参数估计

1、点估计

一般常用θ表示参数，参数θ所有可能取值组成的集合称为参数空间，常用Θ表示。参数估计问题就是根据样本对上述各种未知参数做出估计。

参数估计的形式有两种：点估计与区间估计

在这里如何构造统计量并没有明确的规定，只要它满足一定的合理性即可，这就涉及到两个问题：

（1）是如何给出估计，即估计的方法问题

（2）是如何对不同的估计进行评价，即估计的好坏判断标准

2、无偏性

定义：$\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$是$\theta$的一个估计，$\theta$的参数空间为$\Theta$，若对任意的$\theta \in \Theta$，有$E\left(\hat{\theta}\right) = \theta$（$\theta$是真实值），则称$\hat{\theta}$是$\theta$的无偏估计，否则成为有偏估计

注：无偏性不具有不变性。即若$\hat{\theta}$是θ的无偏估计，一般而言，其函数$g\left(\hat{\theta}\right)$不是$g(\theta)$的无偏估计，除非$g(\theta)$是$\theta$的线性函数

3、有效性

当参数可估计时，其无偏估计可以有很多，如何在无偏估计中进行选择，就涉及到有效性的问题。

定义：设$\widehat{\theta_1}$，$\widehat{\theta_2}$是θ的两个无偏估计，如果对任意的$\theta \in \Theta$，有且至少有一个$\theta \in \Theta$有：$Var(\widehat{\theta_1}) \leq Var(\widehat{\theta_2})$，使得上述不等号严格成立，则称$\widehat{\theta_1}$比$\widehat{\theta_2}$有效

4、替换原理和矩法估计

替换原理是指用样本矩及其函数取替换相应的总体矩及其函数，譬如：

用样本均值估计总计均值$E(X)$，即$\hat{E}(X)=\bar{X}$

用样本方差估计总体方差$Var(X)$，即$\widehat{Var}(X)={S_n}^2$

5、概率函数已知时未知参数的矩估计

设总体具有已知的概率函数$P(x,\theta_1,\dots,\theta_k)$，$x_1,x_2,\dots x_n$是样本，假定总体的$k$阶原点矩存在，则对所有的$j,0<j<k$，都存在，若假设$\theta_1,\dots,\theta_k$，能够表示$\mu_1,\dots\mu_k$的函数$\theta_j=\theta_j(\mu_1,\dots\mu_k)$，则可给出诸$\theta_j$的矩估计：$\widehat{\theta_j}=\theta_j(a_1,\dots a_k),j=1,\dots k$，其中$a_1,\dots a_k$是前$k$阶样本原点矩$a_j=\frac{1}{n}\sum_{i=1}^{n}{x_i}^j$，进一步，如果我们要估计$\theta_1,\dots,\theta_k$的函数$\eta=g(\theta_1,\dots,\theta_k)$，则可直接得到$\mu$的矩估计，$\hat{\mu}=g(\widehat{\theta_1},\dots,\widehat{\theta_k})$

注：$EX,Var(X)$表示的是总体的特征，而$\bar{x},s$表示的是样本的特征，矩估计用的是样本的特征

例 1： 设总体服从指数分布，由于$E(X)=1/\lambda$，即$\lambda=1/E(X)$，故λ的矩法估计为

$$\hat{\lambda}=\frac{1}{\bar{x}}$$

另外，由于$Var(X)=1/\lambda^2$，其反函数为$\lambda=1/\sqrt{Var(X)}$。因此，从替换原理来看，λ的矩法估计也可取为

$$\hat{\lambda}=\frac{1}{s},s\text{为样本标准差}$$

这说明矩估计可能不是唯一的，这是矩法估计的一个缺点，此时通常应该尽量采用低阶矩给出位置参数的估计。

例 2： $x_1.x_2,\dots x_n$是来自(a,b)上的均匀分布$U(a,b)$的样本，a与b军事位置参数，这里$k=2$，由于$EX=\frac{a+b}{2},Var(X)=\frac{(b-a)^2}{12}$

