概率论与数理统计
第七章 参数估计

金玲飞

复旦大学软件学院
Email: lfjin@fudan.edu.cn

2019.12.17

参数估计的基本思想

参数估计：用所获得的样本值去估计参数取值。

参数估计

- **点估计**：用某一数值作为参数的近似值
- **区间估计**：在要求的精度范围内指出参数所在的区间

目录

- 点估计（矩估计，极大似然估计）
- 估计量的评判标准
- 区间估计
- 正态总体均值与方差的区间估计

7.1 点估计

已知分布类型$F(x;\theta)$，但是θ未知。

- 参数空间：Θ，参数$\theta \in \Theta$。

已知分布类型$F(x;\theta)$，但是θ未知。

- 参数空间：Θ，参数$\theta \in \Theta$。
- 估计量(estimator)：用于估计总体参数的随机变量。记$\hat{\theta}$
  - EX: 样本均值就是总体均值μ的一个估计量。
- 估计值(estimated value): 估计参数时计算出来的统计量的具体值。

点估计

定义 (7.1.1)

设总体X的分布函数为$F(x;\theta)$，未知参数$\theta \in \Theta$，$X_1, X_2, \ldots, X_n$是来自总体X的样本，$x_1, x_2, \ldots, x_n$是样本观测值。选取一个统计量$\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$来估计$\theta$。我们称$\hat{\theta}(X_1, X_2, \ldots, X_n)$是$\theta$的点估计量，$\hat{\theta}(x_1, x_2, \ldots, x_n)$ 是θ的点估计值。

点估计

定义 (7.1.1)

设总体X的分布函数为$F(x;\theta)$，未知参数$\theta \in \Theta$，$X_1, X_2, \ldots, X_n$是来自总体X的样本，$x_1, x_2, \ldots, x_n$是样本观测值。选取一个统计量$\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$来估计$\theta$。我们称$\hat{\theta}(X_1, X_2, \ldots, X_n)$是$\theta$的点估计量，$\hat{\theta}(x_1, x_2, \ldots, x_n)$ 是θ的点估计值。

如何选取合适的统计量来估计参数？

矩估计法–1900年由K. Pearson提出

- 设总体X的分布函数为$F(x;\theta_1,\dots,\theta_k)$，有$k$个未知参数$\theta_1,\dots,\theta_k$。$X_1,X_2,\dots,X_n$是来自总体$X$的样本，$x_1,x_2,\dots,x_n$是样本观测值。设总体$X$的矩$E(X^k)$存在，此时所有低阶矩均存在。

矩估计法–1900年由K. Pearson提出

- 设总体X的分布函数为$F(x;\theta_1,\dots,\theta_k)$，有$k$个未知参数$\theta_1,\dots,\theta_k$。$X_1,X_2,\dots,X_n$是来自总体$X$的样本，$x_1,x_2,\dots,x_n$是样本观测值。设总体$X$的矩$E(X^k)$存在，此时所有低阶矩均存在。
- 原理：当n较大是，样本k阶原点矩依概率收敛到总体k阶矩。
- 矩法估计："替换"。用样本矩来代替总体矩，从而得到总体分布中的参数的一种估计。

矩估计法 moment estimate

1. 计算$E(X^j), j=1,\ldots,k$，得到含有未知参数$\theta_1,\ldots,\theta_k$的k个方程。

$$E(X^j)=g_j(\theta_1,\ldots,\theta_k), j=1,\ldots,k$$

2. 近似替换，列出方程组

$$\begin{cases} g_1(\theta_1,\ldots,\theta_k)=\frac{1}{n}\sum_{i=1}^n X_i \\ g_2(\theta_1,\ldots,\theta_k)=\frac{1}{n}\sum_{i=1}^n X_i^2 \\ \cdots\cdots \\ g_k(\theta_1,\ldots,\theta_k)=\frac{1}{n}\sum_{i=1}^n X_i^k \end{cases}$$

3. 解方程组，得到$\theta_1,\ldots,\theta_k$的一组解，$\hat{\theta}_j=\hat{\theta}_j(X_1,X_2,\ldots,X_n)$。

用矩法得到的估计量$\hat{\theta}_j(X_1, X_2, \ldots, X_n)$称为$\hat{\theta}_j$的矩估计量 Moment estimator。估计值$\hat{\theta}_j(x_1, x_2, \ldots, x_n)$ 称为$\hat{\theta}_j$的矩估计值。

用矩法得到的估计量$\hat{\theta}_j(X_1, X_2, \ldots, X_n)$称为$\hat{\theta}_j$的矩估计量 Moment estimator。估计值$\hat{\theta}_j(x_1, x_2, \ldots, x_n)$ 称为$\hat{\theta}_j$的矩估计值。

