

第七章 参数估计



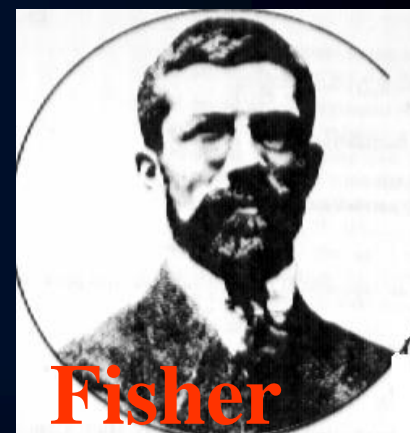
极大似然法

它首先是由德国数学家
高斯在1821年提出的，



然而，这个方法常归功于
英国统计学家费歇。

费歇在1922年重新发现了
这一方法，并首先研究了这
种方法的一些性质。



极大似然法的基本思想

先看一个简单例子：

某位同学与一位猎人一起外出打猎。

一只野兔从前方窜过。

只听一声枪响，野兔应声倒下。



如果要你推测，
是谁打中的呢？
你会如何想呢？

思想方法：一次试验就出现的
事件有较大的概率

猎人

哪个概率大？

同学

一枪打中

猎人命中的概率一般大于这位同学命中的概率. 看来这一枪是猎人射中的.

其数学模型为

令 X 为打一枪的中弹数，则 $X \sim B(1, p)$, p 未知.

设想我们事先知道 p 只有两种可能:

$p=0.9$ 或 $p=0.1$

要估计总体 X 的参数 p 的值

当兔子中弹，即 $\{X=1\}$ 发生了。

若 $p=0.9$ ，则 $P\{X=1\}=0.9$;

若 $p=0.1$ ，则 $P\{X=1\}=0.1$

当兔子不中弹，即 $\{X=0\}$ 发生了。

若 $p=0.9$ ，则 $P\{X=0\}=0.1$;

若 $p=0.1$ ，则 $P\{X=0\}=0.9$

现有样本观测值 $x=1$, 什么样的参数使该样本值出现的可能性最大呢?

总体的分布函数为 $F(x, \theta)$ ，其中 θ 为未知参数，现从该总体抽样，得样本 X_1, X_2, \dots, X_n 要依据该样本对参数 θ 作出估计，

最大似然估计法的基本思想：

一般地，对于给定的样本值 x_1, x_2, \dots, x_n 选择参数 θ 的估计，使得样本在 样本值 x_1, x_2, \dots, x_n 附近出现 的可能性最大

(1).若总体 X 属离散型, 其分布律 $P\{X = x\} = p(x; \theta)$, $\theta \in \Theta$ 的形式为已知, θ 为待估参数, Θ 是 θ 可能取值的范围。

设 X_1, \dots, X_n 是来自 X 的样本, 则 X_1, \dots, X_n 的联合分布律

$$\prod_{i=1}^n p(x_i; \theta)$$

事件 $\{X_1 = x_1, \dots, X_n = x_n\}$ 发生的概率为:

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

它是 θ 的函数, $L(\theta)$ 称为样本的似然函数.

(2).若总体 X 属连续型，其概率密度 $f(x; \theta), \theta \in \Theta$ 的形式已知， θ 为待估参数；

设 X_1, \dots, X_n 是来自 X 的样本，则 X_1, \dots, X_n 的联合密度：

$$\prod_{i=1}^n f(x_i; \theta)$$

设 x_1, \dots, x_n 是相应 X_1, \dots, X_n 的一个样本值，则随机点 (X_1, \dots, X_n) 落在 (x_1, \dots, x_n) 的附近概率

$$\prod_{i=1}^n f(x_i; \theta)$$

考虑函数

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

$L(\theta)$ 称为样本的似然函数.

似然函数

设总体 X 的概率密度（或概率函数）为 $f(x;\theta)$ ， $\theta=(\theta_1,\dots,\theta_m)\in\Theta\subset R^m$ ， X_1,\dots,X_n 为来自总体的样本，则 (X_1,\dots,X_n) 的密度函数（或概率函数）为

$$f(x_1, x_2 \cdots x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

若已知样本观测值 (x_1, \dots, x_n) ，则令

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

称其为样本值 (x_1, \dots, x_n) 的似然函数

注意

1 $L(\theta; x_1, \dots, x_n)$ 中 θ 是自变量, x_1, \dots, x_n 是参变量..

2 似然函数

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

表示在参数 θ 下 样本在 x_1, x_2, \dots, x_n

附近出现的可能性大小

最大似然估计

如果似然函数 $L(\theta; x_1, \cdots, x_n)$
在 $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_m)$ 达到最大值, 即

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

则称 $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_m)$ 为 θ 的最大似然估计。

求极大似然估计(MLE)的一般步骤是:

- (1) 由总体分布导出样本的似然函数;
- (2) 求似然函数 $L(\theta)$ 的最大值点(常常转化为求 $\ln L(\theta)$ 的最大值点), 即 θ 的MLE;
- (3) 在最大值点的表达式中, 用样本值代入就得参数的极大似然估计值.