不难推出，$a=EX-\sqrt{3Var(X)},b=EX+\sqrt{3Var(X)}$

由此即可得到a,b的矩估计：$\hat{a}=\bar{x}-\sqrt{3}s,\hat{b}=\bar{x}+\sqrt{3}s$

6、相合性

点估计是一个统计量，因此它是一个随机变量，在样本量一定的条件下，我们不可能要求他完全等同于参数的真实取值。但是如果我们有足够的观测值，根据格里纹科定理，随着样本量的不断增大，经验分布函数逼近真实分布函数，因此完全可以要求估计量随着样本量的不断增大而逼近参数真值，这就是相合性。严格定义如下：

设$\theta \in \Theta$为未知参数，$\widehat{\theta_n} = \widehat{\theta_n}(x_1, \dots x_n)$是$\theta$的一个估计量，n 是样本容量，若对任何一个$\varepsilon > 0$，有$\lim_{n\to\infty} P(|\widehat{\theta_n} - \theta| \geq \varepsilon) = 0$，则称$\widehat{\theta_n}$为$\theta$的相合估计。

相合性被认为是对估计的一个最基本要求，如果一个估计量在样本量不断增大时，他都不能把被估参数估计到任意指定的精度，那么这个估计是很值得怀疑的。

在判断估计的相合性时下属两个定理是很有用的：

定理 1：设$\widehat{\theta_n} = \widehat{\theta_n}(x_1, \dots x_n)$是$\theta$的一个估计量，若$\lim_{n\to\infty} E\left(\widehat{\theta_n}\right) = \theta$，$\lim_{n\to\infty} Var(\widehat{\theta_n}) = 0$，则称$\widehat{\theta_n}$是$\theta$的相合估计

定理 2：若$\widehat{\theta_{n1}}, \dots, \widehat{\theta_{nk}}$分别是$\theta_1, \dots, \theta_k$的相合估计，$\eta = g(\theta_1, \dots, \theta_k)$是$\theta_1, \dots, \theta_k$的连续函数，则$\widehat{\eta_n} = g(\widehat{\theta_{n1}}, \dots, \widehat{\theta_{nk}})$是$\eta$的相合估计

注：样本均值是总体均值的相合估计

样本标准差是总体标准差的相合估计

样本变异系数是总体变异系数的相合估计

7、最大似然估计(MLE)

引例：设有外形完全相同的两个箱子，甲箱中有 99 个白球和 1 个黑球，乙箱中有 99 个黑球和 1 个白球，今随机地抽取一箱，并从中随机抽取一球，结果取得

白球，问这球是从哪个箱子中取出？
解：不管是哪一个箱子，从箱子中人去一球都有两个可能的结果，A 表示“取出白球”，B 表示“取出黑球”。如果我们取出的是甲箱，则 A 发生的概率是 0.99，而如果取出的是乙箱，则 A 发生的概率为 0.01.现在一次试验中结果 A 发生了，人们的第一印象就是：“此白球（A）最想从甲箱取出的”，或者说，应该认为实验条件对结果 A 出现有利，从而可以推断这球是从甲箱去除的。这个推断很符合人们的经验事实，这里“最像”就是“最大似然”之意。这种想法常称为“最大似然原理”

定义：设总体的概率密度函数为$p(x;\theta),\theta\in\Theta$,其中$\theta$是一个未知参数或几个未知参数组成的参数向量，$\Theta$是参数空间，$x_1,x_2,\dots x_n$是来自总体的样本，将样本的联合概率密度函数看成$\theta$的函数，用$L(\theta,x_1,x_2,\dots x_n)$表示，简记为$L(\theta)$，

$$L(\theta)=L(\theta;x_1,x_2,\dots x_n)=p(x_1,\theta)\cdot p(x_2,\theta)\cdot\dots p(x_n,\theta)$$

