma^2)$.
   于是可假定总体X的分布所在的分布族是正态分布族:

$$\mathcal{F}_1 = \{N(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma > 0\}.$$

这样一来,分布族$\mathcal{F}_1$就成为统计推断的出发点.

1. 在正常情况下，随机误差$\epsilon \sim N(0, \sigma^2)$.
   于是可假定总体X的分布所在的分布族是正态分布族:

$$\mathcal{F}_1 = \{N(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma > 0\}.$$

   这样一来,分布族$\mathcal{F}_1$就成为统计推断的出发点.
2. 假如我们不仅了解ϵ服从正态分布,而且其方差σ^2也知道为σ_0^2(这是可能的,比如知道测量仪器的精度)，那么关于分布族的假定就可缩小为

$$\mathcal{F}_2 = \{N(\mu, \sigma_0^2) : -\infty < \mu < \infty\}.$$

3. 假如我们对随机误差ϵ的信息了解很少，比如根本不知道它的分布是什么类型,只知道它是连续型分布，或者二阶矩存在的分布,那么关于分布族的假定就扩大为
$$\mathcal{F}_3 = \{F(x) : F\text{为连续型分布}\}$$
或
$$\mathcal{F}_4 = \{F(x) : F\text{的二阶矩存在}\}.$$

参数分布族和非参数分布族:

参数分布族和非参数分布族:

参数分布族: 分布族只含有限个未知的实参数. 常表示为

$$\mathcal{F}=\{F(x;\theta):\theta\in\Theta\},$$

θ表示未知参数(或向量),Θ是参数θ可能的取值的范围,常称为参数空间.

参数分布族和非参数分布族:

参数分布族: 分布族只含有限个未知的实参数. 常表示为

$$\mathcal{F} = \{F(x;\theta) : \theta \in \Theta\},$$

θ表示未知参数(或向量),Θ是参数θ可能的取值的范围,常称为参数空间.
从参数分布族出发所得的统计方法称为参数统计方法.
从非参数分布族出发所得的统计方法称为非参数统计方法.

参数分布族和非参数分布族:

参数分布族: 分布族只含有限个未知的实参数. 常表示为

$$\mathcal{F} = \{F(x;\theta) : \theta \in \Theta\},$$

θ表示未知参数(或向量),Θ是参数θ可能的取值的范围,常称为参数空间.
从参数分布族出发所得的统计方法称为参数统计方法.
从非参数分布族出发所得的统计方法称为非参数统计方法.
常用的总体参数分布族有正态分布, 二项分布, 泊松(Poisson)分布, 指数分布, 及Γ(Gamma)分布等.

考察总体的方法:

- 全面调查

考察总体的方法:

- 全面调查
- 抽样调查、随机试验

考察总体的方法:

- 全面调查
- 抽样调查、随机试验

考察总体的方法:

- 全面调查
- 抽样调查、随机试验(采用)

(三)样本(sample)

样本是由从总体中随机地抽取部分个体所组成的.

(三)样本(sample)

样本是由从总体中随机地抽取部分个体所组成的.

(1)抽样: 从总体中根据一定的法则抽取一定数量的个体,这一过程称为“抽样”.

(三)样本(sample)

样本是由从总体中随机地抽取部分个体所组成的.

(1)抽样：从总体中根据一定的法则抽取一定数量的个体,这一过程称为“抽样”.

(2)样本容量：抽样中所抽取的个体的数目.(也称为"样本大小")

(三)样本(sample)

样本是由从总体中随机地抽取部分个体所组成的.

(1)抽样：从总体中根据一定的法则抽取一定数量的个体,这一过程称为“抽样”.

(2)样本容量：抽样中所抽取的个体的数目.(也称为”样本大小”)

(3)由于是随机抽样，因此样本在未被观测前，或作为被研究的一般对象时，也是随机变量.

每一个样本的维数与其总体相同. 如果总体是一维随机变量，那么容量为n的一组样本就可视为n维的随机向量.

注意:样本的双重性

注意:样本的双重性

假定从一个总体中抽取n个个体, 相应的总体变量值(在未被观测前) 通常记为$X_1, X_2, ..., X_n$, 这就是一组容量为n的样本.
从数学的角度来看, $(X_1, X_2, ..., X_n)$构成一个n维随机向量, 当样本被实际观测到时, 其观测值就是一组实际的数据, 通常记为$x_1, x_2, ..., x_n$. 从数学的角度来看, $(x_1, x_2, ..., x_n)$就是n维随机变量$(X_1, X_2, ..., X_n)$在一次观测中的"实现".

