点估计

刘 赪

数学学院

2019年

§1. 矩估计法

矩估计法是英国统计学家K.皮尔逊（K. Pearson）在19世纪末到20世纪初的一系列文章中引进的。其基本思想就是利用样本矩来估计相应的总体矩，下面通过例子来说明。

§1. 矩估计法

矩估计法是英国统计学家K.皮尔逊（K. Pearson）在19世纪末到20世纪初的一系列文章中引进的。其基本思想就是利用样本矩来估计相应的总体矩，下面通过例子来说明。

例1. 已知顾客在某银行的窗口等待服务的时间X服从$Z(\alpha)$，其中参数α未知。总体X的数学期望为$E(X)=\frac{1}{\alpha}$。

§1. 矩估计法

矩估计法是英国统计学家K.皮尔逊（K. Pearson）在19世纪末到20世纪初的一系列文章中引进的。其基本思想就是利用样本矩来估计相应的总体矩，下面通过例子来说明。

例1. 已知顾客在某银行的窗口等待服务的时间X服从$Z(\alpha)$，其中参数α未知。总体X的数学期望为$E(X)=\frac{1}{\alpha}$。

设$X_1,X_2,\cdots,X_n$是来自该总体的一个简单随机样本，其样本均值$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$。

§1. 矩估计法

矩估计法是英国统计学家K.皮尔逊（K. Pearson）在19世纪末到20世纪初的一系列文章中引进的。其基本思想就是利用样本矩来估计相应的总体矩，下面通过例子来说明。

例1. 已知顾客在某银行的窗口等待服务的时间X服从$Z(\alpha)$，其中参数α未知。总体X的数学期望为$E(X)=\frac{1}{\alpha}$。

设$X_1,X_2,\cdots,X_n$是来自该总体的一个简单随机样本，其样本均值$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$。

一个很自然的想法就是，随着样本容量n的逐渐增大，$\overline{X}$也应该是越来越接近于总体均值$\frac{1}{\alpha}$的。因此，

§1. 矩估计法

矩估计法是英国统计学家K.皮尔逊（K. Pearson）在19世纪末到20世纪初的一系列文章中引进的。其基本思想就是利用样本矩来估计相应的总体矩，下面通过例子来说明。

例1. 已知顾客在某银行的窗口等待服务的时间X服从$Z(\alpha)$，其中参数α未知。总体X的数学期望为$E(X)=\frac{1}{\alpha}$。

设$X_1,X_2,\cdots,X_n$是来自该总体的一个简单随机样本，其样本均值$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$。

一个很自然的想法就是，随着样本容量n的逐渐增大，$\overline{X}$也应该是越来越接近于总体均值$\frac{1}{\alpha}$的。因此，

$$E(X)=\frac{1}{\alpha}\hat{=}\overline{X}$$

§1. 矩估计法

矩估计法是英国统计学家K.皮尔逊（K. Pearson）在19世纪末到20世纪初的一系列文章中引进的。其基本思想就是利用样本矩来估计相应的总体矩，下面通过例子来说明。

例1. 已知顾客在某银行的窗口等待服务的时间X服从$Z(\alpha)$，其中参数α未知。总体X的数学期望为$E(X)=\frac{1}{\alpha}$。

设X_1,X_2,$\cdots$,X_n是来自该总体的一个简单随机样本，其样本均值$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$。

一个很自然的想法就是，随着样本容量n的逐渐增大，$\overline{X}$也应该是越来越接近于总体均值$\frac{1}{\alpha}$的。因此，

$$E(X)=\frac{1}{\alpha}\hat{=}\overline{X}$$

由此解得α的矩估计量$\hat{\alpha}=\frac{1}{\overline{X}}$。

将以上过程的基本步骤总结如下:

将以上过程的基本步骤总结如下：

(1) 根据总体X的分布计算其数学期望，将$E(X)$表示为待估参数α的函数$g(\alpha) = \frac{1}{\alpha}$;

将以上过程的基本步骤总结如下:

(1) 根据总体X的分布计算其数学期望，将$E(X)$表示为待估参数α的函数$g(\alpha)=\frac{1}{\alpha}$;

(2) 利用对应的样本一阶原点矩$\overline{X}$估计$E(X)$（总体X的一阶原点矩）;

将以上过程的基本步骤总结如下：

(1) 根据总体X的分布计算其数学期望，将$E(X)$表示为待估参数α的函数$g(\alpha)=\frac{1}{\alpha}$；

(2) 利用对应的样本一阶原点矩$\overline{X}$估计$E(X)$（总体X的一阶原点矩）；

(3) 解方程$g(\alpha)=\overline{X}$，将α用样本$X_1,X_2,\cdots,X_n$表示出来，即为α的矩估计量。

将以上过程的基本步骤总结如下:

(1) 根据总体X的分布计算其数学期望,将$E(X)$表示为待估参数α的函数$g(\alpha)=\frac{1}{\alpha}$;

(2) 利用对应的样本一阶原点矩$\overline{X}$估计$E(X)$(总体X的一阶原点矩);

(3) 解方程$g(\alpha)=\overline{X}$,将α用样本$X_1,X_2,\cdots,X_n$表示出来,即为α的矩估计量。

将以上思路推广至一般情况,如果待估参数有k个,则至少需要构造k个方程。

将以上过程的基本步骤总结如下:

(1) 根据总体X的分布计算其数学期望，将$E(X)$表示为待估参数α的函数$g(\alpha)=\frac{1}{\alpha}$;

