第六章

参数估计与假设检验

统计推断基本问题

参数估计 Parameter estimation

点估计 Point estimation

区间估计 Interval estimation

假设检验 Hypothesis Testing

什么是参数估计?

前提：根据经验可以事先确定总体的分布形式，但对某些参数不清楚。

参数是刻画总体某方面概率特性的数量.

当此数量未知时, 从总体抽出一个***样本***，用某种方法对这个未知参数进行估计就是**参数估计**.

例如，$X \sim N(\mu, \sigma^2)$,

若μ, σ^2未知, 通过构造样本的函数, 给出它的**估计值**或**取值范围**就是参数估计的内容.

点估计 **区间估计**

参数估计的类型

重要

点估计point estimation — 估计未知参数的值

区间估计—

估计未知参数的取值范围，使此范围包含未知参数真值的概率为给定的值.

§ 6.1 参数的点估计

1. 点估计问题

根据样本构造一个统计量$\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$,
用它估计未知参数θ，称为点估计.

称$\hat{\theta}(X_1,\cdots,X_n)$为$\theta$的<u>估计量</u>;

称$\hat{\theta}(x_1,\cdots,x_n)$为$\theta$的<u>估计值</u>.

两种常用的方法:
1. 矩估计 Moment Estimation
2. 最大似然估计Maximum Likelihood Estimation

重要

2. 矩估计法

用相应的样本矩去估计总体矩的估计方法称为矩估计法.

方法

用样本 k 阶矩作为总体 k 阶矩的估计量, 建立含有待估参数的方程, 从而解出待估参数。

重要

设待估计的参数为 $\theta_1,\theta_2,\cdots,\theta_m$

设总体的 k 阶矩存在，记为

$$E(X^k)=\mu_k(\theta_1,\theta_2,\cdots,\theta_m)$$

样本$X_1, X_2,\ldots, X_n$的k阶矩为 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$

样本矩去估计总体矩

令 $\mu_k(\theta_1,\theta_2,\cdots,\theta_m)=A_k \quad k=1,2,\cdots,m$

——含未知参数 $\theta_1,\theta_2,\ldots,\theta_m$ 的方程组

解方程组，得 m 个统计量：

$$\left.\begin{array}{l}\hat{\theta}_1(X_1, X_2, \cdots, X_n) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{\theta}_m(X_1, X_2, \cdots, X_n)\end{array}\right\}$$

未知参数 $\theta_1, \ldots, \theta_m$ 的矩估计量

代入一组样本值得 m 个数：

$$\left.\begin{array}{l}\hat{\theta}_1 = \hat{\theta}_1(x_1, x_2, \cdots, x_n) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{\theta}_m = \hat{\theta}_m(x_1, x_2, \cdots, x_n)\end{array}\right\}$$

未知参数 $\theta_1, \ldots, \theta_m$ 的矩估计值

例6.1.2　一般，不论总体服从什么分布，总体期望 μ 与方差 σ^2 存在，则它们的矩估计量分别为

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

?

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

事实上，按矩法原理，令

$$D(X)=EX^2-E^2X$$

$$\begin{cases}\bar{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i=A_1=EX=\mu\\ \dfrac{1}{n}\sum_{i=1}^{n}X_i^2=A_2=E(X^2)=\mu^2+\sigma^2\end{cases}$$

$$\Rightarrow\begin{cases}\hat{\mu}=\bar{X}\\ \hat{\sigma}^2=\hat{E}(X^2)-\hat{E}^2(X)=A_2-\hat{\mu}^2=\dfrac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2\end{cases}$$

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

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

例 设总体 $X \sim E(\lambda)$, $X_1, X_2, \ldots, X_n$为总体的样本, 求λ的矩法估计量.

解 $E(X) = 1/\lambda$, 令 $\overline{X} = 1/\lambda$.

故 $\hat{\lambda}_{矩} = 1/\overline{X}$.