称为样本的似然函数。如果某统计量$\hat{\theta}=\hat{\theta}(x_1,x_2,\dots x_n)$满足

$$L(\hat{\theta})=\max_{\theta\in\Theta}L(\theta)$$

则称$\hat{\theta}$是θ的最大似然估计，简记为 MLE

例：设一个试验有三种可能结果，其发生概率分别为$p_1=\theta^2,p_2=2\theta(1-\theta),p_3=(1-\theta)^2$，先做了 n 次试验，观测到三种结果发生的次数分别为$n_1,n_2,n_3(n_1+n_2+n_3=n)$,则似然函数为

$$L(\theta)=(\theta^2)^{n_1}[2\theta(1-\theta)]^{n_2}[(1-\theta)^2]^{n_3}=2^{n_2}\cdot\theta^{2n_1+n_2}\cdot(1-\theta)^{2n_3+n_2}$$

其对数似然函数为$Ln\,L(\theta)=(2n_1+n_2)ln\theta+(2n_3+n_2)\ln(1-\theta)+n_2ln2$,
将之关于θ求导，并令其为 0 得到似然方程

$$\frac{2n_1+n_2}{\theta}-\frac{2n_3+n_2}{1-\theta}=0$$

解之，得$\hat{\theta}=\frac{2n_1+n_2}{2(n_1+n_2+n_3)}=\frac{2n_1+n_2}{2n}$,由于$\frac{\partial^2 lnL(\theta)}{\partial\theta^2}=-\frac{2n_1+n_2}{\theta^2}-\frac{2n_3+n_2}{(1-\theta)^2}<0$

所以$\hat{\theta}$是极大值点

极大似然估计的性质

(1)不变性如果$\hat{\theta}$是θ的极大似然估计，则对任一函数$g(\theta)$，其极大似然估计为$g(\hat{\theta})$，该性质称为极大似然估计的不变性。

例：设$x_1,x_2,\dots x_n$是来正态总体$N(\mu,\sigma^2)$的样本，则μ和σ^2的极大似然

估计为$\hat{\mu} = \bar{x}, \hat{\sigma}^2 = s^2$，于是由不变性可得如下参数的极大似然估计，他们是

标准差σ的 MLE 是$\hat{\sigma} = s$

概率$P(X < 3) = \Phi\left(\frac{3-\mu}{\sigma}\right)$的 MLE 是$\Phi\left(\frac{3-\bar{x}}{\sigma}\right)$

（2）渐近正态性，定义：参数θ的相合估计$\widehat{\theta_n}$称为渐进正态的，若存在趋于 0 的非负常数序列$\sigma_n(\theta)$,使得$\frac{\widehat{\theta_n}-\theta}{\sigma_n(\theta)}$依分布收敛于标准正态分布，这时也称$\widehat{\theta_n}$服从监禁正态分布$N(\theta, {\sigma_n}^2(\theta))$，记为$\widehat{\theta_n} \sim AN(\theta, {\sigma_n}^2(\theta))$, ${\sigma_n}^2(\theta)$称为$\widehat{\theta_n}$的渐近方差。

8、均方误差

相合性和渐近正态性是在大样本场合下评价估计好坏的两个重要指标，当小样本时，无偏估计使用方差，有偏估计使用均方误差。

可以得到：均方误差由点估计的方差与偏差$|E\hat{\theta} - \theta|$的平方两部分组成。如果$\hat{\theta}$是θ的无偏估计，则$MSE(\hat{\theta}) = Var(\hat{\theta})$，此时用均方误差评价点估计与用方差是完全一样的。