(2) 利用对应的样本一阶原点矩$\overline{X}$估计$E(X)$（总体X的一阶原点矩）；

(3) 解方程$g(\alpha)=\overline{X}$，将α用样本$X_1,X_2,\cdots,X_n$表示出来，即为α的矩估计量。

将以上思路推广至一般情况，如果待估参数有k个，则至少需要构造k个方程。

例2. 设总体X服从$N\left(\mu,\sigma^2\right)$，其中$\mu$和$\sigma^2$是未知参数。试给出$\mu$和$\sigma^2$的矩估计量。

将以上过程的基本步骤总结如下:

(1) 根据总体X的分布计算其数学期望，将$E(X)$表示为待估参数α的函数$g(\alpha)=\frac{1}{\alpha}$;

(2) 利用对应的样本一阶原点矩$\overline{X}$估计$E(X)$（总体X的一阶原点矩）;

(3) 解方程$g(\alpha)=\overline{X}$，将α用样本$X_1,X_2,\cdots,X_n$表示出来，即为α的矩估计量。

将以上思路推广至一般情况，如果待估参数有k个，则至少需要构造k个方程。

例2. 设总体X服从$N\left(\mu,\sigma^2\right)$，其中$\mu$和$\sigma^2$是未知参数。试给出$\mu$和$\sigma^2$的矩估计量。

Def. 设总体X的分布函数为$F(x;\theta_1,\cdots,\theta_k)$，其中$\theta_1,\cdots,\theta_k$是未知参数。假定总体$X$的$r$阶原点矩$E(X^r)(1\le r\le k)$都存在，一般情况下$E(X^r)$依赖于$\theta_1,\cdots,\theta_k$，记为

将以上过程的基本步骤总结如下:

(1) 根据总体X的分布计算其数学期望，将$E(X)$表示为待估参数α的函数$g(\alpha)=\frac{1}{\alpha}$;

(2) 利用对应的样本一阶原点矩$\overline{X}$估计$E(X)$（总体X的一阶原点矩）；

(3) 解方程$g(\alpha)=\overline{X}$，将α用样本$X_1,X_2,\cdots,X_n$表示出来，即为α的矩估计量。

将以上思路推广至一般情况，如果待估参数有k个，则至少需要构造k个方程。

例2. 设总体X服从$N\left(\mu,\sigma^2\right)$，其中$\mu$和$\sigma^2$是未知参数。试给出$\mu$和$\sigma^2$的矩估计量。

Def. 设总体X的分布函数为$F(x;\theta_1,\cdots,\theta_k)$，其中$\theta_1,\cdots,\theta_k$是未知参数。假定总体$X$的$r$阶原点矩$E(X^r)(1\leq r\leq k)$都存在，一般情况下$E(X^r)$依赖于$\theta_1,\cdots,\theta_k$，记为

$$E(X^r)=g_r(\theta_1,\cdots,\theta_k),\quad r=1,2,\cdots,k \tag{0.1}$$

将以上过程的基本步骤总结如下:

(1) 根据总体X的分布计算其数学期望，将$E(X)$表示为待估参数α的函数$g(\alpha)=\frac{1}{\alpha}$;

(2) 利用对应的样本一阶原点矩$\overline{X}$估计$E(X)$（总体X的一阶原点矩）；

(3) 解方程$g(\alpha)=\overline{X}$，将α用样本$X_1, X_2, \cdots, X_n$表示出来，即为α的矩估计量。

将以上思路推广至一般情况，如果待估参数有k个，则至少需要构造k个方程。

例2. 设总体X服从$N(\mu,\sigma^2)$，其中μ和σ^2是未知参数。试给出μ和σ^2的矩估计量。

Def. 设总体X的分布函数为$F(x;\theta_1,\cdots,\theta_k)$，其中$\theta_1,\cdots,\theta_k$是未知参数。假定总体$X$的$r$阶原点矩$E(X^r)(1\le r\le k)$都存在，一般情况下$E(X^r)$依赖于$\theta_1,\cdots,\theta_k$，记为

$$E(X^r)=g_r(\theta_1,\cdots,\theta_k),\quad r=1,2,\cdots,k \tag{0.1}$$

样本r阶原点矩

$$\overline{X^r} = \frac{1}{n}\sum_{i=1}^{n} X_i^r, \quad r = 1, 2, \cdots, k \tag{0.2}$$

样本r阶原点矩

$$\overline{X^r} = \frac{1}{n}\sum_{i=1}^{n} X_i^r, \quad r = 1, 2, \cdots, k \tag{0.2}$$

令样本r阶原点矩等于总体X的r阶原点矩（$r = 1, 2, \cdots, k$），即

样本r阶原点矩

$$\overline{X^r} = \frac{1}{n}\sum_{i=1}^{n} X_i^r, \quad r = 1, 2, \cdots, k \tag{0.2}$$

令样本r阶原点矩等于总体X的r阶原点矩（$r = 1, 2, \cdots, k$），即

$$\begin{cases} E(X) = g_1(\theta_1, \cdots, \theta_k) & \hat{=} \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} \\ E\left(X^2\right) = g_2(\theta_1, \cdots, \theta_k) & \hat{=} \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \overline{X^2} \\ \vdots \\ E\left(X^k\right) = g_k(\theta_1, \cdots, \theta_k) & \hat{=} \frac{1}{n}\sum_{i=1}^{n} X_i^k = \overline{X^k} \end{cases} \tag{0.3}$$

样本r阶原点矩

$$\overline{X^r} = \frac{1}{n}\sum_{i=1}^{n} X_i^r, \quad r = 1, 2, \cdots, k \tag{0.2}$$