矩法的优点：精髓是替换，计算简便。当样本容量n很大时，矩估计接近被估计参数的真值的概率很大。

用矩法得到的估计量$\hat{\theta}_j(X_1, X_2, \ldots, X_n)$称为$\hat{\theta}_j$的矩估计量 Moment estimator。估计值$\hat{\theta}_j(x_1, x_2, \ldots, x_n)$ 称为$\hat{\theta}_j$的矩估计值。

矩法的优点：精髓是替换，计算简便。当样本容量n很大时，矩估计接近被估计参数的真值的概率很大。

矩法的缺点：

- 要求期望存在。
- 矩估计不唯一(尽量使用低阶矩)。

例子 (7.1.1)

设总体$X \sim B(m, p)$，其中m已知，$p(0 < p < 1)$是未知参数，$X_1, X_2, \ldots, X_n$是来自总体X的样本，求p的矩估计量。

例子 (7.1.1)

设总体$X \sim B(m,p)$，其中m已知，$p(0<p<1)$是未知参数，$X_1,X_2,\ldots,X_n$是来自总体X的样本，求p的矩估计量。

例子 (7.1.2)

设总体$X \sim N(\mu,\sigma^2)$，μ,σ^2是未知参数，$X_1,X_2,\ldots,X_n$是来自总体X的样本，求μ,σ^2的矩估计量。

例子 (7.1.3)

设总体$X \sim U[0, \theta]$，$\theta > 0$是未知参数，$X_1, X_2, \ldots, X_n$是来自总体X的样本，求θ的矩估计量。

例子 (7.1.3)

设总体$X \sim U[0,\theta]$，$\theta > 0$是未知参数，$X_1, X_2, \ldots, X_n$是来自总体X的样本，求θ的矩估计量。

- 若$X \sim U[a,b]$呢？

例子 (7.1.3)

设总体$X \sim U[0,\theta]$，$\theta > 0$是未知参数，$X_1, X_2, \ldots, X_n$是来自总体X的样本，求θ的矩估计量。

- 若$X \sim U[a,b]$呢？

$$\widehat{a} = \overline{X} - \sqrt{3}S, \qquad \widehat{b} = \overline{X} + \sqrt{3}S,$$

最大似然估计法

最大似然估计法: 在总体类型已知的条件下使用的一种参数估计方法。

首先是由高斯(Gauss)在1821年提出。

费希尔(Fisher)在1912年重新发现了这个方法，并证明了一些性质

最大似然原理

最大似然原理: 如果在一次随机试验中结果A发生，而导致A发生的原因有很多。在分析导致A发生的原因时，使结果A发生的概率最大的原因，推断为导致A发生的原因。

最大似然原理

最大似然原理: 如果在一次随机试验中结果A发生，而导致A发生的原因有很多。在分析导致A发生的原因时，使结果A发生的概率最大的原因，推断为导致A发生的原因。

例子 (7.1.4)

设有外形完全相同的两个箱子，甲箱有*99*个白球和*1*个黑球，乙箱有*1*个白球和*99*个黑球，今随机地抽取一箱，然后再从这箱中任取一球，结果发现是白球，问这球是从哪一个箱子中取出的？

设总体X的分布类型已知，但含有未知参数θ。

定义 (7.1.2 似然函数 Likelihood function)

设$X_1, X_2, \ldots, X_n$是来自总体X的样本，$x_1, x_2, \ldots, x_n$是样本观察值。令

$$L(\theta) = L(\theta; x_1, \ldots, x_n) = \begin{cases} \prod_{i=1}^{n} p(x_i; \theta) & \text{离散型} \\ \prod_{i=1}^{n} f(x_i; \theta) & \text{连续型} \end{cases}$$

称$L(\theta)$为似然函数。

设总体X的分布类型已知，但含有未知参数θ。

定义 (7.1.2 似然函数 Likelihood function)

设$X_1, X_2, \ldots, X_n$是来自总体X的样本，$x_1, x_2, \ldots, x_n$是样本观察值。令

$$L(\theta) = L(\theta; x_1, \ldots, x_n) = \begin{cases} \prod_{i=1}^{n} p(x_i; \theta) & \text{离散型} \\ \prod_{i=1}^{n} f(x_i; \theta) & \text{连续型} \end{cases}$$