$(X_1, X_2, \cdots, X_n)$称为容量为n的一组样本，而$(x_1, x_2, ..., x_n)$称为样本观察值.

（“样本空间” $\mathcal{H} = \{(x_1, x_2, ..., x_n)\}$为样本的一切可能取值构成的集合）

(四) 简单随机抽样

抽样的方法有很多. 为了使得抽样所得到的样本能很好地放映总体的信息, 最常用的一种抽样方法是"简单随机抽样",它要求满足下面两点:

- 代表性(即随机性). 总体中的每一个个体都有同等的机会被抽入样本,这意味着样本中每个个体X_k与所考察的总体X同分布.
- 独立性. 样本中每个个体取什么值并不会影响其它个体取什么值. 即$X_1, X_2 \cdots, X_n$相互独立.

由简单随机抽样所得的样本$(X_1, X_2, \cdots, X_n)$称为简单随机样本(simple sample)或 “i.i.d.(independent and identically distributed)样本”，它可用与总体X同分布的n个相互独立的随机变量$X_1, X_2, \cdots, X_n$表示.

由简单随机抽样所得的样本$(X_1, X_2, \cdots, X_n)$称为简单随机样本(simple sample)或"i.i.d.(independent and identically distributed)样本"，它可用与总体X同分布的n个相互独立的随机变量$X_1, X_2, \cdots, X_n$表示.

Definition

设有一总体X(或用分布F表示), $X_1, X_2, \cdots, X_n$为从X中抽取的容量为n的样本, 若

1. (独立性) $X_1,\ X_2,\ \cdots,\ X_n$相互独立;
2. (代表性) $X_1,\ X_2,\ \cdots,\ X_n$分布相同, 与X同分布,

则称$X_1,\ X_2,\ \cdots,\ X_n$为简单随机样本, 简称为简单样本或样本.

假如总体的分布函数是$F(x)$, 则简单随机样本的联合分布函数为

$$F_n(x_1, x_2, \cdots, x_n) = F(x_1)F(x_2)\cdots F(x_n).$$

说明: 本书中,随机变量X的分布函数定义为:

$$F(x) = \mathsf{P}\{X < x\}$$

为左连续函数.

假如总体的分布列(概率密度函数)是$f(x)$，则简单随机样本的联合分布列(联合概率密度函数)为

$$f_n(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n).$$

"有放回抽样"(sampling with replacement)是简单随机抽样.

"有放回抽样"(sampling with replacement)是简单随机抽样.

当总体是无限总体时，从总体中抽取有限个个体并不改变总体的分布，因此对于无限总体，"无放回抽样"(sampling without replacement)[即每次从总体中抽一个个体，观测后不放回总体中，再抽下一个]，或者等价地，从总体中一下子取出n个个体进行观测,也是简单随机抽样.

"有放回抽样"(sampling with replacement)是简单随机抽样.

当总体是无限总体时，从总体中抽取有限个个体并不改变总体的分布，因此对于无限总体，"无放回抽样"(sampling without replacement)[即每次从总体中抽一个个体，观测后不放回总体中，再抽下一个]，或者等价地，从总体中一下子取出n个个体进行观测，也是简单随机抽样.

对于有限总体，在无放回抽样下得到的样本，从理论上说就不再是简单样本了. 但是当总体中个体的数目很多(与样本的大小相比较)时，从总体中抽掉一些个体对总体分布没有太大的影响. 在这种情况下，即使是无放回抽样也可以近似地看成是有放回抽样，因此样本仍可以看成是i.i.d.的，即简单样本.

以后,绝大部分内容是讨论简单样本, 以后在不致引起混淆的场合就简称为样本.

样本分布族—统计模型

如果总体X的分布族为$\mathcal{F}$, 那么从中抽取的简单随机样本$\mathbf{X}=(X_1,\ldots,X_n)$的分布必是

$$\mathfrak{F}=\{F(x_1)\cdots F(x_n): F\in\mathcal{F}\}$$

中的一员, 这个分布族就是样本分布族,它是统计推断的出发点, 又叫做统计模型.

由于简单随机样本的分布族完全由总体分布族确定, 有时也直接把总体分布族叫做统计模型.

当$(X_1, X_2, \cdots, X_n)$不是简单随机样本时，这时样本的分布不能完全由总体的分布确定，样本分布还包含了样本的模型结构信息. 所以一般把样本分布族叫做统计模型.

(五) 从样本认识总体

在统计中，我们的目的是了解总体的某些未知信息(分布信息)，方法是通过能够观察到的样本信息(样本观测值的信息，数据信息)去推断总体的信息.