例6. 1. 3 设总体 X 的概率密度为

$$f(x)=\begin{cases}(\theta+1)x^{\theta}, & 0<x<1\\ 0, & \text{其他}\end{cases}$$

其中 $\theta>-1$ 是未知参数. $X_1, X_2, \cdots, X_n$ 是总体 X 的一个简单样本，求 θ 的矩估计.

$$E(X)=\int_{-\infty}^{+\infty} xf(x)dx=\int_0^1 x(\theta+1)x^{\theta}dx=\frac{\theta+1}{\theta+2},$$

令 $\dfrac{\theta+1}{\theta+2}=\bar{X}$

总体矩

样本矩

得 $\hat{\theta}=\dfrac{2\bar{X}-1}{1-\bar{X}}$

例6. 1. 4 设样本 $X_1, X_2,\ldots, X_n$为来自服从几何分布的总体X, 其概率分布为

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

其中p未知，$0<p<1$，试求p的矩估计量.

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

令 $$\frac{1}{\hat{p}}=\bar{X}$$

故 $$\hat{p}=\frac{1}{\bar{X}}$$

矩法的优点是简单易行,并不需要事先知道总体是什么分布．

缺点是，当总体类型已知时，没有充分利用分布提供的信息．一般场合下,矩估计量不具有唯一性.

其主要原因在于建立矩法方程时，选取那些总体矩用相应样本矩代替带有一定的随意性．

3. 极大似然估计法

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.

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 .

重要

方法思想：一次试验就出现的事件有较大的概率

黑球白球9:1，不知哪种多，有放回抽三次，两次白球，一次黑球.

哪种多？　　　　白球多！

$$\theta = 0.9? \quad \theta = 0.1?$$

最大

$$L(\theta) = P(X_1 = 1, X_2 = 0, X_3 = 1)$$

这种选择一个参数使得实验结果具有最大概率的思想就是极大似然法的基本思想．

重要

求极大似然估计的一般步骤

(1) 构造似然函数$L(\theta)$

若总体X属离散型，其分布律

$$P(X=x)=p(x;\theta),\quad \theta\in\Theta$$

设$X_1,\cdots,X_n$是来自X的样本,

又设$x_1,\cdots,x_n$是$X_1,\cdots,X_n$的一个样本值;

似然函数

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

重要

若总体X属连续型，其概率密度$f(x;\theta),\theta\in\Theta$

$x_1,\cdots,x_n$是$X_1,\cdots,X_n$的一个样本值

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

似然函数

重要

(2) 求似然函数 $L(\theta)$的最大值点

挑选使$L(\theta)$达到最大的参数$\hat{\theta}$，作为θ的估计，即

$$L(x_1,\cdots,x_n;\hat{\theta}) = \max_{\theta\in\Theta} L(x_1,\cdots,x_n;\theta)$$

$\hat{\theta}(x_1,\cdots,x_n)$称为参数$\theta$的<u>极大似然估计值</u>.

$\hat{\theta}(X_1,\cdots,X_n)$称为参数$\theta$的<u>极大似然估计量</u>.

重要

一般，θ 可由下式求得：

$$\frac{dL(\theta)}{d\theta} = 0.$$

又因$L(\theta)$与$\ln L(\theta)$在同一θ处取到极值，因此θ的极大似然估计θ也可从下述方程解得：

$$\frac{d}{d\theta}\ln L(\theta) = 0.$$

对数似然函数

似然方程

重要

注1

未知参数可以不止一个, 如$\theta_1,\dots,\theta_k$

设X 的密度(或分布)为$f(x,\theta_1,\dots,\theta_k)$

则定义似然函数为

$$L(x_1,\cdots,x_n;\theta_1,\cdots,\theta_k)$$

$$=L(\theta_1,\cdots,\theta_k)=\prod_{i=1}^{n} f(x_i,\theta_1,\cdots,\theta_k)$$

可令 $\dfrac{\partial L}{\partial \theta_i}=0, i=1,\cdots,k.$

或 $\dfrac{\partial \ln L}{\partial \theta_i}=0, i=1,\cdots,k.$

似然方程组

解方程组求得 $\theta_1,\cdots,\theta_k$ 的极大似然估计.

