第七章：参数估计

什么是参数

参数通常是刻画总体某些概率特征的量。

分为下面几类：

1. 总体分布中所含的参数 比如指数分布总体 $Exp(\lambda)$ 中的 λ.
2. 总体参数的函数 比如服从正态分布 $N(\mu, \sigma^2)$ 的总体，其取值不超过某个给定常数 a 的概率，即 $P(X \leq a) = \Phi(\frac{a-\mu}{\sigma})$.
3. 分布的数字特征 比如均值 $E[X]$，方差 $[X]$，k 阶原点矩等.

什么是参数估计

当该参数未知时，从总体中抽取一个样本，用某种方法对该未知参数进行估计，这就是参数估计。

例如，假设总体 $X \sim N(\mu, \sigma^2)$，若参数 μ 与 σ^2 未知。先从该总体中抽样得到样本 $X_1, X_2, \ldots, X_n$，然后构造统计量，求出未知参数 μ 与 σ^2 的估计值或取值范围，这就是参数估计。

参数空间

假设总体 $X \sim F(x;\theta_1,\theta_2,\ldots,\theta_m)$，其中分布函数 F 的表达式已知，但参数 $\theta_1,\theta_2,\ldots,\theta_m$ 未知。
若记 $\theta=(\theta_1,\theta_2,\ldots,\theta_m)$，则总体分布可记为：

$$X \sim F(x;\theta)$$

参数 θ 的取值范围称为参数空间，记为 Θ。

参数空间

假设总体 $X \sim F(x;\theta_1,\theta_2,\ldots,\theta_m)$，其中分布函数 F 的表达式已知，但参数 $\theta_1,\theta_2,\ldots,\theta_m$ 未知。
若记 $\theta=(\theta_1,\theta_2,\ldots,\theta_m)$，则总体分布可记为：

$$X \sim F(x;\theta)$$

参数 θ 的取值范围称为参数空间，记为 Θ。

例子：

1. $X \sim B(1,p)$，p 为未知参数
2. $X \sim N(\mu,\sigma^2)$，μ,σ^2 为未知参数

求以上未知参数的参数空间

参数估计的形式

参数估计的形式有下面两种：

1. 点估计
   - 参数 θ 的点估计是寻找合适的统计量 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$，将样本观测值代入后得到的值作为该参数的估计.
2. 区间估计
   - 参数的区间估计是找两个统计量，分别作为一个可能包含该参数的一个随机区间的左右端点.

7.1 参数的点估计

点估计的思想

$X_1, X_2, \ldots, X_n$ 是来自总体 $X \sim F(x; \theta_1, \ldots, \theta_m)$ 的一个样本，$\theta_1, \ldots, \theta_m$ 是未知参数。

构造 m 个统计量：

$$\begin{bmatrix} \hat{\theta}_1(X_1, X_2, \cdots, X_n) \\ \hat{\theta}_2(X_1, X_2, \cdots, X_n) \\ \vdots \\ \hat{\theta}_m(X_1, X_2, \cdots, X_n) \end{bmatrix}$$

点估计的数值

当把样本观测值 $x_1, x_2, \ldots, x_n$ 代入上述统计量，就得到 m 个数值：

$$\begin{bmatrix} \hat{\theta}_1(x_1, x_2, \cdots, x_n) \\ \hat{\theta}_2(x_1, x_2, \cdots, x_n) \\ \vdots \\ \hat{\theta}_m(x_1, x_2, \cdots, x_n) \end{bmatrix}$$

称 $\hat{\theta}_k(X_1, X_2, \cdots, X_n)$ 为 θ_k 的估计量 $(k = 1, 2, \ldots, m)$；称 $\hat{\theta}_k(x_1, x_2, \cdots, x_n)$ 为 θ_k 的估计值 $(k = 1, 2, \ldots, m)$。

问题

1. 如何构造统计量?
2. 如何评价统计量?

点估计方法

常用的点估计方法：

- 矩法估计法
- 极(最)大似然估计法
- 最小二乘估计法
- 贝叶斯方法

矩法估计的思想 I

假设总体 $X \sim F(x;\theta_1,\theta_2,\ldots,\theta_m)$，参数 $\theta_1,\theta_2,\ldots,\theta_m$ 未知。

1. 若总体的 m 阶矩存在：

$$\begin{cases} E[X] = g_1(\theta_1,\theta_2,\cdots,\theta_m), \\ E[X^2] = g_2(\theta_1,\theta_2,\cdots,\theta_m), \\ \qquad\cdots \\ E[X^m] = g_m(\theta_1,\theta_2,\cdots,\theta_m). \end{cases}$$

矩法估计的思想 II

2 设 $X_1, X_2, \ldots, X_n$ 是来自总体 X 的一个样本，则由辛钦大数定律，有：

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} E(X^k) = g_k(\theta_1, \theta_2, \ldots, \theta_m), \quad n \to \infty$$

因此当 n 较大时有：

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \approx g_k(\theta_1, \theta_2, \ldots, \theta_m), \quad (k = 1, 2, \ldots, m)$$

问题：如何估计$\theta_1, \theta_2, \cdots, \theta_m$?