令样本r阶原点矩等于总体X的r阶原点矩（$r = 1, 2, \cdots, k$），即

$$\begin{cases} E(X) = g_1(\theta_1, \cdots, \theta_k) & \hat{=} \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} \\ E\left(X^2\right) = g_2(\theta_1, \cdots, \theta_k) & \hat{=} \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \overline{X^2} \\ \vdots & \\ E\left(X^k\right) = g_k(\theta_1, \cdots, \theta_k) & \hat{=} \frac{1}{n}\sum_{i=1}^{n} X_i^k = \overline{X^k} \end{cases} \tag{0.3}$$

求解该方程组，即可得到参数$\theta_1, \cdots, \theta_k$的矩估计量

$$\hat{\theta}_r = \hat{\theta}_r(X_1, X_2, \cdots, X_n), \quad r = 1, 2, \cdots, k$$

注意:

注意:

（1）定义中选用的是原点矩，也可以用中心矩。只要给定总体矩，采用相应的样本矩估计即可。

注意：

（1）定义中选用的是原点矩，也可以用中心矩。只要给定总体矩，采用相应的样本矩估计即可。

（2）如果要估计的是函数$h(\theta_1,\cdots,\theta_k)$，则用$\hat{h}=h\left(\hat{\theta}_1,\cdots,\hat{\theta}_k\right)$来估计它，称之为矩法估计量。

注意：

（1）定义中选用的是原点矩，也可以用中心矩。只要给定总体矩，采用相应的样本矩估计即可。

（2）如果要估计的是函数$h(\theta_1,\cdots,\theta_k)$，则用$\hat{h}=h\left(\hat{\theta}_1,\cdots,\hat{\theta}_k\right)$来估计它，称之为矩法估计量。

（3）使用矩估计法的前提就是所需要的总体矩必须存在。

注意:

(1) 定义中选用的是原点矩，也可以用中心矩。只要给定总体矩，采用相应的样本矩估计即可。

(2) 如果要估计的是函数$h(\theta_1,\cdots,\theta_k)$，则用$\hat{h}=h\left(\hat{\theta}_1,\cdots,\hat{\theta}_k\right)$来估计它，称之为矩法估计量。

(3) 使用矩估计法的前提就是所需要的总体矩必须存在。

例3.设总体X服从$\pi(\lambda)$，其中$\lambda>0$是未知参数。试求参数λ的矩估计量。

注意：

（1）定义中选用的是原点矩，也可以用中心矩。只要给定总体矩，采用相应的样本矩估计即可。

（2）如果要估计的是函数$h(\theta_1,\cdots,\theta_k)$，则用$\hat{h}=h\left(\hat{\theta}_1,\cdots,\hat{\theta}_k\right)$来估计它，称之为矩法估计量。

（3）使用矩估计法的前提就是所需要的总体矩必须存在。

例3.设总体X服从$\pi(\lambda)$，其中$\lambda>0$是未知参数。试求参数λ的矩估计量。

例4 设总体X 的概率密度为

$$f(x;\theta)=\frac{1}{2\theta}e^{\frac{-|x|}{\theta}},\ \ -\infty<x<+\infty,\theta>0$$

试求参数θ的矩估计量。

§2. 最大似然估计法

最大似然估计（$Maximum\ Likelihood\ Estimation$，简记为MLE）的思想始于高斯的误差理论，由费歇尔（R. A. Fisher）于20世纪20年代将其作为一个一般的估计方法提出来。

§2. 最大似然估计法

最大似然估计（$Maximum\ Likelihood\ Estimation$，简记为MLE）的思想始于高斯的误差理论，由费歇尔（R. A. Fisher）于20世纪20年代将其作为一个一般的估计方法提出来。

一. 似然函数（$Likelihood\ function$）

§2. 最大似然估计法

最大似然估计（$Maximum\ Likelihood\ Estimation$，简记为MLE）的思想始于高斯的误差理论，由费歇尔（R. A. Fisher）于20世纪20年代将其作为一个一般的估计方法提出来。

一. 似然函数（$Likelihood\ function$）

1.离散型总体

§2. 最大似然估计法

最大似然估计（$Maximum\ Likelihood\ Estimation$，简记为MLE）的思想始于高斯的误差理论，由费歇尔（R. A. Fisher）于20世纪20年代将其作为一个一般的估计方法提出来。

一. 似然函数（$Likelihood\ function$）

1.离散型总体

$$L(\theta_1,\cdots,\theta_k)=\prod_{i=1}^{n}p(x_i;\theta_1,\cdots,\theta_k) \tag{0.4}$$

§2. 最大似然估计法

最大似然估计（$Maximum\ Likelihood\ Estimation$，简记为MLE）的思想始于高斯的误差理论，由费歇尔（R. A. Fisher）于20世纪20年代将其作为一个一般的估计方法提出来。

一. 似然函数（$Likelihood\ function$）

1.离散型总体

$$L(\theta_1,\cdots,\theta_k)=\prod_{i=1}^{n}p(x_i;\theta_1,\cdots,\theta_k) \tag{0.4}$$

2.连续型总体

§2. 最大似然估计法

最大似然估计（$Maximum\ Likelihood\ Estimation$，简记为MLE）的思想始于高斯的误差理论，由费歇尔（R. A. Fisher）于20世纪20年代将其作为一个一般的估计方法提出来。

一. 似然函数（$Likelihood\ function$）

1.离散型总体

$$L(\theta_1, \cdots, \theta_k) = \prod_{i=1}^{n} p(x_i; \theta_1, \cdots, \theta_k) \tag{0.4}$$

2.连续型总体

$$L(\theta_1, \cdots, \theta_k) = \prod_{i=1}^{n} f(x_i; \theta_1, \cdots, \theta_k) \tag{0.5}$$