(五) 从样本认识总体

在统计中，我们的目的是了解总体的某些未知信息(分布信息)，方法是通过能够观察到的样本信息(样本观测值的信息，数据信息)去推断总体的信息.

理论根据?

经验分布函数

Definition

定义1.3.2 样本$(X_1, X_2, \cdots, X_n)$的**经验分布函数**(empirical distribution function)定义为

$$F_n(x) = \frac{1}{n}\sharp\{X_i : X_i < x, i = 1, 2, \cdots, n\} \qquad \forall x \in \mathbb{R},$$

其中$\sharp\{\cdot\}$表示集合$\{\cdot\}$中元素的个数.

不难看出，经验分布函数函数也可如下定义：将样本$X_1, X_2, \cdots, X_n$从小到大排列：

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}.$$

则

$$F_n(x) = \begin{cases} 0, & x \leq X_{(1)}, \\ k/n, & X_{(k)} < x \leq X_{(k+1)}, \quad (k = 1, 2, \cdots, n-1) \\ 1, & X_{(n)} < x. \end{cases}$$

对经验分布函数，我们应注意到:

一方面,当给定样本时，它是自变量x的函数,它满足分布函数的一般性质，即：

（1）$F_n(x)$为x的单调非降左连续函数；

（2）$\lim_{x\to-\infty} F_n(x) = 0, \quad \lim_{x\to\infty} F_n(x) = 1.$

对经验分布函数，我们应注意到:

一方面,当给定样本时，它是自变量x的函数,它满足分布函数的一般性质，即：

（1）$F_n(x)$为x的单调非降左连续函数；

（2）$\lim_{x\to-\infty} F_n(x) = 0, \quad \lim_{x\to\infty} F_n(x) = 1.$

另一方面,当给定x时，$F_n(x)$是样本的函数，它的值由样本$X_1, X_2, \cdots, X_n$的观察值唯一确定.

用$F_n(x)$可以来刻画总体分布函数. 为说明这一点，可将$F_n(x)$ 表为另一种形式：

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i<x\}},$$

用$F_n(x)$可以来刻画总体分布函数. 为说明这一点，可将$F_n(x)$ 表为另一种形式：

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i<x\}},$$

因此，如果记$Y_i = I_{\{X_i<x\}}$, $i = 1, ..., n$, 则$Y_i, i = 1, ..., n$, 为i.i.d.随机变量，$E(Y_i) = F(x)$, $Var(Y_i) = F(x)(1 - F(x))$,

用$F_n(x)$可以来刻画总体分布函数. 为说明这一点，可将$F_n(x)$ 表为另一种形式：

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i<x\}},$$

因此，如果记$Y_i = I_{\{X_i<x\}}$, $i = 1, ..., n$, 则$Y_i, i = 1, ..., n$, 为i.i.d.随机变量，$E(Y_i) = F(x)$, $Var(Y_i) = F(x)(1 - F(x))$, 由强大数定律，对给定的$x$有，

$$P(\lim_{n\to\infty} F_n(x) = F(x)) = 1.$$

也就是说，对给定的x, $F_n(x)$ 以概率1收敛到$F(x)$.

直接证明： 记$\xi_i = Y_i - \mathsf{E}Y_i$. 则

$$\mathsf{P}\left(|F_n(x) - F(x)| \geq \epsilon\right) \leq \frac{\mathsf{E}|F_n(x) - F(x)|^4}{\epsilon^4} = \frac{1}{n^4\epsilon^4}\mathsf{E}|\sum_{i=1}^n \xi_i|^4$$

$$= \frac{1}{n^4\epsilon^4}\left\{\sum_{i=1}^n \mathsf{E}\xi_i^4 + \sum_{i\neq j}\mathsf{E}[\xi_i^2\xi_j^2] + \sum_{i\neq j}\mathsf{E}[\xi_i^3\xi_j] + \sum_{i,j,k,l\text{互不相同}}\mathsf{E}[\xi_i\xi_j\xi_k\xi_l]\right\}$$

$$= \frac{n\mathsf{E}\xi_1^4 + n(n-1)(\mathsf{E}\xi_1^2)^2}{n^4\epsilon^4} \leq \frac{n^2}{n^4\epsilon^4} = \frac{1}{n^2\epsilon^4}.$$