用样本k阶矩作为总体k 阶矩的估计量。

矩法估计的基本步骤 I

1. 求解方程组

$$\begin{cases} E[X] = g_1(\theta_1, \theta_2, \cdots, \theta_m), \\ E[X^2] = g_2(\theta_1, \theta_2, \cdots, \theta_m), \\ \cdots \\ E[X^m] = g_m(\theta_1, \theta_2, \cdots, \theta_m). \end{cases}$$

把总体的参数表示成总体各阶原点矩的函数:

$$\begin{cases} \theta_1 = h_1(E[X], E[X^2], \cdots, E[X^m]), \\ \theta_2 = h_2(E[X], E[X^2], \cdots, E[X^m]), \\ \cdots \\ \theta_m = h_m(E[X], E[X^2], \cdots, E[X^m]). \end{cases}$$

矩法估计的基本步骤 II

2 用前 m 阶样本原点矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ 代替 X 的前 m 阶原点矩 $E[X^k]$，$j = 1, 2, \cdots, k$，得到 θ_j 的矩法估计量 $\hat{\theta}_j$：

$$\begin{cases} \hat{\theta}_1 = h_1\left(A_1, A_2, ..., A_m\right), \\ \hat{\theta}_2 = h_2\left(A_1, A_2, ..., A_m\right), \\ \quad \cdots \\ \hat{\theta_m} = h_m\left(A_1, A_2, ..., A_m\right). \end{cases}$$

3 对样本进行一次观测，假设 $(x_1, x_2, \cdots, x_n)$ 是样本的一个观测值，那么将其带入 $\hat{\theta}_j$ 得到的数值 $\hat{\theta}_j(x_1, x_2, \cdots, x_n)$ 就是参数 θ_j 的矩法估计值。

例题

1. **（二项分布）** 设 $X_1, X_2, \ldots, X_n$ 是来自总体 $X \sim B(m, p)$ 的一个样本，m 已知，求未知参数 p 的矩估计量。

例题

1. **（二项分布）** 设 $X_1, X_2, \ldots, X_n$ 是来自总体 $X \sim B(m, p)$ 的一个样本，m 已知，求未知参数 p 的矩估计量。
2. 设总体 X 的分布密度函数为

$$f_X(x,\theta) = \frac{\theta}{2} e^{-\theta|x|}, \quad -\infty < x < \infty, \quad \theta > 0.$$

试求 θ 的矩法估计.

命题

不论总体 X 服从什么分布，若其期望 μ 和方差 σ^2 的存在，则 μ 和 σ^2 的矩估计量分别为：

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

例题

1. **（正态分布）**水泥厂成品打包机装袋的重量服从正态分布，试用矩法估计来估计一台打包机装袋的重量的均值和方差。
2. **（泊松分布）**设总体 X 服从参数为 $\lambda(>0)$ 的泊松分布，求λ 的矩估计值.

例题

1. **（正态分布）**水泥厂成品打包机装袋的重量服从正态分布，试用矩法估计来估计一台打包机装袋的重量的均值和方差。
2. **（泊松分布）**设总体 X 服从参数为 $\lambda(>0)$ 的泊松分布，求λ 的矩估计值.
3. **（均匀分布）**设总体在某一区间$[a,b]$上均匀取值，试用矩法估计该区间的左右端点。

注：随机产生 $U(0,1)$ 的随机数 40 个：

$$0.4387, 0.3816, 0.7655, 0.7952, 0.1869, 0.4898, 0.4456, 0.6463,$$

$$0.7094, 0.7547, 0.2760, 0.6797, 0.6551, 0.1626, 0.1190, 0.4984,$$

$$0.9597, 0.3404, 0.5853, 0.2238, 0.7513, 0.2551, 0.5060, 0.6991,$$

$$0.8909, 0.9593, 0.5472, 0.1386, 0.1493, 0.2575, 0.8127, 0.2543,$$

$$0.8143, 0.2435, 0.9293, 0.3500, 0.1966, 0.2511, 0.6160, 0.4733$$

注：随机产生 $U(0,1)$ 的随机数 40 个：

$$0.4387, 0.3816, 0.7655, 0.7952, 0.1869, 0.4898, 0.4456, 0.6463,$$

$$0.7094, 0.7547, 0.2760, 0.6797, 0.6551, 0.1626, 0.1190, 0.4984,$$

$$0.9597, 0.3404, 0.5853, 0.2238, 0.7513, 0.2551, 0.5060, 0.6991,$$

$$0.8909, 0.9593, 0.5472, 0.1386, 0.1493, 0.2575, 0.8127, 0.2543,$$

$$0.8143, 0.2435, 0.9293, 0.3500, 0.1966, 0.2511, 0.6160, 0.4733$$

计算算得：$\bar{x} = 0.5059275, \quad \hat{s}_n = 0.2573$

计算得到 a, b 的矩估计值：

$$\hat{a} = \bar{x} - \sqrt{3}\hat{s}_n = 0.0602, \quad \hat{b} = \bar{x} + \sqrt{3}\hat{s}_n = 0.9516$$

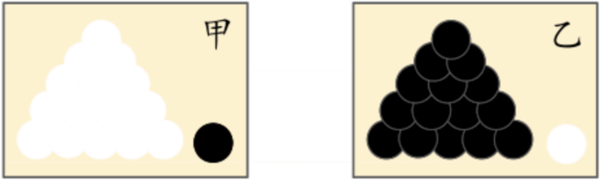
矩估计法小结

1. 原理直观；
2. 只用到总体矩，方法简单.但若总体矩不存在，无法使用矩估计法；
   - 设 $X_1, X_2, \ldots, X_n$ 是来自总体 X 的一个样本，总体 X 服从参数为 θ 的柯西分布，其密度函数为：

     $$f(x;\theta) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}, \quad -\infty < x < \infty$$

     则 θ 的矩估计不存在。
3. 矩估计基于大数定律，所以通常在大样本情况下，才有较好的效果。

最大似然估计



例如:一个箱子里有黑白两种颜色的球，数量为99个和1个，但并不知道到底哪种颜色的球为99个那种颜色的球为1个（含甲乙两种情形）。这时我们随机从箱子里拿出一个球，如果这个球是黑色的，你认为它应该属于哪一种情形?