二. 最大似然估计

二. 最大似然估计

Def.似然函数$L(x_1, x_2, \cdots, x_n; \theta)$，$x_1, x_2, \cdots, x_n$是样本观测值，$\theta = (\theta_1, \cdots, \theta_k)$为未知参数，$\Theta$是参数空间。若存在$\left(\hat{\theta}_1, \cdots, \hat{\theta}_k\right) \in \Theta$，使得

二. 最大似然估计

Def.似然函数$L(x_1,x_2,\cdots,x_n;\theta)$，$x_1,x_2,\cdots,x_n$是样本观测值，$\theta=(\theta_1,\cdots,\theta_k)$为未知参数，$\Theta$是参数空间。若存在$\left(\hat{\theta}_1,\cdots,\hat{\theta}_k\right)\in\Theta$，使得

$$L\left(x_1,x_2,\cdots,x_n;\hat{\theta}_1,\cdots,\hat{\theta}_k\right)=\max_{\theta_1,\cdots,\theta_k}L(x_1,x_2,\cdots,x_n;\theta_1,\cdots,\theta_k)$$

二. 最大似然估计

Def.似然函数$L\left(x_1, x_2, \cdots, x_n; \theta\right)$，$x_1, x_2, \cdots, x_n$是样本观测值，$\theta=\left(\theta_1, \cdots, \theta_k\right)$为未知参数，$\Theta$是参数空间。若存在$\left(\hat{\theta}_1, \cdots, \hat{\theta}_k\right) \in \Theta$，使得

$$L\left(x_1, x_2, \cdots, x_n; \hat{\theta}_1, \cdots, \hat{\theta}_k\right)=\max_{\theta_1, \cdots, \theta_k} L\left(x_1, x_2, \cdots, x_n; \theta_1, \cdots, \theta_k\right)$$

则称$\hat{\theta}_r=\hat{\theta}_r\left(x_1, x_2, \cdots, x_n\right)$（$r=1,2, \cdots, k$）为参数$\theta_1, \cdots, \theta_k$的最大似然估计值，称$\hat{\theta}_r=\hat{\theta}_r\left(X_1, X_2, \cdots, X_n\right)$为最大似然估计量。

二. 最大似然估计

Def.似然函数$L(x_1,x_2,\cdots,x_n;\theta)$，$x_1,x_2,\cdots,x_n$是样本观测值，$\theta=(\theta_1,\cdots,\theta_k)$为未知参数，$\Theta$是参数空间。若存在$\left(\hat\theta_1,\cdots,\hat\theta_k\right)\in\Theta$，使得

$$L\left(x_1,x_2,\cdots,x_n;\hat\theta_1,\cdots,\hat\theta_k\right)=\max_{\theta_1,\cdots,\theta_k} L(x_1,x_2,\cdots,x_n;\theta_1,\cdots,\theta_k)$$

则称$\hat\theta_r=\hat\theta_r(x_1,x_2,\cdots,x_n)$（$r=1,2,\cdots,k$）为参数$\theta_1,\cdots,\theta_k$的最大似然估计值，称$\hat\theta_r=\hat\theta_r(X_1,X_2,\cdots,X_n)$为最大似然估计量。

三. 似然方程/ 对数似然方程

二. 最大似然估计

Def.似然函数$L(x_1,x_2,\cdots,x_n;\theta)$，$x_1,x_2,\cdots,x_n$是样本观测值，$\theta=(\theta_1,\cdots,\theta_k)$为未知参数，$\Theta$是参数空间。若存在$\left(\hat{\theta}_1,\cdots,\hat{\theta}_k\right)\in\Theta$，使得

$$L\left(x_1,x_2,\cdots,x_n;\hat{\theta}_1,\cdots,\hat{\theta}_k\right)=\max_{\theta_1,\cdots,\theta_k}L(x_1,x_2,\cdots,x_n;\theta_1,\cdots,\theta_k)$$

则称$\hat{\theta}_r=\hat{\theta}_r(x_1,x_2,\cdots,x_n)$ $(r=1,2,\cdots,k)$ 为参数$\theta_1,\cdots,\theta_k$的最大似然估计值，称$\hat{\theta}_r=\hat{\theta}_r(X_1,X_2,\cdots,X_n)$为最大似然估计量。

三. 似然方程/ 对数似然方程

$$\frac{\partial L(\theta)}{\partial\theta_i}=0,\quad i=1,2,\cdots,k \tag{0.6}$$

二. 最大似然估计

Def.似然函数$L(x_1, x_2, \cdots, x_n; \theta)$，$x_1, x_2, \cdots, x_n$是样本观测值，$\theta = (\theta_1, \cdots, \theta_k)$为未知参数，$\Theta$是参数空间。若存在$\left(\hat{\theta}_1, \cdots, \hat{\theta}_k\right) \in \Theta$，使得

$$L\left(x_1, x_2, \cdots, x_n; \hat{\theta}_1, \cdots, \hat{\theta}_k\right) = \max_{\theta_1, \cdots, \theta_k} L(x_1, x_2, \cdots, x_n; \theta_1, \cdots, \theta_k)$$

则称$\hat{\theta}_r = \hat{\theta}_r(x_1, x_2, \cdots, x_n)$（$r = 1, 2, \cdots, k$）为参数$\theta_1, \cdots, \theta_k$的最大似然估计值，称$\hat{\theta}_r = \hat{\theta}_r(X_1, X_2, \cdots, X_n)$为最大似然估计量。

三. 似然方程/ 对数似然方程

$$\frac{\partial L(\theta)}{\partial \theta_i} = 0, \quad i = 1, 2, \cdots, k \tag{0.6}$$

$$\frac{\partial \ln L(\theta)}{\partial \theta_i} = 0, \quad i = 1, 2, \cdots, k \tag{0.7}$$

