概率论与数理统计

上 下 回 停

第一节　参数的点估计

- 一、问题的提出
- 二、矩估计法
- 三、最大似然估计

一、点估计问题的提出

在实际中我们经常遇到这样的问题：总体 X 的分布函数 $F(x;\theta)$ 的形式为已知， θ 是未知参数. $X_1,X_2,\cdots,X_n$ 是 X 的一个样本, $x_1,x_2,\cdots,x_n$ 为相应的一个样本值. 我们希望用样本值去估计未知参数 θ，这种问题称为参数估计问题.

例1 已知某电话局在单位时间内收到用户呼唤次数这个总体 X 服从泊松分布 $p(\lambda)$，即 X的分布律

$$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda} \qquad (k=0,1,2,\cdots)$$

的形式已知．利用样本值 $(x_1, x_2, \cdots, x_n)$ 估计 $\lambda=E(X)$的值.

例2　已知某种灯泡的寿命 $X \sim N(\mu, \sigma^2)$，即 X 的分布密度

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

的形式已知，但参数 $\mu,\ \sigma^2$ 未知．利用样本值 $x_1,\ x_2, \cdots,\ x_n$，估计 $\mu = E(X), \sigma^2 = D(X)$.

例3 考虑某厂生产的一批电子元件的寿命这个总体 X；不知道 X 的分布形式，根据样本值 $(x_1, x_2, \cdots, x_n)$ 估计元件的平均寿命和元件寿命的差异程度,即估计总体 X 的均值 $E(X)$ 和方差 $D(X)$.

解决上述参数 θ 的点估计问题的思路是: 设法构造一个合适的统计量 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$, 对 θ 作出合理的估计 .

在数理统计中称统计量 $\hat{\theta}=\hat{\theta}(X_1, X_2,\cdots, X_n)$ 为 θ 的估计量, $\hat{\theta}$ 的观测值 $\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)$ 称为 θ 的估计值 .

点估计常用方法：矩估计和最大似然估计法.

二、矩估计法

矩估计法是由英国统计学家

皮尔逊(K.Pearson)在1894年提出.

矩估计法的基本思想是用样本的 k 阶原点矩

$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ 去估计总体 X 的 k阶原点矩$E(X^k)$;

用样本的 k 阶中心矩 $B_k = \frac{1}{n}\sum_{i=1}^{n}(X - \overline{X})^k$ 去估计总体

的k阶中心矩$E[X - E(X)]^k$;

并由此得到未知参数的估计量 .

设总体X的分布函数为$F\left(x;\ \theta_1,\ \theta_2,\cdots,\ \theta_m\right)$,

$\left(\theta_1,\ \theta_2,\cdots,\ \theta_m\right)$是$m$个待估计的未知参数．设

$\alpha_m = E(X^m)$存在，对任意k，$\left(k=1,2,\cdots,\ m\right)$

$$\alpha_k = E\left(X^k\right) = \int_{-\infty}^{+\infty} x^k dF\left(x;\ \theta_1,\ \theta_2,\cdots,\ \theta_m\right) = \alpha_k\left(\theta_1,\ \theta_2,\cdots,\ \theta_m\right)$$

现用样本矩作为总体矩的估计，即令

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k = \alpha_k\left(\hat{\theta}_1,\hat{\theta}_2,\ldots\hat{\theta}_m\right) \quad \left(k=1,2,\cdots,m\right)$$

这便得到含m个参数$\hat{\theta}_1$，$\hat{\theta}_2,\cdots$，$\hat{\theta}_m$的m个方程组，

解该方程组得

$$\hat{\theta}_k=\hat{\theta}_k(X_1,X_2,\cdots,X_n)\quad(k=1,2,\cdots,m)$$

以$\hat{\theta}_k$作为参数θ_k的估计量. 这种求出估计量的方法

称为矩估计法 .

例4 设总体X服从泊松分布$P(\lambda)$，求参数λ的矩估计量.

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5 求总体 X 的均值 μ 和方差 σ^2 的矩估计.

