

数理统计

第三章

点估计

2026 年 4 月 8 日

1 3.1 引言

- 3.1.1 参数估计问题
- 3.1.2 点估计
- 3.1.3 点估计的优良性准则

2 3.2 矩估计

- 3.2.1 矩法和矩估计量
- 3.2.2 若干例子

3.1.1 参数估计问题

数理统计的任务是用样本去推断总体。参数估计是统计推断的一种重要形式，它是用样本对总体中感兴趣的未知参数作出推断的方法。

3.1.1 参数估计问题

数理统计的任务是用样本去推断总体。参数估计是统计推断的一种重要形式，它是用样本对总体中感兴趣的未知参数作出推断的方法。

设有参数分布族 $\mathcal{F} = \{F_\theta, \theta \in \Theta\}$ ，其中 Θ 是参数空间， F_θ 的分布形式已知，但其分布与未知参数 θ 有关。已知 X_1, \dots, X_n 是从参数分布族 \mathcal{F} 中抽取的简单样本。

3.1.1 参数估计问题

数理统计的任务是用样本去推断总体。参数估计是统计推断的一种重要形式，它是用样本对总体中感兴趣的未知参数作出推断的方法。

设有参数分布族 $\mathcal{F} = \{F_\theta, \theta \in \Theta\}$ ，其中 Θ 是参数空间， F_θ 的分布形式已知，但其分布与未知参数 θ 有关。已知 X_1, \dots, X_n 是从参数分布族 \mathcal{F} 中抽取的简单样本。

那么自然的问题是，样本来自的真实分布是哪一个？

3.1.1 参数估计问题

数理统计的任务是用样本去推断总体。参数估计是统计推断的一种重要形式，它是用样本对总体中感兴趣的未知参数作出推断的方法。

设有参数分布族 $\mathcal{F} = \{F_\theta, \theta \in \Theta\}$ ，其中 Θ 是参数空间， F_θ 的分布形式已知，但其分布与未知参数 θ 有关。已知 X_1, \dots, X_n 是从参数分布族 \mathcal{F} 中抽取的简单样本。

那么自然的问题是，样本来自的真实分布是哪一个？

或者，样本来自的真实分布 F_θ 对应的参数 θ 或其函数 $g(\theta)$ 的值是什么？

3.1.1 参数估计问题

因此，参数估计问题就是如何利用样本对未知参数 θ 或 $g(\theta)$ 作出估计的问题。

3.1.1 参数估计问题

因此，参数估计问题就是如何利用样本对未知参数 θ 或 $g(\theta)$ 作出估计的问题。

例如， $X_1, \dots, X_n \text{ i.i.d. } \sim N(\mu, \sigma^2)$ ，记 $\theta = (\mu, \sigma^2)$ ，希望利用样本对 μ 和 σ^2 或其函数 $g(\theta) = \mu/\sigma^2$ 的值作出估计。

3.1.1 参数估计问题

有时总体分布族 $\mathcal{F} = \{F\}$ 是非参数分布族，其分布形式未知，我们对 F 的信息知之甚少（如只知道是否对称，是否连续等）。

3.1.1 参数估计问题

有时总体分布族 $\mathcal{F} = \{F\}$ 是非参数分布族，其分布形式未知，我们对 F 的信息知之甚少（如只知道是否对称，是否连续等）。

在非参数分布族场合，把参数理解为分布族 \mathcal{F} 上的泛函 $g(F)$ ，如 $g(F)$ 为总体均值、方差或中位数等，希望利用样本对 $g(F)$ 作出估计，这也属于参数估计问题。

3.1.1 参数估计问题

有时总体分布族 $\mathcal{F} = \{F\}$ 是非参数分布族，其分布形式未知，我们对 F 的信息知之甚少（如只知道是否对称，是否连续等）。

在非参数分布族场合，把参数理解为分布族 \mathcal{F} 上的泛函 $g(F)$ ，如 $g(F)$ 为总体均值、方差或中位数等，希望利用样本对 $g(F)$ 作出估计，这也属于参数估计问题。

例如，从某城市居民中抽取一部分，对其年收入作调查，获得样本 X_1, \dots, X_n ，要对该城市居民的年人均收入作出估计，就属于这类问题。

3.1.1 参数估计问题

参数估计 (parameter estimation) 问题常有两类：**点估计**和**区间估计**。

3.1.1 参数估计问题

参数估计 (parameter estimation) 问题常有两类：**点估计**和**区间估计**。

点估计就是用样本函数的一个具体数值去估计一个未知参数。

3.1.1 参数估计问题

参数估计 (parameter estimation) 问题常有两类：**点估计**和**区间估计**。

点估计就是用样本函数的一个具体数值去估计一个未知参数。

区间估计就是用样本函数的两个值构成的区间去估计未知参数的取值范围。

3.1.1 参数估计问题

例如，在某市居民年人均收入的调查中，估计该市居民的年人均收入为82500元，这是一个点估计。

3.1.1 参数估计问题

例如，在某市居民年人均收入的调查中，估计该市居民的年人均收入为82500元，这是一个点估计。

若估计年人均收入在63500元到98500元之间，这就是一个区间估计。

3.1.1 参数估计问题

例如，在某市居民年人均收入的调查中，估计该市居民的年人均收入为82500元，这是一个点估计。

若估计年人均收入在63500元到98500元之间，这就是一个区间估计。

点估计与区间估计是互为补充的参数估计形式。

3.1.2 点估计

定义 (3.1.1 点估计)

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从某个总体 F_θ 中抽取的样本。

3.1.2 点估计

定义 (3.1.1 点估计)

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从某个总体 F_θ 中抽取的样本。

$g(\theta)$ 为总体参数 θ 的实函数，用 $\hat{g}(\mathbf{X})$ 作为 $g(\theta)$ 估计，称为 **点估计** (point estimation)。

3.1.2 点估计

定义 (3.1.1 点估计)

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从某个总体 F_θ 中抽取的样本。

$g(\theta)$ 为总体参数 θ 的实函数，用 $\hat{g}(\mathbf{X})$ 作为 $g(\theta)$ 估计，称为 **点估计** (point estimation)。

此时 $\hat{g}(\mathbf{X})$ 称为 $g(\theta)$ 的 **估计量** (estimator)。

3.1.2 点估计

定义 (3.1.1 点估计)

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从某个总体 F_θ 中抽取的样本。

$g(\theta)$ 为总体参数 θ 的实函数，用 $\hat{g}(\mathbf{X})$ 作为 $g(\theta)$ 估计，称为 **点估计** (point estimation)。

此时 $\hat{g}(\mathbf{X})$ 称为 $g(\theta)$ 的 **估计量** (estimator)。

在一组具体样本值 $\mathbf{x} = (x_1, \dots, x_n)$ 下估计量的值 $\hat{g}(\mathbf{x})$ 称为 $g(\theta)$ 的 **估计值**。

3.1.2 点估计

定义 (3.1.1 点估计)

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从某个总体 F_θ 中抽取的样本。

$g(\theta)$ 为总体参数 θ 的实函数，用 $\hat{g}(\mathbf{X})$ 作为 $g(\theta)$ 估计，称为 **点估计** (point estimation)。

此时 $\hat{g}(\mathbf{X})$ 称为 $g(\theta)$ 的 **估计量** (estimator)。

在一组具体样本值 $\mathbf{x} = (x_1, \dots, x_n)$ 下估计量的值 $\hat{g}(\mathbf{x})$ 称为 $g(\theta)$ 的 **估计值**。

由此定义和样本的两重性知，估计量是样本的函数，是统计量，因而是随机变量。而估计值是估计量的一个实现，是具体数值。

3.1.2 点估计

例 (3.1.1)

对某种环氧树脂片的击穿电压（单位：kV / mm）进行20次观测得到

24.46, 25.61, 26.25, 26.42, 26.66, 27.15, 27.31, 27.54, 27.74, 27.94,
27.98, 28.04, 28.28, 28.49, 28.50, 28.87, 29.11, 29.13, 29.50, 30.88.

3.1.2 点估计

例 (3.1.1)

对某种环氧树脂片的击穿电压（单位：kV / mm）进行20次观测得到

24.46, 25.61, 26.25, 26.42, 26.66, 27.15, 27.31, 27.54, 27.74, 27.94,
27.98, 28.04, 28.28, 28.49, 28.50, 28.87, 29.11, 29.13, 29.50, 30.88.

有证据表明击穿电压值服从均值为 μ 的正态分布。讨论 μ 的点估计问题。

3.1.2 点估计

解

3.1.2 点估计

解

由于正态分布是对称的，均值 μ 也是该分布的中位数。给定的观测值是样本 X_1, \dots, X_{20} 的实现，则基于统计量和正态分布的特点，直观上 μ 的估计可以为

3.1.2 点估计

解

由于正态分布是对称的，均值 μ 也是该分布的中位数。给定的观测值是样本 X_1, \dots, X_{20} 的实现，则基于统计量和正态分布的特点，直观上 μ 的估计可以为

(1) 样本均值 \bar{X} ：估计值为 $\bar{x} = \sum_{i=1}^{20} x_i / 20 = 555.86 / 20 = 27.793$.

3.1.2 点估计

解

由于正态分布是对称的，均值 μ 也是该分布的中位数。给定的观测值是样本 X_1, \dots, X_{20} 的实现，则基于统计量和正态分布的特点，直观上 μ 的估计可以为

(1) 样本均值 \bar{X} ：估计值为 $\bar{x} = \sum_{i=1}^{20} x_i / 20 = 555.86 / 20 = 27.793$.

(2) 样本中位数 $m_{1/2}$ ：估计值为 $m_{1/2} = (27.94 + 27.98) / 2 = 27.960$.

3.1.2 点估计

解

由于正态分布是对称的，均值 μ 也是该分布的中位数。给定的观测值是样本 X_1, \dots, X_{20} 的实现，则基于统计量和正态分布的特点，直观上 μ 的估计可以为

(1) 样本均值 \bar{X} ：估计值为 $\bar{x} = \sum_{i=1}^{20} x_i / 20 = 555.86 / 20 = 27.793$.

(2) 样本中位数 $m_{1/2}$ ：估计值为 $m_{1/2} = (27.94 + 27.98) / 2 = 27.960$.

(3) 样本范围的中心 $(X_{(n)} + X_{(1)}) / 2$ ：估计值为

$$(24.46 + 30.88) / 2 = 27.670,$$

其中 $X_{(1)}$ 和 $X_{(n)}$ 为样本最小和最大次序统计量。

3.1.2 点估计

由此可见，有多种看来合理的参数估计的方法。这就产生了一个问题：
采用哪一个估计量作为 θ 的点估计较好呢？

3.1.2 点估计

由此可见，有多种看来合理的参数估计的方法。这就产生了一个问题：采用哪一个估计量作为 θ 的点估计较好呢？

这就涉及评价一个估计量优劣的标准问题。标准不同，回答也不同。

3.1.2 点估计

由此可见，有多种看来合理的参数估计的方法。这就产生了一个问题：采用哪一个估计量作为 θ 的点估计较好呢？

这就涉及评价一个估计量优劣的标准问题。标准不同，回答也不同。

在经典估计理论中，用来评价估计量好坏的标准有：**无偏性、有效性、相合性和渐近正态性等。**

3.1.2 点估计

点估计是数理统计学中内容很丰富的一个分支。

3.1.2 点估计

点估计是数理统计学中内容很丰富的一个分支。

它主要包括：

3.1.2 点估计

点估计是数理统计学中内容很丰富的一个分支。

它主要包括：

1. 导出估计量的一般方法（矩估计、极大似然法）；

3.1.2 点估计

点估计是数理统计学中内容很丰富的一个分支。

它主要包括：

1. 导出估计量的一般方法（矩估计、极大似然法）；
2. 制定评价估计量优良性的种种合理准则（无偏性、有效性、相合性）；

3.1.2 点估计

点估计是数理统计学中内容很丰富的一个分支。

它主要包括：

1. 导出估计量的一般方法（矩估计、极大似然法）；
2. 制定评价估计量优良性的种种合理准则（无偏性、有效性、相合性）；
3. 寻求某种特定准则下的最优估计量（UMVUE, C-R不等式）。