最大似然估计法的基本步骤如下:

最大似然估计法的基本步骤如下:

(1) 根据总体X的分布，建立似然函数$L(\theta_1, \cdots, \theta_k)$:

最大似然估计法的基本步骤如下:

(1) 根据总体X的分布，建立似然函数$L(\theta_1,\cdots,\theta_k)$:

(2) 给出对数似然函数$\ln L(\theta_1,\cdots,\theta_k)$;

最大似然估计法的基本步骤如下:

(1) 根据总体X的分布，建立似然函数$L(\theta_1, \cdots, \theta_k)$:

(2) 给出对数似然函数$\ln L(\theta_1, \cdots, \theta_k)$;

(3) 建立似然方程组:

$$\frac{\partial L(\theta)}{\partial \theta_i} = 0, \quad i = 1, 2, \cdots, k$$

最大似然估计法的基本步骤如下：

（1）根据总体X的分布，建立似然函数$L\left(\theta_1,\cdots,\theta_k\right)$：

（2）给出对数似然函数$\ln L\left(\theta_1,\cdots,\theta_k\right)$；

（3）建立似然方程组：

$$\frac{\partial L\left(\theta\right)}{\partial\theta_i}=0,\quad i=1,2,\cdots,k$$

或对数似然方程组：

$$\frac{\partial\ln L\left(\theta\right)}{\partial\theta_i}=0,\quad i=1,2,\cdots,k$$

最大似然估计法的基本步骤如下：

（1）根据总体X的分布，建立似然函数$L(\theta_1, \cdots, \theta_k)$：

（2）给出对数似然函数$\ln L(\theta_1, \cdots, \theta_k)$；

（3）建立似然方程组：

$$\frac{\partial L(\theta)}{\partial \theta_i} = 0, \quad i = 1, 2, \cdots, k$$

或对数似然方程组：

$$\frac{\partial \ln L(\theta)}{\partial \theta_i} = 0, \quad i = 1, 2, \cdots, k$$

（4）根据似然方程组或对数似然方程组求解最大似然估计量。

例1. 设总体X服从$Z(\alpha)$，其中$\alpha>0$是未知参数。$X_1, X_2, \cdots, X_n$是来自该总体的一个简单随机样本，试求参数α的最大似然估计量。

例1. 设总体X服从$Z(\alpha)$，其中$\alpha>0$是未知参数。$X_1, X_2, \cdots, X_n$是来自该总体的一个简单随机样本，试求参数α的最大似然估计量。

解：已知$X \sim Z(\alpha)$，其密度函数为

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

例1. 设总体X服从$Z(\alpha)$，其中$\alpha>0$是未知参数。$X_1, X_2, \cdots, X_n$是来自该总体的一个简单随机样本，试求参数α的最大似然估计量。

解：已知$X \sim Z(\alpha)$，其密度函数为

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

（1）建立似然函数：

例1. 设总体X服从$Z(\alpha)$，其中$\alpha>0$是未知参数。$X_1, X_2, \cdots, X_n$是来自该总体的一个简单随机样本，试求参数α的最大似然估计量。

解：已知$X \sim Z(\alpha)$，其密度函数为

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

（1）建立似然函数：

$$\begin{aligned} L(\alpha) &= L(x_1, x_2, \cdots, x_n; \alpha) \\ &= \prod_{i=1}^{n} \alpha e^{-\alpha x_i} \\ &= \alpha^n e^{-\alpha \sum_{i=1}^{n} x_i}, \quad \forall x_i > 0 \end{aligned}$$

例1. 设总体X服从$Z(\alpha)$，其中$\alpha>0$是未知参数。$X_1, X_2, \cdots, X_n$是来自该总体的一个简单随机样本，试求参数α的最大似然估计量。

解：已知$X \sim Z(\alpha)$，其密度函数为

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

（1）建立似然函数：

$$\begin{aligned} L(\alpha) &= L(x_1, x_2, \cdots, x_n; \alpha) \\ &= \prod_{i=1}^{n} \alpha e^{-\alpha x_i} \\ &= \alpha^n e^{-\alpha \sum_{i=1}^{n} x_i}, \quad \forall x_i > 0 \end{aligned}$$

（2）取对数，得到对数似然函数：

例1. 设总体X服从$Z(\alpha)$，其中$\alpha>0$是未知参数。$X_1, X_2, \cdots, X_n$是来自该总体的一个简单随机样本，试求参数α的最大似然估计量。

解：已知$X \sim Z(\alpha)$，其密度函数为

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

（1）建立似然函数：

$$\begin{aligned} L(\alpha) &= L(x_1, x_2, \cdots, x_n; \alpha) \\ &= \prod_{i=1}^{n} \alpha e^{-\alpha x_i} \\ &= \alpha^n e^{-\alpha \sum_{i=1}^{n} x_i}, \quad \forall x_i > 0 \end{aligned}$$

（2）取对数，得到对数似然函数：

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

（3）两边对α求导，建立对数似然方程:

（3）两边对α求导，建立对数似然方程：

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

（3）两边对α求导，建立对数似然方程:

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

（4）求解对数似然方程，得到α的最大似然估计值为

（3）两边对α求导，建立对数似然方程：

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

（4）求解对数似然方程，得到α的最大似然估计值为

$$\hat{\alpha}=\frac{n}{\sum_{i=1}^{n}x_i}=\frac{1}{\overline{x}}$$

（3）两边对α求导，建立对数似然方程：

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

（4）求解对数似然方程，得到α的最大似然估计值为

$$\hat{\alpha}=\frac{n}{\sum_{i=1}^{n}x_i}=\frac{1}{\overline{x}}$$

故其最大似然估计量为$\hat{\alpha}=\frac{1}{\overline{X}}$

（3）两边对α求导，建立对数似然方程：

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

（4）求解对数似然方程，得到α的最大似然估计值为

$$\hat{\alpha}=\frac{n}{\sum_{i=1}^{n}x_i}=\frac{1}{\overline{x}}$$

故其最大似然估计量为$\hat{\alpha}=\frac{1}{\overline{X}}$

例2. 设总体X服从$\pi(\lambda)$，其中$\lambda>0$是未知参数。$X_1,X_2,\cdots,X_n$是来自该总体的一个简单随机样本，试求参数λ的最大似然估计量。

例3. 设总体X服从正态分布$N\left(\mu,\sigma^2\right)$，其中$\mu$和$\sigma^2$是未知参数。已知$X_1,X_2,\cdots,X_n$是来自该总体的一个简单随机样本，试给出$\mu$和$\sigma^2$的最大似然估计量。

例3. 设总体X服从正态分布$N\left(\mu,\sigma^2\right)$，其中$\mu$和$\sigma^2$是未知参数。已知$X_1,X_2,\cdots,X_n$是来自该总体的一个简单随机样本，试给出$\mu$和$\sigma^2$的最大似然估计量。

例4.（接上例）若μ是已知的，而σ^2是未知参数。试给出σ^2的最大似然估计量。

例3. 设总体X服从正态分布$N\left(\mu,\sigma^2\right)$，其中$\mu$和$\sigma^2$是未知参数。已知$X_1$,$X_2$,$\cdots$,$X_n$是来自该总体的一个简单随机样本，试给出$\mu$和$\sigma^2$的最大似然估计量。

例4.（接上例）若μ是已知的，而σ^2是未知参数。试给出σ^2的最大似然估计量。

例5. 设总体X在$[a,b]$上服从均匀分布，a、b未知。试给出a和b的最大似然估计量。

例3. 设总体X服从正态分布$N\left(\mu,\sigma^2\right)$，其中$\mu$和$\sigma^2$是未知参数。已知$X_1,X_2,\cdots,X_n$是来自该总体的一个简单随机样本，试给出$\mu$和$\sigma^2$的最大似然估计量。

例4.（接上例）若μ是已知的，而σ^2是未知参数。试给出σ^2的最大似然估计量。

例5. 设总体X在$[a,b]$上服从均匀分布，a、b未知。试给出a和b的最大似然估计量。

例6. 设总体X的密度函数为

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

参数$\lambda>0$。已知$X_1,X_2,\cdots,X_n$是来自该总体的一个简单随机样本，求参数λ和μ的最大似然估计量。

注意以下几点:

注意以下几点:

(1) 似然函数的给出是使用最大似然估计法的基本前提，因此要求总体X的分布律或密度函数的表达式是已知的;

注意以下几点：

（1）似然函数的给出是使用最大似然估计法的基本前提，因此要求总体X的分布律或密度函数的表达式是已知的；

（2）多数情况下，可以通过求导或求偏导，建立方程或方程组来求解最大似然估计量；

注意以下几点:

(1) 似然函数的给出是使用最大似然估计法的基本前提，因此要求总体X的分布律或密度函数的表达式是已知的;

(2) 多数情况下，可以通过求导或求偏导，建立方程或方程组来求解最大似然估计量;

(3) 当待估参数是函数的分段点时，通常可以利用顺序统计量确定最大似然估计量。

注意以下几点:

(1) 似然函数的给出是使用最大似然估计法的基本前提，因此要求总体X的分布律或密度函数的表达式是已知的;

(2) 多数情况下，可以通过求导或求偏导，建立方程或方程组来求解最大似然估计量;

(3) 当待估参数是函数的分段点时，通常可以利用顺序统计量确定最大似然估计量。

命题. θ是总体X的分布的未知参数，$u = u(\theta)$ ($\theta \in \Theta$) 是θ的函数，具有单值反函数$\theta = \theta(u)$，且设$\hat{\theta}$是θ的最大似然估计，则$\hat{u} = u\left(\hat{\theta}\right)$是$u(\theta)$的最大似然估计。

注意以下几点：

(1) 似然函数的给出是使用最大似然估计法的基本前提，因此要求总体X的分布律或密度函数的表达式是已知的；

(2) 多数情况下，可以通过求导或求偏导，建立方程或方程组来求解最大似然估计量；

(3) 当待估参数是函数的分段点时，通常可以利用顺序统计量确定最大似然估计量。

命题. θ是总体X的分布的未知参数，$u = u(\theta)$ $(\theta \in \Theta)$ 是θ的函数，具有单值反函数$\theta = \theta(u)$，且设$\hat{\theta}$是θ的最大似然估计，则$\hat{u} = u\left(\hat{\theta}\right)$是$u(\theta)$的最大似然估计。

上述性质也称为最大似然估计的不变性，在一定情况下可以用来估计参数的函数。

§3. 估计量的评选标准

§ 3. 估计量的评选标准

$$
\begin{aligned}
MSE\left(\hat{\theta}\right) &= E\left[\left(\hat{\theta}-\theta\right)^{2}\right] \\
&= E\left[\left(\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)+\left(E\left(\hat{\theta}\right)-\theta\right)\right)^{2}\right] \\
&= E\left[\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)^{2}\right]+\left(E\left(\hat{\theta}\right)-\theta\right)^{2} \\
&= D\left(\hat{\theta}\right)+\left(E\left(\hat{\theta}\right)-\theta\right)^{2}
\end{aligned}
$$

