第7.1节　　参数的点估计

一、点估计问题的提法

二、估计量的求法

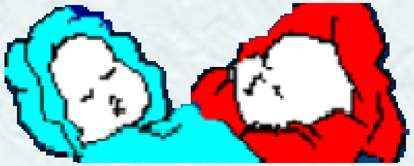
三、小结

现在我们来介绍一类重要的统计推断问题

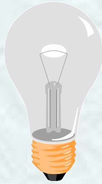
参数估计

参数估计问题是利用从总体抽样得到的信息来估计总体的某些参数或者参数的某些函数.

估计新生儿的平均体重

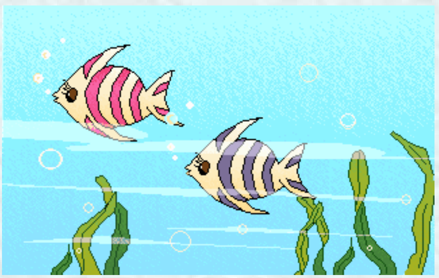


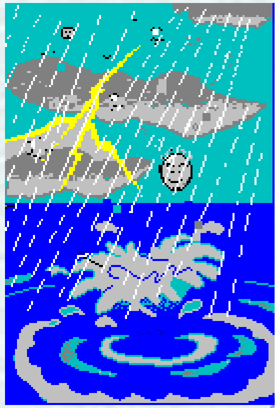
估计废品率



估计平均降雨量

估计湖中鱼数





…

…

在参数估计问题中，假定总体分布形式已知，未知的仅仅是一个或几个参数.

即当随机变量总体分布类型已知的情况下，从抽样得到的样本特征数值来有效推断总体未知参数的问题

参数估计问题的一般提法

即设有一个统计总体，总体的分布函数为 $F(x;\theta)$ 其中 θ 为未知参数，先从该总体抽样得 $X_1 \cdots X_n$，依样本对参数 θ 作出估计，或者估计关于参数 θ 的某个已知函数 $g(\theta)$，此类问题称为参数估计。

$$\text{参数估计}\begin{cases}\text{点估计}\\ \\ \text{区间估计}\end{cases}$$

例如我们要估计某队男生的平均身高.

（假定身高服从正态分布 $N(\mu, 0.1^2)$）

现从该总体选取容量为5的样本，我们的任务是要根据选出的样本（5个数）求出总体均值 μ 的估计. 而全部信息就由这5个数组成 .

设这5个数是:

1.65 1.67 1.68 1.78 1.69

估计 μ 为1.68， 这是点估计.

估计 μ 在区间 [1.57, 1.84] 内，这是区间估计.

一、点估计

设总体X的分布函数形式已知, 但它的一个或多个参数为未知, 借助于总体X的一个样本来估计总体未知参数的问题称为**点估计问题**.

点估计问题就是要构造一个适当的统计量 $\hat{\theta}(X_1, X_2, \cdots, X_n)$,用它的观察值 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 来估计未知参数 θ.

$$\left.\begin{array}{l}\hat{\theta}(X_1, X_2, \cdots, X_n)\text{称为}\theta\text{的估计量.}\\ \hat{\theta}(x_1, x_2, \cdots, x_n)\text{称为}\theta\text{的估计值.}\end{array}\right\}\begin{array}{l}\text{通称估计,}\\ \text{简记为}\hat{\theta}.\end{array}$$

注：估计量为样本的函数，样本不同，估计量不同。

例1 在某纺织厂细纱机上的断头次数X是一个随机变量,假设它服从以$\lambda>0$为参数的泊松分布,参数λ为未知,现检查了150只纱锭在某一时间段内断头的次数,数据如下,试估计参数λ.

断头次数k	0	1	2	3	4	5	6	
断头k次的纱锭数n_k	45	60	32	9	2	1	1	150

解 先确定一个统计量 $\overline{X}$,再计算出$\overline{X}$的观察值$\overline{x}$,把$\overline{x}$作为参数λ的估计值.

