



西北工业大学
NORTHWESTERN POLYTECHNICAL UNIVERSITY



概率论与数理统计





第一节 基本概念



一、问题的提出



二、总体与个体



三、随机样本的定义



四、统计量



一、问题的提出

概率论中的一个最基本的假设就是：
研究对象的分布已知.

而在实际中，我们往往不知道随机变量的确切分布，这就是数理统计所讨论问题的应用背景，它需要用已有的部分信息去推断整体情况.

数理统计研究内容十分广泛，其中一类重要的问题便是统计推断.

统计推断是利用实验数据对研究对象的性质作出推断，其中的一个重要方面就是参数估计问题。

例如：某车站在旅游旺季的旅客到来人数服从Piosson分布，这样车站就需要知道该分布的具体参数，从而合理的安排车辆调度。但是在实际中，车站往往不知道分布的具体参数情况，仅仅知道一些往年旺季的车辆调度数据。



对于这个问题，由概率论的知识可知， $\text{Poisson}(\lambda)$ 分布的参数 λ 是其均值，那么一个直观粗略的想法就是计算出往年旺季汽车调度的均值，以其均值来代替 λ ，从而得到汽车的调度分布。

类似上述这类问题被称为——参数估计。



统计推断的另一个重要问题是**假设检验**。

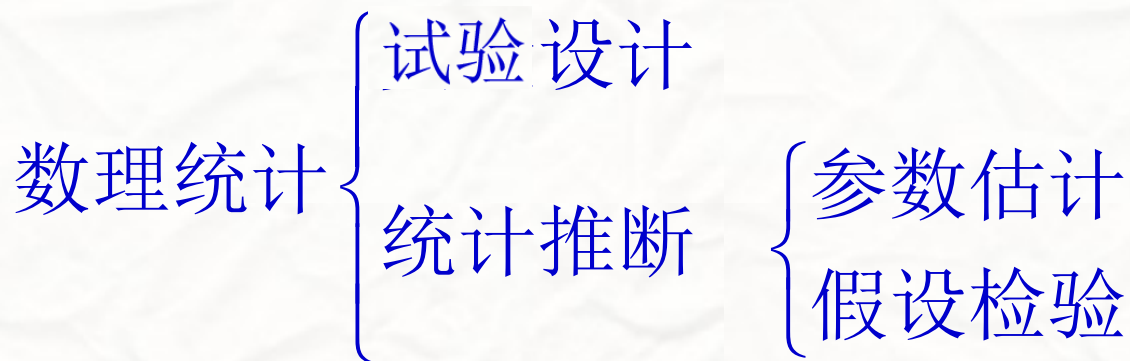
例如某电子元件企业为了提高该厂元器件的装箱出厂率，需要对元器件的合格率加以确定，而实际中决策人员只知道平时的次品率在1%附近，而且可以知道最近一批产品的装箱合格率.而我们现在想确认次品率是否为1%？




对于这个问题，我们的一个想法就是在货物中随机提取若干箱，然后对每箱的元器件进行逐箱检验，如果次品率小于等于1%的箱数占抽样总箱数的比例非常高，那么在实际中我们就认为该厂的元器件次品率为1%.

类似的问题被称为——假设检验.

概括地讲，数理统计研究以有效的方式采集、整理和分析受到随机因素影响的数据，并对所考察的问题做出推断和预测，直至提供依据和建议。



 **统计推断：** 研究如何加工、处理数据，从而对所考察对象的性质做出尽可能精确和可靠的推断。

二、总体与个体

总体： 在数理统计中，把研究对象的全体称为总体（或母体）。

个体： 总体中每个研究对象称为个体。

例如，在考察我校某届本科生学习质量时，该届本科生的全体称为一个**总体**，每一个本科生称为一个**个体**。

例1 考察某批电视机的寿命质量时，这批电视机的全体称为一个总体，每台电视机称为一个个体.



总体



个体

在实际中，我们并不关心总体的各个方面，而往往关心它的某项或几项数量指标。

当我们说到总体，就是指一个具有确定概率分布的随机变量。这时，每个个体具有的数量指标的全体就是总体。

通常，我们用随机变量 $X, Y, Z \dots$ 等
表示总体.

因此，在理论上可以把总体与概率分布等
同起来.