§ 3. 估计量的评选标准

$$
\begin{aligned}
MSE\left(\hat{\theta}\right) &= E\left[\left(\hat{\theta}-\theta\right)^2\right] \\
&= E\left[\left(\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)+\left(E\left(\hat{\theta}\right)-\theta\right)\right)^2\right] \\
&= E\left[\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)^2\right]+\left(E\left(\hat{\theta}\right)-\theta\right)^2 \\
&= D\left(\hat{\theta}\right)+\left(E\left(\hat{\theta}\right)-\theta\right)^2
\end{aligned}
$$

一. 无偏性（unbiasedness）

§ 3. 估计量的评选标准

$$
\begin{aligned}
MSE\left(\hat{\theta}\right) &= E\left[\left(\hat{\theta}-\theta\right)^2\right] \\
&= E\left[\left(\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)+\left(E\left(\hat{\theta}\right)-\theta\right)\right)^2\right] \\
&= E\left[\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)^2\right]+\left(E\left(\hat{\theta}\right)-\theta\right)^2 \\
&= D\left(\hat{\theta}\right)+\left(E\left(\hat{\theta}\right)-\theta\right)^2
\end{aligned}
$$

一. 无偏性（unbiasedness）

Def.设$\hat{\theta}=\hat{\theta}\left(X_1,X_2,\cdots,X_n\right)$是未知参数$\theta$的一个估计量，若

§ 3. 估计量的评选标准

$$
\begin{aligned}
MSE\left(\hat{\theta}\right) &= E\left[\left(\hat{\theta}-\theta\right)^2\right] \\
&= E\left[\left(\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)+\left(E\left(\hat{\theta}\right)-\theta\right)\right)^2\right] \\
&= E\left[\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)^2\right]+\left(E\left(\hat{\theta}\right)-\theta\right)^2 \\
&= D\left(\hat{\theta}\right)+\left(E\left(\hat{\theta}\right)-\theta\right)^2
\end{aligned}
$$

一. 无偏性（unbiasedness）

Def.设$\hat{\theta}=\hat{\theta}\left(X_1,X_2,\cdots,X_n\right)$是未知参数$\theta$的一个估计量，若

$$E\left(\hat{\theta}\right)=\theta \tag{0.8}$$

§ 3. 估计量的评选标准

$$
\begin{aligned}
MSE\left(\hat{\theta}\right) &= E\left[\left(\hat{\theta}-\theta\right)^2\right] \\
&= E\left[\left(\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)+\left(E\left(\hat{\theta}\right)-\theta\right)\right)^2\right] \\
&= E\left[\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)^2\right]+\left(E\left(\hat{\theta}\right)-\theta\right)^2 \\
&= D\left(\hat{\theta}\right)+\left(E\left(\hat{\theta}\right)-\theta\right)^2
\end{aligned}
$$

一. 无偏性（unbiasedness）

Def.设$\hat{\theta}=\hat{\theta}\left(X_1,X_2,\cdots,X_n\right)$是未知参数$\theta$的一个估计量，若

$$E\left(\hat{\theta}\right)=\theta \tag{0.8}$$

则称$\hat{\theta}=\hat{\theta}\left(X_1,X_2,\cdots,X_n\right)$是$\theta$的无偏估计量。如果$\hat{\theta}$不是无偏的，

其差

$$b_n = E\left(\hat{\theta}\right) - \theta \tag{0.9}$$

称为$\hat{\theta}$的偏差。

其差

$$b_n = E\left(\hat{\theta}\right) - \theta \tag{0.9}$$

称为$\hat{\theta}$的偏差。

特别，若$\lim\limits_{n\to\infty} b_n = 0$，则称$\hat{\theta}$是$\theta$的渐近无偏估计量。

其差
$$b_n = E\left(\hat{\theta}\right) - \theta \tag{0.9}$$
称为$\hat{\theta}$的偏差。

特别，若$\lim_{n\to\infty} b_n = 0$，则称$\hat{\theta}$是$\theta$的渐近无偏估计量。

例1.设$X_1,X_2,\cdots,X_n$是来自总体X的样本，若$D(X) = \sigma^2$是有限的。试证：样本二阶中心矩S_n^2是σ^2的有偏估计量。

其差

$$b_n = E\left(\hat{\theta}\right) - \theta \tag{0.9}$$

称为$\hat{\theta}$的偏差。

特别，若 $\lim\limits_{n\to\infty} b_n = 0$，则称$\hat{\theta}$是$\theta$的渐近无偏估计量。

例1.设$X_1, X_2, \cdots, X_n$是来自总体X的样本，若$D(X) = \sigma^2$是有限的。试证：样本二阶中心矩S_n^2是σ^2的有偏估计量。

命题. 设$X_1, X_2, \cdots, X_n$是来自总体X的简单随机样本。无论总体X服从什么分布，只要其数学期望μ和方差σ^2存在且有限，则

其差

$$b_n = E\left(\hat{\theta}\right) - \theta \tag{0.9}$$

称为$\hat{\theta}$的偏差。

特别，若$\lim\limits_{n\to\infty} b_n = 0$，则称$\hat{\theta}$是$\theta$的渐近无偏估计量。

例1.设$X_1,X_2,\cdots,X_n$是来自总体X的样本，若$D(X) = \sigma^2$是有限的。试证：样本二阶中心矩S_n^2是σ^2的有偏估计量。

