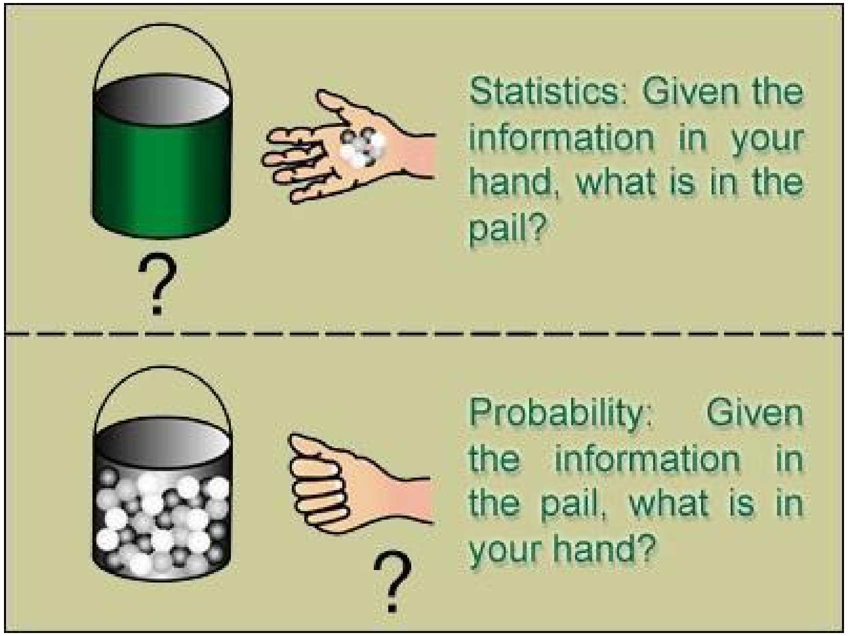
概率论 VS 数理统计

概率论：演绎推理（由模型推测数据）

统计：归纳（由数据预估模型）



数理统计：以概率论为理论基础，根据试验所观察到的数据，来研究随机现象. 通过统计分析对研究对象的客观规律做出合理的估计和推断.

统计的基本思想是从总体中合理抽出一部分个体（样本），对样本数据进行整理分析，从而根据样本信息来估计和推测总体的性质.

数理统计：对数据进行收集、整理、分析与推断

- 数理统计
  - 采集样本（合理、有效）
  - 统计推断
    - 统计估计
      - 参数估计
        - 点估计（第七章）
        - 区间估计（第七章）
      - 非参数估计（不讲）
    - 假设检验
      - 参数假设检验（第八章）
      - 非参数（分布）假设检验（不讲）

第六章　样本及样本函数的分布

一、总体与样本

二、直方图与样本分布函数

三、样本函数及其概率分布

四、χ^2分布

五、t分布

六、F分布

第一节　总体与样本

一、　总体

二、　简单随机样本

三、　小结

一、总体与个体

1. 总体

研究对象的全体称为总体或母体.

2. 个体　　总体中的每个元素称为个体.

实例1　考察某工厂生产的灯泡的寿命

总体**--**该厂生产的所有灯泡

个体**--**每一只灯泡

总体　**研究对象的数量指标X（随机变量）**

个体　**总体X的每个可能取值**

例　考察某工厂生产的灯泡的寿命

总体--该厂生产的所有灯泡的使用寿命X

个体--每一只灯泡的使用寿命，即X的每一个可能取值

3. 有限总体和无限总体

总体中所包含的个体总数叫做总体容量.

总体中包含有限个个体的叫有限总体,
包含无限个个体的叫无限总体。

当有限总体包含的个体的总数很大时, 可近似地将它看成是无限总体.