3.1.3 点估计的优良性准则

1. 无偏性

3.1.3 点估计的优良性准则

1. 无偏性

在评价估计量好坏时，一般总希望参数 θ 的估计量 $\hat{\theta}(\mathbf{X})$ 的平均值与 θ 越接近越好，即 $E[\hat{\theta}(\mathbf{X}) - \theta] = E[\hat{\theta}(\mathbf{X})] - \theta$ 越小越好。

3.1.3 点估计的优良性准则

1. 无偏性

在评价估计量好坏时，一般总希望参数 θ 的估计量 $\hat{\theta}(\mathbf{X})$ 的平均值与 θ 越接近越好，即 $E[\hat{\theta}(\mathbf{X}) - \theta] = E[\hat{\theta}(\mathbf{X})] - \theta$ 越小越好。

由于估计量 $\hat{\theta}(\mathbf{X})$ 是随机变量， $\hat{\theta}(\mathbf{X})$ 的值有时比 θ 的真值大，有时比 θ 的真值小，我们自然希望 $\hat{\theta}(\mathbf{X})$ 在大量重复使用时，在平均意义下 $\hat{\theta}(\mathbf{X})$ 与 θ 的偏差越小越好。

3.1.3 点估计的优良性准则

1. 无偏性

在评价估计量好坏时，一般总希望参数 θ 的估计量 $\hat{\theta}(\mathbf{X})$ 的平均值与 θ 越接近越好，即 $E[\hat{\theta}(\mathbf{X}) - \theta] = E[\hat{\theta}(\mathbf{X})] - \theta$ 越小越好。

由于估计量 $\hat{\theta}(\mathbf{X})$ 是随机变量， $\hat{\theta}(\mathbf{X})$ 的值有时比 θ 的真值大，有时比 θ 的真值小，我们自然希望 $\hat{\theta}(\mathbf{X})$ 在大量重复使用时，在平均意义下 $\hat{\theta}(\mathbf{X})$ 与 θ 的偏差越小越好。

当期望值 $E[\hat{\theta}(\mathbf{X}) - \theta] = 0$ 时就得到无偏性的概念。将其一般化，用 $g(\theta)$ 代替 θ ，用 $\hat{g}(\mathbf{X})$ 代替 $\hat{\theta}(\mathbf{X})$ ，得到如下定义。

3.1.3 点估计的优良性准则

定义 (3.1.2 (无偏估计))

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 $\{F_\theta, \theta \in \Theta\}$ 中抽取的样本, $g(\theta)$ 是定义于参数空间 Θ 上的已知函数。 $\hat{g}(\mathbf{X}) = \hat{g}(X_1, \dots, X_n)$ 是 $g(\theta)$ 的一个估计量, 如果

3.1.3 点估计的优良性准则

定义 (3.1.2 (无偏估计))

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 $\{F_\theta, \theta \in \Theta\}$ 中抽取的样本, $g(\theta)$ 是定义于参数空间 Θ 上的已知函数。 $\hat{g}(\mathbf{X}) = \hat{g}(X_1, \dots, X_n)$ 是 $g(\theta)$ 的一个估计量, 如果

$$E_\theta[\hat{g}(\mathbf{X})] = g(\theta), \theta \in \Theta,$$

则称 $\hat{g}(\mathbf{X})$ 为 $g(\theta)$ 的一个**无偏估计**。

3.1.3 点估计的优良性准则

定义 (3.1.2 (无偏估计))

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 $\{F_\theta, \theta \in \Theta\}$ 中抽取的样本, $g(\theta)$ 是定义于参数空间 Θ 上的已知函数。 $\hat{g}(\mathbf{X}) = \hat{g}(X_1, \dots, X_n)$ 是 $g(\theta)$ 的一个估计量, 如果

$$E_\theta[\hat{g}(\mathbf{X})] = g(\theta), \theta \in \Theta,$$

则称 $\hat{g}(\mathbf{X})$ 为 $g(\theta)$ 的一个**无偏估计**。

记 $\hat{g}(\mathbf{X}) = \hat{g}_n(\mathbf{X})$, 若 $E_\theta[\hat{g}_n(\mathbf{X})] \neq g(\theta)$, 但

3.1.3 点估计的优良性准则

定义 (3.1.2 (无偏估计))

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 $\{F_\theta, \theta \in \Theta\}$ 中抽取的样本, $g(\theta)$ 是定义于参数空间 Θ 上的已知函数。 $\hat{g}(\mathbf{X}) = \hat{g}(X_1, \dots, X_n)$ 是 $g(\theta)$ 的一个估计量, 如果

$$E_\theta[\hat{g}(\mathbf{X})] = g(\theta), \theta \in \Theta,$$

则称 $\hat{g}(\mathbf{X})$ 为 $g(\theta)$ 的一个**无偏估计**。

记 $\hat{g}(\mathbf{X}) = \hat{g}_n(\mathbf{X})$, 若 $E_\theta[\hat{g}_n(\mathbf{X})] \neq g(\theta)$, 但

$$\lim_{n \rightarrow \infty} E_\theta[\hat{g}_n(\mathbf{X})] = g(\theta), \theta \in \Theta,$$

则称 $\hat{g}_n(\mathbf{X})$ 为 $g(\theta)$ 的**渐近无偏估计**。

3.1.3 点估计的优良性准则

注意：参数空间 Θ 中的每个 θ 对应的分布 F_θ 都可能是产生样本的“真实”分布，因此在无偏估计的定义中要求等式对所有的参数 $\theta \in \Theta$ 成立。

3.1.3 点估计的优良性准则

无偏性的含义有两个：

3.1.3 点估计的优良性准则

无偏性的含义有两个：

(a) **第一个含义是无系统性偏差**。由于样本的随机性， $\hat{g}(X)$ 是样本 X 的函数，因此它是一个随机变量。

3.1.3 点估计的优良性准则

无偏性的含义有两个：

(a) **第一个含义是无系统性偏差**。由于样本的随机性， $\hat{g}(X)$ 是样本 X 的函数，因此它是一个随机变量。

用估计量 $\hat{g}(X)$ 去估计 $g(\theta)$ ，对某些样本 x ， $\hat{g}(x)$ 与 $g(\theta)$ 相比偏低；对另一些样本 x ， $\hat{g}(x)$ 偏高。

3.1.3 点估计的优良性准则

无偏性的含义有两个：

(a) **第一个含义是无系统性偏差**。由于样本的随机性， $\hat{g}(X)$ 是样本 X 的函数，因此它是一个随机变量。

用估计量 $\hat{g}(X)$ 去估计 $g(\theta)$ ，对某些样本 x ， $\hat{g}(x)$ 与 $g(\theta)$ 相比偏低；对另一些样本 x ， $\hat{g}(x)$ 偏高。

无偏性表明，把这些正负偏差在概率上平均起来，其值为0。

3.1.3 点估计的优良性准则

例如，用一杆秤去称东西，**误差的来源**有两个：

3.1.3 点估计的优良性准则

例如，用一杆秤去称东西，**误差的来源**有两个：

(1) 这杆秤自身结构上有问题，用它称东西总是倾向于偏高或总是倾向于偏低，这属于**系统误差**；

3.1.3 点估计的优良性准则

例如，用一杆秤去称东西，**误差的来源**有两个：

(1) 这杆秤自身结构上有问题，用它称东西总是倾向于偏高或总是倾向于偏低，这属于**系统误差**；

(2) 另一种误差是随机误差，由不可控制的因素产生，如温度、湿度和工作人员心理波动等影响造成的，这属于**随机误差**。

3.1.3 点估计的优良性准则

例如，用一杆秤去称东西，**误差的来源**有两个：

(1) 这杆秤自身结构上有问题，用它称东西总是倾向于偏高或总是倾向于偏低，这属于**系统误差**；

(2) 另一种误差是随机误差，由不可控制的因素产生，如温度、湿度和工作人员心理波动等影响造成的，这属于**随机误差**。

无偏性相当于要求无系统性误差。而随机误差总是存在的，大量重复使用无偏估计量，误差有时为正，有时为负。但是，随机误差可以在大量重复使用中正负相抵消。

3.1.3 点估计的优良性准则

(b) **无偏性的另一个含义是：**要求估计量大量重复使用，在多次重复使用下给出接近真值 $g(\theta)$ 的估计（**大样本性质**）。

3.1.3 点估计的优良性准则

(b) **无偏性的另一个含义是**：要求估计量大量重复使用，在多次重复使用下给出接近真值 $g(\theta)$ 的估计（**大样本性质**）。

设想这样一种情况：每天抽样对 $g(\theta)$ 进行估计，第 i 天的样本为 $\mathbf{X}^{(i)} = (X_1^{(i)}, \dots, X_n^{(i)})$ ，估计值为 $\hat{g}(\mathbf{X}^{(i)})$ ，一共作了 m 天。

3.1.3 点估计的优良性准则

(b) **无偏性的另一个含义是**：要求估计量大量重复使用，在多次重复使用下给出接近真值 $g(\theta)$ 的估计（**大样本性质**）。

设想这样一种情况：每天抽样对 $g(\theta)$ 进行估计，第 i 天的样本为 $\mathbf{X}^{(i)} = (X_1^{(i)}, \dots, X_n^{(i)})$ ，估计值为 $\hat{g}(\mathbf{X}^{(i)})$ ，一共作了 m 天。

设 $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}$ 是独立同分布的样本，如 $\hat{g}(\mathbf{X})$ 有无偏性，按大数定律有

3.1.3 点估计的优良性准则

(b) **无偏性的另一个含义是**：要求估计量大量重复使用，在多次重复使用下给出接近真值 $g(\theta)$ 的估计（**大样本性质**）。

设想这样一种情况：每天抽样对 $g(\theta)$ 进行估计，第 i 天的样本为 $\mathbf{X}^{(i)} = (X_1^{(i)}, \dots, X_n^{(i)})$ ，估计值为 $\hat{g}(\mathbf{X}^{(i)})$ ，一共作了 m 天。

设 $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}$ 是独立同分布的样本，如 $\hat{g}(\mathbf{X})$ 有无偏性，按大数定律有

$$P \left(\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \hat{g}(\mathbf{X}^{(i)}) = g(\theta) \right) = 1.$$

3.1.3 点估计的优良性准则

也就是说，尽管一次估计结果 $\hat{g}(X^{(i)})$ 不一定恰好等于 $g(\theta)$ ，但在大量重复使用时，多次估计的算术平均值，可以任意接近 $g(\theta)$ 。

3.1.3 点估计的优良性准则

也就是说，尽管一次估计结果 $\hat{g}(\mathbf{X}^{(i)})$ 不一定恰好等于 $g(\theta)$ ，但在大量重复使用时，多次估计的算术平均值，可以任意接近 $g(\theta)$ 。

如果这一估计量 $\hat{g}(\mathbf{X})$ 只使用一次，无偏性这个概念就失去意义。

3.1.3 点估计的优良性准则

不妨设想一个例子。某工厂生产一种产品，从长期来看废品率稳定在 p_0 上。设商店每日从该工厂进货一批，每批 N 件，每件单价 a 元。

3.1.3 点估计的优良性准则

不妨设想一个例子。某工厂生产一种产品，从长期来看废品率稳定在 p_0 上。设商店每日从该工厂进货一批，每批 N 件，每件单价 a 元。

双方议定：若某日抽样中 p_0 的估计值为 \hat{p} ，则商店付给工厂 $N(1 - \hat{p})a$ 元。

3.1.3 点估计的优良性准则

不妨设想一个例子。某工厂生产一种产品，从长期来看废品率稳定在 p_0 上。设商店每日从该工厂进货一批，每批 N 件，每件单价 a 元。

双方议定：若某日抽样中 p_0 的估计值为 \hat{p} ，则商店付给工厂 $N(1 - \hat{p})a$ 元。

这时，就一日的情况而言， \hat{p} 相对 p_0 可能偏高，也可能偏低，因而有一方要吃一点亏。

3.1.3 点估计的优良性准则

但从长远看，若 \hat{p} 为 p_0 的无偏估计，则平均来说哪一方也不吃亏。

3.1.3 点估计的优良性准则

但从长远看，若 \hat{p} 为 p_0 的无偏估计，则平均来说哪一方也不吃亏。

无偏性保证从长远来看是公平的。如果应用中这种经常的重复使用性不存在，无偏性也就失去意义。

3.1.3 点估计的优良性准则

例 (3.1.2)