命题. 设$X_1,X_2,\cdots,X_n$是来自总体X的简单随机样本。无论总体X服从什么分布，只要其数学期望μ和方差σ^2存在且有限，则

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

分别是μ和σ^2的无偏估计量。

其差

$$b_n = E\left(\hat{\theta}\right) - \theta \tag{0.9}$$

称为$\hat{\theta}$的偏差。

特别，若$\lim\limits_{n\to\infty} b_n = 0$，则称$\hat{\theta}$是$\theta$的渐近无偏估计量。

例1. 设$X_1,X_2,\cdots,X_n$是来自总体X的样本，若$D\left(X\right) = \sigma^2$是有限的。试证：样本二阶中心矩$S_n^2$是$\sigma^2$的有偏估计量。

命题. 设$X_1,X_2,\cdots,X_n$是来自总体X的简单随机样本。无论总体X服从什么分布，只要其数学期望μ和方差σ^2存在且有限，则

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

分别是μ和σ^2的无偏估计量。

注意：当$\hat{\theta}$是θ的无偏估计量时，$g\left(\hat{\theta}\right)$并不一定是$g\left(\theta\right)$的无偏估计量。

例2. 设总体X服从指数分布$Exp\left(\frac{1}{\theta}\right)$，其密度函数为

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}}, & x\geq 0\\ 0, & x<0\end{cases}$$

其中θ是未知参数。$X_1,X_2,\cdots,X_n$是简单随机样本，$X_{(1)}\leq X_{(2)}\leq\cdots\leq X_{(n)}$是相应的顺序统计量。

试证明$\overline{X}$和$nX_{(1)}$都是θ的无偏估计量。

例2. 设总体X服从指数分布$Exp\left(\frac{1}{\theta}\right)$，其密度函数为

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}}, & x\geq 0\\ 0, & x<0\end{cases}$$

其中θ是未知参数。$X_1,X_2,\cdots,X_n$是简单随机样本，$X_{(1)}\leq X_{(2)}\leq\cdots\leq X_{(n)}$是相应的顺序统计量。

试证明$\overline{X}$和$nX_{(1)}$都是θ的无偏估计量。

二. 有效性（efficiency）

例2. 设总体X服从指数分布$Exp\left(\frac{1}{\theta}\right)$，其密度函数为

$$f\left(x;\theta\right)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}}, & x\geq 0\\ 0, & x<0\end{cases}$$

其中θ是未知参数。X_1,X_2,$\cdots$,X_n是简单随机样本，$X_{(1)}\leq X_{(2)}\leq\cdots\leq X_{(n)}$是相应的顺序统计量。

试证明$\overline{X}$和$nX_{(1)}$都是θ的无偏估计量。

二. 有效性（efficiency）

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

$$D\left(\hat{\theta}_1\right)<D\left(\hat{\theta}_2\right) \tag{0.10}$$

则称$\hat{\theta}_1$比$\hat{\theta}_2$有效。

例3.设总体X在$[0,\theta]$上服从均匀分布，$\theta > 0$是未知参数。

例3.设总体X在$[0,\theta]$上服从均匀分布，$\theta>0$是未知参数。

证明$\hat{\theta}_1=2\overline{X}$和$\hat{\theta}_2=\frac{n+1}{n}X_{(n)}$都是$\theta$的无偏估计量。并说明$\hat{\theta}_1$和$\hat{\theta}_2$哪一个更有效?

例3.设总体X在$[0,\theta]$上服从均匀分布，$\theta>0$是未知参数。

证明$\hat{\theta}_1=2\overline{X}$和$\hat{\theta}_2=\frac{n+1}{n}X_{(n)}$都是$\theta$的无偏估计量。并说明$\hat{\theta}_1$和$\hat{\theta}_2$哪一个更有效?

三. 相合性（consistency）

例3.设总体X在$[0,\theta]$上服从均匀分布，$\theta>0$是未知参数。

证明$\hat{\theta}_1=2\overline{X}$和$\hat{\theta}_2=\frac{n+1}{n}X_{(n)}$都是$\theta$的无偏估计量。并说明$\hat{\theta}_1$和$\hat{\theta}_2$哪一个更有效?

三. 相合性（consistency）

Def. 设$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$是未知参数$\theta$的估计量，若对于任意的$\theta\in\Theta$，当$n\to\infty$时，$\hat{\theta}$依概率收敛于$\theta$，则称$\hat{\theta}$是$\theta$的相合估计量，或一致估计量。即$\forall\varepsilon>0$，有

$$\lim_{n\to\infty}P\left\{|\hat{\theta}-\theta|<\varepsilon\right\}=1 \tag{0.11}$$

记为$\hat{\theta}\xrightarrow{P}\theta$。

例3.设总体X在$[0,\theta]$上服从均匀分布，$\theta>0$是未知参数。

证明$\hat{\theta}_1=2\overline{X}$和$\hat{\theta}_2=\frac{n+1}{n}X_{(n)}$都是$\theta$的无偏估计量。并说明$\hat{\theta}_1$和$\hat{\theta}_2$哪一个更有效?

三. 相合性（consistency）

Def. 设$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$是未知参数$\theta$的估计量，若对于任意的$\theta\in\Theta$，当$n\to\infty$时，$\hat{\theta}$依概率收敛于$\theta$，则称$\hat{\theta}$是$\theta$的相合估计量，或一致估计量。即$\forall\varepsilon>0$，有

$$\lim_{n\to\infty}P\left\{|\hat{\theta}-\theta|<\varepsilon\right\}=1 \tag{0.11}$$

记为$\hat{\theta}\xrightarrow{P}\theta$。

相合性是评价估计量优劣的一个重要标准，也是对估计量的一个基本要求。如果一个估计量不具有相合性，那么随着样本容量n的增大，对未知参数估计的精度并不一定能提高，估计量甚至会明显偏离被估参数。