三、随机样本的定义

1. 样本的定义

从总体 X 中，随机地抽取 n 个个体：

$$X_1, X_2, \dots, X_n$$

称为总体 X 的一个**样本**，记为

$$(X_1, X_2, \dots, X_n)$$

样本中所包含个体的总数 n 称为**样本容量**。

注 样本 (X_1, X_2, \dots, X_n) 是一个 n 维随机变量。

例2 为了了解数学专业本科毕业生的月薪情况，调查了某地区100名2020届数学专业的本科生的月薪情况，试问

- (1) 什么是总体？
- (2) 什么是样本？
- (3) 样本容量是多少？

解 (1) 总体是该地区2020届数学本科毕业生的月薪；

(2) 样本是被调查的100名2020届数学本科毕业生的月薪；

(3) 样本容量是100.

2. 样本值

每一次抽取 X_1, X_2, \dots, X_n 所得到的 n 个确定的具体数值, 记为

$$(x_1, x_2, \dots, x_n)$$

称为样本

$$(X_1, X_2, \dots, X_n)$$

的一个**样本值**(观察值).



- 数理统计的基本任务是: 根据从总体中抽取的样本, 利用样本的信息推断总体的性质.

3. 简单随机样本

若来自总体 X 的样本 (X_1, X_2, \dots, X_n) 具有下列两个特征:

(1) 代表性: X_1, X_2, \dots, X_n 中每一个与总体 X 有相同的分布.

(2) 独立性: X_1, X_2, \dots, X_n 是相互独立的随机变量.

则称 (X_1, X_2, \dots, X_n) 为 n 的简单随机样本.

获得简单随机样本的抽样方法称为简单随机抽样.

例3 设人的身高服从正态分布，现考察某地区1000人的身高。其身高记为 $X_1, X_2, \dots, X_{1000}$ 则称 $X_1, X_2, \dots, X_{1000}$ 为来自于正态总体容量为1000的简单随机样本。

又如, 某实验室对200只灯泡作寿命实验, 并且规定当有60只灯泡失效时停止实验, 如此取得的寿命样本就不是简单随机样本。

总体和样本的严格数学定义：

定义5.1 一个随机变量 X 或者其相应的分布函数 $F(x)$ 称为一个总体.

定义5.2 设随机变量 X 的分布函数为 $F(x)$, 若 X_1, X_2, \dots, X_n 是具有同一分布函数 $F(x)$, 且相互独立的随机变量, 则称 X_1, X_2, \dots, X_n 为来自总体 X 的容量为 n 的简单随机样本, 简称样本.

4. 样本的分布

定理5.1 设 (X_1, X_2, \dots, X_n) 为来自总体 X 的样本.

(1)若总体 X 的分布函数为 $F(x)$,则样本

(X_1, X_2, \dots, X_n) 的分布函数为 $\prod_{i=1}^n F(x_i)$

(2)若总体 X 的分布密度为 $p(x)$,则样本

(X_1, X_2, \dots, X_n) 的分布密度为 $\prod_{i=1}^n p(x_i)$

(3)若总体 X 的分布律为 $P(X = x_i) = p(x_i)(i = 1, \cdots, n)$

则样本 (X_1, X_2, \cdots, X_n) 的分布律为 $\prod_{i=1}^n p(x_i)$.

注 熟记以上三种不同情形
下样本的联合概率分布的具
体形式是解决问题的关键!!



例4 设 X_k 来自于参数为 λ_k 的指数分布,
 $k = 1, 2, \dots, n$, 且相互独立, 求 (X_1, X_2, \dots, X_n)
的联合分布函数.

解 由于 X_k 来自于参数为 λ_k 的指数分布,

$$\text{所以 } F_{X_k}(x_k) = 1 - e^{-\lambda_k x_k}, x_k \geq 0$$

由独立性有:

$$\begin{aligned} F_{(X_1, \dots, X_n)}(x_1, x_2, \dots, x_n) &= \prod_{k=1}^n F_{X_i}(x_i) \\ &= \prod_{k=1}^n (1 - e^{-\lambda_k x_k}), \quad x_k \geq 0, k = 1, 2, \dots, n. \end{aligned}$$

例5 设 X_1, X_2, X_3 是来自柯西分布的样本, 其密度函数为 $p(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}, -\infty < x < +\infty$, 试求 (X_1, X_2, X_3) 的联合概率密度.

解 因为 X_i 自于柯西分布, 所以 X_i 的密度函数是
$$p_{X_i}(x_i) = \frac{1}{\pi} \cdot \frac{1}{1+x_i^2}, \quad -\infty < x_i < +\infty \quad (i=1,2,3)$$

所以 (X_1, X_2, X_3) 的联合概率密度是:

$$\begin{aligned} P_{(X_1, X_2, X_3)}(x_1, x_2, x_3) &= \prod_{i=1}^3 p_{X_i}(x_i) \\ &= \frac{1}{\pi^3} \cdot \frac{1}{(1+x_1^2)(1+x_2^2)(1+x_3^2)}, \quad x_1, x_2, x_3 \in R \end{aligned}$$

例6 设总体 X 服从参数为 λ 的Poisson分布
 (X_1, X_2, \dots, X_n) 是来自于总体的样本, 求此样本
的联合分布律.