设 X_1, \dots, X_n 是取自期望为 μ , 方差为 σ^2 的总体的一个简单样本。显然样本均值 \bar{X} 是 μ 的无偏估计。证明样本方差

$$S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n - 1)$$

是 σ^2 的无偏估计。

3.1.3 点估计的优良性准则

证

由题知, $E(X_1) = E(\bar{X}) = \mu$, $Var(X_1) = \sigma^2$,

$$Var(\bar{X}) = Var\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{Var(\sum_{i=1}^n X_i)}{n^2} = \frac{Var(X_1)}{n} = \frac{\sigma^2}{n}, \text{ 所以有}$$

3.1.3 点估计的优良性准则

证

由题知, $E(X_1) = E(\bar{X}) = \mu$, $Var(X_1) = \sigma^2$,

$Var(\bar{X}) = Var\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{Var(\sum_{i=1}^n X_i)}{n^2} = \frac{Var(X_1)}{n} = \frac{\sigma^2}{n}$, 所以有

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

=

3.1.3 点估计的优良性准则

证

由题知, $E(X_1) = E(\bar{X}) = \mu$, $Var(X_1) = \sigma^2$,

$Var(\bar{X}) = Var\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{Var(\sum_{i=1}^n X_i)}{n^2} = \frac{Var(X_1)}{n} = \frac{\sigma^2}{n}$, 所以有

$$\begin{aligned} E(S^2) &= E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{1}{n-1} E\left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2)\right) = \frac{1}{n-1} (nE(X_1^2) - nE(\bar{X}^2)) \end{aligned}$$

3.1.3 点估计的优良性准则

$$= \frac{n}{n-1} [(Var(X_1) + (EX_1)^2) - (Var(\bar{X}) + (E\bar{X})^2)]$$
$$=$$

3.1.3 点估计的优良性准则

$$\begin{aligned} &= \frac{n}{n-1} [(Var(X_1) + (EX_1)^2) - (Var(\bar{X}) + (E\bar{X})^2)] \\ &= \frac{n}{n-1} [(\sigma^2 + \mu^2) - (\sigma^2/n + \mu^2)] = \sigma^2. \end{aligned}$$

3.1.3 点估计的优良性准则

$$\begin{aligned} &= \frac{n}{n-1} [(Var(X_1) + (EX_1)^2) - (Var(\bar{X}) + (E\bar{X})^2)] \\ &= \frac{n}{n-1} [(\sigma^2 + \mu^2) - (\sigma^2/n + \mu^2)] = \sigma^2. \end{aligned}$$

故样本方差 S^2 是 σ^2 的无偏估计。

□

3.1.3 点估计的优良性准则

2. 有效性

3.1.3 点估计的优良性准则

2. 有效性

在应用中，同一个参数的无偏估计常常不止一个，那么选用哪一个无偏估计更好呢？

3.1.3 点估计的优良性准则

2. 有效性

在应用中，同一个参数的无偏估计常常不止一个，那么选用哪一个无偏估计更好呢？

为了解决好这一问题，就要讨论估计量的有效性。

3.1.3 点估计的优良性准则

2. 有效性

在应用中，同一个参数的无偏估计常常不止一个，那么选用哪一个无偏估计更好呢？

为了解决好这一问题，就要讨论估计量的有效性。

设 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 为 θ 的两个无偏估计，由无偏性可知它们的一阶原点矩相等，比较它们的二阶中心距——方差，方差越小越好。

3.1.3 点估计的优良性准则

例如，到商店购买电视机，看中了其中的两种品牌，它们分别由甲、乙两厂生产，外观、音质和画面都不错。

3.1.3 点估计的优良性准则

例如，到商店购买电视机，看中了其中的两种品牌，它们分别由甲、乙两厂生产，外观、音质和画面都不错。

根据市场调查，甲、乙两厂生产的这两种电视机平均使用寿命相同，都是20年。

3.1.3 点估计的优良性准则

例如，到商店购买电视机，看中了其中的两种品牌，它们分别由甲、乙两厂生产，外观、音质和画面都不错。

根据市场调查，甲、乙两厂生产的这两种电视机平均使用寿命相同，都是20年。

甲厂生产的电视机质量较稳定，最低使用寿命有18年，最高可使用22年；而乙厂生产的电视机稳定性要差一些，最差的使用10年就坏了，但最好的可使用30年。

3.1.3 点估计的优良性准则

例如，到商店购买电视机，看中了其中的两种品牌，它们分别由甲、乙两厂生产，外观、音质和画面都不错。

根据市场调查，甲、乙两厂生产的这两种电视机平均使用寿命相同，都是20年。

甲厂生产的电视机质量较稳定，最低使用寿命有18年，最高可使用22年；而乙厂生产的电视机稳定性要差一些，最差的使用10年就坏了，但最好的可使用30年。

应该购买哪一个厂家生产的电视机呢？

3.1.3 点估计的优良性准则

若将电视机的使用寿命视为随机变量，甲、乙两厂生产的电视机使用寿命的均值相等，但乙厂的质量不稳定，即方差较大。

3.1.3 点估计的优良性准则

若将电视机的使用寿命视为随机变量，甲、乙两厂生产的电视机使用寿命的均值相等，但乙厂的质量不稳定，即方差较大。

从稳健的角度出发，显然愿意购买甲厂生产的电视机，其风险较小，即方差较小，质量稳定。

3.1.3 点估计的优良性准则

定义 (3.1.3 (有效性))

设 $\hat{g}_1(\mathbf{X}) = \hat{g}_1(X_1, \dots, X_n)$ 和 $\hat{g}_2(\mathbf{X}) = \hat{g}_2(X_1, \dots, X_n)$ 为 $g(\theta)$ 的两个不同无偏估计量, 若

$$\text{Var}_\theta(\hat{g}_1(\mathbf{X})) \leq \text{Var}_\theta(\hat{g}_2(\mathbf{X})), \text{ 一切 } \theta \in \Theta,$$

且至少存在一个 $\theta \in \Theta$, 使得严格不等号成立, 则称估计量 $\hat{g}_1(\mathbf{X})$ 比 $\hat{g}_2(\mathbf{X})$ 有效。

3.1.3 点估计的优良性准则

从这个定义出发可以看出，在均值相等的条件下，方差越小的估计量越有效。

3.1.3 点估计的优良性准则

从这个定义出发可以看出，在均值相等的条件下，方差越小的估计量越有效。

例如， X_1, \dots, X_n 是取自总体 F 的一个简单样本，设总体均值 μ 和总体方差 σ^2 都存在，则 $\hat{\theta}_1 = X_1$ 和 $\hat{\theta}_2 = \bar{X}$ 都是总体均值 μ 的无偏估计量，它们的方差分别是

3.1.3 点估计的优良性准则

从这个定义出发可以看出，在均值相等的条件下，方差越小的估计量越有效。

例如， X_1, \dots, X_n 是取自总体 F 的一个简单样本，设总体均值 μ 和总体方差 σ^2 都存在，则 $\hat{\theta}_1 = X_1$ 和 $\hat{\theta}_2 = \bar{X}$ 都是总体均值 μ 的无偏估计量，它们的方差分别是

$$\text{Var}(\hat{\theta}_1) = \sigma^2, \text{Var}(\hat{\theta}_2) = \frac{1}{n}\sigma^2.$$

3.1.3 点估计的优良性准则

从这个定义出发可以看出，在均值相等的条件下，方差越小的估计量越有效。

例如， X_1, \dots, X_n 是取自总体 F 的一个简单样本，设总体均值 μ 和总体方差 σ^2 都存在，则 $\hat{\theta}_1 = X_1$ 和 $\hat{\theta}_2 = \bar{X}$ 都是总体均值 μ 的无偏估计量，它们的方差分别是

$$\text{Var}(\hat{\theta}_1) = \sigma^2, \text{Var}(\hat{\theta}_2) = \frac{1}{n}\sigma^2.$$

后者方差更小，可见 \bar{X} 比 X_1 更有效。且 n 越大， \bar{X} 对 μ 的估计就越有效。

3.1.3 点估计的优良性准则

从这个定义出发可以看出，在均值相等的条件下，方差越小的估计量越有效。

例如， X_1, \dots, X_n 是取自总体 F 的一个简单样本，设总体均值 μ 和总体方差 σ^2 都存在，则 $\hat{\theta}_1 = X_1$ 和 $\hat{\theta}_2 = \bar{X}$ 都是总体均值 μ 的无偏估计量，它们的方差分别是

$$\text{Var}(\hat{\theta}_1) = \sigma^2, \text{Var}(\hat{\theta}_2) = \frac{1}{n}\sigma^2.$$

后者方差更小，可见 \bar{X} 比 X_1 更有效。且 n 越大， \bar{X} 对 μ 的估计就越有效。这就是在物体的称重问题中，为什么要将物体称 n 次，用其平均值作为物重的理由。

3.1.3 点估计的优良性准则

3. 相合性

3.1.3 点估计的优良性准则

3. 相合性

大量实践表明，随着样本容量 n 的增加，估计量 $\hat{g}(\mathbf{X}) = \hat{g}(X_1, \dots, X_n)$ 与被估计参数 $g(\theta)$ 的偏差越来越小，这是一个良好估计量应具有的性质。

3.1.3 点估计的优良性准则

3. 相合性

大量实践表明，随着样本容量 n 的增加，估计量 $\hat{g}(\mathbf{X}) = \hat{g}(X_1, \dots, X_n)$ 与被估计参数 $g(\theta)$ 的偏差越来越小，这是一个良好估计量应具有的性质。

试想，若不然，无论做多少次试验，也不能把 $g(\theta)$ 估计到任意指定的精确程度，这样的估计量显然是不可取的。

3.1.3 点估计的优良性准则

定义 (3.1.4 (相合性))

设对每个自然数 n , $\hat{g}_n(\mathbf{X}) = \hat{g}_n(X_1, \dots, X_n)$ 是 $g(\theta)$ 一个估计量, 若 $\hat{g}_n(\mathbf{X})$ 依概率收敛到 $g(\theta)$, 即对任何 $\theta \in \Theta$ 及 $\epsilon > 0$ 有

3.1.3 点估计的优良性准则

定义 (3.1.4 (相合性))

设对每个自然数 n , $\hat{g}_n(\mathbf{X}) = \hat{g}_n(X_1, \dots, X_n)$ 是 $g(\theta)$ 一个估计量, 若 $\hat{g}_n(\mathbf{X})$ 依概率收敛到 $g(\theta)$, 即对任何 $\theta \in \Theta$ 及 $\epsilon > 0$ 有

$$\lim_{n \rightarrow \infty} P_{\theta}(|\hat{g}_n(\mathbf{X}) - g(\theta)| \geq \epsilon) = 0,$$

3.1.3 点估计的优良性准则

定义 (3.1.4 (相合性))

设对每个自然数 n , $\hat{g}_n(\mathbf{X}) = \hat{g}_n(X_1, \dots, X_n)$ 是 $g(\theta)$ 一个估计量, 若 $\hat{g}_n(\mathbf{X})$ 依概率收敛到 $g(\theta)$, 即对任何 $\theta \in \Theta$ 及 $\epsilon > 0$ 有

$$\lim_{n \rightarrow \infty} P_{\theta}(|\hat{g}_n(\mathbf{X}) - g(\theta)| \geq \epsilon) = 0,$$

则称 $\hat{g}_n(\mathbf{X})$ 为 $g(\theta)$ 的**弱相合估计**。

3.1.3 点估计的优良性准则

定义 (3.1.4 (相合性) 续)

若对任何 $\theta \in \Theta$ 有

3.1.3 点估计的优良性准则

定义 (3.1.4 (相合性) 续)