最大似然估计

Fisher最大似然原理

在随机试验中，许多事件都有可能发生，概率大的事件发生的概率也大。若只进行一次试验，事件 A 发生了，则我们有理由认为 A 比其他事件发生的概率都大。

问题

如何将 Fisher 的最大似然思想应用于参数估计?

问题

如何将 Fisher 的最大似然思想应用于参数估计?

通过若干次试验，观察其结果，利用试验结果得到某个参数值能够使样本出现的概率为最大

1. 离散型总体 联合概率分布在观测点处概率最大.
2. 连续型总体 联合概率密度函数在观测点处取值最大.

似然函数

1. 总体 X 服从某种离散型分布，含有参数 θ，某次观察值为 $(x_1, x_2, \cdots, x_n)$，其似然函数为

$$L(x_1, \cdots, x_n; \theta) = P(X_1 = x_1, \cdots, X_n = x_n; \theta) = \prod_{i=1}^{n} P\{X_i = x_i; \theta\},$$

2. 总体 X 服从某种连续型分布，含有参数 θ，某次观察值为 $(x_1, x_2, \cdots, x_n)$，其似然函数为

$$L(x_1, \cdots, x_n; \theta) = f(X_1 = x_1, \cdots, X_n = x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

最大似然估计量

最大似然估计量

如上定义似然函数 $L(x_1, x_2, \cdots, x_n; \theta)$，如果某统计量 $\hat{\theta}_L = \hat{\theta}_L(X_1, X_2, \cdots, X_n) \in \Theta$，使得

$$L(x_1, x_2, \cdots, x_n; \hat{\theta}_L(x_1, x_2, \cdots, x_n)) = \max_{\theta} L(x_1, x_2, \cdots, x_n; \theta),$$

就称 $\hat{\theta}_L(X_1, X_2, \cdots, X_n)$ 为 θ 的最大似然估计量.

最大似然估计求解的一般步骤

1. 写出总体的概率/密度函数$f(x)$
2. 写出似然函数$L(x_1, x_2, \cdots, x_n; \theta)$
3. 对似然函数取ln, 计算对数似然函数lnL
4. 对各个参数θ求导（偏导），并令导数为0

指数分布的最大似然估计量

例题

对于 $X \sim \mathrm{Exp}(\lambda)$，有样本观测值$(x_1, x_2, \cdots, x_n)$，求参数 λ 的最大似然估计。

指数分布的最大似然估计量 I

1. 写出总体的概率/密度函数$f(x)$

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

2. 写出似然函数$L(x_1, x_2, \cdots, x_n; \theta)$

$$L(x_1, x_2, \cdots, x_n, \lambda) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right).$$

3. 对似然函数取ln, 计算lnL

$$G(x_1, x_2, \cdots, x_n, \lambda) = \ln L(x_1, x_2, \cdots, x_n, \lambda) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i.$$

指数分布的最大似然估计量 II

4 对参数θ求导，并令导数为0

对上式求其驻点，即求使得$\dfrac{\mathrm{d}G}{\mathrm{d}\lambda}=0$ 的 λ， 得到

$$\lambda^*=\frac{n}{\sum_{i=1}^n x_i}=\frac{1}{\overline{X}}$$

验证下

$$\frac{\mathrm{d}^2G}{\mathrm{d}\lambda^2}|_{\lambda^*}=-\frac{n}{\lambda^2}|_{\lambda^*}=-\frac{(\sum_{i=1}^n x_i)^2}{n}<0,$$

所以 $\lambda^\star$ 确实为 G（从而是 L）的极大值点

例题

1. 对于 $X \sim \text{Poisson}(\lambda)$，有样本观测值 $(x_1, x_2, \cdots, x_n)$，求参数 λ 的最大似然估计。
2. 对于 $X \sim \text{N}(\mu, \sigma^2)$，有样本观测值 $(x_1, x_2, \cdots, x_n)$，求参数 μ, σ^2 的最大似然估计。

例题

1. 对于 $X \sim \mathrm{Poisson}(\lambda)$，有样本观测值$(x_1, x_2, \cdots, x_n)$，求参数 λ 的最大似然估计。
2. 对于 $X \sim \mathrm{N}(\mu, \sigma^2)$，有样本观测值$(x_1, x_2, \cdots, x_n)$，求参数 μ, σ^2 的最大似然估计。
3. 对于 $X \sim \mathrm{U}[a, b]$，有样本观测值$(x_1, x_2, \cdots, x_n)$，求参数 a, b 的最大似然估计。

均匀分布的最大似然估计

似然函数为

$$L(x_1, x_2, \cdots, x_n, \theta) = \begin{cases} \dfrac{1}{(b-a)^n}, & x_1, x_2, \cdots, x_n \in [a, b], \\ 0, & \text{其他}. \end{cases}$$

对于给定的观测值 $(x_1, x_2, \cdots, x_n)$，要使得 L 取到最大值，则

$$a = \min\{x_1, x_2, \cdots, x_n\}$$

$$b = \max\{x_1, x_2, \cdots, x_n\}$$

问题：未知参数的最大似然估计唯一吗?

问题：未知参数的最大似然估计唯一吗?