解 总体 X 的分布律为

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, \dots$$

因为 X_1, X_2, \dots, X_n 独立同分布,
所以, (X_1, X_2, \dots, X_n) 的联合分布律为

$$P(X_1 = k_1, \cdots, X_n = k_n) = \prod_{i=1}^n P(X_i = k_i)$$
$$= \prod_{i=1}^n \frac{\lambda^{k_i}}{k_i!} e^{-\lambda} = \frac{\lambda^{k_1 + \cdots + k_n}}{k_1! k_2! \cdots k_n!} e^{-n\lambda}$$

$$k_i = 0, 1, 2, \cdots, i = 1, 2, \cdots, n$$

此即样本的联合分布律

四、统计量

由样本推断总体情况,需要对样本值进行“[加工](#)”,这就需要构造一些样本的函数,它把样本中所含的信息集中起来.

1. 统计量

定义5.3 设 (X_1, X_2, \dots, X_n) 是来自总体 X 的一个样本, $f(X_1, X_2, \dots, X_n)$ 是 X_1, X_2, \dots, X_n 的函数, 若 f 中不含任何关于总体 X 的未知参数, 则称 $f(X_1, X_2, \dots, X_n)$ 是一个统计量.

设 (x_1, x_2, \dots, x_n) 是样本 (X_1, X_2, \dots, X_n)
的观察值
则称 $f(x_1, x_2, \dots, x_n)$ 是 $f(X_1, X_2, \dots, X_n)$
的观察值

- 注** 1° 统计量 $f(X_1, X_2, \dots, X_n)$ 是随机变量;
- 2° 统计量用于统计推断, 故不应含任何关于总体 X 的未知参数;
- 3° 统计量是样本的函数, 它是一个随机变量, 统计量的分布称为**抽样分布**.

例7 设总体 X 服从参数为 λ 的Poisson分布
 X_1, X_2, \dots, X_n 为来自总体的样本, 判断下列那些
是统计量?

$T_1 = \min(X_1, \dots, X_n)$ 是

$T_2 = \begin{cases} 1, & \sum_{i=1}^n X_i > \lambda^{-1} \\ 0, & \text{其他} \end{cases}$ 不是

$T_3 = X_1 / X_n$ 是

解 直接利用统计量定义判别



统计量是样本的函数, 完全由样本决定, 不含任何关于总体 X 的未知参数.


2. 几个常用统计量的定义

(1) 样本矩

设 X_1, X_2, \dots, X_n 是来自总体的一个样本,
 x_1, x_2, \dots, x_n 是这一样本的观察值.

1) **样本均值** $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i;$

其观察值 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$



它反映了
总体均值
的信息

可用于推断: $E(X)$.

2) 样本方差

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

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

它反映了总体方差
的信息

可用于推断: $D(X)$.

其观察值

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

3) 样本标准差

$$S_n = \sqrt{S_n^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2};$$

其观察值

$$s_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

4) 修正样本方差

$$S_n^{*2} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right).$$

其观察值

$$s_n^{*2} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right).$$

样本方差与修正样本方差的关系：

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n} S_n^{*2}.$$

注 1° 当 n 较大时， S_n^{*2} 与 S_n^2 差别微小；

2° 当 n 较小时， S_n^{*2} 比 S_n^2 有更好的统计性质。

5) 样本 k 阶(原点)矩

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k, k = 1, 2, \dots; \text{特例: } A_1 = \bar{X}$$

其观察值 $a_k = \frac{1}{n} \sum_{i=1}^n x_i^k, k = 1, 2, \dots.$

6) 样本 k 阶中心矩

特例: $B_2 = S_n^2$

$$B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k, k = 2, 3, \dots;$$

其观察值 $b_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k, k = 2, 3, \dots.$

样本矩具有下列性质:

性质5.1 设总体 X 的期望 $EX = \mu$, 方差 $DX = \sigma^2$,
 (X_1, X_2, \dots, X_n) 为来自总体 X 的样本, 则有

$$(1) \quad E(\bar{X}) = \mu$$

$$(2) \quad D(\bar{X}) = \frac{1}{n} \sigma^2;$$

$$(3) \quad E(S_n^2) = \frac{n-1}{n} \sigma^2;$$

$$(4) \quad E(S_n^{*2}) = \sigma^2.$$

请熟记此结论
!!!