称$L(\theta)$为似然函数。

注：似然函数与联合分布一样吗？

最大似然法

- 将θ 看成原因， $(x_1, \ldots, x_n)$ 看成结果。
- 最大似然法： 寻找参数值$\hat{\theta}(x_1, \ldots, x_n)$使得当$\theta = \hat{\theta}(x_1, \ldots, x_n)$时出现样本观察值$(x_1, \ldots, x_n)$的可能性最大。

最大似然法

- 将θ 看成原因，$(x_1, \ldots, x_n)$ 看成结果。
- 最大似然法： 寻找参数值$\hat{\theta}(x_1, \ldots, x_n)$使得当$\theta = \hat{\theta}(x_1, \ldots, x_n)$时出现样本观察值$(x_1, \ldots, x_n)$的可能性最大。

定义 (7.1.3 Maximum likelihood estimate)

设$L(\theta) = L(\theta; x_1, \ldots, x_n)$是似然函数，若存在$\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$，使得

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

则称$\hat{\theta}(x_1, \ldots, x_n)$是未知参数$\theta$的最大似然估计值，称$\hat{\theta}(X_1, \ldots, X_n)$ 是未知参数θ的最大似然估计量。

如何求$L(\theta)$的最大值点？

如何求$L(\theta)$的最大值点？

如何求$L(\theta)$的最大值点？

如何求$L(\theta)$的最大值点？

- 若似然函数$L(\theta)$关于θ可导，令

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

解此方程得θ的最大似然估计值$\hat{\theta}(x_1, x_2, \ldots, x_n)$。

如何求$L(\theta)$的最大值点？

如何求$L(\theta)$的最大值点？

- 若似然函数$L(\theta)$关于θ可导，令

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

解此方程得θ的最大似然估计值$\hat{\theta}(x_1, x_2, \ldots, x_n)$。

- $L(\theta)$与$lnL(\theta)$具有相同的最大值点，令

$$\frac{dlnL(\theta)}{d\theta} = 0$$

解此方程得θ的最大似然估计值$\hat{\theta}(x_1, x_2, \ldots, x_n)$。

最大似然估计的一般步骤归纳如下：

(1) 求似然函数$L(\theta)$；

(2) 求出$lnL(\theta)$并整理；

(3)求似然方程组$\frac{\partial}{\partial\theta}lnL(\theta)=0$；

(3) 解上述方程得最大似然估计值$\hat{\theta}(x_1,x_2,\ldots,x_n)$或最大似然估计量$\hat{\theta}(X_1,X_2,\ldots,X_n)$。

例子 (7.1.5)

设总体$X \sim B(m, p)$，其中m已知，p为未知参数，$X_1, \ldots, X_n$是来自X的样本，$x_1, \ldots, x_n$是样本观察值，求p的最大似然估计量。

例子 (7.1.5)

设总体$X \sim B(m, p)$，其中m已知，p为未知参数，$X_1, \ldots, X_n$是来自X的样本，$x_1, \ldots, x_n$是样本观察值，求p的最大似然估计量。

p的最大似然估计量是$\frac{\overline{X}}{m}$。

例子 (7.1.7)

设总体$X \sim N(\mu, \sigma^2)$，$X_1, \ldots, X_n$是来自X的样本，$x_1, x_2, \ldots, x_n$是样本观察值，求μ, σ^2的最大似然估计量。

例子 (7.1.7)

设总体$X \sim N(\mu, \sigma^2)$，$X_1, \ldots, X_n$是来自X的样本，$x_1, x_2, \ldots, x_n$是样本观察值，求μ, σ^2的最大似然估计量。

例子 (7.1.8)

设总体X服从$[0, \theta]$上的均匀分布，$\theta > 0$是未知参数，$X_1, \ldots, X_n$是来自X的样本，$x_1, \ldots, x_n$是样本观察值，求θ的最大似然估计值。

例子 (7.1.7)

设总体$X \sim N(\mu, \sigma^2)$，$X_1, \ldots, X_n$是来自X的样本，$x_1, x_2, \ldots, x_n$是样本观察值，求μ, σ^2的最大似然估计量。

例子 (7.1.8)

设总体X服从$[0, \theta]$上的均匀分布，$\theta > 0$是未知参数，$X_1, \ldots, X_n$是来自X的样本，$x_1, \ldots, x_n$是样本观察值，求θ的最大似然估计值。

θ的最大似然估计值是$\widehat{\theta} = \max\{x_1, \cdots, x_n\}$

- 最大似然估计不一定唯一。

不变性

- 若$\hat{\theta}$是总体分布中未知参数θ的最大似然估计，则$g(\hat{\theta})$是否仍是$g(\theta)$的最大似然估计？

不变性

- 若$\hat{\theta}$是总体分布中未知参数θ的最大似然估计，
则$g(\hat{\theta})$是否仍是$g(\theta)$的最大似然估计？

性质 (7.1.4 最大似然估计的不变性)