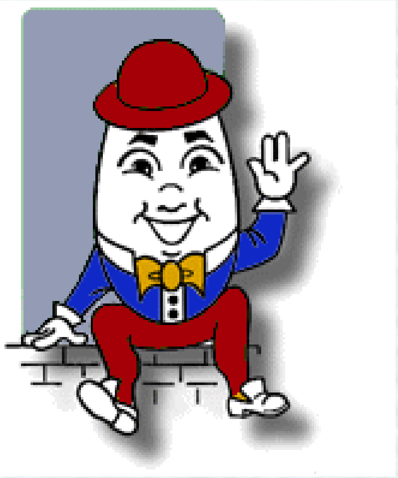
$\overline{x}=1.133$. λ的估计值为 1.133.

二、估计量的求法

由于估计量是样本的函数, 是随机变量, 故对不同的样本值, 得到的参数值往往不同, 求估计量的问题是关键问题.

估计量的求法: (两种)

矩估计法和极大似然估计法.



一、 矩估计法

它是基于一种简单的“替换”思想建立起来的一种估计方法 .

是英国统计学家*K*.皮尔逊最早提出的 .

矩估计法是通过参数与总体矩的关系，解出参数，并用样本矩替代总体矩而得到的参数估计方法。（由大数定理可知样本矩依概率收敛于总体矩，且许多分布所含参数都是矩的函数）

下面我们考虑总体为连续型随机变量的情况：

设X为连续型随机变量，其概率密度为$f(x,\theta_1,\cdots,\theta_k)$，其中$\theta_1,\cdots,\theta_k$为待估计的未知参数，假设$X$的$j$阶原点矩存在，则

$$a_j = E(X^j) = \int_{-\infty}^{+\infty} x^j f(x,\theta_1,\cdots,\theta_k)dx = a_j(\theta_1,\cdots,\theta_k),$$ 也是$\theta_1,\cdots,\theta_k$的函数。

假定方程组$\begin{cases} a_1(\theta_1,\cdots,\theta_k)=a_1 \\ a_2(\theta_1,\cdots,\theta_k)=a_2 \\ \vdots \\ a_k(\theta_1,\cdots,\theta_k)=a_k \end{cases}$，则可求出$\begin{cases} \theta_1=\theta_1(a_1,\cdots,a_k) \\ \theta_2=\theta_2(a_1,\cdots,a_k) \\ \vdots \\ \theta_k=\theta_k(a_1,\cdots,a_k) \end{cases}$

假定方程组 $\begin{cases} a_1(\theta_1,\cdots,\theta_k)=a_1 \\ a_2(\theta_1,\cdots,\theta_k)=a_2 \\ \vdots \\ a_k(\theta_1,\cdots,\theta_k)=a_k \end{cases}$,则可求出 $\begin{cases} \theta_1=\theta_1(a_1,\cdots,a_k) \\ \theta_2=\theta_2(a_1,\cdots,a_k) \\ \vdots \\ \theta_k=\theta_k(a_1,\cdots,a_k) \end{cases}$

则$X_1,\ldots,X_n$为X的样本时，可用$A_j=\frac{1}{n}\sum_{i=1}^{n}X_i^j$替代$a_j$，$j=1,\cdots,k$,

然后再用$\hat{\theta}_i=\theta_i(A_i,\cdots,A_k)$来估计$\theta_i(i=1,2,\cdots,k)$.

我们称$\hat{\theta}_i$为矩估计量，矩估计量的观察值为矩估计值。

矩估计法的具体步骤:

(1).求出$\alpha_l = E(X^l) = \alpha(\theta_1, \theta_2, \cdots, \theta_k) \quad l = 1,2,\cdots k$

(2).令 $\alpha_l = A_l, \quad A_l = \frac{1}{n}\sum_{i=1}^{n} X_i^l; l = 1,2,\cdots,k$

这是一个包含 k 个未知参数 $\theta_1, \theta_2, \cdots, \theta_k$ 的方程组.

(3).解出其中$\theta_1, \theta_2, \cdots, \theta_k$,用 $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k$ 表示.

(4).用方程组的解$\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k$分别作为$\theta_1, \theta_2, \cdots, \theta_k$的估计量,这个估计量称为矩估计量.

例2 设总体X服从泊松分布 $P(\lambda)$， 求参数λ的估计量.

解：设 $X_1, X_2, \cdots X_n$ 是总体X的一个样本,
由于 $E(X)=\lambda$,可得

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

例3 设总体 X 在 $[a,b]$ 上服从均匀分布，其中 a, b 未知，$(X_1, X_2, \cdots, X_n)$ 是来自总体 X 的样本，求 a, b 的矩估计量.