例6. 1. 6 在6.1.3中：设总体 X 的概率密度为

$$f(x)=\begin{cases}(\theta+1)x^{\theta}, & 0<x<1\\ 0, & \text{其他}\end{cases}$$

其中 $\theta>-1$ 是未知参数. $X_1, X_2, \cdots, X_n$ 是总体 X 的一个简单样本，求 θ 的极大似然估计.

解：似然函数为

$$L(\theta)=\prod_{i=1}^{n}[(\theta+1)x_i^{\theta}]=(\theta+1)^n\left(\prod_{i=1}^{n}x_i\right)^{\theta}$$

$$(0<x_i<1)$$

对数似然函数为

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

求导并令其为0

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

解得 $\hat{\theta} = -\dfrac{n}{\sum_{i=1}^{n}\ln x_i} - 1$ θ的极大似然估计

例6.1.7 设总体 $X \sim N(\mu,\sigma^2)$, $x_1, x_2,\dots, x_n$ 是 X的样本值, 求 μ, σ^2 的极大似然估计.

解 $L(x_1, x_2,\dots, x_n;\mu,\sigma^2)$

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

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

似然方程组为

$$\begin{cases} \left(\dfrac{\partial}{\partial \mu}\ln L\right) = \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0 \\ \left(\dfrac{\partial}{\partial(\sigma^2)}\ln L\right) = \dfrac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2-\dfrac{n}{2(\sigma^2)}=0 \end{cases}$$

$$\Longrightarrow \begin{cases} \hat{\mu}_{mle} = \dfrac{1}{n}\sum_{i=1}^{n}x_i = \bar{x} \\ \hat{\sigma^2}_{mle} = \dfrac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \end{cases}$$

μ, σ^2的极大似然估计量分别为

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

重要

注2

用上述方法求参数的极大似然估计值有时行不通，这时要用极大似然原则来求.

$L(\theta)$

无驻点

不可导

例6.1.8 设某种元件使用寿命 X 的概率密度为 重要

$$f(x)=\begin{cases}2e^{-2(x-\theta)}, & x\ge\theta\\ 0, & x<\theta\end{cases}$$

其中 $\theta>0$ 是未知参数. 设 $x_1,,\cdots,x_n$ 是样本观测值，求 θ 的极大似然估计.

解：似然函数为

$$L(\theta)=\begin{cases}\prod_{i=1}^{n}[2e^{-2(x_i-\theta)}], & x_i\ge\theta\\ 0, & \text{其它}\end{cases}$$

$$=\begin{cases}2^n e^{-2\sum(x_i-\theta)}, & x_i\ge\theta\\ 0, & \text{其它}\end{cases}$$

$$当 x_i \geq \theta 时, L(\theta) > 0$$

$$\therefore \ln L(\theta) = n \ln 2 - 2\sum_{i=1}^{n}(x_i - \theta)$$

$$\because \frac{d \ln L(\theta)}{d\theta} = 2n > 0$$

$\therefore L(\theta)$单调增加，无极值点

f(x)= $\begin{cases} 2e^{-2(x-\theta)}, & x \geq \theta \\ 0, & x < \theta \end{cases}$

∴回头分析似然函数定义

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

$\because x_i < \theta$时$f(x_i) = 0$

不是最小的

$\because x_i \geq \theta$时$\hat{\theta} \neq \min\{x_1, x_2, \ldots, x_n\}$时，$\exists f(x_i) = 0$ $\therefore L(\hat{\theta}) = 0$

是最小的

$\therefore$ iff $\hat{\theta} = \min\{x_1, x_2, \ldots, x_n\}$时，有$\forall f(x_i) \neq 0$

$\therefore L(\hat{\theta}) > 0$取得极值

$\therefore$ θ的极大似然估计

$$\hat{\theta} = \min\{x_1, x_2, \cdots, x_n\}$$

例6.1.9 设$X_1, X_2, \ldots, X_n$是来自参数为λ的泊松分布总体的一个样本，求 λ的极大似然估计.