设$g(\theta)$是单射，若$\hat{\theta}$是总体分布中未知参数θ的最大似然估计，则$g(\hat{\theta})$是$g(\theta)$的最大似然估计。

表 7.1　常见分布中参数的估计量

分布	矩估计	最大似然估计	性质
两点分布 $B(1,p)$	$\hat{p}=\overline{X}$	$\hat{p}=\overline{X}$	有效估计,相合估计
二项分布 $B(N,p)$	$\hat{p}=\frac{\overline{X}}{N}$	$\hat{p}=\frac{\overline{X}}{N}$	无偏估计,相合估计
泊松分布 (参数λ)	$\hat{\lambda}=\overline{X}$	$\hat{\lambda}=\overline{X}$	有效估计,相合估计
正态分布 $N(\mu,\sigma^2)$	$\hat{\mu}=\overline{X}$ $\hat{\sigma}^2=\frac{n-1}{n}S^2$ $\hat{\sigma}=\sqrt{\frac{n-1}{n}}S$	$\hat{\mu}=\overline{X}$ $\hat{\sigma}^2=\frac{n-1}{n}S^2$ $\hat{\sigma}=\sqrt{\frac{n-1}{n}}S$	$\hat{\mu}=\overline{X}$是μ的有效、相合估计;S^2为σ^2的渐近有效估计,一致最小方差无偏估计,相合估计;S为σ的相合估计,有偏估计
正态分布$N(\mu,\sigma_0^2)$ (σ_0^2已知)	$\hat{\mu}=\overline{X}$	$\hat{\mu}=\overline{X}$	有效估计,相合估计
正态分布$N(\mu_0,\sigma^2)$ (μ_0已知)	$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)^2$	$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)^2$	有效估计,相合估计
指数分布(参数λ) (记$\theta=\frac{1}{\lambda}$)	$\hat{\lambda}=\frac{1}{\overline{X}}$ $\hat{\theta}=\overline{X}$	$\hat{\lambda}=\frac{1}{\overline{X}}$ $\hat{\theta}=\overline{X}$	$\hat{\lambda}=\frac{1}{\overline{X}}$是$\lambda$的相合估计,$\frac{1}{\overline{X}}-\frac{1}{n\overline{X}}$是$\lambda$的一致最小方差无偏估计,相合估计;$\hat{\theta}=\overline{X}$是$\theta$的有效、相合估计
均匀分布 $U(0,\theta)$	$\hat{\theta}=2\overline{X}$	$\hat{\theta}=\max_{i\leqslant n}X_i$	$2\overline{X}$是θ的无偏、相合估计;$\frac{n+1}{n}\max_{i\leqslant n}X_i$是$\theta$的一致最小方差无偏估计,相合估计
均匀分布 $U(a,b)$	$\hat{a}=\overline{X}-\sqrt{3}\sqrt{\frac{n-1}{n}}S$ $\hat{b}=\overline{X}+\sqrt{3}\sqrt{\frac{n-1}{n}}S$	$\hat{a}=\min_{i\leqslant n}X_i$ $\hat{b}=\max_{i\leqslant n}X_i$	相合估计,有偏估计

7.2 估计量的优良性准则

估计量的优良性准则

一个“好”的估计量应该具有如下的条件：

- 无偏性：$E(\hat{\theta}) = \theta$
- 有效性：无偏估计量的方差尽可能小。
- 相合性：当样本容量越来越大时，$\hat{\theta}$应越来越接近θ的真值。

无偏性

反映估计量相对待估参数有无系统偏差。

定义 (7.2.1 无偏性)

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

$$E(\hat{\theta}) = \theta, \forall \theta \in \Theta$$

则称$\hat{\theta}$是θ的无偏估计量 *unbiased estimator*。若

$$\lim_{n \to \infty} E(\hat{\theta}) = \theta, \forall \theta \in \Theta$$

则称$\hat{\theta}$是θ的渐近无偏估计量。

无偏性

反映估计量相对待估参数有无系统偏差。

定义 (7.2.1 无偏性)