解 $\alpha_1 = E(X) = \dfrac{a+b}{2}$,

$$\alpha_2 = E(X^2) = D(X) + [E(X)]^2 = \frac{(a-b)^2}{12} + \frac{(a+b)^2}{4},$$

令 $\dfrac{a+b}{2} = A_1 = \dfrac{1}{n}\sum_{i=1}^{n} X_i$,

$$\frac{(a-b)^2}{12} + \frac{(a+b)^2}{4} = A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2,$$

即 $\begin{cases} a+b=2A_1 \\ b-a=\sqrt{12(A_2-A_1{}^2)} \end{cases}$

解方程组得到***a*, *b***的矩估计量分别为

$$\hat{a}=A_1-\sqrt{3(A_2-A_1{}^2)}=\overline{X}-\sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2},$$

$$\hat{b}=A_1+\sqrt{3(A_2-A_1{}^2)}=\overline{X}+\sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2},$$

例4 设总体 X服从几何分布,即有分布律

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

其中$p\ (0<p<1)$未知,$(X_1,X_2,\cdots,X_n)$是来自总体X的样本,求p的矩估计量.

解 $$\mu=E(X)=\sum_{k=1}^{\infty}kp(1-p)^{k-1}=\frac{1}{p},$$

令 $$\frac{1}{\hat{p}}=A_1=\overline{X},$$

$$\therefore \hat{p}=\frac{1}{\overline{X}}$$ 为所求p的矩估计量.

例5 设总体 X 的均值 μ 和方差 σ^2 都存在,且有 $\sigma^2 > 0$, 但 μ 和 σ^2 均为未知, 又设 $X_1, X_2, \cdots, X_n$ 是一个样本, 求 μ 和 σ^2 的矩估计量.

解 $\alpha_1 = E(X) = \mu,$

$$\alpha_2 = E(X^2) = D(X) + [E(X)]^2 = \sigma^2 + \mu^2,$$

令 $\begin{cases} \mu = A_1 \\ \sigma^2 + \mu^2 = A_2 \end{cases}$

解方程组得到矩估计量分别为 $\hat{\mu} = A_1 = \overline{X},$

$$\hat{\sigma}^2 = A_2 - A_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

上例表明:

总体均值与方差的矩估计量的表达式，不因不同的总体分布而异.

例 $X \sim N(\mu,\sigma^2)$, μ,σ^2未知,即得μ,σ^2的矩估计量

$$\hat{\mu} = \overline{X}, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

一般地:

用样本均值$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$作为总体$X$的均值的矩估计,

用样本二阶中心矩 $B_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$作为总体

X的方差的矩估计.

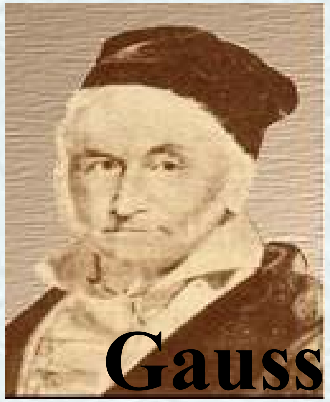
矩法的优点是简单易行,

缺点是矩估计值的精度一般比较差。

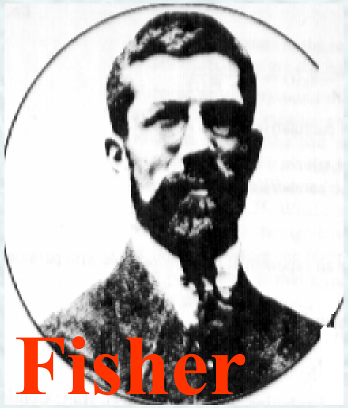
二、 最大（极大）似然估计法

极大似然法是在总体类型已知条件下使用的一种参数估计方法．

它首先是由德国数学家高斯在1821年提出的 ,然而，这个方法常归功于英国统计学家费歇．



费歇在1922年重新发现了这一方法，并首先研究了这种方法的一些性质．



- 极大似然估计是在已知总体分布形式的情形下的点估计。
- 极大似然估计的基本思路：根据样本的具体情况来选择未知参数θ的估计量　　，使样本发生的可能性最大。
  - 例：某试验有若干可能结果**A,B,C,…**，若在一次试验中，结果**A**出现，则一般认为试验条件对**A**有利，即**A**的概率最大。
  - 例：一名初学者和一名获得射击金牌者一同进行实弹射击，两人同打一个靶子，各打一发，最后仅中一发，则我们估计是获得金牌者射中。