直接证明： 记$\xi_i = Y_i - \mathsf{E}Y_i$. 则

$$\mathsf{P}\left(|F_n(x) - F(x)| \geq \epsilon\right) \leq \frac{\mathsf{E}|F_n(x) - F(x)|^4}{\epsilon^4} = \frac{1}{n^4\epsilon^4}\mathsf{E}|\sum_{i=1}^n \xi_i|^4$$

$$= \frac{1}{n^4\epsilon^4}\left\{\sum_{i=1}^n \mathsf{E}\xi_i^4 + \sum_{i\neq j}\mathsf{E}[\xi_i^2\xi_j^2] + \sum_{i\neq j}\mathsf{E}[\xi_i^3\xi_j] + \sum_{i,j,k,l\text{互不相同}}\mathsf{E}[\xi_i\xi_j\xi_k\xi_l]\right\}$$

$$= \frac{n\mathsf{E}\xi_1^4 + n(n-1)(\mathsf{E}\xi_1^2)^2}{n^4\epsilon^4} \leq \frac{n^2}{n^4\epsilon^4} = \frac{1}{n^2\epsilon^4}.$$

从而

$$\sum_{n=1}^\infty \mathsf{P}\left(|F_n(x) - F(x)| \geq \epsilon\right) < \infty, \quad \forall \epsilon > 0.$$

因此

$$F_n(x) \to F(x) \quad a.s. \quad \forall x.$$

因此

$$F_n(x) \to F(x) \ \ a.s. \ \ \forall x.$$

不仅如此，我们有如下更强的结论:

Theorem

定理1.3.1 *(格里汶科)(Glivenko-Cantelli)*

设样本$X_1, ..., X_n$, $i.i.d. \sim F$, $F_n(x)$为经验分布函数, 则

$$P(\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| = 0) = 1.$$

Theorem

定理1.3.1 *(格里汶科)(Glivenko-Cantelli)*
设样本$X_1, ..., X_n$, $i.i.d. \sim F$, $F_n(x)$为经验分布函数, 则

$$P(\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| = 0) = 1.$$

定理1.3.1的的含义是：$F_n(x)$在整个实轴上以概率1一致收敛到$F(x)$. 此结论强于强大数定律所给的结论. 因此, 当样本容量足够大时, $F_n(x)$能够很好地拟合总体分布$F(x)$. 而总体的信息全部包含在它的分布$F(x)$中. 这样自然可以由样本去认识总体.

定理的证明:

在概率论中我们已经学过: $F(x)$ 有广义反函数$F^{-1}(y)$ 使得$y < F(x) \Leftrightarrow F^{-1}(y) < x$.

设$U_1, U_2, \cdots, U_n$ 为i.i.d. $U[0,1]$均匀变量, 则X_i与$F^{-1}(U_i)$同分布, 所以可以不妨设$X_i = F^{-1}(U_i)$.

定理的证明:

在概率论中我们已经学过: $F(x)$ 有广义反函数$F^{-1}(y)$ 使得$y < F(x) \Leftrightarrow F^{-1}(y) < x$.

设$U_1, U_2, \cdots, U_n$ 为i.i.d. $U[0,1]$均匀变量, 则X_i与$F^{-1}(U_i)$同分布, 所以可以不妨设$X_i = F^{-1}(U_i)$.

记$G_n(t) = \frac{1}{n}\sum_{i=1}^{n} I\{U_i < t\}$. 则

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\{F^{-1}(U_i) < x\} = \frac{1}{n}\sum_{i=1}^{n} I\{(U_i) < F(x)\} = G_n(F(x)).$$

定理的证明:

在概率论中我们已经学过: $F(x)$ 有广义反函数$F^{-1}(y)$ 使得$y < F(x) \Leftrightarrow F^{-1}(y) < x$.

设$U_1, U_2, \cdots, U_n$ 为i.i.d. $U[0,1]$均匀变量, 则X_i与$F^{-1}(U_i)$同分布, 所以可以不妨设$X_i = F^{-1}(U_i)$.

记$G_n(t) = \frac{1}{n}\sum_{i=1}^n I\{U_i < t\}$. 则

$$F_n(x) = \frac{1}{n}\sum_{i=1}^n I\{F^{-1}(U_i) < x\} = \frac{1}{n}\sum_{i=1}^n I\{(U_i) < F(x)\} = G_n(F(x)).$$

所以

$$\sup_x |F_n(x) - F(x)| = \sup_x |G_n(F(x)) - F(x)| \le \sup_{0\le t\le 1} |G_n(t) - t|.$$