若对任何 $\theta \in \Theta$ 有

$$P_{\theta} \left(\lim_{n \rightarrow \infty} \hat{g}_n(\mathbf{X}) = g(\theta) \right) = 1,$$

3.1.3 点估计的优良性准则

定义 (3.1.4 (相合性) 续)

若对任何 $\theta \in \Theta$ 有

$$P_{\theta} \left(\lim_{n \rightarrow \infty} \hat{g}_n(\mathbf{X}) = g(\theta) \right) = 1,$$

则称 $\hat{g}_n(\mathbf{X})$ 为 $g(\theta)$ 的**强相合估计**。

3.1.3 点估计的优良性准则

定义 (3.1.4 (相合性) 续)

若 $r > 0$ 和对任何 $\theta \in \Theta$, 有

3.1.3 点估计的优良性准则

定义 (3.1.4 (相合性) 续)

若 $r > 0$ 和对任何 $\theta \in \Theta$, 有

$$\lim_{n \rightarrow \infty} E_{\theta} |\hat{g}_n(\mathbf{X}) - g(\theta)|^r = 0,$$

3.1.3 点估计的优良性准则

定义 (3.1.4 (相合性) 续)

若 $r > 0$ 和对任何 $\theta \in \Theta$, 有

$$\lim_{n \rightarrow \infty} E_{\theta} |\hat{g}_n(\mathbf{X}) - g(\theta)|^r = 0,$$

则称 $\hat{g}_n(\mathbf{X})$ 为 $g(\theta)$ 的 r 阶矩相合估计。当 $r = 2$ 时称为均方相合估计。

3.1.3 点估计的优良性准则

估计量的相合性是对大样本问题提出的要求，是估计量的一种大样本性质。

3.1.3 点估计的优良性准则

估计量的相合性是对大样本问题提出的要求，是估计量的一种**大样本性质**。

由概率论中关于这几种收敛性的关系，可知上述三种相合性有如下关系：

3.1.3 点估计的优良性准则

估计量的相合性是对大样本问题提出的要求，是估计量的一种**大样本性质**。

由概率论中关于这几种收敛性的关系，可知上述三种相合性有如下关系：

(1) 强相合 \Rightarrow 弱相合，反之不必对；

3.1.3 点估计的优良性准则

估计量的相合性是对大样本问题提出的要求，是估计量的一种**大样本性质**。

由概率论中关于这几种收敛性的关系，可知上述三种相合性有如下关系：

- (1) 强相合 \Rightarrow 弱相合，反之不必对；
- (2) 对任何 $r > 0$ 有： r 阶矩相合 \Rightarrow 弱相合，反之不必对；

3.1.3 点估计的优良性准则

估计量的相合性是对大样本问题提出的要求，是估计量的一种**大样本性质**。

由概率论中关于这几种收敛性的关系，可知上述三种相合性有如下关系：

- (1) 强相合 \Rightarrow 弱相合，反之不必对；
- (2) 对任何 $r > 0$ 有： r 阶矩相合 \Rightarrow 弱相合，反之不必对；
- (3) 强相合与 r 阶矩相合之间没有包含关系。

3.1.3 点估计的优良性准则

例 (3.1.3)

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从均匀分布 $U(0, \theta)$ 中抽取的简单样本, θ 为未知参数。证明 $T(\mathbf{X}) = (\prod_{i=1}^n X_i)^{1/n}$ 是 $g(\theta) = \theta e^{-1}$ 的强相合估计。

3.1.3 点估计的优良性准则

例 (3.1.3)

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从均匀分布 $U(0, \theta)$ 中抽取的简单样本, θ 为未知参数。证明 $T(\mathbf{X}) = (\prod_{i=1}^n X_i)^{1/n}$ 是 $g(\theta) = \theta e^{-1}$ 的强相合估计。

证

已知 X_i 独立同分布于均匀分布 $U(0, \theta)$, 则 X_i 的密度函数为

$$f(x; \theta) = \frac{1}{\theta} I_{\{x < \theta\}}.$$

3.1.3 点估计的优良性准则

令 $Y_i = \ln X_i$, $i = 1, \dots, n$, 则 $X_i = e^{Y_i}$, $|J| = \left| \frac{\partial x_i}{\partial y_i} \right| = e^{y_i}$,

且 Y_1, \dots, Y_n 相互独立, 具有下列共同的分布

3.1.3 点估计的优良性准则

令 $Y_i = \ln X_i$, $i = 1, \dots, n$, 则 $X_i = e^{Y_i}$, $|J| = \left| \frac{\partial x_i}{\partial y_i} \right| = e^{y_i}$,
且 Y_1, \dots, Y_n 相互独立, 具有下列共同的分布

$$f(y; \theta) = \frac{1}{\theta} |J| I_{(-\infty, \ln \theta)}(y) = \frac{1}{\theta} e^y I_{(-\infty, \ln \theta)}(y).$$

3.1.3 点估计的优良性准则

令 $Y_i = \ln X_i$, $i = 1, \dots, n$, 则 $X_i = e^{Y_i}$, $|J| = \left| \frac{\partial x_i}{\partial y_i} \right| = e^{y_i}$,
且 Y_1, \dots, Y_n 相互独立, 具有下列共同的分布

$$f(y; \theta) = \frac{1}{\theta} |J| I_{(-\infty, \ln \theta)}(y) = \frac{1}{\theta} e^y I_{(-\infty, \ln \theta)}(y).$$

Y_i 的期望为

3.1.3 点估计的优良性准则

令 $Y_i = \ln X_i$, $i = 1, \dots, n$, 则 $X_i = e^{Y_i}$, $|J| = \left| \frac{\partial x_i}{\partial y_i} \right| = e^{y_i}$,
且 Y_1, \dots, Y_n 相互独立, 具有下列共同的分布

$$f(y; \theta) = \frac{1}{\theta} |J| I_{(-\infty, \ln \theta)}(y) = \frac{1}{\theta} e^y I_{(-\infty, \ln \theta)}(y).$$

Y_i 的期望为

$$E(Y_1) = \frac{1}{\theta} \int_{-\infty}^{\ln \theta} y e^y dy = \frac{1}{\theta} \left[y e^y \Big|_{-\infty}^{\ln \theta} - \int_{-\infty}^{\ln \theta} e^y dy \right] = \ln \theta - 1.$$

3.1.3 点估计的优良性准则

故由Kolmogorov强大数定律可知

3.1.3 点估计的优良性准则

故由Kolmogorov强大数定律可知

$$\ln T(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n \ln X_i = \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y} \xrightarrow{a.s.} E(Y_1) = \ln \theta - 1.$$

3.1.3 点估计的优良性准则

故由Kolmogorov强大数定律可知

$$\ln T(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n \ln X_i = \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y} \xrightarrow{a.s.} E(Y_1) = \ln \theta - 1.$$

因此有

3.1.3 点估计的优良性准则

故由Kolmogorov强大数定律可知

$$\ln T(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n \ln X_i = \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y} \xrightarrow{a.s.} E(Y_1) = \ln \theta - 1.$$

因此有

$$T(\mathbf{X}) = e^{\bar{Y}} \xrightarrow{a.s.} e^{\ln \theta - 1} = \theta \times e^{-1}.$$

3.1.3 点估计的优良性准则

故由Kolmogorov强大数定律可知

$$\ln T(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n \ln X_i = \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y} \xrightarrow{a.s.} E(Y_1) = \ln \theta - 1.$$

因此有

$$T(\mathbf{X}) = e^{\bar{Y}} \xrightarrow{a.s.} e^{\ln \theta - 1} = \theta \times e^{-1}.$$

所以 $T(\mathbf{X}) = (\prod_{i=1}^n X_i)^{1/n}$ 是 $g(\theta) = \theta e^{-1}$ 的强相合估计。 □

1 3.1 引言

- 3.1.1 参数估计问题
- 3.1.2 点估计
- 3.1.3 点估计的优良性准则

2 3.2 矩估计

- 3.2.1 矩法和矩估计量
- 3.2.2 若干例子

3.2.1 矩法和矩估计量

设 X_1, \dots, X_n 是从总体 F 中抽取的简单样本。这时，样本矩可用来估计总体 F 相应的总体矩。

3.2.1 矩法和矩估计量

设 X_1, \dots, X_n 是从总体 F 中抽取的简单样本。这时，样本矩可用来估计总体 F 相应的总体矩。

即样本 k 阶原点矩

3.2.1 矩法和矩估计量

设 X_1, \dots, X_n 是从总体 F 中抽取的简单样本。这时，样本矩可用来估计总体 F 相应的总体矩。

即样本 k 阶原点矩

$$a_{nk} = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k = 1, 2, \dots$$

3.2.1 矩法和矩估计量

设 X_1, \dots, X_n 是从总体 F 中抽取的简单样本。这时，样本矩可用来估计总体 F 相应的总体矩。

即样本 k 阶原点矩

$$a_{nk} = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k = 1, 2, \dots$$

是总体 k 阶原点矩

$$a_k = E(X^k)$$

的“自然”的矩估计量。

3.2.1 矩法和矩估计量

特别的，总体均值

$$a_1 = E(X)$$

的“自然”的矩估计量是样本均值

$$a_{n1} = \bar{X}.$$

3.2.1 矩法和矩估计量

样本 k 阶中心距

3.2.1 矩法和矩估计量

样本 k 阶中心距

$$m_{nk} = \frac{1}{n} \sum_{i=1}^k (X_i - \bar{X})^k, \quad k = 2, 3, \dots$$

3.2.1 矩法和矩估计量

样本 k 阶中心距

$$m_{nk} = \frac{1}{n} \sum_{i=1}^k (X_i - \bar{X})^k, \quad k = 2, 3, \dots$$

是总体 k 阶中心矩

$$\mu_k = E[X - E(X)]^k$$

的“自然”的矩估计量。

3.2.1 矩法和矩估计量

特别的，总体方差

$$\mu_2 = E[X - E(X)]^2$$

的“自然”的矩估计量是样本二阶中心矩

3.2.1 矩法和矩估计量

特别的，总体方差

$$\mu_2 = E[X - E(X)]^2$$

的“自然”的矩估计量是样本二阶中心矩

$$m_{n2} = S_n^2 = \sum_{i=1}^n (X_i - \bar{X})^2/n,$$

3.2.1 矩法和矩估计量

特别的，总体方差

$$\mu_2 = E[X - E(X)]^2$$

的“自然”的矩估计量是样本二阶中心矩

$$m_{n2} = S_n^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / n,$$

它与样本方差 $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$ 只差一个常数因子。

3.2.1 矩法和矩估计量

用 a_{nk} , m_{nk} 分别估计 α_k 和 μ_k 是一种基于直观的方法, 它的依据是 a_{nk} 是 α_k 的无偏估计, 即

3.2.1 矩法和矩估计量

用 a_{nk} , m_{nk} 分别估计 α_k 和 μ_k 是一种基于直观的方法, 它的依据是 a_{nk} 是 α_k 的无偏估计, 即

$$E(a_{nk}) = \frac{1}{n} \sum_{i=1}^n E(X_i^k) = \frac{1}{n} \sum_{i=1}^n \alpha_k = \alpha_k.$$

3.2.1 矩法和矩估计量

用 a_{nk} , m_{nk} 分别估计 α_k 和 μ_k 是一种基于直观的方法, 它的依据是 a_{nk} 是 α_k 的无偏估计, 即

$$E(a_{nk}) = \frac{1}{n} \sum_{i=1}^n E(X_i^k) = \frac{1}{n} \sum_{i=1}^n \alpha_k = \alpha_k.$$

但 m_{nk} 一般不是 μ_k 的无偏估计。

3.2.1 矩法和矩估计量

用 a_{nk} , m_{nk} 分别估计 α_k 和 μ_k 是一种基于直观的方法, 它的依据是 a_{nk} 是 α_k 的无偏估计, 即

$$E(a_{nk}) = \frac{1}{n} \sum_{i=1}^n E(X_i^k) = \frac{1}{n} \sum_{i=1}^n \alpha_k = \alpha_k.$$