- 例**3.**袋中有许多黑球和白球，已知他们的数目之比为**3:1**，但不知道哪种球多，试通过试验估计抽到黑球的概率。
- 解：显然抽到黑球的概率为**1/4**或**3/4**，设有放回球的前提下从袋中抽取**3**次球，其中抽到黑球的概率

$$X \sim B(3, p),$$

$$P\{X = k\} = C_3^k p^k (1-p)^{3-k}, (k = 0, 1, 2, 3; p = \frac{1}{4} 或 p = \frac{3}{4})$$

p \ x	0	1	2	3
1/4	27/64	27/64	9/64	1/64
3/4	1/64	9/64	27/64	27/64

若已知样本中黑球数x=0，则由对应于p=1/4的P{x=0}大于p=3/4的P{x=0},说明样本{x=0}来自于p=1/4的总体的可能性比来自p=3/4高。

因此，根据x＝0，用 $\hat{p}=1/4$ 来做p的估计较合理。

同理，x＝1时，$\hat{p}=1/4$；x＝2时，$\hat{p}=3/4$；

x＝3时，$\hat{p}=3/4$.

所以

$$\hat{p}=\begin{cases}1/4, x=0,1\\ 3/4, x=2,3\end{cases}$$

- 若总体X属连续型，其概率密度$f(x,\theta)$，θ为待估参数，样本$X_1,\cdots,X_n$的联合概率密度为

$$f(x_1,\cdots,x_n)=\prod_{i=1}^{n} f(x_i,\ \theta),$$ 对于确定的$x_1,\cdots,x_n$，

它是θ的函数，记为$L(x_1,\cdots,x_n;\theta)=\prod_{i=1}^{n} f(x_i,\theta)$，

并称其为<u>似然函数</u>，记为$L(\theta)$。

- 注：似然函数的概念并不仅限于连续随机变量，对于离散型随机变量，用$P\{X=x\}=p(x,\theta)$替代
- $f(x,\theta)$即可。

- 设总体X的分布形式已知，且只含一个未知参数θ，

若存在θ的一个值$\hat{\theta}$，使得当$\theta=\hat{\theta}$时，

$L(x_1,\cdots,x_n;\hat{\theta})=\max L(x_1,\cdots,x_n;\theta)$，则称$\hat{\theta}(x_1,\cdots,x_n)$为$\theta$

的极大似然估计值，称$\hat{\theta}(X_1,\cdots,X_n)$为$\theta$的极大似然估计量。

- **注：由于求未知参数θ的极大似然估计量 $\hat{\theta}$的问题，即求 $L(x_1,\cdots,x_n;\hat{\theta})$的极大值问题。**
- **一般步骤：**

当L关于θ可微时，L在$\dfrac{dL(\theta)}{d\theta}=0$处取得极值，

解方程即可求θ的极大似然估计$\hat{\theta}$；

又因为$L(\theta)$为连乘积形式，所以为方便运算，取对数。且$L(\theta)$与$\ln L(\theta)$在同一θ处取到极值，因此θ的极大似然估计$\hat{\theta}$也可从下述方程解得：$\frac{d}{d\theta}\ln L(\theta)=0.$

当待估参数不仅一个时，例如有k个$\theta_1,\cdots,\theta_k$时，L为$\theta_1,\cdots,\theta_k$的函数，记为$L(\theta_1,\cdots,\theta_k)$,则解下面方程组

$$\begin{cases}\dfrac{\partial \ln L}{\partial \theta_1}=0\\ \dfrac{\partial \ln L}{\partial \theta_2}=0\\ \vdots\\ \dfrac{\partial \ln L}{\partial \theta_k}=0\end{cases}$$
,可得$\theta_1,\cdots,\theta_k$的极大似然估计。

例1 设 $X \sim B(1, p)$, $X_1, X_2, \cdots, X_n$ 是来自 X 的一个样本, 求 p 的最大似然估计量.