因此，只要证明

$$\sup_{0\le t\le 1} |G_n(t) - t| \to 0 \quad a.s.$$

因此，只要证明

$$\sup_{0\le t\le 1} |G_n(t)-t| \to 0 \quad a.s.$$

对任意的$\epsilon>0$，取$0=t_0<t_1<\cdots<t_k=1$，使得$t_i-t_{i-1}<\epsilon$.
当$t_{i-1}\le t\le t_i$ 时，

$$\begin{aligned} G_n(t)-t \le & G_n(t_i)-t_{i-1} \\ \le & G_n(t_i)-t_i+t_i-t_{i-1} \le \max_{i\le k}|G_n(t_i)-t_i|+\epsilon. \end{aligned}$$

同理，$t-G_n(t)\le \max_{i\le k}|G_n(t_i)-t_i|+\epsilon$. 所以，

$$\sup_{0\le t\le 1}|G_n(t)-t| \le \max_{i\le k}|G_n(t_i)-t_i|+\epsilon.$$

所以

$$
\begin{aligned}
&\mathsf{P}\left(\sup_{0\le t\le 1}|G_n(t)-t|\ge 2\epsilon\right)\le \mathsf{P}\left(\max_{i\le k}|G_n(t_i)-t_i|\ge\epsilon\right)\\
\le&\sum_{i=1}^{k}\mathsf{P}\left(|G_n(t_i))-t_i|\ge\epsilon\right)\le\sum_{i=1}^{k}\frac{\mathsf{E}|G_n(t_i))-t_i|^4}{\epsilon^4}\le\frac{k}{\epsilon^4}\frac{1}{n^2}.
\end{aligned}
$$

所以

$$P\left(\sup_{0\le t\le 1}|G_n(t)-t|\ge 2\epsilon\right)\le P\left(\max_{i\le k}|G_n(t_i)-t_i|\ge\epsilon\right)$$

$$\le\sum_{i=1}^{k}P\left(|G_n(t_i))-t_i|\ge\epsilon\right)\le\sum_{i=1}^{k}\frac{E|G_n(t_i))-t_i|^4}{\epsilon^4}\le\frac{k}{\epsilon^4}\frac{1}{n^2}.$$

所以

$$\sum_{n=1}^{\infty}P\left(\sup_{0\le t\le 1}|G_n(t)-t|\ge 2\epsilon\right)<\infty,\quad\forall\epsilon>0.$$

故

$$\sup_{0\le t\le 1}|G_n(t)-t|\to 0\ \ a.s.$$

定理的证明(方法二): 将x的取值分段:对任意的正整数r，记$x_{r,k}$表示满足下述不等式的最小的x:

$$F(x-0)=F(x)\leq \frac{k}{r}\leq F(x+0),$$

$k=1,2,\cdots,r.$

定理的证明(方法二): 将x的取值分段:对任意的正整数r，记$x_{r,k}$表示满足下述不等式的最小的x:

$$F(x-0)=F(x)\leq\frac{k}{r}\leq F(x+0),$$

$k=1,2,\cdots,r$. 由Borel强大数律

$$P(\lim_{n\to\infty}F_n(x_{r,k})=F(x_{r,k}))=1.$$

类似地，有

$$P(\lim_{n\to\infty}F_n(x_{r,k}+0)=F(x_{r,k}+0))=1.$$

定义事件

$$
\begin{aligned}
A_k^r =&\{\lim_{n\to\infty} F_n(x_{r,k}) = F(x_{r,k})\} = \{\lim_{n\to\infty} |F_n(x_{r,k}) - F(x_{r,k})| = 0\},\\
B_k^r =&\{\lim_{n\to\infty} F_n(x_{r,k}+0) = F(x_{r,k}+0)\}\\
=&\{\lim_{n\to\infty} |F_n(x_{r,k}+0) - F(x_{r,k}+0)| = 0\},
\end{aligned}
$$

定义事件

$$A_k^r = \{\lim_{n\to\infty} F_n(x_{r,k}) = F(x_{r,k})\} = \{\lim_{n\to\infty} |F_n(x_{r,k}) - F(x_{r,k})| = 0\},$$

$$\begin{aligned} B_k^r &= \{\lim_{n\to\infty} F_n(x_{r,k}+0) = F(x_{r,k}+0)\} \\ &= \{\lim_{n\to\infty} |F_n(x_{r,k}+0) - F(x_{r,k}+0)| = 0\}, \end{aligned}$$

$$A^r = \bigcap_{k=1}^{r} (A_k^r \bigcap B_k^r) = \{\lim_{n\to\infty} \max_{1\le k\le r} \{\max(|F_n(x_{r,k}) - F(x_{r,k})|,$$