例子: 设 $X_1, X_2, \ldots, X_n$ 是来自总体 $X \sim U(\theta - 1, \theta + 1)$ 的样本，$x_1, x_2, \ldots, x_n$ 是样本观测值。试求 θ 的最大似然估计。

最大似然估计的不变性

设 $\hat{\theta}$ 是 θ 的极大似然估计，$u = u(\theta)$ 是 θ 的函数，且有单值反函数：

$$\theta = \theta(u)$$

则 $u(\hat{\theta})$ 是 u 的最大似然估计。

例题

对于 $X \sim \text{Poisson}(\lambda)$，有样本$(x_1, x_2, \cdots, x_n)$，求$P(X = 0)$的最大似然估计。

7.2 估计量优劣性的评价

估计量的优劣性评价

对同一个参数，不同方法得到的估计量可能不同。

1. 应该选用哪一个估计量?
2. 用什么标准来评价一个估计量的好坏?

常用标准：

1. 无偏性
2. 有效性
3. 相合性

无偏估计量

定义 7.2.1

设总体 $X \sim F_X(\cdot,\theta)$，$\theta \in \Theta$，g 为 θ 的函数，$T(X_1, X_2, \cdots, X_n)$ 是 $g(\theta)$ 的一个估计量，如果

$$E[T(X_1, X_2, \cdots, X_n)] = g(\theta),$$

称 $T(X_1, X_2, \cdots, X_n)$ 为 $g(\theta)$ 的无偏估计量. 如果

$$\lim_{n\to\infty} E[T(X_1, X_2, \cdots, X_n)] = g(\theta),$$

则称 $T(X_1, X_2, \cdots, X_n)$ 为 $g(\theta)$ 的渐进无偏估计量.

无偏估计量

对任意分布总体 X有 $EX = \mu, DX = \sigma^2$，若$(X_1, \cdots, X_n)$为其样本。

① 样本均值 $\overline{X}$ 是 μ 的无偏估计，即 $E\overline{X} = \mu$。

② 样本方差 $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

$$\lim_{n\to\infty} ES_n^2 = \lim_{n\to\infty} \frac{n-1}{n}\sigma^2 = \sigma^2$$

是 σ^2 的渐进无偏估计。

③ 修正样本方差 $S_n^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 是 σ^2 的无偏估计，即 $ES_n^{*2} = \sigma^2$。

例题

$X \sim U[0,\theta]$，其区间右端点的矩估计量 $\hat{\theta}_M$ 与最大似然估计量 $\hat{\theta}_L$ 是否为 θ 的无偏估计量?

无偏与渐进无偏估计量 I

1. 对于矩法估计量

$$\hat{\theta}_M = 2\overline{X},$$

计算

$$E_\theta[\hat{\theta}_M] = 2E_\theta[\overline{X}] = \frac{2}{n}\sum_{i=1}^{n} E_\theta[X_i]$$
$$= \frac{2}{n} \cdot n \cdot \frac{\theta}{2} = \theta.$$

所以 $\hat{\theta}_M$ 是 θ 的无偏估计量.

无偏与渐进无偏估计量 II

2. 对于最大似然估计量

$$\hat{\theta}_L = X_{(n)},$$

由

$$f_{X_{(n)}}(x) = n[F_X(x)]^{n-1} f_X(x),$$

并且

$$f_X(x) = \begin{cases} \dfrac{1}{\theta}, & x \in [0,\theta], \\ 0, & \text{其他}. \end{cases} \quad F_X(x) = \begin{cases} 0, & x < 0, \\ \dfrac{x}{\theta}, & x \in [0,\theta], \\ 1, & \text{其他}. \end{cases}$$

有

$$f_{X_{(n)}}(x) = \begin{cases} n\left(\dfrac{x}{\theta}\right)^{n-1} \dfrac{1}{\theta}, & x \in [0,\theta], \\ 0, & \text{其他}. \end{cases}$$

无偏与渐进无偏估计量 III

所以

$$E_\theta[\hat{\theta}_L] = E_\theta[X_{(n)}] = \int_0^\theta xn\frac{x^{n-1}}{\theta^n}\,\mathrm{d}x = \frac{n}{n+1}\theta.$$

所以 $\hat{\theta}_L$ 不是 θ 的无偏估计量，而是 θ的渐进无偏估计量. 为此，若我们令

$$\hat{\theta} = \frac{n+1}{n}\hat{\theta}_L,$$

那么 $\hat{\theta}$ 就是 θ 的无偏估计量.

对于同一参数，可能存在几个无偏估计量。

无偏与渐进无偏估计量 IV

若 $\hat{\theta}$ 是 θ 的无偏估计量，那么 $g(\hat{\theta})$ 一般不是 $g(\theta)$ 的无偏估计量，除非 $g(x)$ 是线性函数.

例如: S_n^{*2} 是 σ^2 的无偏估计量，但是 S_n^* 却不是 σ 的无偏估计量.

无偏估计量

无偏估计量有可能具有很大的偏差

如: 对总体 $X \sim \text{Poisson}(\lambda)$，参数 $\lambda > 0$，$(X_1, X_2, \cdots, X_n)$ 为其样本，那么

$$T(X_1, X_2, \cdots, X_n) = (-2)^{X_1}$$

是 $e^{-3\lambda}$ 的无偏估计量，即

$$E[T(X_1, X_2, \cdots, X_n)] = e^{-3\lambda}.$$

但 $e^{-3\lambda} > 0$，$(-2)^{X_1}$ 却在 X_1 为奇数的时候取负值，可见偏差很大.