4. 总体的分布

总体X的分布函数

$$F(x) = P\{X \le x\}, -\infty < x < +\infty$$

叫做总体的分布函数;

总体X的数字特征叫做总体的数字特征。

若X是离散型随机变量, 其概率分布叫做离散型总体的概率分布;

若X是连续型随机变量, 其概率密度叫做连续型总体的概率密度;

将它们和总体X的分布函数统称为总体的分布。

二、 简单随机样本

从总体中随机抽取若干个个体的过程称为抽样.

从总体中随机抽取的待测个体组成的集合称为样本.

样本中所含的个体的数目称为样本容量.

从总体X中随机抽取的容量为n的样本常记为X_1，X_2，…, X_n，其中每个X_i都是随机变量.

样本X_1，X_2，…, X_n的确切数值 x_1，x_2，…, x_n 称为样本观测值.

例6.1.1 从某工厂生产的一批100个灯泡中, 随机抽测了 10个灯泡的使用寿命(小时)为: 1200, 1312, 1200, 1231, 1200, 1344, 894, 384, 1577, 1114.

灯泡使用寿命的全体是一个总体, 每一个灯泡的寿命是一个个体, 这10个灯泡的寿命就是一个样本.

简单随机抽样

为使样本能很好的反映总体的情况，从总体中抽取样本，必须满足下述两个条件：

1. **随机性** 为了使测试到的数据能很好地反映总体的情况，要求总体中每一个个体被抽到的可能性是相等的.即每个个体与总体同分布。

2. **独立性** 各次抽取必须是独立的，即每次的抽样结果之间互不影响.

这种随机的、独立的抽样方法称为简单随机抽样。
由此得到的样本称为简单随机样本。

从总体 X 中抽取 的n 个简单随机样本X_1，X_2，…，X_n 是与 X 具有相同分布的随机变量，且X_1，X_2，…，X_n 相互独立.

简单随机样本

若总体X的样本$X_1, X_2, \cdots, X_n$满足:

(1) $X_1, X_2, \cdots, X_n$与X有相同的分布;

(2) $X_1, X_2, \cdots, X_n$相互独立,

则称$X_1, X_2, \cdots, X_n$为来自总体X的简单随机样本. 简称样本.

样本X_1，X_2，…，X_n所有可能取值的全体称为样本空间,

样本观察值 x_1，x_2，…，x_n是样本空间中的一个点.

样本X_1，X_2，…，X_n所有可能取值的全体称为样本空间，样本观察值 x_1，x_2，…，x_n是样本空间中的一个点.

如果从总体X 中抽取样本, 得到观测值 $x_1, x_2, \ldots, x_n$, 则可认为相互独立事件 $\{X_1 = x_1\},\{X_2 = x_2\},\cdots\{X_n = x_n\}$ 同时发生.

若总体X的分布函数为$F(x)$, 则样本$X_1,\cdots,X_n$的联合分布函数为

$$F(x_1,\cdots,x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) = \prod_{i=1}^{n} F(x_i)$$

- 若总体 X为连续型随机变量，其概率密度为 $f(x)$，则样本$X_1, X_2, \ldots, X_n$的联合概率密度函数为

$$f(x_1,\cdots,x_n)=\prod_{i=1}^{n} f_{X_i}(x_i)=\prod_{i=1}^{n} f(x_i)$$

- 若总体X为离散型随机变量且分布律为$P\{X=x\}=p(x)$

则 $X_1,\cdots,X_n$ 的分布律为

$$P\{X_1=x_1,\cdots,X_n=x_n\}=\prod_{i=1}^{n} P\{X_i=x_i\}=\prod_{i=1}^{n} p(x_i).$$

例6.1.2 若$X_1,\cdots,X_n$是参数为λ的泊松分布总体X的样本，

求$(X_1,\cdots,X_n)$的分布律．

解：总体 X 的分布律为

$$p(x)=P\{X=x\}=\frac{\lambda^x}{x!}e^{-\lambda},x=0,1,\cdots$$

所以（$X_1,\cdots,X_n$）的分布律为

$$P\{X_1=x_1,\cdots,X_n=x_n\}=\prod_{i=1}^{n}p(x_i)$$

$$=\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}=\frac{\lambda^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}e^{-n\lambda},\quad x_i=0,1,\cdots.$$

例6.1.3 设总体 X 服从参数为 $\lambda\,(\lambda>0)$ 的指数分布, $(X_1,X_2,\cdots,X_n)$ 是来自总体的样本, 求样本 $(X_1,X_2,\cdots,X_n)$ 的概率密度.