解 设 $x_1, x_2, \cdots, x_n$ 为相应于样本 $X_1, X_2, \cdots, X_n$ 的一个样本值,

X 的分布律为 $P\{X = x\} = p^x (1-p)^{1-x}, x = 0,1$

似然函数 $L(p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$

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

$$\ln L(p) = \left(\sum_{i=1}^{n} x_i\right)\ln p + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-p),$$

令 $$\frac{\mathrm{d}}{\mathrm{d}p}\ln L(p) = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = 0,$$

解得 p 的最大似然估计值 $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$

p 的最大似然估计量为 $\boxed{\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}.}$

这一估计量与矩估计量是相同的.

例2 设总体 $X \sim N(\mu,\sigma^2)$, μ,σ^2为未知参数, $x_1, x_2, \cdots, x_n$ 是来自 X 的一个样本值, 求 μ 和 σ^2 的最大似然估计量.

解 X的概率密度为

$$p(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

X 的似然函数为

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}},$$

$$\ln L(\mu,\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2,$$

令 $$\begin{cases} \dfrac{\partial}{\partial\mu}\ln L(\mu,\sigma^2) = 0 \\ \dfrac{\partial}{\partial\sigma^2}\ln L(\mu,\sigma^2) = 0 \end{cases}$$

$$\begin{cases} \dfrac{1}{\sigma^2}\left[\displaystyle\sum_{i=1}^{n}x_i - n\mu\right] = 0, \\ -\dfrac{n}{2\sigma^2} + \dfrac{1}{2(\sigma^2)^2}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2 = 0, \end{cases}$$

由$\frac{1}{\sigma^2}\left[\sum_{i=1}^{n} x_i - n\mu\right] = 0$解得 $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$,

由$-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$解得

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

故μ和σ^2的最大似然估计量分别为

$\hat{\mu} = \bar{X}$, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. 它们与相应的矩估计量相同.

说明:

用上述求导方法求参数的最大似然估计有时行不通，这时要用极大似然原则来求.

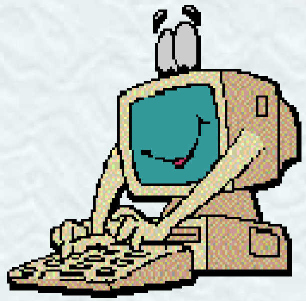
例3 设总体 X在$[a,b]$上服从均匀分布,其中a, b未知, $x_1,x_2,\cdots,x_n$是来自总体 X的一个样本值, 求a,b的最大似然估计量.

解 记 $x_{(1)}=\min(x_1,x_2,\cdots,x_n)$,

$$x_{(n)}=\max(x_1,x_2,\cdots,x_n),$$

X的概率密度为

$$p(x;a,b)=\begin{cases}\dfrac{1}{b-a}, & a\le x\le b\\ 0, & \text{其它}\end{cases}$$



因为$a \le x_1, x_2, \cdots, x_n \le b$等价于$a \le x_{(1)}, x_{(n)} \le b$,

作为a, b的函数的似然函数为

$$L(a,b)=\begin{cases}\dfrac{1}{(b-a)^n}, & a \le x_{(1)},\ b \ge x_{(n)}\\ 0, & \text{其它}\end{cases}$$

于是对于满足条件 $a \le x_{(1)},\ b \ge x_{(n)}$的任意$a,b$有

$$L(a,b)=\frac{1}{(b-a)^n} \le \frac{1}{(x_{(n)}-x_{(1)})^n},$$

即似然函数 $L(a,b)$ 在 $a = x_{(1)}$, $b = x_{(n)}$ 时

取到最大值 $(x_{(n)} - x_{(1)})^{-n}$,

a, b 的最大似然估计值

$$\hat{a} = x_{(1)} = \min_{1 \le i \le n} x_i, \qquad \hat{b} = x_{(n)} = \max_{1 \le i \le n} x_i,$$

a, b 的最大似然估计量

$$\hat{a} = \min_{1 \le i \le n} X_i, \qquad \hat{b} = \max_{1 \le i \le n} X_i .$$

三、小结

两种求点估计的方法: $\begin{cases} \text{矩估计法} \\ \text{最大似然估计法} \end{cases}$

在统计问题中往往先使用最大似然估计法, 在最大似然估计法使用不方便时, 再用矩估计法.

$$\text{似然函数 } L(\theta) = L(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} p(x_i; \theta)$$

费希尔资料

Ronald Aylmer Fisher

Born: 17 Feb 1890 in London, England

Died: 29 July 1962 in Adelaide, Australia