但 m_{nk} 一般不是 μ_k 的无偏估计。

当样本大小 n 较大时, 偏差不显著, 且必要时可作一些修正, 使之成为无偏估计。

3.2.1 矩法和矩估计量

例 (3.2.1)

设 $\mu_2 = \sigma^2$ 是总体 X 的方差，令 X_1, \dots, X_n *i.i.d.* $\sim X$ ，则 $S_n^2 = m_{n2}$ 不是 σ^2 的无偏估计。

3.2.1 矩法和矩估计量

例 (3.2.1)

设 $\mu_2 = \sigma^2$ 是总体 X 的方差, 令 X_1, \dots, X_n *i.i.d.* $\sim X$, 则 $S_n^2 = m_{n2}$ 不是 σ^2 的无偏估计。

证

由例3.1.2可知

3.2.1 矩法和矩估计量

例 (3.2.1)

设 $\mu_2 = \sigma^2$ 是总体 X 的方差, 令 X_1, \dots, X_n *i.i.d.* $\sim X$, 则 $S_n^2 = m_{n2}$ 不是 σ^2 的无偏估计。

证

由例3.1.2可知

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

=

3.2.1 矩法和矩估计量

例 (3.2.1)

设 $\mu_2 = \sigma^2$ 是总体 X 的方差, 令 X_1, \dots, X_n *i.i.d.* $\sim X$, 则 $S_n^2 = m_{n2}$ 不是 σ^2 的无偏估计。

证

由例3.1.2可知

$$\begin{aligned} E(S_n^2) &= E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{n-1}{n} E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) \\ &= \frac{n-1}{n} E(S^2) = \frac{n-1}{n} \sigma^2. \end{aligned}$$

3.2.1 矩法和矩估计量

因而 m_{n2} 不是 σ^2 的无偏估计，且是系统性地偏低。将其修正，只需用

3.2.1 矩法和矩估计量

因而 m_{n2} 不是 σ^2 的无偏估计，且是系统性地偏低。将其修正，只需用

$$m_{n2}^* = \frac{n}{n-1} m_{n2} = \frac{n}{n-1} S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = S^2$$

代替 m_{n2} ，就得到 $E(m_{n2}^*) = E(S^2) = \sigma^2$ ，即 S^2 为总体方差的无偏估计。

□

3.2.1 矩法和矩估计量

一般的，样本的 k 阶中心矩可以用样本的原点矩表示（令 $a_{n0} = 1$ ）。

3.2.1 矩法和矩估计量

一般的，样本的 k 阶中心矩可以用样本的原点矩表示（令 $a_{n0} = 1$ ）。

$$m_{nk} = \frac{1}{n} \sum_{i=1}^n (X_i - a_{n1})^k$$

=

3.2.1 矩法和矩估计量

一般的，样本的 k 阶中心矩可以用样本的原点矩表示（令 $a_{n0} = 1$ ）。

$$\begin{aligned}m_{nk} &= \frac{1}{n} \sum_{i=1}^n (X_i - a_{n1})^k \\&= \frac{1}{n} \sum_{i=1}^n \sum_{r=0}^k \binom{k}{r} X_i^r (-1)^{k-r} a_{n1}^{k-r} \\&= \end{aligned}$$

3.2.1 矩法和矩估计量

一般的，样本的 k 阶中心矩可以用样本的原点矩表示（令 $a_{n0} = 1$ ）。

$$\begin{aligned}m_{nk} &= \frac{1}{n} \sum_{i=1}^n (X_i - a_{n1})^k \\&= \frac{1}{n} \sum_{i=1}^n \sum_{r=0}^k \binom{k}{r} X_i^r (-1)^{k-r} a_{n1}^{k-r} \\&= \sum_{r=0}^k (-1)^{k-r} \binom{k}{r} \left(\underbrace{\frac{1}{n} \sum_{i=1}^n X_i^r}_{a_{nr}} \right) a_{n1}^{k-r} \\&= \end{aligned}$$

3.2.1 矩法和矩估计量

一般的，样本的 k 阶中心矩可以用样本的原点矩表示（令 $a_{n0} = 1$ ）。

$$\begin{aligned}m_{nk} &= \frac{1}{n} \sum_{i=1}^n (X_i - a_{n1})^k \\&= \frac{1}{n} \sum_{i=1}^n \sum_{r=0}^k \binom{k}{r} X_i^r (-1)^{k-r} a_{n1}^{k-r} \\&= \sum_{r=0}^k (-1)^{k-r} \binom{k}{r} \left(\underbrace{\frac{1}{n} \sum_{i=1}^n X_i^r}_{a_{nr}} \right) a_{n1}^{k-r} \\&= \sum_{r=0}^k (-1)^{k-r} \binom{k}{r} \underbrace{a_{nr}} a_{n1}^{k-r}.\end{aligned}\tag{1}$$

3.2.1 矩法和矩估计量

定义 (3.2.1 (矩估计))

设有总体分布族 $\{f(x, \theta), \theta \in \Theta\}$, Θ 是参数空间, $g(\theta)$ 是定义在 Θ 上参数 θ 的函数, 它可以表示为总体分布的某些矩的函数, 即

3.2.1 矩法和矩估计量

定义 (3.2.1 (矩估计))

设有总体分布族 $\{f(x, \theta), \theta \in \Theta\}$, Θ 是参数空间, $g(\theta)$ 是定义在 Θ 上参数 θ 的函数, 它可以表示为总体分布的某些矩的函数, 即

$$g(\theta) = G(\alpha_1, \alpha_2, \dots, \alpha_k; \mu_2, \dots, \mu_s). \quad (2)$$

3.2.1 矩法和矩估计量

定义 (3.2.1 (矩估计))

设有总体分布族 $\{f(x, \theta), \theta \in \Theta\}$, Θ 是参数空间, $g(\theta)$ 是定义在 Θ 上参数 θ 的函数, 它可以表示为总体分布的某些矩的函数, 即

$$g(\theta) = G(\alpha_1, \alpha_2, \dots, \alpha_k; \mu_2, \dots, \mu_s). \quad (2)$$

设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从上述分布族中抽取的简单样本, 将式 (2) 中的 α_i 和 μ_j 分别用它们“自然”矩估计量 a_{ni} 和 m_{nj} 代替,

3.2.1 矩法和矩估计量

定义 (3.2.1 (矩估计) 续)

得

$$\hat{g}(\mathbf{X}) = G(a_{n1}, \dots, a_{nk}; m_{n2}, \dots, m_{ns}), \quad (3)$$

3.2.1 矩法和矩估计量

定义 (3.2.1 (矩估计) 续)

得

$$\hat{g}(\mathbf{X}) = G(a_{n1}, \dots, a_{nk}; m_{n2}, \dots, m_{ns}), \quad (3)$$

则 $\hat{g}(\mathbf{X})$ 作为 $g(\theta)$ 的估计量, 称为 $g(\theta)$ 的**矩估计量**。这种求矩估计量的方法称为**矩法**。

3.2.2 若干例子

例 (3.2.2)

设总体 X 服从具有成功概率为 θ 的两点分布 $B(1, \theta)$ 。

$\mathbf{X} = (X_1, \dots, X_n)$ 是从总体 X 中抽取的简单样本，

求 $g_1(\theta) = \theta$ 和 $g_2(\theta) = \theta(1 - \theta)$ 的矩估计量。

解

3.2.2 若干例子

例 (3.2.2)

设总体 X 服从具有成功概率为 θ 的两点分布 $B(1, \theta)$ 。

$\mathbf{X} = (X_1, \dots, X_n)$ 是从总体 X 中抽取的简单样本，

求 $g_1(\theta) = \theta$ 和 $g_2(\theta) = \theta(1 - \theta)$ 的矩估计量。

解

由于 $X \sim B(1, \theta)$ ，则有 $P(X = x) = \theta^x(1 - \theta)^{1-x}$, $x = 0, 1$ 。

故 $E(X) = \alpha_1 = \theta$ ，所以按定义 $g_1(\theta) = \alpha_1$ 的矩估计量为

3.2.2 若干例子

例 (3.2.2)

设总体 X 服从具有成功概率为 θ 的两点分布 $B(1, \theta)$ 。

$\mathbf{X} = (X_1, \dots, X_n)$ 是从总体 X 中抽取的简单样本，

求 $g_1(\theta) = \theta$ 和 $g_2(\theta) = \theta(1 - \theta)$ 的矩估计量。

解

由于 $X \sim B(1, \theta)$ ，则有 $P(X = x) = \theta^x(1 - \theta)^{1-x}$ ， $x = 0, 1$ 。

故 $E(X) = \alpha_1 = \theta$ ，所以按定义 $g_1(\theta) = \alpha_1$ 的矩估计量为

$$\hat{g}_1(X_1, \dots, X_n) = \hat{\theta}(X_1, \dots, X_n) = \bar{X};$$

3.2.2 若干例子

按定义 $g_2(\theta) = \alpha_1(1 - \alpha_1)$ 的矩估计量为

3.2.2 若干例子

按定义 $g_2(\theta) = \alpha_1(1 - \alpha_1)$ 的矩估计量为

$$\hat{g}(X_1, \dots, X_n) = \bar{X}(1 - \bar{X}).$$



3.2.2 若干例子

例 (3.2.3)

设总体 X 服从均匀分布 $U(\theta_1, \theta_2)$, $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 X 中抽取的简单样本, 令 $\theta = (\theta_1, \theta_2)$, 其中 $-\infty < \theta_1 < \theta_2 < \infty$. 求 θ_1 和 θ_2 的矩估计量。

解

3.2.2 若干例子

例 (3.2.3)

设总体 X 服从均匀分布 $U(\theta_1, \theta_2)$, $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 X 中抽取的简单样本, 令 $\theta = (\theta_1, \theta_2)$, 其中 $-\infty < \theta_1 < \theta_2 < \infty$. 求 θ_1 和 θ_2 的矩估计量。

解

由于 $X \sim U(\theta_1, \theta_2)$, 由均匀分布的性质可知

3.2.2 若干例子

例 (3.2.3)

设总体 X 服从均匀分布 $U(\theta_1, \theta_2)$, $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 X 中抽取的简单样本, 令 $\theta = (\theta_1, \theta_2)$, 其中 $-\infty < \theta_1 < \theta_2 < \infty$. 求 θ_1 和 θ_2 的矩估计量。

解

由于 $X \sim U(\theta_1, \theta_2)$, 由均匀分布的性质可知

$$E(X) = \alpha_1 = (\theta_1 + \theta_2)/2,$$

$$Var(X) = \mu_2 = (\theta_2 - \theta_1)^2/12.$$

3.2.2 若干例子

解此方程组得

3.2.2 若干例子

解此方程组得

$$\theta_1 = \alpha_1 - \sqrt{3\mu_2},$$

$$\theta_2 = \alpha_1 + \sqrt{3\mu_2}.$$

3.2.2 若干例子

解此方程组得

$$\theta_1 = \alpha_1 - \sqrt{3\mu_2},$$

$$\theta_2 = \alpha_1 + \sqrt{3\mu_2}.$$

将上式中 α_1 和 μ_2 分别用 \bar{X} 和 m_{n2} 代入得

3.2.2 若干例子

解此方程组得

$$\theta_1 = \alpha_1 - \sqrt{3\mu_2},$$

$$\theta_2 = \alpha_1 + \sqrt{3\mu_2}.$$

将上式中 α_1 和 μ_2 分别用 \bar{X} 和 m_{n2} 代入得

$$\hat{\theta}_1(X_1, \dots, X_n) = \bar{X} - \sqrt{3m_{n2}} = \bar{X} - \sqrt{3}S_n,$$

$$\hat{\theta}_2(X_1, \dots, X_n) = \bar{X} + \sqrt{3}S_n,$$

3.2.2 若干例子

解此方程组得

$$\theta_1 = \alpha_1 - \sqrt{3\mu_2},$$

$$\theta_2 = \alpha_1 + \sqrt{3\mu_2}.$$

将上式中 α_1 和 μ_2 分别用 \bar{X} 和 m_{n2} 代入得

$$\hat{\theta}_1(X_1, \dots, X_n) = \bar{X} - \sqrt{3m_{n2}} = \bar{X} - \sqrt{3}S_n,$$

$$\hat{\theta}_2(X_1, \dots, X_n) = \bar{X} + \sqrt{3}S_n,$$

其中 $S_n^2 = m_{n2}$ (亦可以用 S 代替 S_n)。 □

3.2.2 若干例子

例 (3.2.4)