证 (1) $E(\bar{X}) = \mu$

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \mu$$

(2) $D(\bar{X}) = \frac{1}{n} \sigma^2$

$$\begin{aligned} D(\bar{X}) &= D\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n D(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n} \sigma^2. \end{aligned}$$

$$(3) E(S_n^2) = \frac{n-1}{n} \sigma^2$$

$$\begin{aligned} E(S_n^2) &= E\left[\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2\right] = \frac{1}{n} \sum_{i=1}^n E(X_i^2) - E(\bar{X}^2) \\ &= \frac{1}{n} \sum_{i=1}^n [D(X_i) + (E(X_i))^2] - [D(\bar{X}) + (E(\bar{X}))^2] \\ &= \frac{1}{n} \sum_{i=1}^n (\sigma^2 + \mu^2) - \left(\frac{1}{n} \sigma^2 + \mu^2\right) = \frac{n-1}{n} \sigma^2. \end{aligned}$$

$$(4) E(S_n^{*2}) = E\left(\frac{n}{n-1} S_n^2\right) = \frac{n}{n-1} E(S_n^2) = \sigma^2$$

例8 设对总体 X 服从两点分布 $B(1, p)$
 (X_1, X_2, \dots, X_n) 是来自于总体分布的样本,
 \bar{X} 和 S^2 是样本均值与修正样本方差, 试计算:
 $E\bar{X}, D\bar{X}$ 和 ES^2 .

解 由两点分布知 $EX = p, DX = p(1 - p)$

利用样本矩的性质得

$$E\bar{X} = EX = p$$

$$D\bar{X} = \frac{1}{n} DX = \frac{p(1 - p)}{n}$$

$$ES^2 = DX = p(1 - p).$$

性质5.2 若总体 X 的 k 阶矩 $E(X^k) = \mu_k$ 存在,

则当 $n \rightarrow \infty$ 时, $A_k \rightarrow \mu_k, k = 1, 2, \dots$

证 因为 X_1, X_2, \dots, X_n 独立且与 X 同分布,
所以 $X_1^k, X_2^k, \dots, X_n^k$ 独立且与 X^k 同分布,
故有 $E(X_1^k) = E(X_2^k) = \dots = E(X_n^k) = \mu_k$.

再根据第四章**辛钦定理**, 即

若*r.v* $X_1 \cdots X_n$ 独立同分布, 且 $EX_k = \mu$, 则

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right\} = 0$$

由上述定理可得

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k \xrightarrow{P} \mu_k, \quad (n \rightarrow \infty), \quad k = 1, 2, \dots;$$

由第四章关于依概率收敛的序列的性质知

$$g(A_1, A_2, \dots, A_k) \xrightarrow{P} g(\mu_1, \mu_2, \dots, \mu_k),$$

其中 g 是连续函数.

注 性质5.2是下一章矩估计法的理论根据.

(2) 次序统计量

设 (X_1, X_2, \dots, X_n) 是从总体 X 中抽取的一个样本,
 (x_1, x_2, \dots, x_n) 是其一个观测值,将观测值按由小
到大的次序重新排列为

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

当 (X_1, X_2, \dots, X_n) 取值为 (x_1, x_2, \dots, x_n) 时,定义

$X_{(k)}$ 取值为 $x_{(k)}$ ($k = 1, 2, \dots, n$),由此得到

$$(X_{(1)}, X_{(2)}, \dots, X_{(n)})$$

称为样本 (X_1, X_2, \dots, X_n) 的**次序统计量**.

对应的 $(x_{(1)}, x_{(2)}, \cdots, x_{(n)})$ 称为其观测值.

$X_{(k)}$: 样本 (X_1, X_2, \cdots, X_n) 的第 k 个次序统计量.

特别地, $X_{(1)} = \min_{1 \leq i \leq n} X_i$ 称为最小次序统计量.

$X_{(n)} = \max_{1 \leq i \leq n} X_i$ 称为最大次序统计量.

注 由于每个 $X_{(k)}$ 都是样本 (X_1, X_2, \cdots, X_n) 的函数, 所以, $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ 也是随机变量, 但它们一般不相互独立.