若 Θ 是开集, L 是 θ 的可微函数,

解似然方程组

$$\frac{\partial}{\partial \theta_r} L(\theta; x_1, x_2, \dots, x_n) = 0 \quad r = 1, 2, \dots, k$$

可求得未知参数的极大似然估计 $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$

例1 设 X_1, X_2, \dots, X_n 是取自总体 $X \sim B(1, p)$ 的一个样本，求参数 p 的极大似然估计。

解： $P(X = x) = p^x (1 - p)^{1-x}, x = 0, 1$

x_1, x_2, \dots, x_n 似然函数为：

$$\begin{aligned} L(p) &= \prod_{i=1}^n P(X_i = x_i) \\ &= \prod_{i=1}^n p^{x_i} (1 - p)^{1-x_i} \\ &= p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i} \end{aligned}$$

$$X_i \sim \begin{Bmatrix} 0 & 1 \\ 1-p & p \end{Bmatrix}$$

$$x_i \in \{0, 1\}, i = 1, 2, \dots, n$$

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

对数似然函数为:

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

对数似然方程为

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

得 $\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n$

容易验证 $\hat{p} = \bar{x}$ 为 $L(p)$ 的最大值点 即为 p 的MLE.

例2 设 X_1, X_2, \dots, X_n 独立同分布，都服从 *Poisson* 分布 $P(\lambda)$ ，给定观测数据 x_1, x_2, \dots, x_n ，试求参数 λ 的最大似然估计。

解： X 的分布列为

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}, x = 0, 1, \dots$$

因此似然函数为

$$\begin{aligned} L(\lambda) &= P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \frac{\lambda^{x_2}}{x_2!} e^{-\lambda} \dots \frac{\lambda^{x_n}}{x_n!} e^{-\lambda} \end{aligned}$$

对数似然函数为：

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

令

$$\frac{dl(\lambda)}{d\lambda} = \frac{1}{\lambda} \sum_{i=1}^n x_i - n = 0$$

得 λ 的最大似然估计为

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n$$

例3. 设 $X \sim N(\mu, \sigma^2)$, μ, σ^2 为未知参数,

x_1, \dots, x_n 是来自 X 的一个样本值,

求: μ, σ^2 的最大似然估计.

解: X 的概率密度为:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\}$$

似然函数为:

$$L(\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu)^2\right\}$$

对数似然函数为：

$$l(\mu, \sigma^2) = \ln L(\mu, \sigma^2)$$

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

$$\text{令} \begin{cases} \frac{\partial l}{\partial \mu} = 0 \\ \frac{\partial l}{\partial \sigma^2} = 0 \end{cases}$$

$$\text{即:} \begin{cases} \frac{1}{\sigma^2} [\sum_{i=1}^n x_i - n\mu] = 0 \\ -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - \mu)^2 = 0 \end{cases}$$

解得 μ 和 σ^2 的最大似然估计为

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

例4 设 X_1, X_2, \dots, X_n 是取自总体 X 的一个样本

$$X \sim f(x) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1 \\ 0, & \text{其它} \end{cases} \quad \text{其中 } \theta > 0,$$

求 θ 的极大似然估计.

解: x_1, x_2, \dots, x_n 似然函数为

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \theta x_i^{\theta-1} \\ &= \theta^n \left(\prod_{i=1}^n x_i \right)^{\theta-1} \quad (0 < x_i < 1) \quad 1 \leq i \leq n \end{aligned}$$

对数似然函数为

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

对数似然函数为

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

似然函数为

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

从中解得

$$\hat{\theta} = -n / \sum_{i=1}^n \ln x_i \quad \text{即为 } \theta \text{ 的MLE.}$$

例5 设总体 X 的概率分布为

X	0	1	2
P	θ	$1-2\theta$	θ

其中 $0 < \theta < \frac{1}{2}$ 为未知参数。今对 X 进行观测，得如下样本值

0, 1, 2, 0, 2, 1

求 θ 的最大似然估计值。

X	0	1	2
P	θ	$1-2\theta$	θ

解：样本值 0, 1, 2, 0, 2, 1 似然函数为

$$L(\theta) = P(X_1 = 0)P(X_2 = 1)P(X_3 = 2)$$

$$P(X_4 = 0)P(X_5 = 2)P(X_6 = 1)$$

$$= \theta^4 (1 - 2\theta)^2,$$

对数似然函数为

$$\ln L(\theta) = 4 \ln \theta + 2 \ln(1 - 2\theta)$$

似然方程为

$$\frac{d \ln L(\theta)}{d \theta} = \frac{4}{\theta} - \frac{4}{1 - 2\theta} = 0$$

从中解得 $\hat{\theta} = \frac{1}{3}$

即为 θ 的 MLE .