设总体 X 有概率密度

$$f(x, \theta) = \begin{cases} 2\sqrt{\frac{\theta}{\pi}} \exp\{-\theta x^2\}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

其中 $\theta > 0$ 为未知参数。这个分布称为Maxwell分布，在气体分子动力学中有应用。设 $\mathbf{X} = (X_1, \dots, X_n)$ 为抽自此总体的简单样本，求 $g(\theta) = 1/\theta$ 的矩估计量。

3.2.2 若干例子

解

3.2.2 若干例子

解

由于 $X \sim f(x; \theta)$, 由方程

3.2.2 若干例子

解

由于 $X \sim f(x; \theta)$, 由方程

$$\alpha_1 = E(X) = 2\sqrt{\frac{\theta}{\pi}} \int_0^{\infty} x e^{-\theta x^2} dx = \frac{1}{\sqrt{\pi\theta}},$$

3.2.2 若干例子

解

由于 $X \sim f(x; \theta)$, 由方程

$$\alpha_1 = E(X) = 2\sqrt{\frac{\theta}{\pi}} \int_0^{\infty} x e^{-\theta x^2} dx = \frac{1}{\sqrt{\pi\theta}},$$

解得 $g(\theta) = 1/\theta = \pi\alpha_1^2$ 。

3.2.2 若干例子

解

由于 $X \sim f(x; \theta)$, 由方程

$$\alpha_1 = E(X) = 2\sqrt{\frac{\theta}{\pi}} \int_0^{\infty} x e^{-\theta x^2} dx = \frac{1}{\sqrt{\pi\theta}},$$

解得 $g(\theta) = 1/\theta = \pi\alpha_1^2$ 。将 α_1 用 \bar{X} 代替, 得

3.2.2 若干例子

解

由于 $X \sim f(x; \theta)$, 由方程

$$\alpha_1 = E(X) = 2\sqrt{\frac{\theta}{\pi}} \int_0^{\infty} x e^{-\theta x^2} dx = \frac{1}{\sqrt{\pi\theta}},$$

解得 $g(\theta) = 1/\theta = \pi\alpha_1^2$ 。将 α_1 用 \bar{X} 代替, 得

$$\hat{g}_1(\mathbf{X}) = \pi\bar{X}^2.$$

3.2.2 若干例子

另一方面，由另一方程

3.2.2 若干例子

另一方面，由另一方程

$$\alpha_2 = E_{\theta}(X^2) = 2\sqrt{\frac{\theta}{\pi}} \int_0^{\infty} x^2 e^{-\theta x^2} dx = \frac{1}{2\theta},$$

3.2.2 若干例子

另一方面，由另一方程

$$\alpha_2 = E_\theta(X^2) = 2\sqrt{\frac{\theta}{\pi}} \int_0^\infty x^2 e^{-\theta x^2} dx = \frac{1}{2\theta},$$

解得 $g(\theta) = 1/\theta = 2\alpha_2$ 。

3.2.2 若干例子

另一方面，由另一方程

$$\alpha_2 = E_\theta(X^2) = 2\sqrt{\frac{\theta}{\pi}} \int_0^\infty x^2 e^{-\theta x^2} dx = \frac{1}{2\theta},$$

解得 $g(\theta) = 1/\theta = 2\alpha_2$ 。将 α_2 用 $a_{n2} = \sum_{i=1}^n X_i^2/n$ 代入，得 $g(\theta)$ 的矩估计

3.2.2 若干例子

另一方面，由另一方程

$$\alpha_2 = E_\theta(X^2) = 2\sqrt{\frac{\theta}{\pi}} \int_0^\infty x^2 e^{-\theta x^2} dx = \frac{1}{2\theta},$$

解得 $g(\theta) = 1/\theta = 2\alpha_2$ 。将 α_2 用 $a_{n2} = \sum_{i=1}^n X_i^2/n$ 代入，得 $g(\theta)$ 的矩估计

$$\hat{g}_2(\mathbf{X}) = 2a_{n2} = \frac{2}{n} \sum_{i=1}^n X_i^2.$$

□

3.2.2 若干例子

由例3.2.4可得到 $g(\theta) = 1/\theta$ 的矩估计有两个：

3.2.2 若干例子

由例3.2.4可得到 $g(\theta) = 1/\theta$ 的矩估计有两个:

矩估计1: $\hat{g}_1(\mathbf{X}) = \pi \bar{X}^2$ 。

矩估计2: $\hat{g}_2(\mathbf{X}) = \frac{2}{n} \sum_{i=1}^n X_i^2$ 。

3.2.2 若干例子

由例3.2.4可得到 $g(\theta) = 1/\theta$ 的矩估计有两个:

矩估计1: $\hat{g}_1(\mathbf{X}) = \pi \bar{X}^2$ 。

矩估计2: $\hat{g}_2(\mathbf{X}) = \frac{2}{n} \sum_{i=1}^n X_i^2$ 。

由此可见矩估计不唯一。

3.2.2 若干例子

由例3.2.4可得到 $g(\theta) = 1/\theta$ 的矩估计有两个:

矩估计1: $\hat{g}_1(\mathbf{X}) = \pi \bar{X}^2$ 。

矩估计2: $\hat{g}_2(\mathbf{X}) = \frac{2}{n} \sum_{i=1}^n X_i^2$ 。

由此可见矩估计不唯一。

这两个矩估计量 $\hat{g}_1(\mathbf{X})$ 和 $\hat{g}_2(\mathbf{X})$ 中哪一个更好?

3.2.2 若干例子

可以证明基于 a_{n2} 的估计量 $\hat{g}_2(\mathbf{X})$ ，在 $g(\theta)$ 一切无偏估计类中是方差最小者。基于 a_{n1} 的估计 $\hat{g}_1(\mathbf{X})$ 不是 $g(\theta)$ 的无偏估计。

证明如下：

3.2.2 若干例子

可以证明基于 a_{n2} 的估计量 $\hat{g}_2(\mathbf{X})$ ，在 $g(\theta)$ 一切无偏估计类中是方差最小者。基于 a_{n1} 的估计 $\hat{g}_1(\mathbf{X})$ 不是 $g(\theta)$ 的无偏估计。

证明如下：

矩估计量 $\hat{g}_2(\mathbf{X}) = \frac{2}{n} \sum_{i=1}^n X_i^2$ 的期望为

3.2.2 若干例子

可以证明基于 a_{n2} 的估计量 $\hat{g}_2(\mathbf{X})$, 在 $g(\theta)$ 一切无偏估计类中是方差最小者。基于 a_{n1} 的估计 $\hat{g}_1(\mathbf{X})$ 不是 $g(\theta)$ 的无偏估计。

证明如下:

矩估计量 $\hat{g}_2(\mathbf{X}) = \frac{2}{n} \sum_{i=1}^n X_i^2$ 的期望为

$$\begin{aligned} E(\hat{g}_2(\mathbf{X})) &= E\left(\frac{2}{n} \sum_{i=1}^n X_i^2\right) \\ &= \frac{2}{n} n E(X^2) = \frac{2}{n} \frac{n}{2\theta} = \frac{1}{\theta} = g(\theta). \end{aligned}$$

3.2.2 若干例子

可以证明基于 a_{n2} 的估计量 $\hat{g}_2(\mathbf{X})$, 在 $g(\theta)$ 一切无偏估计类中是方差最小者。基于 a_{n1} 的估计 $\hat{g}_1(\mathbf{X})$ 不是 $g(\theta)$ 的无偏估计。

证明如下:

矩估计量 $\hat{g}_2(\mathbf{X}) = \frac{2}{n} \sum_{i=1}^n X_i^2$ 的期望为

$$\begin{aligned} E(\hat{g}_2(\mathbf{X})) &= E\left(\frac{2}{n} \sum_{i=1}^n X_i^2\right) \\ &= \frac{2}{n} n E(X^2) = \frac{2}{n} \frac{n}{2\theta} = \frac{1}{\theta} = g(\theta). \end{aligned}$$

故 $\hat{g}_2(\mathbf{X})$ 是 $g(\theta)$ 的无偏估计。

3.2.2 若干例子

Fisher信息为

3.2.2 若干例子

Fisher信息为

$$I_1(\theta) = -E\left(\frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2}\right) = -E\left(\frac{\partial^2 \ln[\sqrt{\theta} \exp\{-\theta X^2\}]}{\partial \theta^2}\right)$$

=

3.2.2 若干例子

Fisher信息为

$$\begin{aligned} I_1(\theta) &= -E\left(\frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2}\right) = -E\left(\frac{\partial^2 \ln[\sqrt{\theta} \exp\{-\theta X^2\}]}{\partial \theta^2}\right) \\ &= -E\left(\frac{\partial^2 \frac{1}{2} \ln \theta - \theta X^2}{\partial \theta^2}\right) = -E\left(\frac{\partial \frac{1}{2\theta} - X^2}{\partial \theta}\right) \\ &= \end{aligned}$$

3.2.2 若干例子

Fisher信息为

$$\begin{aligned} I_1(\theta) &= -E\left(\frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2}\right) = -E\left(\frac{\partial^2 \ln[\sqrt{\theta} \exp\{-\theta X^2\}]}{\partial \theta^2}\right) \\ &= -E\left(\frac{\partial^2 \frac{1}{2} \ln \theta - \theta X^2}{\partial \theta^2}\right) = -E\left(\frac{\partial \frac{1}{2\theta} - X^2}{\partial \theta}\right) \\ &= -E\left(-\frac{1}{2\theta^2}\right) = \frac{1}{2\theta^2}. \end{aligned}$$

3.2.2 若干例子

则C-R下界为

3.2.2 若干例子

则C-R下界为

$$\frac{[g'(\theta)]^2}{nI_1(\theta)} = \frac{\left(-\frac{1}{\theta^2}\right)^2}{n\frac{1}{2\theta^2}} = \frac{\frac{1}{\theta^4}}{n\frac{1}{2\theta^2}} = \frac{2\theta^2}{n\theta^4} = \frac{2}{n\theta^2}.$$

3.2.2 若干例子

其方差为

3.2.2 若干例子

其方差为

$$\begin{aligned} \text{Var}(\hat{g}_2(\mathbf{X})) &= \text{Var}\left(\frac{2}{n} \sum_{i=1}^n X_i^2\right) = \frac{4}{n^2} \sum_{i=1}^n \text{Var} X_i^2 \\ &= \frac{4}{n^2} \sum_{i=1}^n [E X_i^4 - (E X_i^2)^2] \\ &= \end{aligned}$$

3.2.2 若干例子

其方差为

$$\begin{aligned} \text{Var}(\hat{g}_2(\mathbf{X})) &= \text{Var}\left(\frac{2}{n} \sum_{i=1}^n X_i^2\right) = \frac{4}{n^2} \sum_{i=1}^n \text{Var} X_i^2 \\ &= \frac{4}{n^2} \sum_{i=1}^n [E X_i^4 - (E X_i^2)^2] \\ &= \frac{4}{n^2} \sum_{i=1}^n \left[\int_0^\infty 2\sqrt{\frac{\theta}{\pi}} x^4 e^{-\theta x^2} dx - \left(\frac{1}{2\theta}\right)^2 \right] \\ &= \frac{4}{n^2} \sum_{i=1}^n \left[\int_0^\infty -\frac{1}{\sqrt{\pi\theta}} x^3 de^{-\theta x^2} - \left(\frac{1}{2\theta}\right)^2 \right] \end{aligned}$$