定理5.2 设总体 X 的分布密度为 $p(x)$ (或分布函数为 $F(x)$), $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ 为总体 X 的样本 (X_1, X_2, \dots, X_n) 的次序统计量. 则有

(1) 最大次序统计量 $X_{(n)}$ 的分布密度为

$$p_{X_{(n)}}(x) = n[F(x)]^{n-1} p(x).$$

(2) 最小次序统计量 $X_{(1)}$ 的分布密度为

$$p_{X_{(1)}}(x) = n[1 - F(x)]^{n-1} p(x).$$

证 (1) $F_{X_{(n)}}(x) = P\{X_{(n)} \leq x\}$

$$= P\{\max_{1 \leq i \leq n} X_i \leq x\}$$

$$= P\{X_1 \leq x, X_2 \leq x, \dots, X_n \leq x\}$$

$$= P\{X_1 \leq x\} \cdot P\{X_2 \leq x\} \cdots P\{X_n \leq x\}$$

$$= F^n(x)$$

$$\therefore p_{X_{(n)}}(x) = \frac{d F_{X_{(n)}}(x)}{d x} = n F^{n-1}(x) \cdot p(x)$$

$$(2) \quad F_{X_{(1)}}(x) = P\{X_{(1)} \leq x\}$$

$$= P\{\min_{1 \leq i \leq n} X_i \leq x\} = 1 - P\{\min_{1 \leq i \leq n} X_i > x\}$$

$$= 1 - P\{X_1 > x, X_2 > x, \dots, X_n > x\}$$

$$= 1 - [1 - F(x)]^n$$

$$\therefore p_{X_{(1)}}(x) = \frac{d F_{X_{(1)}}(x)}{d x}$$

$$= -n[1 - F(x)]^{n-1} \cdot [-F'(x)] = n[1 - F(x)]^{n-1} \cdot p(x).$$

例9 设总体 X 服从指数分布, 其密度函数为
 $p(x) = \lambda e^{-\lambda x}, x \geq 0$, 试求 $X_{(1)}, X_{(n)}$ 的分布.

解 由总体 X 的密度函数得, X 的分布函数为

$$F(x) = 1 - e^{-\lambda x}, x \geq 0$$

由定理得 $X_{(1)}$ 的分布密度为

$$\begin{aligned} p_{X_{(1)}}(x) &= n[1 - F(x)]^{n-1} p(x) \\ &= n\lambda e^{-n\lambda x}, x \geq 0 \end{aligned}$$

同样可得 $X_{(n)}$ 的分布密度为

$$\begin{aligned} p_{X_{(n)}}(x) &= n[F(x)]^{n-1} p(x) \\ &= n\lambda e^{-\lambda x} [1 - e^{-\lambda x}]^{n-1}, x \geq 0 \end{aligned}$$

(3) 经验分布函数

定义5.5 设 X_1, X_2, \dots, X_n 是总体 X 的一个样本, $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ 为总体 X 的样本 (X_1, X_2, \dots, X_n) 的次序统计量.

$(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ 为其观测值, 设 x 是任一实数, 称函数

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \frac{k}{n}, & x_{(k)} \leq x < x_{(k+1)}, \\ 1, & x \geq x_{(n)}. \end{cases}$$

为总体 X 的**经验分布函数**，即对于任何实数

x 经验分布函数 $F_n(x)$ 为样本值中不超过 x 的个数再除以 n ，亦即

$$F_n(x) = \frac{\mu_n(x)}{n}$$

其中 $\mu_n(x)$ ($-\infty < x < +\infty$) 表示 x_1, x_2, \dots, x_n 中不超过 x 的个数.

注 1° k 为样本中不超过 x 的样本的最大个数,
即在 n 次重复独立试验中, 事件
 $\{X \leq x\}$ 发生的次数.

($\because x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(k)} \leq x$, 有 k 个样品的取值 $\leq x$)

2° $F_n(x)$ 为事件 $\{X \leq x\}$ 的频率.

事实上, 令 $\mu_n(x) = \sum_{i=1}^n I_i$, 其中

$$I_i = \begin{cases} 1, & \{X_i \leq x\} \text{发生} \\ 0, & \{X_i \leq x\} \text{不发生} \end{cases}, \quad \text{则} \quad F_n(x) = \frac{\mu_n(x)}{n}.$$

性质

(1) 对于给定的一组样本值 (x_1, x_2, \dots, x_n) , $F_n(x)$ 满足分布函数的特征, 是一个分布函数.