解

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

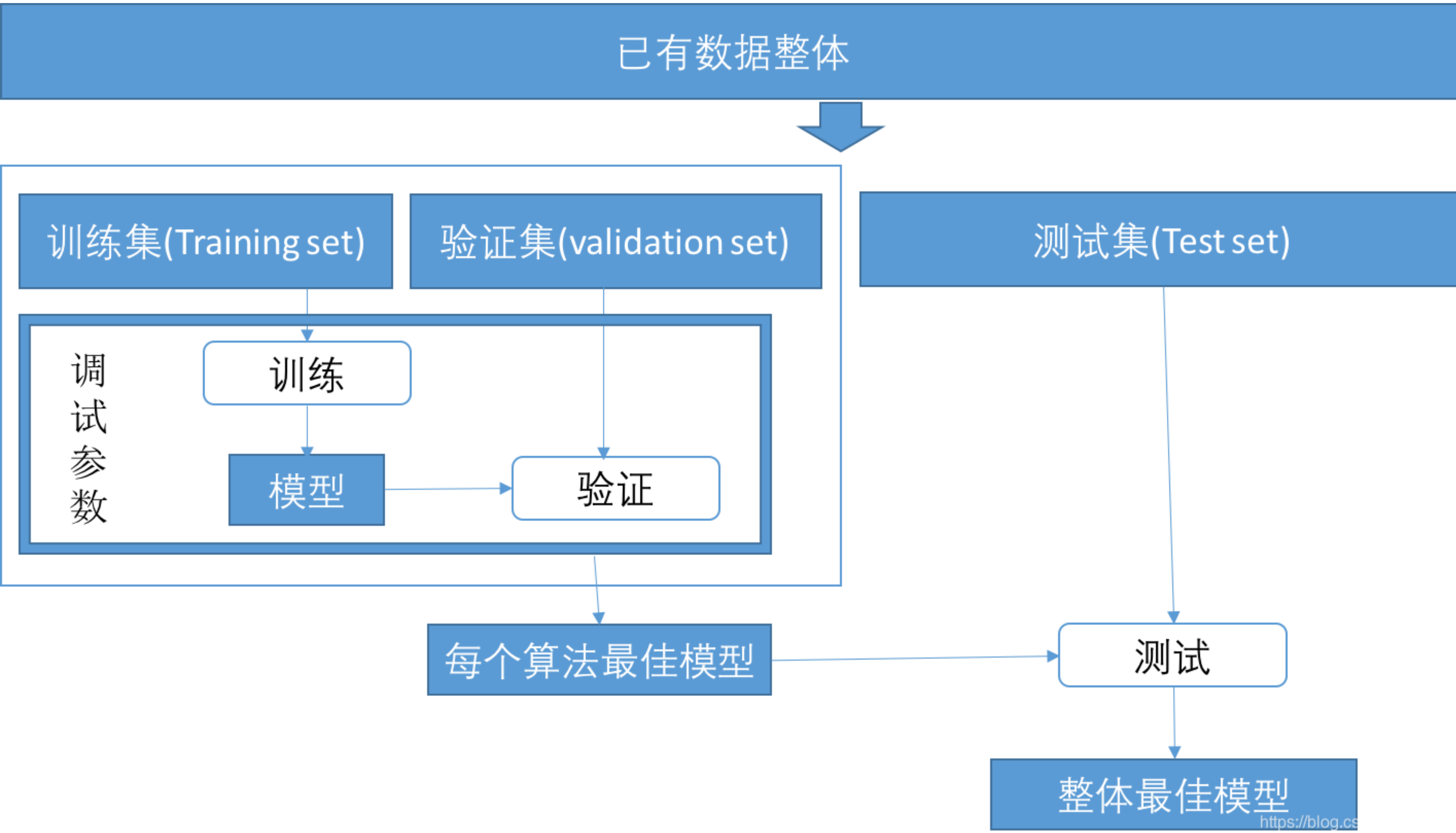
$$L=\prod_{i=1}^{n}\left(\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}\right)=\frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}$$

$$\ln L=-n\lambda+\sum_{i=1}^{n}x_i\ln\lambda-\ln\prod_{i=1}^{n}x_i!$$

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

$\therefore \hat{\lambda}=\dfrac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$ 最大似然估计量为$\hat{\lambda}=\bar{X}$