解 设 $X_1, X_2, \cdots, X_n$ 是总体 X 的一个样本,

由于

$$\begin{cases} E(X) = \mu \\ E(X^2) = D(X) + (EX) = \sigma^2 + \mu^2 \end{cases}$$

故令

$$\begin{cases} \overline{X} = \mu \\ \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 = \sigma^2 + \mu^2 \end{cases}$$

解得

$$\begin{cases} \hat{\mu} = \overline{X} \\ \hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = S_n^2 \end{cases}$$

例6 设总体 X 服从区间上 $[\theta_1,\theta_2]$ 的均匀分布,

求参数 θ_1,θ_2 的矩估计量.

解 设 $X_1, X_2, \cdots, X_n$ 是总体 X 的一个样本,

容易求得

$$\begin{cases} E(X)=\dfrac{\theta_1+\theta_2}{2} \\ D(X)=\dfrac{(\theta_2-\theta_1)^2}{12} \end{cases}$$

故令

$$\begin{cases} \bar{X} = \dfrac{\theta_1 + \theta_2}{2} \\ S_n^2 = \dfrac{(\theta_2 - \theta_1)^2}{12} \end{cases}$$

解得θ_1和 θ_2 的矩估计量为

$$\hat{\theta}_1 = \bar{X} - \sqrt{3} S_n$$

$$\hat{\theta}_2 = \bar{X} + \sqrt{3} S_n$$

例7 设总体X的分布 密度为

$$p(x;\ \theta)=\frac{1}{2\theta}e^{-\frac{|x|}{\theta}}\quad(-\infty<x<+\infty,\ \theta>0)$$

$X_1, X_2,\cdots, X_n$为总体X的一个样本，求参数θ的矩估计量.

解 由于$p(x;\ \theta)$只含有一个未知参数θ,一般只需求出$E(X)$便能得到θ的矩估计量，但是

$$E(X)=\int_{-\infty}^{+\infty}x\cdot\frac{1}{2\theta}e^{-\frac{|x|}{\theta}}dx=0$$

即$E(X)$不含有θ,故不能由此得到θ的矩估计量.

$$E\left(X^2\right)=\int_{-\infty}^{+\infty}x^2\frac{1}{2\theta}e^{-\frac{|x|}{\theta}}dx=\frac{1}{\theta}\int_0^{+\infty}x^2e^{-\frac{x}{\theta}}dx=2\theta^2$$

于是解得θ的矩估计量为

$$\hat{\theta}=\sqrt{\frac{1}{2n}\sum_{i=1}^{n}X_i^2}$$

本例θ的矩估计量也可以这样求得

$$E|X|=\int_{-\infty}^{+\infty}|x|\frac{1}{2\theta}e^{-\frac{|x|}{\theta}}dx=\frac{1}{\theta}\int_0^{+\infty}xe^{-\frac{x}{\theta}}=\theta$$

故令

$$\frac{1}{n}\sum_{i=1}^{n}|X_i| = \theta$$

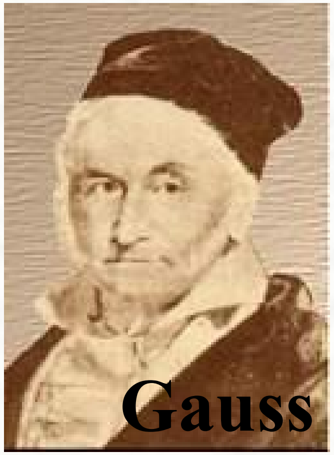
即 θ 的矩估计量为

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}|X_i|$$

该例表明参数的矩估计量不唯一.

三、最大似然估计

最大似然估计作为一种点估计方法最初是由德国数学家高斯(Gauss)于1821年提出，英国统计学家费歇尔(R.A.Fisher)在1922年作了进一步发展使之成为数理统计中最重要应用最广泛的方法之一.