解 总体 X 的概率密度为 $f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0,\\ 0, & x\le 0,\end{cases}$

因为 $X_1,X_2,\cdots,X_n$ 相互独立, 且与 X 有相同的分布,

所以 $(X_1,X_2,\cdots,X_n)$ 的概率密度为

$$f(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}f(x_i)=\begin{cases}\lambda^n e^{-\lambda\sum\limits_{i=1}^{n}x_i}, & x_i>0,\\ 0, & \text{其他}.\end{cases}$$

例6.1.4 设总体 X 服从两点分布 $B(1,p)$, 其中$0<p<1$, $(X_1,X_2,\cdots,X_n)$是来自总体的样本, 求样本 $(X_1,X_2,\cdots,X_n)$ 的分布律.

解 总体 X 的分布律为

$$P\{X=i\}=p^i(1-p)^{1-i} \qquad (i=0,1)$$

因为 $X_1,X_2,\cdots,X_n$相互独立,

且与 X 有相同的分布,

所以$(X_1,X_2,\cdots,X_n)$的分布律为

$$P\{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\}$$

$$= P\{X_1 = x_1\}P\{X_2 = x_2\}\cdots P\{X_n = x_n\}$$

$$= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

其中 $x_1, x_2, \cdots, x_n$ 在集合 $\{0,1\}$ 中取值.

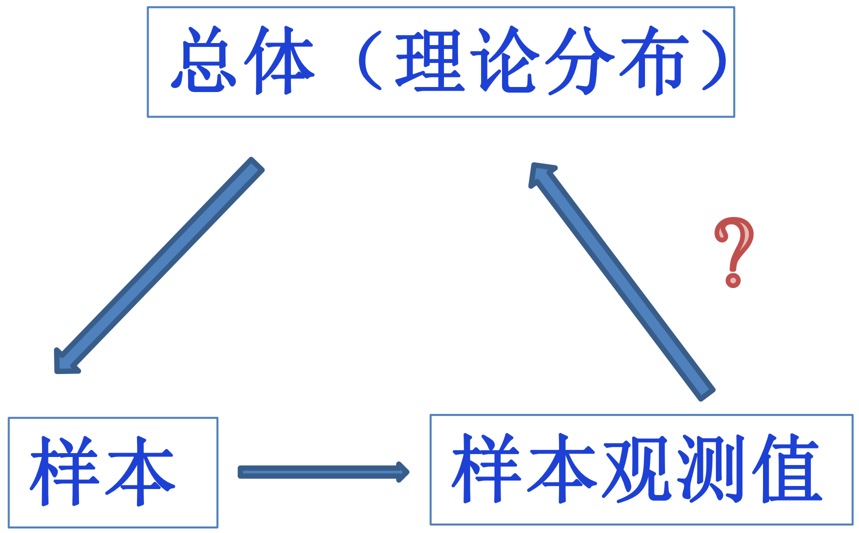
三、小结

数理统计：对数据进行收集、整理、分析与推断

基本概念：个体　总体 $\begin{cases}\text{有限总体}\\ \text{无限总体}\end{cases}$　简单随机样本

说明1　一个总体对应一个随机变量X, 我们将不区分总体和相应的随机变量, 统称为总体X.

说明2　在实际中遇到的总体往往是有限总体, 它对应一个离散型随机变量; 当总体中包含的个体的个数很大时, 在理论上可认为它是一个无限总体.



数理统计：对数据进行收集、整理、分析与推断