统计机器学习一般过程



已有数据整体
训练集(Training set)
验证集(validation set)
测试集(Test set)
调试参数
训练
模型
验证
每个算法最佳模型
测试
整体最佳模型
https://blog.cs

高斯判别模型

$x \in \mathbb{R}^n$ 为连续值随机变量，标签 $y \in \{0,1\}$ ，训练集的大小为 m 。利用先验的强假设，高斯判别分析模型对 $p(x \mid y)$ 和 $p(y)$ 进行建模：

$$y \sim \text{ Bernoulli } (\phi)$$
$$x|y=0 \sim \mathcal{N}\left(\mu_0, \Sigma\right)$$
$$x|y=1 \sim \mathcal{N}\left(\mu_1, \Sigma\right)$$

即假设标签 y 服从伯努利分布， $p(x \mid y)$ 服从多元正态分布.

高斯判别模型

$$p(y) = \phi^y (1-\phi)^{1-y}$$

$$p(x|y=0) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_0)^T \Sigma^{-1} (x-\mu_0)\right)$$

$$p(x|y=1) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_1)^T \Sigma^{-1} (x-\mu_1)\right)$$

模型的参数为 $\phi, \Sigma, \mu_0, \mu_1$ 。注意，此处 $p(x|y=0)$ 与 $p(x|y=1)$ 服从的多元高斯分布拥有相同的协方差矩阵 Σ 。

高斯判别模型

$$p(y) = \phi^y (1-\phi)^{1-y}$$

$$p(x|y=0) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_0)^T \Sigma^{-1} (x-\mu_0)\right)$$

$$p(x|y=1) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_1)^T \Sigma^{-1} (x-\mu_1)\right)$$

模型的参数为 $\phi, \Sigma, \mu_0, \mu_1$ 。注意，此处 $p(x|y=0)$ 与 $p(x|y=1)$ 服从的多元高斯分布拥有相同的协方差矩阵 Σ 。

参数学习

为了使模型的分类效果达到最佳，需要对模型的参数 $\phi, \Sigma, \mu_0, \mu_1$ 进行优化。此处通过极大似然估计来选取参数 $\phi, \Sigma, \mu_0, \mu_1$ 的最优值。先列出参数的对数似然函数 $\ell(\phi, \mu_0, \mu_1, \Sigma)$ ：

$$\ell(\phi, \mu_0, \mu_1, \Sigma) = \log \prod_{i=1}^{m} p\left(x^{(i)}, y^{(i)}; \phi, \mu_0, \mu_1, \Sigma\right)$$

参数学习

利用 $p(x,y)=p(x\mid y)p(y)$ 将上述对数似然展开：

$$\ell\left(\phi,\mu_0,\mu_1,\Sigma\right)=\log\prod_{i=1}^{m}p\left(x^{(i)}|y^{(i)};\mu_0,\mu_1,\Sigma\right)p\left(y^{(i)};\phi\right)$$