定义：设有样本$x_1, x_2, \dots x_n$，对待参数θ，设有一个估计类，称$\hat{\theta}(x_1, x_2, \dots x_n)$是该估计类中的一致最小均方误差估计，如果对该估计类中另外任意一个θ的估计$\hat{\theta}$，在参数空间Θ上都有

$$MSE_\theta(\hat{\theta}) \le MSE_\theta(\tilde{\theta})$$

9、一致最小方差无偏估计

对参数估计问题，设$\hat{\theta}$是θ的一个无偏估计，如果对另外任意一个θ的无偏估计$\tilde{\theta}$，在参数空间Θ上都有

$$Var_\theta(\hat{\theta}) \le Var_\theta(\tilde{\theta})$$

则称$\hat{\theta}$是θ的一致最小方差无偏估计，简记 UMVUE

定理(Rao-Blackwell 定理)：设总体概率函数是$p(x,\theta)$, $x_1,x_2,\dots x_n$是其样本，$T=T(x_1,x_2,\dots x_n)$是θ的充分统计量，则对θ的任一无偏估计$\hat{\theta}=\hat{\theta}(x_1,x_2,\dots x_n)$,令$\tilde{\theta}=E(\hat{\theta}|T)$，则$\tilde{\theta}$也是$\theta$的无偏估计，且

$$Var_\theta(\tilde{\theta})\le Var_\theta(\hat{\theta})$$

定理表明：如果无偏估计不是充分统计量的函数，则将之对充分统计量求条件期望可以得到一个新的无偏估计，该估计的方差比原来的估计的方差要小，从而降低了无偏估计的方差。换言之，考虑θ的估计问题只需要在基于充分统计量的函数中进行即可，该说法对所有的统计推断问题都是真骨企鹅的，这便是所谓的充分性原则。

10.贝叶斯估计

贝叶斯学派的基本观点：任一未知量θ都可看作随机变量，可用一个概率分布去描述，这个分布成为**先验分布**；在获得样本之后，总体分布，样本与先验分布通过贝叶斯公式结合起来得到一个关于未知量θ新的分布——**后验分布**；任何关于θ的统计推断都应该基于θ的后验分布进行。

贝叶斯公式的密度函数形式

（1）总体依赖于参数θ的概率函数在贝叶斯统计中记为$P(x|\theta)$，他表示在随机变量θ取某个给定值时总体的条件概率函数；

（2）根据参数θ的先验信息可确定先验分布$\pi(\theta)$

（3）从贝叶斯观点看，样本$x_1,x_2,\dots x_n$的产生分两步进行，首先从先验分布$\pi(\theta)$产生一个样本θ_0，然后从$P(x|\theta_0)$中产生一组样本。

这时样本的联合条件概率函数为$p(x_1,\dots x_n|\theta_0)=\prod_{i=1}^{m}p(x_i|\theta_0)$

这个分布综合了总体信息和样本信息

（4）θ_0是未知的，他是按先验分布$\pi(\theta)$产生的。为把先验信息综合进去,不能只考虑θ_0,对θ的其他值发生的可能性也要加以考虑，故要用$\pi(\theta)$进行综合。这样一来，样本$x_1,x_2,\dots x_n$和参数θ的联合分布为$h(x_1,x_2,\dots x_n,\theta)=p(x_1,x_2,\dots x_n|\theta)\pi(\theta)$，这个联合分布把总体信息、样本信息和先验信息三种可用信息都综合进去了；

（5）我们的目的是要对未知参数作θ统计推断，在没有样本信息时，我们只能依据先验分布对θ做出推断，在有了样本观测值$x_1,x_2,\dots x_n$之后，则应依据$h(x_1,x_2,\dots x_n,\theta)$对$\theta$作出推断，若把$h(x_1,x_2,\dots x_n,\theta)$作如下分解：

$$h(x_1,x_2,\dots x_n,\theta)=\pi(\theta|x_1,x_2,\dots x_n)m(x_1,x_2,\dots x_n)$$

其中$m(x_1,x_2,\dots x_n)$是$x_1,x_2,\dots x_n$的边际概率函数

$$m(x_1,x_2,\dots x_n)=\int_\theta h(x_1,x_2,\dots x_n,\theta)\,d\theta=\int_\theta p(x_1,x_2,\dots x_n|\theta)\,\pi(\theta)d\theta$$