目 上 下 回 停 续

1. 似然函数

设总体 X的分布律为$P(X=x)=p(x;\ \theta)$或分布密度为$p(x;\ \theta)$，其中$\theta=(\theta_1,\theta_2,\ldots,\theta_m)$是未知参数，$(X_1,\ X_2,\cdots,\ X_n)$的分布律(或分布密度)为$\prod_{i=1}^{n}p(x_i;\ \theta)$，当给定样本值 $(x_1,\ x_2,\cdots,\ x_n)$后，它只是参数θ的函数，记为$L(\theta)$，即

$$L(\theta)=\prod_{i=1}^{n}p(x_i;\ \theta)$$

则称 $L(\theta)$为似然函数，似然函数实质上是样本的分布律或分布密度.

2. 最大似然估计法

最大似然原理的直观想法：在试验中概率最大的事件最有可能出现．一个试验如有若干个可能结果 A，$B,\cdots,$若在一次试验中，结果A出现，则认为A出现的概率最大.

例8 假定一个盒中黑球和白球两种球的数目之比为 3:1，但不知哪种球多， p表示从盒中任取一球是黑球的概率，那么$p=1/4$或$3/4$，现在有放回地从盒中抽3个球，试根据样本中的黑球数 X 来估计参数p.

解 随机变量 $X \sim B(3,\ p)$，即

$$P\{X=x\}=C_3^x p^x (1-p)^{3-x} \quad (X=0,1,2,3)$$

估计 p 只需在 $p=1/4$ 和 $p=3/4$ 之间作出选择.

计算这两种情况下 X 的分布律:

X	0	1	2	3
$p=1/4, P\{X=x\}$	27/64	27/64	9/64	1/64
$p=3/4, P\{X=x\}$	1/64	9/64	27/64	27/64

p 的估计

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

目 上 下 回 停 续

设总体X 的分布律为$P\{X=x\}=p(x;\theta)$. $(x_1,x_2,\cdots,x_n)$是样本的一个观测值, 则样本$(X_1, X_2,\cdots, X_n)$取值$(x_1,x_2,...,x_n)$的概率为

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

$$=\prod_{i=1}^{n}P\{X=x_i\}=\prod_{i=1}^{n}p(x_i;\theta)=L(\theta)$$

既然在一次试验中得到的样本值$(x_1,x_2,...,x_n)$,

那么样本取该样本值的概率应较大，所以选取使

这似然函数 $L(\theta)$达到最大的参数值作为估计值，

称为最大似然估计法.

目 上 下 回 停 续

定义6.1 设总体 X的分布密度(或分布律)为$p(x;\ \theta)$,

其中$\theta=(\theta_1,\theta_{2,},...,\theta_m)$为未知参数．又设$x_1,x_2,...,x_n$

是总体X 的一个样本值，如果似然函数

$$L(\theta)=\prod_{i=1}^{n}p(x_i;\ \theta)$$

在$\hat{\theta}=(\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_m)$处达到最大，则称$\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_m$

分别为$\theta_1,\theta_2,\cdots,\theta_m$的最大似然估计量.

由于

$$\ln L(\theta)=\sum_{i=1}^{n}\ln p(x_i;\ \theta)$$

$\ln L(\theta)$ 与 $L(\theta)$ 有相同的最大值点 .因此, $\hat{\theta}$ 为最大似然估计的必要条件为

$$\left.\frac{\partial \ln L(\theta)}{\partial \theta_i}\right|_{\theta=\hat{\theta}}=0 \qquad (i=1,2,\cdots,\ m)$$

称它为似然方程, 其中 $\theta=(\theta_1,\theta_2,\ldots,\theta_m)$.

求最大似然估计量的一般步骤为：

1° 求似然函数 $L(\theta)$；

2° 求出 $\ln L(\theta)$ 及似然方程

$$\left.\frac{\partial \ln L(\theta)}{\partial \theta_i}\right|_{\theta=\hat{\theta}} = 0 \quad (i = 1,2,\cdots, m)$$

3° 解似然方程得到最大似然估计值

$$\hat{\theta}_i = \hat{\theta}_i(x_1, x_2, \cdots, x_m) \quad (i = 1,2,\cdots, m)$$

4° 最后得到最大似然估计量

$$\hat{\theta}_i = \hat{\theta}_i(X_1, X_2, \cdots, X_m) \quad (i = 1,2,\cdots, m)$$

例9 设总体X服从泊松分布$P(\lambda)$,其中λ为未知参数，试求参数λ的最大似然估计量.