继续将等式右边展开：

$$\ell\left(\phi,\mu_0,\mu_1,\Sigma\right)=\sum_{k=1}^{m}\log p\left(x^{(i)}|y^{(i)};\mu_0,\mu_1,\Sigma\right)+\sum_{i=1}^{m}\log p\left(y^{(i)};\phi\right)$$

$$=\sum_{i=1}^{m}\log\left(p\left(x^{(i)}|y^{(i)}=1;\mu_1,\Sigma\right)^{y^{(i)}}\cdot p\left(x^{(i)}|y^{(i)}=0;\mu_0,\Sigma\right)^{1-y^{(i)}}\right)+\sum_{i=1}^{m}\log p\left(y^{(i)};\phi\right)$$

$$=\sum_{i=1}^{m}y^{(i)}\log p\left(x^{(i)}|y^{(i)}=1;\mu_1,\Sigma\right)+\sum_{i=1}^{m}\left(1-y^{(i)}\right)\log p\left(x^{(i)}|y^{(i)}=0;\mu_0,\Sigma\right)+\sum_{i=1}^{m}\log p\left(y^{(i)};\phi\right)$$

求解参数 ϕ,Σ,μ_0,μ_1 使此处的对数似然函数 $\ell\left(\phi,\mu_0,\mu_1,\Sigma\right)$ 最大化的过程被称为参数的极大似然估计。先求对数似然函数 $\ell\left(\phi,\mu_0,\mu_1,\Sigma\right)$ 对各个参数的偏导数，通过令各偏导数为0取得使对数似然函数 $\ell\left(\phi,\mu_0,\mu_1,\Sigma\right)$ 最大的参数最优值。

重要

4. 点估计的评价标准

对于同一个未知参数,不同的方法得到的估计量可能不同,于是提出问题:

- 应该选用哪一种估计量?
- 用何标准来评价一个估计量的好坏?

常用标准

(1) 无偏性 unbiased

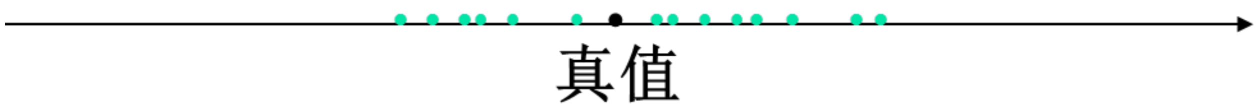
(2) 有效性quality

(3) 一致性(相合性) consistent

重要

(1) 无偏性unbiased estimation

估计量是随机变量，对于不同的样本值会得到不同的估计值 . 我们希望估计值在未知参数真值附近摆动，如果估计值的期望值等于未知参数的真值. 这就导致无偏性这个标准 .



设$\hat{\theta}(X_1,\cdots,X_n)$是未知参数 θ的估计量，若

$$E(\hat{\theta})=\theta$$

则称 $\hat{\theta}$ 为θ的<u>无偏估计(**unbiased estimation**)</u> .

例 设总体 X 的 k 阶矩 $\mu_k=E(X^k)$ 存在，

$X_1, X_2,\ldots,X_n$ 是总体 X 的样本，

证明：不论 X 服从什么分布(但期望存在)，

则 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$ 是 μ_k 的无偏估计量.

证 由于 $E(X_i^k)=\mu_k \quad i=1,2,\cdots,n$

因而 $E(A_k)=E(\frac{1}{n}\sum_{i=1}^{n}X_i^k)=\frac{1}{n}\sum_{i=1}^{n}E(X_i^k)$

$$=\frac{1}{n}\cdot n\cdot\mu_k=\mu_k$$

重要

特别地

样本均值 $\bar{X}$ 是总体期望 $E(X)$ 的无偏估计量

样本二阶原点矩 $A_2=\frac{1}{n}\sum_{i=1}^{n}X_i^2$ 是总体二阶原点矩 $\mu_2=E(X^2)$ 的无偏估计量

例 设总体 X 的期望与方差存在，X 的样本为 $(X_1, X_2, \cdots, X_n)$，

(1) $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 不是 $D(X)$ 的无偏估计;

样本方差

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

(2) $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 是 $D(X)$ 的无偏估计.