有效估计量

对于总体 X 的参数 $g(\theta)$ 的两个无偏估计量 $T_1(X_1, X_2, \cdots, X_n)$，$T_2(X_1, X_2, \cdots, X_n)$，若有

$$Var[T_1] < Var[T_2],$$

则称 T_1 比 T_2 有效.

由于方差反映随机变量取值的波动程度，因此认为波动较小的统计量更有效是合理的.

均匀分布例子续

判断总体 $X \sim U[0, \theta]$ 的区间右端点的矩估计量和校正的最大似然估计量哪一个更有效?

均匀分布例子续 I

$$
\begin{aligned}
Var[\hat{\theta}_M] &= Var[2\overline{X}] = Var\left[\frac{2}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{4}{n^2}\sum_{i=1}^{n} Var[X_i] = \frac{4}{n^2} n \frac{\theta^2}{12} = \frac{\theta^2}{3n}.
\end{aligned}
$$

$$
\begin{aligned}
Var[\hat{\theta}] = Var\left[\frac{n+1}{n} X_{(n)}\right] &= \frac{(n+1)^2}{n^2}\left[E[X_{(n)}^2] - \left(E[X_{(n)}]\right)^2\right] \\
&= \frac{(n+1)^2}{n^2}\left\{\int_0^{\theta} x^2 n \frac{x^{n-1}}{\theta^n}\,\mathrm{d}x - \frac{n^2}{(n+1)^2}\theta^2\right\} \\
&= \frac{(n+1)^2}{n^2}\left\{\frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2\right\} = \frac{\theta^2}{n(n+2)}
\end{aligned}
$$

均匀分布例子续 II

显然的，当 $n > 1$，就有

$$Var[\hat{\theta}] < Var[\hat{\theta}_M],$$

即 $\hat{\theta}$ 比 $\hat{\theta}_M$ 更有效.

一致最小方差无偏估计量

定义 7.2.2

设总体 $X \sim F_X(\cdot, \theta)$，$\theta \in \Theta$，若 $T_0(X_1, X_2, \cdots, X_n)$ 是$g(\theta)$ 的无偏估计量，且对 $g(\theta)$ 的任意无偏估计量 $T(X_1, X_2, \cdots, X_n)$ 都有

$$Var[T_0] \leq Var[T], \quad \forall \theta \in \Theta,$$

则称 T_0 是 $g(\theta)$ 的一致最小方差无偏估计量.

对于指数类分布，比如 $\mathsf{Exp}(\lambda)$， $\mathsf{Pois}(\lambda)$，$B(n, p)$ 和 $N(\mu, \sigma^2)$，其参数的最大似然估计量（也是矩法估计量）是一致最小方差无偏估计量.

例题

设总体 X 的密度函数为

$$f(x;\theta)=\begin{cases}\frac{3x^2}{\theta^3} & 0\le x\le\theta \\ 0 & \text{其他}\end{cases},$$

其中，$\theta>0$ 为未知参数，X_1,X_2,X_3 是取自总体 X 的样本。

(1) 证明

$$T_1=\frac{4}{9}(X_1+X_2+X_3)$$

和

$$T_2=\frac{10}{9}\max(X_1,X_2,X_3)$$

都是 θ 的无偏估计量；

(2) 比较 T_1,T_2 的方差并指出哪个更有效。

相合估计量

定义 7.2.3

设总体 $X \sim F_X(\cdot,\theta)$，$\theta \in \Theta$，并且设 $T(X_1, X_2, \cdots, X_n)$ 是 $g(\theta)$ 的估计量，如果对于任意的 $\epsilon > 0$，有

$$\lim_{n\to\infty} P(|T(X_1, X_2, \cdots, X_n) - g(\theta)| \geq \epsilon) = 0, \quad \forall\theta \in \Theta,$$

则称 $T(X_1, X_2, \cdots, X_n)$ 为 $g(\theta)$ 的相合估计量.

相合估计量的判定依据

若 $\hat{\theta}$ 为 θ 的相合估计，且 $\eta = g(\theta)$ 为 θ 的连续函数，那么 $\hat{\eta} = g(\hat{\theta})$ 为 η 的相合估计量.

1. $\overline{X^k}$ 是 $E[X^k]$ 的相合估计量.
2. S_n^2 和 S_n^{*2} 都是 σ^2 的相合估计量.
3. S_n 和 S_n^* 都是 σ 的相合估计量.

例题

设总体 X 服从区间 $[\theta, \theta+1]$ 上的均匀分布，其中 θ 是未知参数，设 $X_1, X_2, \ldots, X_n$ 是来自总体 X 的一个样本；证明：$\hat{\theta} = \bar{X} - \frac{1}{2}$ 是参数 θ 的相合估计量。

7.3 参数的区间估计

点估计方法的局限

设 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ 是未知参数 θ 的估计量。

1. 用 $\hat{\theta}$ 估计 θ，有多高的精度?
2. 用 $\hat{\theta}$ 估计 θ，有多高的可信度?
3. 未知参数 θ 落在什么范围内?

点估计方法的局限

设 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ 是未知参数 θ 的估计量。

1. 用 $\hat{\theta}$ 估计 θ，有多高的精度?
2. 用 $\hat{\theta}$ 估计 θ，有多高的可信度?
3. 未知参数 θ 落在什么范围内?