解 设样本$(X_1, X_2, \cdots, X_n)$的一个观测值为$(x_1, x_2, \ldots, x_m)$,由于总体 $X \sim P(\lambda)$,故有

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

似然函数为

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{\lambda^{\sum\limits_{i=1}^{n} x_i}}{\prod\limits_{i=1}^{n} x_i!} e^{-n\lambda}$$

取对数

$$\ln L(\lambda)=\left(\sum_{i=1}^{n}x_i\right)\ln\lambda-\ln\prod_{i=1}^{n}x_i!-n\lambda$$

$$\frac{d\ln L(\lambda)}{d\lambda}=\frac{1}{\lambda}\sum_{i=1}^{n}x_i-n=0$$

即 $$\hat{\lambda}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$$

所以 λ 的最大似然估计量为 $\hat{\lambda}=\bar{X}$.

例10 设总体$X \sim N\left(\mu,\sigma^2\right)$，求参数$\mu$，$\sigma^2$的最大似然估计量 .

解 设$\left(X_1, X_2,\cdots, X_n\right)$是总体 X的样本，其观测值为$\left(x_1,x_2,\ldots,x_n\right)$, 由总体 $X \sim N\left(\mu,\sigma^2\right)$，分布密度为

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

目 上 下 例 停 续

似然函数

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

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

$$\frac{\partial\ln L(\theta)}{\partial\mu}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0$$

$$\frac{\partial\ln L(\theta)}{\partial\sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2=0$$

解似然方程得

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x},$$

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

最大似然估计量为$\hat{\mu}=\bar{X},\hat{\sigma}^2=S_n^2.$

例11 设总体 X服从区间$[0, \theta]$上的均匀分布,试求参数θ矩估计量和最大似然估计量.

解 设$(X_1, X_2, \cdots, X_n)$ 是总体 X 的样本,其观测值为$(x_1, x_2, \ldots, x_n)$,

$$E(X) = \frac{\theta}{2}$$

故 $$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\theta}{2}$$

即 θ的矩估计量为$\hat{\theta} = \frac{2}{n}\sum_{i=1}^{n} X_i = 2\bar{X}$

总体 X 的分布密度为

$$p(x;\ \theta)=\begin{cases}\dfrac{1}{\theta} & (0\le x\le\theta)\\ 0 & \text{其他}\end{cases}$$

则似然函数为

$$L(\theta)=\prod_{i=1}^{n}p(x;\ \theta)=\begin{cases}\dfrac{1}{\theta^n} & (0\le x_1,x_2,\ldots,x_n\le\theta)\\ 0, & \text{其他}\end{cases}$$

$$=\begin{cases}\dfrac{1}{\theta^n}, & \max\limits_{1\le i\le n}x_i\le\theta<+\infty,\min x_i\ge 0\\ 0, & \text{其他}\end{cases}$$

当 $\theta = \max\limits_{1 \le i \le n} x_i$ 时 $L(\theta)$ 达到最大, 故 θ 的最大似然

估计量为

$$\theta = \max_{1 \le i \le n} X_i = X_{(n)}$$

例12 设总体 X 在 $[\theta,\theta+1]$ 上服从均匀分布,其中 θ 未知, $x_1,x_2,\cdots,x_n$ 是来自总体 X 的一个样本值,求 θ 的最大似然估计量.