样本修正方差

$$E\left(S^2\right) = \sigma^2$$

证：$E(S_n^2)=\dfrac{n-1}{n}\sigma^2$

$$E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right)=E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i^2-2X_i\overline{X}+\overline{X}^2)\right)$$

$$=\frac{1}{n}E\left(\sum_{i=1}^{n}X_i^2-2\overline{X}\sum_{i=1}^{n}X_i+\sum_{i=1}^{n}\overline{X}^2\right)=\frac{1}{n}E\left(\sum_{i=1}^{n}X_i^2-2n\overline{X}^2+n\overline{X}^2\right)$$

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

$$=\frac{1}{n}\sum_{i=1}^{n}E(X_i^2)-\left[D\left(\overline{X}\right)+E^2\left(\overline{X}\right)\right]=\frac{1}{n}\sum_{i=1}^{n}(D(X_i)+E(X_i)^2)-\left[D\left(\overline{X}\right)+E^2\left(\overline{X}\right)\right]$$

设$E(X)=\mu, D(X)=\sigma^2$ 则 $E\left(\overline{X}\right)=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\mu$ $D\left(\overline{X}\right)=\frac{1}{n}\sigma^2$

$$=\sigma^2+\mu^2-\left(\frac{1}{n}\sigma^2+\mu^2\right)=\frac{n-1}{n}\sigma^2$$

例6.1.10 设随机变量$X_1, X_2, \ldots, X_n$是来自正态分布$N(\mu,\sigma^2)$的一个样本，适当选择常数c，使 $c\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2$ 为σ^2的无偏估计

解 令$Y_i = X_{i+1} - X_i$，则$Y_i \sim N(0, 2\sigma^2)$,

正态分布性质

由 $E(c\sum_{i=1}^{n-1}Y_i^2) = c\sum_{i=1}^{n-1}E(Y_i^2)$

$$= c(n-1)\{D(Y_i)+E^2(Y_i)\}$$

$$= c(n-1)2\sigma^2$$

$\because$ 无偏估计，$\therefore c(n-1)2\sigma^2 = \sigma^2$

$$\therefore c = \frac{1}{2(n-1)}$$

一个参数往往有不止一个无偏估计，若$\hat{\theta}_1$和$\hat{\theta}_2$都是参数θ的无偏估计量，我们可以比较$E(\hat{\theta}_1-\theta)^2$ 和 $E(\hat{\theta}_2-\theta)^2$的大小来决定二者谁更优.

D(X)=E[X-E(X)]^2

由于 $$D(\hat{\theta}_1)=E(\hat{\theta}_1-\theta)^2$$

$$D(\hat{\theta}_2)=E(\hat{\theta}_2-\theta)^2$$

所以无偏估计以方差小者为好，这就引进了有效性这一概念.

重要

(2) 有效性quality

设$\hat{\theta}_1=\hat{\theta}_1(X_1,\cdots,X_n)$和$\hat{\theta}_2=\hat{\theta}_2(X_1,\cdots,X_n)$ 都是参数θ的无偏估计量，若有

$$D(\hat{\theta}_1)<D(\hat{\theta}_2)$$

则称$\hat{\theta}_1$较$\hat{\theta}_2$有效.

重要

在数理统计中常用到最小方差无偏估计.

设$X_1,\ldots,X_n$是取自总体X的一个样本，

$\hat{\theta}(X_1,\cdots,X_n)$ 是未知参数 θ的无偏估计量，

$$\boxed{D(\hat{\theta}) \le D(\hat{\theta}^*)}$$

$\hat{\theta}^*$是θ的任一无偏估计.

则称 $\hat{\theta}$ 为θ的最小方差无偏估计.

（也称最佳无偏估计）

例如 $X \sim N(\mu, \sigma^2)$，样本是X_1, X_2, X_3

$$E(\hat{\mu}_1) = E(\hat{\mu}_2) = \mu$$

$$\left.\begin{aligned} \hat{\mu}_1 &= \frac{1}{3}(X_1 + X_2 + X_3) \\ \hat{\mu}_2 &= \frac{1}{3}X_1 + \frac{1}{4}X_2 + \frac{5}{12}X_3 \end{aligned}\right\}$$ 都是μ的无偏估计量

哪一个更有效?