$$|F_n(x_{r,k}+0) - F(x_{r,k}+0)|)\} = 0\},$$

$$A = \bigcap_{r=1}^{\infty} A^r.$$

那么$P(A_k^r)=P(B_k^r)=1$. 注意到

$$P(\overline{A^r})=P(\bigcup_{k=1}^{r}(\overline{A_k^r}\cup\overline{B_k^r}))$$

$$\leq\sum_{k=1}^{r}(P(\overline{A_k^r})+P(\overline{B_k^r}))=0$$

那么$P(A_k^r) = P(B_k^r) = 1$. 注意到

$$
\begin{aligned}
P(\overline{A^r}) &= P(\bigcup_{k=1}^{r}(\overline{A_k^r} \cup \overline{B_k^r})) \\
&\le \sum_{k=1}^{r}(P(\overline{A_k^r}) + P(\overline{B_k^r})) = 0
\end{aligned}
$$

故

$$
P(\overline{A}) = P(\bigcup_{r=1}^{\infty}\overline{A^r}) = \lim_{n\to\infty} P(\bigcup_{r=1}^{n}\overline{A^r}) \le \lim_{n\to\infty}\sum_{r=1}^{n}P(\overline{A^r}) = 0.
$$

即

$$P(A) = 1.$$

记

$$E = \{\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| = 0\}.$$

剩下只要证

$$A \subset E.$$

现对任何满足$x_{r,k} < x \le x_{r,k+1} (k = 1, 2, \cdots, r-1)$的$x$，有

$$F_n(x_{r,k} + 0) \le F_n(x) \le F_n(x_{r,k+1}),$$

$$F(x_{r,k} + 0) \le F(x) \le F(x_{r,k+1}).$$

现对任何满足$x_{r,k} < x \le x_{r,k+1} (k = 1, 2, \cdots, r-1)$的$x$，有

$$F_n(x_{r,k} + 0) \le F_n(x) \le F_n(x_{r,k+1}),$$

$$F(x_{r,k} + 0) \le F(x) \le F(x_{r,k+1}).$$

由此可得，当$k = 1, 2, \cdots, r-1$时

$$\begin{aligned} F_n(x) - F(x) \le & F_n(x_{r,k+1}) - F(x_{r,k} + 0) \\ = & F_n(x_{r,k+1}) - F(x_{r,k+1}) + F(x_{r,k+1}) - F(x_{r,k} + 0) \\ \le & \max_k |F_n(x_{r,k}) - F(x_{r,k})| + \frac{1}{r}. \end{aligned}$$

同理

$$
\begin{aligned}
& F(x) - F_n(x) \\
\le & F(x_{r,k+1}) - F_n(x_{r,k}+0) \\
= & F(x_{r,k+1}) - F(x_{r,k}+0) + F(x_{r,k}+0) - F_n(x_{r,k}+0) \\
\le & \frac{1}{r} + \max_k |F(x_{r,k}+0) - F_n(x_{r,k}+0)|.
\end{aligned}
$$

同理

$$
\begin{aligned}
& F(x) - F_n(x) \\
\le & F(x_{r,k+1}) - F_n(x_{r,k}+0) \\
= & F(x_{r,k+1}) - F(x_{r,k}+0) + F(x_{r,k}+0) - F_n(x_{r,k}+0) \\
\le & \frac{1}{r} + \max_k |F(x_{r,k}+0) - F_n(x_{r,k}+0)|.
\end{aligned}
$$

另外，类似地可证：当$x \le x_{r,1}$时，有

$$|F_n(x) - F(x)| \le |F_n(x_{r,1}) - F(x_{r,1})| + \frac{1}{r};$$

而当$x > x_{r,r}$时，有

$$|F_n(x) - F(x)| \le |F_n(x_{r,r}+0) - F(x_{r,r}+0)| + \frac{1}{r}.$$

综合上述得:

$$
\begin{aligned}
&\sup_{-\infty<x<\infty} |F_n(x) - F(x)| \\
\leq & \max_{1\leq k\leq r} \{\max\{|F_n(x_{r,k}) - F(x_{r,k})|, |F_n(x_{r,k}+0) - F(x_{r,k}+0)|\}\} \\
& + \frac{1}{r}
\end{aligned}
$$

由于当事件A发生，则事件A^r发生$(\forall r)$. 从而此时，有

$$
\begin{aligned}
&\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| \\
\le &\lim_{n\to\infty} \max_{1\le k\le r} \{\max\{|F_n(x_{r,k}) - F(x_{r,k})|, \\
&\qquad |F_n(x_{r,k}+0) - F(x_{r,k}+0)\}\} + \frac{1}{r} \\
= &0 + \frac{1}{r}.
\end{aligned}
$$