它与θ无关,不含θ的任何信息,因此能用来对θ作出推断的仅是条件分布$\pi(\theta|x_1,x_2,\dots x_n)$，它的计算公式是

$$\pi(\theta|x_1,x_2,\dots x_n)=\frac{h(x_1,x_2,\dots x_n,\theta)}{m(x_1,x_2,\dots x_n)}=\frac{p(x_1,x_2,\dots x_n|\theta)\pi(\theta)}{\int_\theta p(x_1,x_2,\dots x_n|\theta)\,\pi(\theta)d\theta}$$

这个条件分布称为θ的后验分布,它集中了总体、样本和先验中有关θ的一切信息

11、区间估计

定义 1：设θ是总体的一个参数，其参数空间为Θ，$x_1,x_2,\dots x_n$是来自

该总体的样本，对给定一个$\alpha(0<\alpha<1)$,若有两个统计量$\widehat{\theta_L}=\widehat{\theta_L}(x_1,x_2,\dots x_n)$和$\widehat{\theta_U}=\widehat{\theta_U}(x_1,x_2,\dots x_n)$，若对任意的$\theta\in\Theta$，有$P_\theta\left(\widehat{\theta_L}\le\theta\le\widehat{\theta_U}\right)\ge1-\alpha$,则称随机区间$[\widehat{\theta_L},\widehat{\theta_U}]$为$\theta$的置信水平为$1-\alpha$的置信区间，或简称$[\widehat{\theta_L},\widehat{\theta_U}]$是$\theta$的$1-\alpha$的置信区间，$\widehat{\theta_L}$置信下限，$\widehat{\theta_U}$置信上限

定义 2: 沿用定义 1 的记号, 如对给定的$\alpha(0<\alpha<1)$, 对任意$\theta\in\Theta$, 有$P_\theta\left(\widehat{\theta_L}\le\theta\le\widehat{\theta_U}\right)=1-\alpha$，称$[\widehat{\theta_L},\widehat{\theta_U}]$为$\theta$的置信水平为$1-\alpha$同等置信区间。

定义 3：设$\widehat{\theta_L}=\widehat{\theta_L}(x_1,x_2,\dots x_n)$是统计量，对给定的$\alpha\in(0,1)$和任意的$\theta\in\Theta$，有$P_\theta\left(\widehat{\theta_L}\le\theta\right)\ge1-\alpha,\forall\theta\in\Theta$，则称$\widehat{\theta_L}$为$\theta$的置信水平为$1-\alpha$的(单侧)置信下限。假如等号对一切的$\theta\in\Theta$成立，则$\widehat{\theta_L}$为$\theta$的置信水平为$1-\alpha$的同等置信下限。

定义 4：设$\widehat{\theta_U}=\widehat{\theta_U}(x_1,x_2,\dots x_n)$是统计量，对给定的$\alpha\in(0,1)$和任意的$\theta\in\Theta$, 有$P_\theta\left(\widehat{\theta_U}\ge\theta\right)\ge1-\alpha,\forall\theta\in\Theta$, 则称$\widehat{\theta_U}$为$\theta$的置信水平为$1-\alpha$的(单侧)置信下限。假如等号对一切的$\theta\in\Theta$成立，则$\widehat{\theta_U}$为$\theta$的置信水平为$1-\alpha$的同等置信下限。

12、枢轴量法

构造未知参数θ的置信区间的最常用的方法是枢轴量法，其步骤可以概括为如下三个步骤:

（1）设法构造一个样本和θ的函数$G=G(x_1,x_2,\dots x_n,\theta)$使得$G$的分布不依赖于未知参数，一般称具有这种性质的$G$为枢轴量

（2）适当地选择两个常数c,d，使对给定的$\alpha(0<\alpha<1)$有

$$P(c < G < d) = 1 - \alpha$$

（3）假如能将$c \le G \le d$进行不等式等价变形化为$\widehat{\theta_L} \le \theta \le \widehat{\theta_U}$，则$[\widehat{\theta_L}, \widehat{\theta_U}]$为$\theta$的置信水平为$1 - \alpha$同等置信区间。