(2) 由于 $F_n(x)$ 是样本的函数, 故 $F_n(x)$ 是随机变量. 可以证明 $nF_n(x) \sim B(n, F(x))$, 所以

$$E[F_n(x)] = F(x), \quad D[F_n(x)] = \frac{F(x)[1 - F(x)]}{n}$$

(3) $F_n(x)$ 依概率收敛于 $F(x)$. 即

$$\lim_{n \rightarrow \infty} P\{|F_n(x) - F(x)| < \varepsilon\} = 1 \quad (\forall \varepsilon > 0)$$

格里汶科定理5.3

格里汶科

对于任一实数 x , 当 $n \rightarrow \infty$ 时, $F_n(x)$ 以概率 1 一致收敛于分布函数 $F(x)$, 即

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

对于任一实数 x 当 n 充分大时, 经验分布函数的任一个观察值 $F_n(x)$ 与总体分布函数 $F(x)$ 只有微小的差别, 从而在实际上可当作 $F(x)$ 来使用.

例10 设从总体 X 中取得一个容量为5的样本，样本观测值为 -2, -1, 2.5, 3.1, 3.7, 试求此样本经验分布函数 $F(x)$.

解 由经验分布函数的定义知

$$F(x) = \begin{cases} 0 & x < -2 \\ 1/5 & -2 \leq x < -1 \\ 2/5 & -1 \leq x < 2.5 \\ 3/5 & 2.5 \leq x < 3.1 \\ 4/5 & 3.1 \leq x < 3.7 \\ 1 & 3.7 \leq x \end{cases}$$

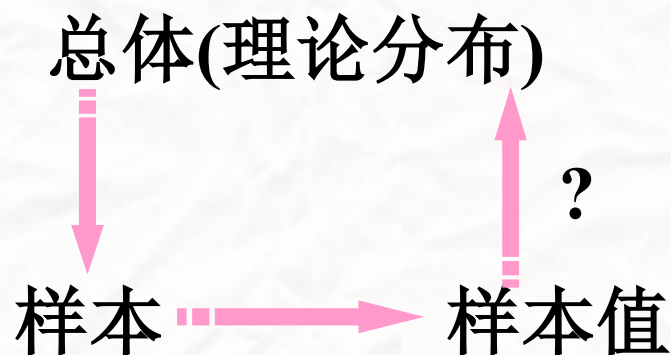
内容小结

基本概念: 个体 总体 { 有限总体
无限总体 随机样本

说明 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 .

统计推断 { 估计 — 参数估计 (第六章)
非参数估计
检验 — 参数假设检验 (第七章)
非参数检验 (第七章)

总体、样本、样本值的相互关系:



数理统计是从手中已有的资料--样本值,去推断总体的情况--总体的分布 $F(x)$ 的性质.

样本是联系二者的**桥梁**.

总体分布决定了样本取值的概率规律,也就是样本取到样本值的规律,因而可以由样本值去推断总体.

两个最重要的统计量:

样本均值

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

样本方差

$$S_n^{*2} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

再见



备用题

例1-1 考察某地区居民的人均收入或某批轿车的质量时，每个居民平均收入的全体或每辆轿车质量的全体就构成了总体。



某地区人均收入



某批轿车质量

例1-2 考察一批电子器件的寿命，从中任意抽取100件作完全寿命试验，则称这100只灯泡的寿命为来自总体的一个样本，100为其相应的样本容量；

寿命试验完成后，所测得的寿命数据称为样本观察值.

例2-1 某生产厂商想了解自己产品的使用情况，于是委托咨询公司进行了一次售后访查。

(1) 该项调查的总体是什么？

(2) 该项调查的样本是什么？

解 (1) 该项调查的总体是所有使用该产品的消费者；

(2) 该项调查的样本是所有被调查的消费者。

例2-2调查人员在某地区一高档住宅随机采访了230位居民的收入情况，宣布该地区的人均年收入为15万元，试用总体与样本的关系来分析此事？

答 该高档住宅的人均收入是该地区人均收入的一个特殊群体的特殊样本，它只能反映这个特殊群体的特征，并不能代表该地区的人均收入，即不能反映该地区的人均年收入情况，故这种说法是不真实！

例4-1 设 (X_1, X_2, \dots, X_n) 是来自于 $U(0, 1)$ 的一个样本, 求 (X_1, X_2, \dots, X_n) 的联合概率密度.

解 总体 X 的概率密度为

$$f(x) = 1, x \in (0, 1)$$

所以, (X_1, X_2, \dots, X_n) 的联合概率密度为

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= \prod_{i=1}^n f(x_i) \\ &= \prod_{i=1}^n 1 = 1, x_i \in (0, 1), i = 1, \dots, n \end{aligned}$$

例4-2 若 X_1, \dots, X_n 是总体 $X \sim N(\mu, \sigma^2)$ 的样本, 求 (X_1, \dots, X_n) 的联合概率密度.