由于当事件A发生，则事件A^r发生$(\forall r)$. 从而此时，有

$$
\begin{aligned}
&\lim_{n\to\infty}\sup_{-\infty<x<\infty}|F_n(x)-F(x)|\\
\le&\lim_{n\to\infty}\max_{1\le k\le r}\{\max\{|F_n(x_{r,k})-F(x_{r,k})|,\\
&\qquad|F_n(x_{r,k}+0)-F(x_{r,k}+0)\}\}+\frac{1}{r}\\
=&0+\frac{1}{r}.
\end{aligned}
$$

由r的任意性，可得

$$
\lim_{n\to\infty}\sup_{-\infty<x<\infty}|F_n(x)-F(x)|=0.
$$

$A\subset E$. 结论得证.

§1.3 Data reduction– Statistics 统计量

统计分析就是通过数据认识总体。

§1.3 Data reduction– Statistics 统计量

统计分析就是通过数据认识总体。

目的概括地说是要了解总体分布的统计特性，

在参数模型中就归结为了解一些刻画模型的参数.

§1.3 Data reduction– Statistics 统计量

统计分析就是通过数据认识总体。
目的概括地说是要了解总体分布的统计特性,
在参数模型中就归结为了解一些刻画模型的参数.

统计分析的出发点(或依据)就是样本, 因为样本中包含了总体分布的信息.

但因为样本是n维的，这些信息是分散到样本的每个分量上的.
因而，直接从样本出发来推断总体分布是不方便的.
为此，需要对样本进行加工，将样本中分散的信息浓缩集中起来，这种对样本进行加工，浓缩集中信息的过程所产生的结果就是得到适当的样本函数.

但因为样本是n维的，这些信息是分散到样本的每个分量上的.
因而，直接从样本出发来推断总体分布是不方便的.
为此，需要对样本进行加工，将样本中分散的信息浓缩集中起来，这种对样本进行加工，浓缩集中信息的过程所产生的结果就是得到适当的样本函数.

统计量(statistics)就是样本的函数.

打个比方，样本就好比"原材料"，经过加工后得到的"产品"就是统计量.

Definition

定义1.3.1 设有一个分布族$\mathcal{F}=\{F\}$, $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$是从其中某个分布抽取的一个样本. 假如样本的实值函数(Borel可测函数)

$$T=T(\boldsymbol{X})=T(X_1,X_2,\cdots,X_n)$$

不依赖于分布族$\mathcal{F}$, 则称T为此分布族的统计量.

当$\mathcal{F}$为参数分布族$\{F(x;\theta):\theta\in\Theta\}$时, 则称不依赖于未知参数$\theta$的实值函数$T$为$\mathcal{F}$的统计量.

Definition

类似地, 假如样本的实向量函数(Borel可测函数)

$$\widetilde{T} = \widetilde{T}(\boldsymbol{X}) = (T_1(\boldsymbol{X}), T_2(\boldsymbol{X}), \cdots, T_k(\boldsymbol{X}))$$

不依赖于分布族$\mathcal{F}$, 则称T为此分布族的向量统计量.

Example

$X \sim N(a, \sigma^2)$, $X_1, \ldots, X_n$ 样本.

- $\sum_{i=1}^n X_i$;
- $\sum_{i=1}^n X_i^2$;
- $\sum_{i=1}^n (X_i - a)$;
- $\sum_{i=1}^n X_i^2/\sigma^2$;
- X_1,
- 100.

说明:

- 统计量可从解析式上进行判断;
- $T=T(\boldsymbol{X})$为样本的函数, 由于样本具有随机性, 故$T=T(\boldsymbol{X})$仍然为随机变量;(统计量也具有双重性, 当样本观测值$\boldsymbol{x}$得到时, $T=T(\boldsymbol{x})$就是统计量的(观测)值)
- $T=T(\boldsymbol{X})$的分布可以依赖于分布族;
- 当样本取定后, 统计量的值也应确定.

抽样分布

指的是统计量$T(\boldsymbol{X})$的分布. 若记为$F_T(t)$, 则

$$
\begin{aligned}
F_T(t) &= \mathsf{P}\{T(\boldsymbol{X}) < t\} \\
&= \mathsf{P}\{(X_1, X_2, \cdots, X_n) \in B\} \\
&\qquad (\text{其中} B = \{(x_1, x_2, \cdots, x_n) : T(x_1, x_2, \cdots, x_n) < t\}) \\
&= \int \cdots \int_B dF(x_1) dF(x_2) \cdots dF(x_n),
\end{aligned}
$$

其中$F(x)$是总体X的分布函数.