解 X 的概率密度为

$$f(x;\theta)=\begin{cases}1, & \theta\le x\le\theta+1\\ 0, & \text{其它}\end{cases}$$

$$L(\theta)=\begin{cases}1, & \theta\le x_{(1)},\ x_{(n)}\le\theta+1,\\ 0, & \text{其它}\end{cases}$$

似然函数在不为零的区域上是常数。

那么，只要θ不超过$x_{(1)}$, $\theta+1$不小于$x_{(n)}$，都可使L达到最大。

也即$x_{(n)}-1\le\theta\le x_{(1)}$时, 都可使L达到最大。

故θ的最大似然估计是区间$[X_{(n)}-1, X_{(1)}]$中任一值。

此例说明最大似然估计有时不唯一。

内容小结

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.

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

目 上 下 回 停 续


再见

备用题

例9-1 一罐中装有白球和黑球，有放回地抽取一个容量为n的样本，其中有 k 个白球，求罐中黑球与白球之比 R 的最大似然估计.

解

$$X_i = \begin{cases} 1, & \text{取到白球} \\ 0, & \text{取到黑球} \end{cases} \qquad i = 1, \ldots, n$$

则$X_1, X_2, \ldots, X_n$是取自$B(1, p)$的样本，p是每次抽取时取到白球的概率，p未知 .

先求 p 的最大似然估计:

目 上 下 回 停 续

我们容易求得

p的最大似然估计为 $\hat{p}=\dfrac{k}{n}$

由前述最大似然估计的性质不难求得

$R=\dfrac{1-p}{p}$ 的最大似然估计是

$$\hat{R}=\frac{1-\hat{p}}{\hat{p}}=\frac{n}{k}-1$$

例10-1 设总体 X服从区间$[0, \theta]$上的均匀分布,试求参数θ矩估计量和最大似然估计量.

解 设$(X_1, X_2, \cdots, X_n)$ 是总体 X 的样本,

其观测值为$(x_1, x_2, ..., x_n)$,

$$E(X) = \frac{\theta}{2}$$

故 $$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\theta}{2}$$

即 θ的矩估计量为$\hat{\theta} = \dfrac{2}{n}\sum_{i=1}^{n} X_i = 2\bar{X}$

总体 X 的分布密度为

$$p(x;\ \theta)=\begin{cases}\dfrac{1}{\theta} & (0\le x\le\theta)\\ 0 & 其他\end{cases}$$

则似然函数为

$$L(\theta)=\prod_{i=1}^{n}p(x;\ \theta)=\begin{cases}\dfrac{1}{\theta}, & (0\le x_1,x_2,\dots,x_n\le\theta)\\ 0, & 其他\end{cases}$$

$$=\begin{cases}\dfrac{1}{\theta^n}, & \max\limits_{1\le i\le n} x_i\le\theta<+\infty,\min x_i\ge 0\\ 0, & 其他\end{cases}$$

目 上 下 回 停 续

当 $\theta = \max\limits_{1\le i\le n} x_i$ 时 $L(\theta)$ 达到最大, 故 θ 的最大似然

估计量为

$$\theta = \max_{1\le i\le n} X_i = X_{(n)}$$

目 上 下 回 停 续

例10-2 设总体 X 服从 $N(\mu,\sigma^2)$, 对于容量为 n 的样本, 求使得 $\int_A^{+\infty} f(x;\mu;\sigma^2)dx = 0.05$ 的点 A 的最大似然估计.

解 设 $x_1, x_2, \cdots, x_n$ 为来自总体 X 的一个样本, 可求得 μ 与 σ^2 的最大似然估计分别为

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

由 $\int_{A}^{+\infty} f\left(x;\mu;\sigma^{2}\right)dx = P(x > A)$

$$= P\left(\frac{x-\mu}{\sigma} > \frac{A-\mu}{\sigma}\right) = 0.05$$

则 $P\left(\frac{x-\mu}{\sigma} \le \frac{A-\mu}{\sigma}\right) = \Phi\left(\frac{A-\mu}{\sigma}\right) = 0.95$

查表得 $\frac{A-\mu}{\sigma} = 1.645$

故 A 的最大似然估计 $\hat{A} = \hat{\mu} + 1.645\hat{\sigma}$.