解 总体 X 的概率密度为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty.$$

所以 (X_1, \dots, X_n) 的联合概率密度为

$$\begin{aligned} f(x_1, \dots, x_n) &= \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\ &= (2\pi)^{-\frac{n}{2}} (\sigma)^{-n} e^{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}} \quad (-\infty < x_i < \infty, i = 1, \dots, n) \end{aligned}$$

例4-3 设总体 X 在 $[a, b]$ 上服从均匀分布,
 X_1, X_2, \dots, X_n 是来自该总体的样本,
求样本的联合概率密.

解 由已知条件, 总体的概率密度为

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0 & \text{其他} \end{cases}$$

X_i 与 X 同分布, 即

$$X_i \sim f(x_i) = \begin{cases} \frac{1}{b-a}, & a \leq x_i \leq b \\ 0 & \text{其他} \end{cases} \quad i = 1, 2, \dots, n$$

则 X_1, X_2, \dots, X_n 联合概率密度为:

$$f(x_1, x_2, \dots, x_n) = \begin{cases} \frac{1}{(b-a)^n}, & a \leq x_i \leq b, i = 1, \dots, n \\ 0 & \text{其他} \end{cases}$$

例5-1 设总体 X 服从两点分布 $B(1, p)$, $0 < p < 1$,
 (X_1, X_2, \dots, X_n) 是来自总体的样本, 求样本 (X_1, X_2, \dots, X_n) 的分布律

解 总体 X 的分布律为

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

因为 X_1, X_2, \dots, X_n 相互独立,

且与 X 有相同的分布,

所以 (X_1, X_2, \dots, 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.

例5-2 设 X_1, X_2, \dots, X_n 是来自二项分布 $B(N, p)$ 的样本, 试求 (X_1, X_2, \dots, X_n) 的联合分布律.

解 由于样本来自二项分布, 所以 X 分布律为

$$P\{X_k = x_k\} = C_N^{x_k} p^{x_k} (1-p)^{N-x_k}, k = 1, 2, \dots, N$$

所以, (X_1, X_2, \dots, X_n) 的联合分布律为

$$\begin{aligned} P_{(X_1, X_2, \dots, X_n)}(x_1, x_2, \dots, x_n) &= \prod_{k=1}^n P\{X_k = x_k\} \\ &= C_N^{x_1} \cdots C_N^{x_n} p^{\sum_{k=1}^n x_k} (1-p)^{N-\sum_{k=1}^n x_k} \end{aligned}$$

例7-1 设 X_1, X_2, X_3 是来自总体 $N(\mu, \sigma^2)$ 的一个样本, 其中 μ 为已知, σ^2 为未知, 判断下列各式哪些是统计量, 哪些不是?

$$T_1 = X_1,$$

$$T_4 = \max(X_1, X_2, X_3),$$

$$T_2 = X_1 + X_2 e^{X_3}, \quad T_5 = X_1 + X_2 - 2\mu,$$

$$T_3 = \frac{1}{3}(X_1 + X_2 + X_3),$$

是

$$T_6 = \frac{1}{\sigma^2}(X_1^2 + X_2^2 + X_3^2).$$

不是

例7-2 设总体 X 服从两点 $B(1, p)$, 其中 p 是未知参数
 X_1, X_2, \dots, X_n 是来自总体的简单随机样本
试指出下列那些是统计量:

$$X_1 + X_2, (X_4 - X_n)^2, \min_{1 \leq i \leq n} X_i, \max_{1 \leq i \leq n} X_i?$$

解 根据统计量的定义, 它完全由样本决定,
而与未知参数无关, 所以

$$X_1 + X_2, (X_4 - X_n)^2, \min_{1 \leq i \leq n} X_i, \max_{1 \leq i \leq n} X_i$$

而 $X_3 + p$ 不是统计量.

例9-1 设总体 X 服从区间 $[0, \theta]$ 上的均匀分布,

(X_1, X_2, \dots, X_n) 为总体 X 的样本, 试求 $X_{(1)}$ 和 $X_{(n)}$ 的分布.