$$D(\hat{\mu}_1) = \frac{1}{3}\sigma^2 < D(\hat{\mu}_2) = \frac{25}{72}\sigma^2$$

$\hat{\mu}_1$ 最有效

推广

结论(P163.8) 设 $E(X)=\mu$，$D(X)=\sigma^2$

$(X_1, X_2, \ldots, X_n)$为总体 X 的一个样本

(1) 设常数 c_i，$i=1,2,\cdots,n.$ $\sum_{i=1}^{n} c_i = 1.$

证明 $\hat{\mu} = \sum_{i=1}^{n} c_i X_i$ 是 μ 的无偏估计量

(2) 证明 $\hat{\mu} = \sum_{i=1}^{n} c_i X_i$ 中 $\bar{X}$ 最有效

证 (1) $E(\hat{\mu}) = \sum_{i=1}^{n} c_i E(X_i) = \sum_{i=1}^{n} c_i \mu = \mu$

(2) $$D(\hat{\mu})=D(\sum_{i=1}^{n}c_iX_i)=\sum_{i=1}^{n}c_i^2D(X_i)=\sigma^2\sum_{i=1}^{n}c_i^2$$

完全平方公式

$$1=\left(\sum_{i=1}^{n}c_i\right)^2=\sum_{i=1}^{n}c_i^2+\sum_{1\le i<j\le n}2c_ic_j$$

$$\le\sum_{i=1}^{n}c_i^2+\sum_{1\le i<j\le n}(c_i^2+c_j^2)=n\sum_{i=1}^{n}c_i^2$$

$$\Rightarrow\ \sum_{i=1}^{n}c_i^2\ge\frac{1}{n}\ \Rightarrow\ D(\bar{X})=\frac{1}{n}\sigma^2\le D(\hat{\mu})$$

结论 算术均值$\bar{X}$比加权均值$\sum_{i=1}^{n}c_iX_i$更有效.

例6.1.11 设Y_1和Y_2是θ的两个相互独立的无偏估计，若$D(Y_1)$=2 $D(Y_2)$。求常数a，b，使得aY_1+bY_2为此种线性组合中有最小方差的无偏估计

解

$$\because E(Y_1)=\theta, E(Y_2)=\theta,$$

$$\therefore E(aY_1+bY_2)=(a+b)\theta$$

$\because$ 无偏估计 $\therefore (a+b)\theta=\theta$

因此 $a+b=1$

$$D(aY_1+bY_2)=a^2D(Y_1)+b^2D(Y_2)$$

$$=2a^2D(Y_2)+b^2D(Y_2)$$

$$=[2a^2+(1-a)^2]D(Y_2)$$

$$=(3a^2-2a+1)D(Y_2)$$

令 $\dfrac{dD}{da}=6a-2=0$

则 $a=\dfrac{1}{3}, b=\dfrac{2}{3}$

$\because \dfrac{d^2D}{da^2}=6>0 \therefore D$在$a=\dfrac{1}{3}, b=\dfrac{2}{3}$时取最小值

$\therefore \dfrac{1}{3}Y_1+\dfrac{2}{3}Y_2$为$\theta$的最小方差无偏估计

(3) 一致性(相合性) Consistent

定义 设$\hat{\theta}=\hat{\theta}(X_1, X_2, \cdots, X_n)$是总体参数$\theta$的估计量. 若$n\to\infty$时, $\hat{\theta}$ 依概率收敛于θ,

即$\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P(\left|\hat{\theta}-\theta\right| \geq \varepsilon) = 0$$

则称 $\hat{\theta}$ 是参数θ的一致(或相合)估计量.

一致性估计量仅在样本容量**n**足够大时, 才显示其优越性.

关于一致性的常用结论

$\left.\begin{array}{l}\text{样本 } k \text{ 阶矩是总体 } k \\ \text{阶矩的一致性估计量}\end{array}\right\}$ 由大数定律证明

矩法得到的估计量一般为一致估计量

在一定条件下, 极大似然估计具有一致性