3.2.2 若干例子

$$\begin{aligned} &= \frac{4}{n^2} \sum_{i=1}^n \left[-\frac{1}{\sqrt{\pi\theta}} x^3 e^{-\theta x^2} \Big|_0^\infty + \int_0^\infty \frac{1}{\sqrt{\pi\theta}} 3x^2 e^{-\theta x^2} dx - \left(\frac{1}{2\theta}\right)^2 \right] \\ &= \frac{4}{n^2} \sum_{i=1}^n \left[\int_0^\infty -\frac{3}{2\sqrt{\pi}\theta^{3/2}} de^{-\theta x^2} - \left(\frac{1}{2\theta}\right)^2 \right] \\ &= \end{aligned}$$

3.2.2 若干例子

$$\begin{aligned} &= \frac{4}{n^2} \sum_{i=1}^n \left[-\frac{1}{\sqrt{\pi\theta}} x^3 e^{-\theta x^2} \Big|_0^\infty + \int_0^\infty \frac{1}{\sqrt{\pi\theta}} 3x^2 e^{-\theta x^2} dx - \left(\frac{1}{2\theta}\right)^2 \right] \\ &= \frac{4}{n^2} \sum_{i=1}^n \left[\int_0^\infty -\frac{3}{2\sqrt{\pi\theta^{3/2}}} de^{-\theta x^2} - \left(\frac{1}{2\theta}\right)^2 \right] \\ &= \frac{4}{n^2} \sum_{i=1}^n \left[-\frac{3}{2\sqrt{\pi\theta^{3/2}}} x e^{-\theta x^2} \Big|_0^\infty + \int_0^\infty \frac{3}{2\sqrt{\pi\theta^{3/2}}} e^{-\theta x^2} dx - \left(\frac{1}{2\theta}\right)^2 \right] \\ &= \frac{4}{n^2} \sum_{i=1}^n \left[\int_0^\infty \frac{3}{2\theta^2} \frac{1}{\sqrt{2\pi} \frac{1}{\sqrt{2\theta}}} \exp\left\{-\frac{x^2}{2(1/\sqrt{2\theta})^2}\right\} dx - \left(\frac{1}{2\theta}\right)^2 \right] \end{aligned}$$

3.2.2 若干例子

$$= \frac{4}{n^2} \sum_{i=1}^n \left[\frac{3}{2\theta^2} \times \frac{1}{2} = \frac{3}{4\theta^2} - \left(\frac{1}{2\theta} \right)^2 \right] = \frac{4}{n^2} \left[\frac{3n}{4\theta^2} - \frac{n}{4\theta^2} \right] = \frac{2}{n\theta^2}.$$

3.2.2 若干例子

$$= \frac{4}{n^2} \sum_{i=1}^n \left[\frac{3}{2\theta^2} \times \frac{1}{2} = \frac{3}{4\theta^2} - \left(\frac{1}{2\theta} \right)^2 \right] = \frac{4}{n^2} \left[\frac{3n}{4\theta^2} - \frac{n}{4\theta^2} \right] = \frac{2}{n\theta^2}.$$

故矩估计量 $\hat{g}_2(\mathbf{X}) = \frac{2}{n} \sum_{i=1}^n X_i^2$ 是 $g(\theta) = 1/\theta$ 的无偏估计且达到了C-R下界，是一致最小方差无偏估计（UMVUE）。

3.2.2 若干例子

矩估计量 $\hat{g}_1(\mathbf{X})$ 的期望为

$$E(\hat{g}_1(\mathbf{X})) = E(\pi \bar{X}^2) = \pi[\text{Var}(\bar{X}) + (E\bar{X})^2]$$

=

3.2.2 若干例子

矩估计量 $\hat{g}_1(\mathbf{X})$ 的期望为

$$\begin{aligned} E(\hat{g}_1(\mathbf{X})) &= E(\pi \bar{X}^2) = \pi [Var(\bar{X}) + (E\bar{X})^2] \\ &= \pi \left[Var \left(\frac{\sum_{i=1}^n X_i}{n} \right) + (E\bar{X})^2 \right] \\ &= \end{aligned}$$

3.2.2 若干例子

矩估计量 $\hat{g}_1(\mathbf{X})$ 的期望为

$$\begin{aligned} E(\hat{g}_1(\mathbf{X})) &= E(\pi \bar{X}^2) = \pi [Var(\bar{X}) + (E\bar{X})^2] \\ &= \pi \left[Var \left(\frac{\sum_{i=1}^n X_i}{n} \right) + (EX)^2 \right] \\ &= \pi \left[\frac{\sum_{i=1}^n Var(X_i)}{n^2} + (EX)^2 \right] = \pi \left(\frac{Var(X_i)}{n} + (EX)^2 \right) \end{aligned}$$

3.2.2 若干例子

$$= \pi \left(\frac{\frac{1}{2\theta} - \frac{1}{\pi\theta}}{n} + \frac{1}{\pi\theta} \right) = \frac{\pi}{2\theta n} - \frac{1}{\theta n} + \frac{1}{\theta} \neq g(\theta) = \frac{1}{\theta}$$

3.2.2 若干例子

$$= \pi \left(\frac{\frac{1}{2\theta} - \frac{1}{\pi\theta}}{n} + \frac{1}{\pi\theta} \right) = \frac{\pi}{2\theta n} - \frac{1}{\theta n} + \frac{1}{\theta} \neq g(\theta) = \frac{1}{\theta}$$

故矩估计量 $\hat{g}_2(\mathbf{X})$ 不是 $g(\theta)$ 的无偏估计。

□

3.2.2 若干例子

例 (3.2.5)

设 X_1, \dots, X_n 为从总体 X 中抽取的 *i.i.d.* 样本, 求总体 X 的变异系数、偏度和峰度

$$\nu = \frac{\sqrt{\mu_2}}{\alpha_1}, \beta_1 = \frac{\mu_3}{\mu_2^{3/2}}, \beta_2 = \frac{\mu_4}{\mu_2^2} - 3$$

的矩估计量。

3.2.2 若干例子

解

3.2.2 若干例子

解

将上述表达式中的 α_1 , μ_2 , μ_3 和 μ_4 分别用它们的矩估计 a_{n1} , m_{n2} , m_{n3} 和 m_{n4} 代替, 得到 ν , β_1 , β_2 的矩估计, 也称为样本变异系数、样本偏度和样本峰度, 它们的表达式分别为

3.2.2 若干例子

解

将上述表达式中的 α_1 , μ_2 , μ_3 和 μ_4 分别用它们的矩估计 a_{n1} , m_{n2} , m_{n3} 和 m_{n4} 代替, 得到 ν , β_1 , β_2 的矩估计, 也称为样本变异系数、样本偏度和样本峰度, 它们的表达式分别为

$$\begin{aligned}\hat{\nu}(X_1, \dots, X_n) &= \frac{\sqrt{m_{n2}}}{a_{n1}}, \\ \hat{\beta}_1(X_1, \dots, X_n) &= \frac{m_{n3}}{m_{n2}^{3/2}}, \\ \hat{\beta}_2(X_1, \dots, X_n) &= \frac{m_{n4}}{m_{n2}^2} - 3.\end{aligned}$$



3.2.2 若干例子

例 (3.2.6)

设总体 X 有概率密度

$$f(x, \theta) = \begin{cases} \frac{\theta_2}{\Gamma((1+\theta_1)/\theta_2)} x^{\theta_1} \exp\{-x^{\theta_2}\}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

参数 $\theta = (\theta_1, \theta_2)$ 的变化范围是 $-1 < \theta_1 < \infty, \theta_2 > 0$ 。

3.2.2 若干例子

例 (3.2.6)

设总体 X 有概率密度

$$f(x, \theta) = \begin{cases} \frac{\theta_2}{\Gamma((1+\theta_1)/\theta_2)} x^{\theta_1} \exp\{-x^{\theta_2}\}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

参数 $\theta = (\theta_1, \theta_2)$ 的变化范围是 $-1 < \theta_1 < \infty$, $\theta_2 > 0$ 。

设 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ 为抽自此总体的简单样本, 求 θ_1 和 θ_2 的矩估计量。

3.2.2 若干例子

解 由伽马函数可知 $\Gamma(s) = \int_0^{\infty} x^{s-1} e^{-x} dx, x > 0$, 经计算可得总体 X 的一阶原点矩

3.2.2 若干例子

解 由伽马函数可知 $\Gamma(s) = \int_0^{\infty} x^{s-1} e^{-x} dx$, $x > 0$, 经计算可得总体 X 的一阶原点矩

$$\alpha_1 = \int_0^{\infty} x f(x; \boldsymbol{\theta}) dx = \int_0^{\infty} x \frac{\theta_2}{\Gamma((1 + \theta_1)/\theta_2)} x^{\theta_1} \exp\{-x^{\theta_2}\} dx$$

3.2.2 若干例子

解 由伽马函数可知 $\Gamma(s) = \int_0^{\infty} x^{s-1} e^{-x} dx$, $x > 0$, 经计算可得总体 X 的一阶原点矩

$$\begin{aligned}\alpha_1 &= \int_0^{\infty} x f(x; \boldsymbol{\theta}) dx = \int_0^{\infty} x \frac{\theta_2}{\Gamma((1 + \theta_1)/\theta_2)} x^{\theta_1} \exp\{-x^{\theta_2}\} dx \\ &\stackrel{x^{\theta_2}=t}{=} \int_0^{\infty} \frac{\theta_2}{\Gamma(\frac{1+\theta_1}{\theta_2})} t^{\frac{\theta_1+1}{\theta_2}} \exp\{-t\} \frac{1}{\theta_2} t^{\frac{1}{\theta_2}-1} dt\end{aligned}$$

3.2.2 若干例子

解 由伽马函数可知 $\Gamma(s) = \int_0^{\infty} x^{s-1} e^{-x} dx$, $x > 0$, 经计算可得总体 X 的一阶原点矩

$$\begin{aligned}\alpha_1 &= \int_0^{\infty} x f(x; \boldsymbol{\theta}) dx = \int_0^{\infty} x \frac{\theta_2}{\Gamma((1 + \theta_1)/\theta_2)} x^{\theta_1} \exp\{-x^{\theta_2}\} dx \\ &\stackrel{x^{\theta_2}=t}{=} \int_0^{\infty} \frac{\theta_2}{\Gamma(\frac{1+\theta_1}{\theta_2})} t^{\frac{\theta_1+1}{\theta_2}} \exp\{-t\} \frac{1}{\theta_2} t^{\frac{1}{\theta_2}-1} dt \\ &= \int_0^{\infty} \frac{1}{\Gamma(\frac{1+\theta_1}{\theta_2})} t^{\frac{\theta_1+2}{\theta_2}-1} \exp\{-t\} dt\end{aligned}$$

3.2.2 若干例子

解 由伽马函数可知 $\Gamma(s) = \int_0^{\infty} x^{s-1} e^{-x} dx$, $x > 0$, 经计算可得总体 X 的一阶原点矩

$$\begin{aligned}\alpha_1 &= \int_0^{\infty} x f(x; \boldsymbol{\theta}) dx = \int_0^{\infty} x \frac{\theta_2}{\Gamma((1 + \theta_1)/\theta_2)} x^{\theta_1} \exp\{-x^{\theta_2}\} dx \\ &\stackrel{x^{\theta_2}=t}{=} \int_0^{\infty} \frac{\theta_2}{\Gamma(\frac{1+\theta_1}{\theta_2})} t^{\frac{\theta_1+1}{\theta_2}} \exp\{-t\} \frac{1}{\theta_2} t^{\frac{1}{\theta_2}-1} dt \\ &= \int_0^{\infty} \frac{1}{\Gamma(\frac{1+\theta_1}{\theta_2})} t^{\frac{\theta_1+2}{\theta_2}-1} \exp\{-t\} dt \\ &= \Gamma\left(\frac{2 + \theta_1}{\theta_2}\right) / \Gamma\left(\frac{1 + \theta_1}{\theta_2}\right)\end{aligned}\tag{4}$$

3.2.2 若干例子

同理可得总体 X 的二阶原点矩为

3.2.2 若干例子

同理可得总体 X 的二阶原点矩为

$$\alpha_2 = \Gamma\left(\frac{3 + \theta_1}{\theta_2}\right) / \Gamma\left(\frac{1 + \theta_1}{\theta_2}\right).$$