区间估计：希望根据所给的样本确定一个随机区间，使得该随机区间包含参数的概率满足给定的条件。

参数的区间估计

定义 7.3.1

设总体 $X \sim F_X(\cdot,\theta)$，$\theta \in \Theta$，$(X_1, X_2, \cdots, X_n)$ 为样本. 若有统计量 $T_1(X_1, X_2, \cdots, X_n)$ 和 $T_2(X_1, X_2, \cdots, X_n)$，使得对给定的$\alpha(0 < \alpha < 1)$ 有

$$P(T_1(X_1, X_2, \cdots, X_n) \le \theta \le T_2(X_1, X_2, \cdots, X_n)) = 1 - \alpha$$

则称 $[T_1(X_1, X_2, \cdots, X_n), T_2(X_1, X_2, \cdots, X_n)]$ 是 θ 的置信度为 $1-\alpha$ 的区间估计，或叫做**置信区间**。其中，$1-\alpha$ 称为**置信度**或**置信水平**，$T_1(X_1, X_2, \cdots, X_n), T_2(X_1, X_2, \cdots, X_n)$ 分别称为**置信下限**和**置信上限**。

区间估计的几点说明

1. 置信区间的区间长度（置信上限-置信下限）反映了估计的精度。区间长度越小，估计精度越高。
2. $1-\alpha$ 反映了估计的可信度。$1-\alpha$ 越大，估计的可信度越高；但通常会导致置信区间的区间长度增大，从而导致估计的精度降低。
3. α 给定后，置信区间的选取不唯一，通常选取置信区间的区间长度最小的区间。

问题

对于正态分布总体 $X \sim N(\mu, \sigma_0^2)$，其中 σ_0 已知，有样本$(X_1, X_2, ...X_n)$，求 μ 的置信度为 $1-\alpha$ 的区间估计.

求解置信区间的一般过程 I

1. 构造样本的一个函数：

$$T = T(X_1, X_2, \cdots, X_n; \theta)$$

   它含有待估参数 θ，不含其它未知参数，其分布已知，且分布不依赖于待估参数（常由 θ 的点估计出发考虑）。

   - 例如 $\bar{X} \sim N(\mu, \frac{\sigma_0^2}{n})$

$$\Longrightarrow T(X_1, X_2, \cdots, X_n; \mu) = \frac{\bar{X} - \mu}{\sigma_0/\sqrt{n}} \sim N(0, 1)$$

2. 对给定的置信度 $1-\alpha$，确定 $T(X_1, X_2, \cdots, X_n; \theta)$ 的分布的两个分位点：a,b，使得

$$P\{a \le T(X_1, X_2, \cdots, X_n; \theta) \le b\} = 1 - \alpha$$

求解置信区间的一般过程 II

- 一般分位数选取$a=x_{\alpha/2}$，$b=x_{1-\alpha/2}$
- 例如

$$P(|T|\le u_{1-\frac{\alpha}{2}})=1-\alpha \Rightarrow P\left(\left|\frac{\overline{X}-\mu}{\sqrt{\frac{\sigma_0^2}{n}}}\right|\le u_{1-\frac{\alpha}{2}}\right)=1-\alpha.$$

3. 根据 $a\le Y\le b$ 反解出 θ 的范围.

$$T_1(X_1,X_2,\cdots,X_n)\le\theta\le T_2(X_1,X_2,\cdots,X_n),$$

即有了 θ的$1-\alpha$ 置信区间.

- 例如

$$P\left(\overline{X}-u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_0^2}{n}}\le\mu\le\overline{X}+u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_0^2}{n}}\right)=1-\alpha.$$

正态分布总体均值的区间估计–方差 σ^2 已知

对于正态分布总体 $X \sim N(\mu, \sigma_0^2)$，其中 σ_0 已知，求 μ 的置信度为 $1-\alpha$ 的区间估计.

正态分布总体均值的区间估计–方差 σ^2 已知

假设 $(X_1, X_2, \cdots, X_n)$ 为其样本，由 $\overline{X} \sim N(\mu, \frac{\sigma_0^2}{n})$, 知

$$Y = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma_0^2}{n}}} \sim N(0, 1).$$

$$P(|Y| \leq u_{1-\frac{\alpha}{2}}) = 1 - \alpha \Rightarrow P\left(\left|\frac{\overline{X} - \mu}{\sqrt{\frac{\sigma_0^2}{n}}}\right| \leq u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

即

$$P\left(\overline{X} - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_0^2}{n}} \leq \mu \leq \overline{X} + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_0^2}{n}}\right) = 1 - \alpha.$$

所以我们有 μ 的置信度为 $1 - \alpha$ 的区间估计为

$$\left[\overline{X} - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_0^2}{n}}, \overline{X} + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_0^2}{n}}\right].$$

正态分布总体均值的区间估计–方差 σ^2 未知

对于正态分布总体 $X \sim N(\mu, \sigma^2)$，其中 σ 未知，求 μ 的置信度为 $1-\alpha$ 的区间估计.