解 总体 X 的分布密度为 $p(x) = \begin{cases} \frac{1}{\theta}, & 0 \leq x \leq \theta \\ 0, & \text{其他} \end{cases}$

X 的分布函数为

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{\theta}, & 0 \leq x \leq \theta. \\ 1, & x > \theta \end{cases}$$

由定理5.2得 $X_{(1)}$ 的分布密度为

$$\begin{aligned} p_{X_{(1)}}(x) &= n[1 - F(x)]^{n-1} p(x) \\ &= \begin{cases} \frac{n}{\theta} \left(1 - \frac{x}{\theta}\right)^{n-1}, & 0 \leq x \leq \theta \\ 0, & \text{其他} \end{cases} \end{aligned}$$

而 $X_{(n)}$ 的分布密度为

$$p_{X_{(n)}}(x) = nF^{n-1}(x)p(x) = \begin{cases} \frac{n}{\theta^n} x^{n-1}, & 0 \leq x \leq \theta \\ 0, & \text{其他} \end{cases}$$

例9-2 设总体 X 服从柯西分布, 密度函数 $p(x)$

$$= \frac{1}{\pi} \cdot \frac{1}{1+x^2}, -\infty < x < +\infty, \quad X_1, X_2, \dots, X_{2n-1} \text{ 是}$$

来自总体的简单随机样本, 试求样本中位数的概率分布.

解 因为共有 $2n-1$ 个样本, 所以样本中位数为 $X_{(n)}$

由总体 的密度函数可得, X 的分布函数是

$$F(x) = \frac{1}{\pi} \cdot [\operatorname{arctg}(x) + \frac{\pi}{2}], -\infty < x < +\infty$$

所以，由定理得样本中位数的分布密度为

$$p_{X_{(n)}}(x) = (2n-1)C_{2n-2}^{n-1}[F(x)]^{n-1}[1-F(x)]^{n-1}p(x)$$

$$= C_{2n-2}^{n-1} \cdot \frac{2n-1}{\pi(1+x^2)}$$

$$\cdot \left[\frac{1}{\pi} (\operatorname{arctg}(x) + \frac{\pi}{2}) \right]^{n-1}$$

$$\cdot \left[\frac{1}{2} - \frac{1}{\pi} \operatorname{arctg}(x) \right]^{n-1}, -\infty < x < +\infty.$$

例9-3 设总体 X 的分布密度为 $f(x) = 2x, 0 < x < 1$
 (X_1, X_2, \dots, X_n) 是来自于总体的样本, 试求次序
统计量 $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ 的联合密度函数.

解 由次序统计量的联合密度函数性质知

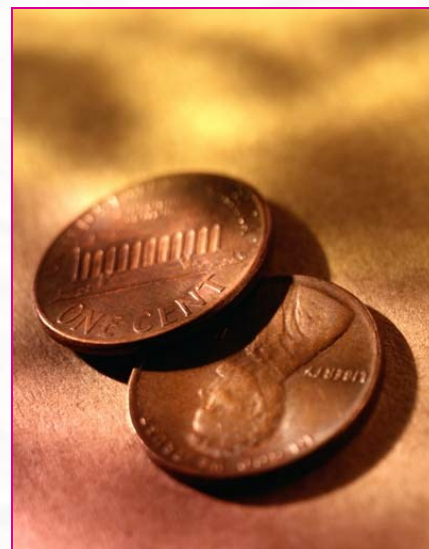
$$f_{(X_{(1)}, X_{(2)}, \dots, X_{(n)})}(x_{(1)}, x_{(2)}, \dots, x_{(n)})$$

$$= \begin{cases} n! \prod_{i=1}^n f(x_{(i)}), & x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)} \\ 0, & \text{其他} \end{cases}$$

$$= \begin{cases} 2^n n! \prod_{i=1}^n x_{(i)}, & x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)} \\ 0, & \text{其他} \end{cases}$$

例10-1 设总体 X 具有一个样本值1, 1, 2,
则经验分布函数 $F_3(x)$ 的观察值为

$$F_3(x) = \begin{cases} 0, & x < 1 \\ \frac{2}{3}, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$



例10-2 设总体 $X \sim F(x)$, 其经验分布函数为 $F_n(x)$ 试证 $E[F_n(x)] = F(x)$

$$D[F_n(x)] = \frac{1}{n} F(x)[1 - F(x)]$$

解 因为 $P\{nF_n(x) = k\} = C_n^k F^k(x)[1 - F(x)]^{n-k}$
所以 $E[nF_n(x)] = nF(x)$

$$D[nF_n(x)] = nF(x)[1 - F(x)]$$

从而 $E[F_n(x)] = F(x)$

$$D[F_n(x)] = \frac{1}{n^2} D[nF_n(x)] = \frac{1}{n} F(x)[1 - F(x)].$$

例10-3 设总体 X 具有一个样本值1, 2, 3,

则经验分布函数
 $F_3(x)$ 的观察值为 $F_3(x) = \begin{cases} 0, & x < 1, \\ \frac{1}{3}, & 1 \leq x < 2, \\ \frac{2}{3}, & 2 \leq x < 3, \\ 1, & x \geq 3. \end{cases}$