常用统计量

- 样本均值(sample mean) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
  ———反映总体均值$\mu = \mathsf{E}X$.

常用统计量

- 样本均值(sample mean) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

  ———反映总体均值$\mu = \mathsf{E}X$.

- 样本方差(sample variance) $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

  (有时采用$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$, 注意区别!)

  ———反映总体方差$\sigma^2 = \mathsf{Var}(X)$.

常用统计量

- 样本均值(sample mean) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
  ———反映总体均值$\mu = \mathsf{E}X$.
- 样本方差(sample variance) $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
  (有时采用$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$, 注意区别!)
  ———反映总体方差$\sigma^2 = \mathsf{Var}(X)$.
- 样本k阶(原点)矩(k-th sample moment): $a_{n,k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$
  ———反映总体k阶原点矩$\mathsf{E}X^k$.

常用统计量

- 样本均值(sample mean) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

  ——反映总体均值$\mu = \mathsf{E}X$.

- 样本方差(sample variance) $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

  (有时采用$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$, 注意区别!)

  ——反映总体方差$\sigma^2 = \mathsf{Var}(X)$.

- 样本k阶(原点)矩(k-th sample moment): $a_{n,k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

  ——反映总体k阶原点矩$\mathsf{E}X^k$.

- 样本k阶中心矩: $m_{n,k} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$

  ——反映总体k阶中心矩$\mathsf{E}(X - \mathsf{E}X)^k$.

- 样本变异系数: $S_n/\overline{X}$. 它反映了总体的变异系数$\sqrt{Var(X)}/E(X)$的信息. X的变异系数反映了随机变量X在以它的均值为单位时,取值的离散程度.

- 样本变异系数: $S_n/\overline{X}$. 它反映了总体的变异系数$\sqrt{Var(X)}/E(X)$的信息. X的变异系数反映了随机变量X在以它的均值为单位时,取值的离散程度.
- 样本偏度(sample skewness):

$$\frac{\frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})^3}{\left(\frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})^2\right)^{3/2}} = \frac{\sqrt{n}\sum_{i=1}^n (X_i-\overline{X})^3}{\left(\sum_{i=1}^n (X_i-\overline{X})^2\right)^{3/2}}.$$

它反映了总体偏度$\frac{\mathsf{E}(X-\mathsf{E}X)^3}{\sigma^3}$的信息.

- 样本变异系数: $S_n/\overline{X}$. 它反映了总体的变异系数$\sqrt{Var(X)}/E(X)$的信息. X的变异系数反映了随机变量X在以它的均值为单位时,取值的离散程度.
- 样本偏度(sample skewness):

$$\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right)^{3/2}}=\frac{\sqrt{n}\sum_{i=1}^{n}(X_i-\overline{X})^3}{\left(\sum_{i=1}^{n}(X_i-\overline{X})^2\right)^{3/2}}.$$

它反映了总体偏度$\frac{\mathsf{E}(X-\mathsf{E}X)^3}{\sigma^3}$的信息.

- 样本峰度(sample kurtosis):

$$\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^4}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right]^2}-3=\frac{n\sum_{i=1}^{n}(X_i-\overline{X})^4}{\left[\sum_{i=1}^{n}(X_i-\overline{X})^2\right]^2}-3.$$

它反映了总体峰度$\frac{\mathsf{E}(X-\mathsf{E}X)^4}{\sigma^4}-3$的信息.

- 次序统计量(待续)

- 次序统计量(待续)
- 经验分布函数

- 次序统计量(待续)
- 经验分布函数
- 二维随机向量的样本协方差(sample covariance)

$$S_{XY} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}).$$

统计量的独立性与可积性

Definition

定义 设T_1和T_2是分布族$\mathcal{F} = \{F\}$上的两个统计量,假如对每一个分布F而言, T_1与T_2都是相互独立的,则称这两个统计量是独立的.

统计量的独立性与可积性

Definition

定义 设T_1和T_2是分布族$\mathcal{F} = \{F\}$上的两个统计量,假如对每一个分布F而言,T_1与T_2都是相互独立的,则称这两个统计量是独立的.

Definition

定义 设T是分布族$\mathcal{F} = \{F\}$上的统计量, 假如对每一个分布F而言, T是可积的,即数学期望存在,则称这个统计量是可积的.