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

$$E(\hat{\theta}) = \theta, \forall \theta \in \Theta$$

则称$\hat{\theta}$是θ的无偏估计量 *unbiased estimator*。若

$$\lim_{n \to \infty} E(\hat{\theta}) = \theta, \forall \theta \in \Theta$$

则称$\hat{\theta}$是θ的渐近无偏估计量。

称$E(\hat{\theta}) - \theta$为系统偏差。

- 样本均值是总体均值的无偏估计
- 样本矩是总体矩的无偏估计
- 样本方差是总体方差的无偏估计。

- 样本均值是总体均值的无偏估计
- 样本矩是总体矩的无偏估计
- 样本方差是总体方差的无偏估计。

- 样本标准差S是总体标准差σ的无偏估计吗？

例子 (7.2.1)

设$X_1, \ldots, X_n$是来自总体$X \sim N(\mu, 1)$的样本，$x_1, \ldots, x_n$是样本观察值，求μ的最大似然估计量$\hat{\mu}$，并说明$\hat{\mu}$是无偏估计量。

例子 (7.2.1)

设$X_1, \ldots, X_n$是来自总体$X \sim N(\mu, 1)$的样本，$x_1, \ldots, x_n$是样本观察值，求μ的最大似然估计量$\hat{\mu}$，并说明$\hat{\mu}$是无偏估计量。

$$\widehat{\mu} = \overline{X}, E(\widehat{\mu}) = E(\overline{X}) = \mu$$

例子 (7.2.1)

设$X_1, \ldots, X_n$是来自总体$X \sim N(\mu, 1)$的样本，$x_1, \ldots, x_n$是样本观察值，求μ的最大似然估计量$\hat{\mu}$，并说明$\hat{\mu}$是无偏估计量。

$$\widehat{\mu} = \overline{X}, E(\widehat{\mu}) = E(\overline{X}) = \mu$$

例子 (7.2.2)

设总体X服从$[0, \theta]$上的均匀分布，试证：θ的矩估计量$\hat{\theta}_1 = 2\overline{X}$是$\theta$的无偏估计量，$\theta$的最大似然估计量$\hat{\theta}_2 = X_{(n)}$是$\theta$的渐近无偏估计量。

(1) 无偏估计一般不唯一；如均匀分布

(1) 无偏估计一般不唯一；如均匀分布

(2) 无偏估计没有不变性（除非$g(\theta)$是θ的线性函数）；

(1) 无偏估计一般不唯一；如均匀分布

(2) 无偏估计没有不变性（除非$g(\theta)$是θ的线性函数）；

(3) 只有大量重复使用同一估计量时，无偏性才有意义；

(4) 无偏估计不一定存在；

(5) 无偏性的要求可能导致不合理的要求；

例子

设$x_1, \cdots, x_n$是来自二点分布$b(1, p)$的一个样本，

(1) 寻求$p(1-p)$的无偏估计；

(2) 证明$\frac{1}{p}$的无偏估计不存在。

有效性

在无偏估计中，方差小的估计量较好。

定义 (7.2.2 有效性)

设$\hat{\theta}_1 = \hat{\theta}_1(X_1, \ldots, X_n)$和$\hat{\theta}_2 = \hat{\theta}_2(X_1, \ldots, X_n)$都是未知参数$\theta$的无偏估计量，若

$$D(\hat{\theta}_1) \leq D(\hat{\theta}_2), \quad \forall \theta \in \Theta$$

且至少存在一个$\theta \in \Theta$，使得上述不等号严格成立，则称$\hat{\theta}_1$比$\hat{\theta}_2$有效。

例子 (7.2.3)

根据例子*7.2.2*，如果总体X服从$[0,\theta]$上的均匀分布，则我们有θ的两个无偏估计量

$$\hat{\theta}_1 = 2\overline{X}$$

$$\hat{\theta}_3 = \frac{n+1}{n} X_{(n)}$$

试比较它们的有效性。

相合性

当样本容量$n \to \infty$

定义 (7.2.4 Consistent estimator)

设$\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$是未知参数$\theta$的估计量，若对任意正数$\epsilon$，有

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

则称$\hat{\theta}$是θ的相合估计量。

相合性

当样本容量$n \to \infty$

定义 (7.2.4 Consistent estimator)

设$\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$是未知参数$\theta$的估计量，若对任意正数$\epsilon$，有

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

则称$\hat{\theta}$是θ的相合估计量。

一个估计量必须具有相合性。

- 样本均值是总体均值的相合估计
- 样本矩是总体矩的相合估计
- 样本方差是总体方差的相合估计。

- 样本均值是总体均值的相合估计
- 样本矩是总体矩的相合估计
- 样本方差是总体方差的相合估计。

相合估计量不变性：若$\hat{\theta}$是θ的相合估计，对于θ的某个连续函数$g(\theta)$，则$g(\hat{\theta})$ 是$g(\theta)$ 的相合估计。