正态分布总体均值的区间估计–方差 σ^2 未知

由

$$\frac{\overline{X}-\mu}{S_n}\sqrt{n-1}\sim t(n-1)\text{或}\frac{\overline{X}-\mu}{S_n^*}\sqrt{n}\sim t(n-1).$$

有

$$P\left(\left|\frac{\overline{X}-\mu}{S_n}\sqrt{n-1}\right|\leq t_{1-\frac{\alpha}{2}}(n-1)\right)=1-\alpha,$$

故 μ 的置信度为 $1-\alpha$ 的区间估计为

$$\left[\overline{X}-t_{1-\frac{\alpha}{2}}(n-1)\frac{S_n}{\sqrt{n-1}},\overline{X}+t_{1-\frac{\alpha}{2}}(n-1)\frac{S_n}{\sqrt{n-1}}\right].$$

或

$$\left[\overline{X}-t_{1-\frac{\alpha}{2}}(n-1)\frac{S_n^*}{\sqrt{n}},\overline{X}+t_{1-\frac{\alpha}{2}}(n-1)\frac{S_n^*}{\sqrt{n}}\right].$$

例题

灯泡厂从某天生产的一批灯泡中随机抽取 10 只进行寿命测试，测得数据$\bar{x} = 1147, s_{10}^{*} = 87.0568$ 长期实践表明灯泡寿命服从正态分布 $N(\mu, \sigma^2)$，σ^2 未知，求总体均值 μ 的置信度为 0.95 的置信区间。

正态分布总体方差的区间估计–均值 μ 已知

对于正态分布总体 $X \sim N(\mu_0, \sigma^2)$，其中 μ_0 已知，求 σ^2 的置信度为 $1-\alpha$ 的区间估计.

正态分布总体方差的区间估计–均值 μ 已知 I

由于总体与样本独立同分布，

$$Y_i = \frac{X_i - \mu_0}{\sigma} \sim N(0,1) \Rightarrow Z = \sum_{i=1}^{n} Y_i^2 \sim \chi^2(n).$$

所以

$$P\left(\chi^2_{\frac{\alpha}{2}}(n) \le Z \le \chi^2_{1-\frac{\alpha}{2}}(n)\right) = 1 - \alpha.$$

即

$$P\left(\frac{1}{\chi^2_{1-\frac{\alpha}{2}}(n)} \le \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \mu_0)^2} \le \frac{1}{\chi^2_{\frac{\alpha}{2}}(n)}\right) = 1 - \alpha.$$

正态分布总体方差的区间估计–均值 μ 已知 II

得到 σ^2 置信度为 $1-\alpha$ 的区间估计为

$$\left[\frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)}, \frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\chi^2_{\frac{\alpha}{2}}(n)}\right].$$

正态分布总体方差的区间估计–均值 μ 未知

对于正态分布总体 $X \sim N(\mu, \sigma^2)$，其中 μ 未知，求 σ^2 的置信度为 $1-\alpha$ 的区间估计.

由抽样分布基本定理，

$$\frac{nS_n^2}{\sigma^2} \sim \chi^2(n-1),$$

令

$$P\left(\chi^2_{\frac{\alpha}{2}}(n-1) \leq \frac{nS_n^2}{\sigma^2} \leq \chi^2_{1-\frac{\alpha}{2}(n-1)}\right) = 1-\alpha,$$

所以 σ^2 的置信度 $1-\alpha$ 的区间估计为

$$\left[\frac{nS_n^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \frac{nS_n^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right].$$

或

$$\left[\frac{(n-1)S_n^{*2}}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \frac{(n-1)S_n^{*2}}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right].$$

例题

设某次考试学生的考试成绩X服从正态分布$N(\mu,\sigma^2)$，其中μ,σ^2均未知。现从中随机抽取容量为25的一个样本，测得样本均值$\bar{x}=66.5$，样本方差$s^2=100$。求总体方差σ^2的置信度为0.90的置信区间。

两个正态总体参数的区间估计

设 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，且 X，Y 相互独立，样本分别为 $(X_1,X_2,\cdots,X_m)$，$(Y_1,Y_2,\cdots,Y_n)$. 其中总体 X 与 Y 的样本均值与样本方差(修正样本方差)分别为

$$\overline{X},\ \overline{Y},\ S_{1m}^2,\ S_{2n}^2(S_{1m}^{*2},\ S_{2n}^{*2}).$$

问题

$\mu_1-\mu_2$ 和 $\dfrac{\sigma_1}{\sigma_2}$ 的区间估计量分别是什么?

两正态总体 $\mu_1 - \mu_2$ 的区间估计–σ_1^2 和 σ_2^2 已知

σ_1^2 和 σ_2^2 已知，求 $\mu_1 - \mu_2$ 的区间估计.

两正态总体 $\mu_1-\mu_2$ 的区间估计–σ_1^2 和 σ_2^2 已知

由 $\overline{X}\sim N(\mu_1,\frac{\sigma_1^2}{m}),\overline{Y}\sim N(\mu_2,\frac{\sigma_2^2}{n})$,

$$\Rightarrow \quad \frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\sim N(0,1).$$

令

$$P\left(\left|\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\right|\le u_{1-\frac{\alpha}{2}}\right)=1-\alpha,$$

得 u_1-u_2 的置信度为 $1-\alpha$ 区间估计为

$$\left[(\overline{X}-\overline{Y})-u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}},(\overline{X}-\overline{Y})+u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}\right].$$

两正态总体 $\mu_1 - \mu_2$ 的区间估计–$\sigma_1 = \sigma_2$ 已知

$\sigma_1 = \sigma_2 = \sigma$，但是具体的值未知，求 $\mu_1 - \mu_2$的区间估计.