3.2.2 若干例子

同理可得总体 X 的二阶原点矩为

$$\alpha_2 = \Gamma\left(\frac{3 + \theta_1}{\theta_2}\right) / \Gamma\left(\frac{1 + \theta_1}{\theta_2}\right).$$

按矩估计方法，用 a_{n1} 和 a_{n2} 分别代替式（4）和式（5）中的 α_1 和 α_2 ，用 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 分别代替 θ_1 和 θ_2 ，得到如下的方程组：

3.2.2 若干例子

同理可得总体 X 的二阶原点矩为

$$\alpha_2 = \Gamma\left(\frac{3 + \theta_1}{\theta_2}\right) / \Gamma\left(\frac{1 + \theta_1}{\theta_2}\right).$$

按矩估计方法，用 a_{n1} 和 a_{n2} 分别代替式（4）和式（5）中的 α_1 和 α_2 ，用 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 分别代替 θ_1 和 θ_2 ，得到如下的方程组：

$$\begin{aligned} a_{n1} &= \Gamma\left(\frac{2 + \hat{\theta}_1}{\hat{\theta}_2}\right) / \Gamma\left(\frac{1 + \hat{\theta}_1}{\hat{\theta}_2}\right), \\ a_{n2} &= \Gamma\left(\frac{3 + \hat{\theta}_1}{\hat{\theta}_2}\right) / \Gamma\left(\frac{1 + \hat{\theta}_1}{\hat{\theta}_2}\right). \end{aligned}$$

3.2.2 若干例子

其解就是 θ_1 和 θ_2 的矩估计。但此处得不出 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 的简单解析表达式，而只能用数值方法。 □

3.2.2 若干例子

其解就是 θ_1 和 θ_2 的矩估计。但此处得不出 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 的简单解析表达式，而只能用数值方法。 □

此例说明不是所有的矩估计都有解析表达式。

3.2.2 若干例子

例 (3.2.7)

设 (X_i, Y_i) , $i = 1, 2, \dots, n$ 为从一个二维总体中抽取的简单样本, 求总体分布的协方差 σ_{12} 和相关系数 ρ 的矩估计。

3.2.2 若干例子

例 (3.2.7)

设 (X_i, Y_i) , $i = 1, 2, \dots, n$ 为从一个二维总体中抽取的简单样本, 求总体分布的协方差 σ_{12} 和相关系数 ρ 的矩估计。

解

总体分布的协方差 $\sigma_{12} = E(X - E(X))(Y - E(Y))$ 的矩估计量为样本协方差

3.2.2 若干例子

例 (3.2.7)

设 (X_i, Y_i) , $i = 1, 2, \dots, n$ 为从一个二维总体中抽取的简单样本, 求总体分布的协方差 σ_{12} 和相关系数 ρ 的矩估计。

解

总体分布的协方差 $\sigma_{12} = E(X - E(X))(Y - E(Y))$ 的矩估计量为样本协方差

$$m_{12} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}),$$

3.2.2 若干例子

例 (3.2.7)

设 (X_i, Y_i) , $i = 1, 2, \dots, n$ 为从一个二维总体中抽取的简单样本, 求总体分布的协方差 σ_{12} 和相关系数 ρ 的矩估计。

解

总体分布的协方差 $\sigma_{12} = E(X - E(X))(Y - E(Y))$ 的矩估计量为样本协方差

$$m_{12} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}),$$

其中 $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$, $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$ 。

3.2.2 若干例子

$\rho = \frac{Cov(X,Y)}{SD(X)SD(Y)}$ 的矩估计量是样本相关系数，它的矩估计量为

3.2.2 若干例子

$\rho = \frac{Cov(X,Y)}{SD(X)SD(Y)}$ 的矩估计量是样本相关系数，它的矩估计量为

$$r = \frac{m_{12}}{\sqrt{m_{n2,x}m_{n2,y}}} = \frac{\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n}}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}}}$$

=

3.2.2 若干例子

$\rho = \frac{\text{Cov}(X,Y)}{SD(X)SD(Y)}$ 的矩估计量是样本相关系数，它的矩估计量为

$$\begin{aligned} r &= \frac{m_{12}}{\sqrt{m_{n2,x}m_{n2,y}}} = \frac{\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n}}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}}} \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{Q_1^2 \cdot Q_2^2}}, \end{aligned}$$

3.2.2 若干例子

$\rho = \frac{\text{Cov}(X,Y)}{SD(X)SD(Y)}$ 的矩估计量是样本相关系数，它的矩估计量为

$$\begin{aligned} r &= \frac{m_{12}}{\sqrt{m_{n2,x}m_{n2,y}}} = \frac{\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n}}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}}} \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{Q_1^2 \cdot Q_2^2}}, \end{aligned}$$

此处 $Q_1^2 = \sum_{i=1}^n (X_i - \bar{X})^2$, $Q_2^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2$.

□

3.2.2 若干例子

例 (3.2.8)

在第二次世界大战期间，德军的豹式坦克给盟军造成了很大的损失。估计德军拥有该坦克的数量对确保诺曼底登录的成功非常重要。

3.2.2 若干例子

例 (3.2.8)

在第二次世界大战期间，德军的豹式坦克给盟军造成了很大的损失。估计德军拥有该坦克的数量对确保诺曼底登录的成功非常重要。

对此具体有两种可供选择的做法：第一种做法是使用传统的情报手段；第二种做法是使用统计学方法来进行估计。

3.2.2 若干例子

例 (3.2.8 (续))

盟军最初通过传统情报手段得出的结论是从1940年6月到1942年9月期间每月生产1400辆坦克。

3.2.2 若干例子

例 (3.2.8 (续))

盟军最初通过传统情报手段得出的结论是从1940年6月到1942年9月期间每月生产1400辆坦克。

但这个数字显然与事实不符。因为在长达8个月的斯大林格勒战役中，德、奥、意使用了1200辆坦克，也就是说1400的数字高得离谱。

3.2.2 若干例子

例 (3.2.8 (续))

盟军最初通过传统情报手段得出的结论是从1940年6月到1942年9月期间每月生产1400辆坦克。

但这个数字显然与事实不符。因为在长达8个月的斯大林格勒战役中，德、奥、意使用了1200辆坦克，也就是说1400的数字高得离谱。

最终盟军找到了关键线索：序列号。

3.2.2 若干例子

例 (3.2.8 (续))

盟军发现每辆被俘坦克上都有一个独一无二的序列号，这些序列号主要是变速箱的号码。

3.2.2 若干例子

例 (3.2.8 (续))

盟军发现每辆被俘坦克上都有一个独一无二的序列号，这些序列号主要是变速箱的号码。

通过仔细观测，他们发现从这些序列号可以推算出坦克的生产量。

3.2.2 若干例子

例 (3.2.8 (续))

盟军发现每辆被俘坦克上都有一个独一无二的序列号，这些序列号主要是变速箱的号码。

通过仔细观测，他们发现从这些序列号可以推算出坦克的生产量。

假设坦克是按从1到N顺序产生的，那么如何使用统计方法估算坦克总数呢？

3.2.2 若干例子

解 在这个问题中，我们观测到 k 辆坦克。直观上可以使用观测到的最大序列号 m 来估计 N 。

3.2.2 若干例子

解 在这个问题中，我们观测到 k 辆坦克。直观上可以使用观测到的最大序列号 m 来估计 N 。

在 N 辆坦克中随机抽查 k 辆，记观测到最大序列号为 $X_{(k)}$ ，
则 $\{X_{(k)} = m\}$ 的概率为

3.2.2 若干例子

解 在这个问题中，我们观测到 k 辆坦克。直观上可以使用观测到的最大序列号 m 来估计 N 。

在 N 辆坦克中随机抽查 k 辆，记观测到最大序列号为 $X_{(k)}$ ，
则 $\{X_{(k)} = m\}$ 的概率为

$$P(X_{(k)} = m) = \binom{m-1}{k-1} / \binom{N}{k}.$$

3.2.2 若干例子

解 在这个问题中，我们观测到 k 辆坦克。直观上可以使用观测到的最大序列号 m 来估计 N 。

在 N 辆坦克中随机抽查 k 辆，记观测到最大序列号为 $X_{(k)}$ ，则 $\{X_{(k)} = m\}$ 的概率为

$$P(X_{(k)} = m) = \binom{m-1}{k-1} / \binom{N}{k}.$$

给定总数 N 和样本量 k ，利用 $\sum_{m=k}^N \binom{m}{k} = \binom{N+1}{k+1}$ ，

3.2.2 若干例子

则

$$E[X_{(k)}] = \sum_{m=k}^N mP(X_{(k)} = m)$$

=

3.2.2 若干例子

则

$$\begin{aligned} E[X_{(k)}] &= \sum_{m=k}^N mP(X_{(k)} = m) \\ &= \sum_{m=k}^N m \binom{m-1}{k-1} / \binom{N}{k} \\ &= \end{aligned}$$

3.2.2 若干例子

则

$$\begin{aligned} E[X_{(k)}] &= \sum_{m=k}^N mP(X_{(k)} = m) \\ &= \sum_{m=k}^N m \binom{m-1}{k-1} / \binom{N}{k} \\ &= \frac{k(N+1)}{k+1} = \mu. \end{aligned}$$

3.2.2 若干例子

则

$$\begin{aligned} E[X_{(k)}] &= \sum_{m=k}^N mP(X_{(k)} = m) \\ &= \sum_{m=k}^N m \binom{m-1}{k-1} / \binom{N}{k} \\ &= \frac{k(N+1)}{k+1} = \mu. \end{aligned}$$

因此，可以得到

3.2.2 若干例子

则

$$\begin{aligned} E[X_{(k)}] &= \sum_{m=k}^N mP(X_{(k)} = m) \\ &= \sum_{m=k}^N m \binom{m-1}{k-1} / \binom{N}{k} \\ &= \frac{k(N+1)}{k+1} = \mu. \end{aligned}$$

因此，可以得到

$$N = \mu(1 + k^{-1}) - 1.$$

3.2.2 若干例子

于是，由矩估计方法，采用 $\bar{X}_{(k)} = X_{(k)}$ 代替上式中的 μ ，得到 N 的矩估计量为

3.2.2 若干例子

于是，由矩估计方法，采用 $\bar{X}_{(k)} = X_{(k)}$ 代替上式中的 μ ，得到 N 的矩估计量为

$$\hat{N} = X_{(k)}(1 + k^{-1}) - 1 = m(1 + k^{-1}) - 1.$$

3.2.2 若干例子

经计算

$$E(X_{(k)}^2) = \frac{k^2(N+1)^2 + N(N+1)k}{(k+2)(k+1)},$$

3.2.2 若干例子

经计算

$$E(X_{(k)}^2) = \frac{k^2(N+1)^2 + N(N+1)k}{(k+2)(k+1)},$$

$$\begin{aligned} \text{Var}(X_{(k)}^2) &= E(X_{(k)}^2) - [E(X_{(k)})]^2 \\ &= \frac{(N+1)k(N-k)}{(k+2)(k+1)^2}. \end{aligned}$$

3.2.2 若干例子

矩估计 \hat{N} 的方差为

$$\begin{aligned} \text{Var}(\hat{N}) &= (1 + k^{-1})^2 \text{Var}(X_{(k)}) \\ &= \frac{1}{k} \frac{(N - k)(N + 1)}{k + 2} \\ &\approx \frac{N^2}{k^2}, \text{ 当 } k \ll N \text{ 时.} \end{aligned}$$

从而其标准差约为 N/k 。

3.2.2 若干例子

盟军使用上述统计方法来确定坦克的生产速度，得到结论：

3.2.2 若干例子

盟军使用上述统计方法来确定坦克的生产速度，得到结论：
1940年夏天到1942年秋天，德军每月生产坦克246辆。

3.2.2 若干例子

盟军使用上述统计方法来确定坦克的生产速度，得到结论：

1940年夏天到1942年秋天，德军每月生产坦克246辆。

而根据第二次世界大战后德军内部数据，德军坦克实际每月产量为245辆，只比估计值少一辆。