两正态总体 $\mu_1-\mu_2$ 的区间估计–$\sigma_1=\sigma_2$ 已知

由

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}\sim t(m+n-2).$$

其中 $S_w=\sqrt{\frac{mS_{1m}^2+nS_{2n}^2}{m+n-2}}$ 以及 t 分布关于 y 轴对称的性质，得到

$$P\left(\left|\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}\right|\leq t_{1-\frac{\alpha}{2}}(m+n-2)\right)=1-\alpha.$$

知 $\mu_1-\mu_2$ 的置信度为 $1-\alpha$ 的区间估计量为

$$\left[(\overline{X}-\overline{Y})-t_{1-\frac{\alpha}{2}}(m+n-2)S_w\sqrt{\frac{1}{m}+\frac{1}{n}},(\overline{X}-\overline{Y})+t_{1-\frac{\alpha}{2}}(m+n-2)S_w\sqrt{\frac{1}{m}+\frac{1}{n}}\right].$$

例题

为了比较两位职员为顾客办理个人话费缴纳业务的速度，分别记录了两位职员为12名顾客办理业务所需的时间（单位：min），经过计算其样本均值与修正样本方差分别为：

职员甲

$$\bar{x} = 13, s_1^{*2} = 14.32, n_1 = 12.$$

职员乙

$$\bar{y} = 16, s_2^{*2} = 15.98, n_2 = 12.$$

假定每位职员办理业务所需时间都服从正态分布，且方差相同，试求甲、乙两位职员所办理业务平均所需时间之差$\mu_{甲} - \mu_{乙}$的90%的置信区间。

两正态总体 $\frac{\sigma_1^2}{\sigma_2^2}$ 的区间估计–μ_1，μ_2 已知

μ_1，μ_2 已知，求 $\frac{\sigma_1^2}{\sigma_2^2}$ 的区间估计.

两正态总体 $\frac{\sigma_1^2}{\sigma_2^2}$ 的区间估计–μ_1，μ_2 已知 I

$$\frac{1}{\sigma_1^2}\sum_{i=1}^{m}(X_i-\mu_1)^2 \sim \chi^2(m),\ \frac{1}{\sigma_2^2}\sum_{i=1}^{n}(Y_i-\mu_2)^2 \sim \chi^2(n),$$

根据 F 分布的定义有

$$\frac{n}{m}\frac{\sum_{i=1}^{m}(X_i-\mu_1)^2}{\sum_{i=1}^{n}(Y_i-\mu_2)^2}\frac{\sigma_2^2}{\sigma_1^2} \sim F(m,n).$$

即

$$P\left(F_{\frac{\alpha}{2}}(m,n) \le \frac{n}{m}\frac{\sum_{i=1}^{m}(X_i-\mu_1)^2}{\sum_{i=1}^{n}(Y_i-\mu_2)^2}\frac{\sigma_2^2}{\sigma_1^2} \le F_{1-\frac{\alpha}{2}}(m,n)\right) = 1-\alpha.$$

由此可得 $\frac{\sigma_1^2}{\sigma_2^2}$ 的置信度为 $1-\alpha$ 的区间估计

$$\left[\frac{1}{F_{1-\frac{\alpha}{2}}(m,n)}\frac{n\sum_{i=1}^{m}(X_i-\mu_1)^2}{m\sum_{i=1}^{n}(Y_i-\mu_2)^2}, \frac{1}{F_{\frac{\alpha}{2}}(m,n)}\frac{n\sum_{i=1}^{m}(X_i-\mu_1)^2}{m\sum_{i=1}^{n}(Y_i-\mu_2)^2}\right].$$

两正态总体 $\frac{\sigma_1^2}{\sigma_2^2}$ 的区间估计–μ_1，μ_2 未知

μ_1，μ_2 未知，求 $\frac{\sigma_1^2}{\sigma_2^2}$ 的区间估计.

两正态总体 $\frac{\sigma_1^2}{\sigma_2^2}$ 的区间估计–μ_1，μ_2 未知 I

由 $\frac{mS_{1m}^2}{nS_{2n}^2}\frac{\sigma_2^2}{\sigma_1^2}\frac{n-1}{m-1} \sim F(m-1, n-1)$. 有

$$P\left(F_{\frac{\alpha}{2}}(m-1, n-1) \le \frac{mS_{1m}^2}{nS_{2n}^2}\frac{\sigma_2^2}{\sigma_1^2}\frac{n-1}{m-1} \le F_{1-\frac{\alpha}{2}}(m-1, n-1)\right) = 1-\alpha.$$

可得 $\frac{\sigma_1^2}{\sigma_2^2}$ 的置信度为 $1-\alpha$ 的区间估计量为

$$\left[\frac{1}{F_{1-\frac{\alpha}{2}}(m-1, n-1)}\frac{m(n-1)S_{1m}^2}{n(m-1)S_{2n}^2}, \frac{1}{F_{\frac{\alpha}{2}}(m-1, n-1)}\frac{m(n-1)S_{1m}^2}{n(m-1)S_{2n}^2}\right].$$

或

$$\left[\frac{1}{F_{1-\frac{\alpha}{2}}(m-1, n-1)}\frac{S_{1m}^{*2}}{S_{2n}^{*2}}, \frac{1}{F_{\frac{\alpha}{2}}(m-1, n-1)}\frac{S_{1m}^{*2}}{S_{2n}^{*2}}\right].$$

例题

为了比较两位职员为顾客办理个人话费缴纳业务的速度，分别记录了两位职员为12名顾客办理业务所需的时间（单位：min），经过计算其样本均值与修正样本方差分别为：

职员甲

$$\bar{x} = 13, s_1^{*2} = 14.32, n_1 = 12.$$

职员乙

$$\bar{y} = 16, s_2^{*2} = 15.98, n_2 = 12.$$

假定每位职员办理业务所需时间都服从正态分布，试求甲、乙两位职员所办理业务平均所需时间方差比的95%的置信区间。