若

L 不是 θ 的可微函数, 需用其它方法求极大似然估计.

例6 设 X_1, X_2, \dots, X_n 是取自总体 $X \sim U(a, b)$ 的一个样本，求参数 a, b 的最大似然估计.

解 X 的密度函数为

$$f(x; a, b) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{其它} \end{cases}$$

似然函数为

$$L(a, b) = \begin{cases} \frac{1}{(b-a)^n}, & a < x_i < b, \\ & i = 1, 2, \dots, n \\ 0, & \text{其它} \end{cases}$$

$$\begin{aligned}\ln L(a, b) &= \ln \frac{1}{(b-a)^n} = -\ln(b-a)^n \\ &= -n \ln(b-a)\end{aligned}$$

$$\begin{cases} \frac{\partial \ln L(a, b)}{\partial a} = \frac{n}{b-a} = 0 \\ \frac{\partial \ln L(a, b)}{\partial b} = -\frac{n}{b-a} = 0 \end{cases}$$

不能求解。

$$L(a,b) = \begin{cases} \frac{1}{(b-a)^n}, & a \leq x_i \leq b, \\ 0, & i = 1, 2, \dots, n \\ \text{其它} \end{cases}$$

令 $x_{(1)} = \min \{x_1, x_2, \dots, x_n\}$
 $x_{(n)} = \max \{x_1, x_2, \dots, x_n\}$

因为 $a \leq x_1, \dots, x_n \leq b$, 等价于 $a \leq x_{(1)}, x_{(n)} \leq b$,

$$L(a,b) = \begin{cases} \frac{1}{(b-a)^n}, & a \leq x_{(1)}, b \geq x_{(n)}; \\ 0, & \text{其它} \end{cases}$$

似然函数 a 越大, b 越小, L 越大.

$$\text{取 } \hat{a} = x_{(1)}, \hat{b} = x_{(n)}$$

则对满足 $a \leq x_{(1)} \leq x_{(n)} \leq b$ 的一切 a, b , 都有

$$\frac{1}{(b-a)^n} \leq \frac{1}{(x_{(n)} - x_{(1)})^n}$$

故 $\hat{a} = x_{(1)}, \hat{b} = x_{(n)}$

是 a, b 的最大似然估计值.

练习

设 X_1, X_2, \dots, X_n 是取自总体 $X \sim U(0, \theta)$ 的一个样本，求参数 θ 的最大似然估计.

最大似然估计的不变性

设 θ 的函数 $u = u(\theta)$, $\theta \in \Theta$. 假设 $\hat{\theta}$ 是 X 的概率分布中参数 θ 的最大似然估计. 则 $\hat{u} = u(\hat{\theta})$ 是 $u(\theta)$ 的最大似然估计。

例：正态总体X服从 $N(\mu, \sigma^2)$,

$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ 是 σ^2 的最大似然估计

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

是标准差 σ 的最大似然估计

练习：求 $\theta = \frac{\sigma}{\mu}$ 最大似然估计

故 $\hat{\theta} = \frac{\hat{\sigma}}{\hat{\mu}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} / \bar{X}$

三 估计量的评选标准

对于同一个未知参数, 不同的方法得到的估计量可能不同, 于是提出问题

- ▲ 应该选用哪一种估计量?
- ▲ 用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) 一致性

1. 无偏性

定义 设 (X_1, X_2, \dots, X_n) 是总体 X 的样本
 $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ 是总体参数 θ 的估计量
 $E(\hat{\theta})$ 存在, 且对于任意 $\theta \in \Theta$ 都有

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

则称 $\hat{\theta}$ 是 θ 的无偏估计量.

无偏估计量仅在多次重复使用时才显示其优越性.

关于无偏性的常用结论

设总体 X 的期望 $E(X)$ 与方差 $D(X)$ 存在,

(X_1, X_2, \dots, X_n) 是 X 的一个样本, $n > 1$

(1) 样本均值 \bar{X} 是总体期望 $E(X)$ 的无偏估计量

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

例1 设 X_1, X_2, \dots, X_n 是来自总体 X 的样本，且 $E(X)=\mu$ 。以下两个估计是否为 μ 的无偏估计

$$\hat{\mu}_1 = \frac{1}{2} X_1 + \frac{3}{8} X_2 + \frac{1}{8} X_3$$

$$\hat{\mu}_2 = \frac{1}{3} X_1 + \frac{3}{4} X_2 - \frac{1}{12} X_3$$

例2 设总体 $X \sim N(\mu, \sigma^2)$, 其中参数 μ, σ^2 未知, 试用极大似然估计法求 μ, σ^2 的估计量, 并问是否是无偏估计? 若不是, 请修正它成为无偏估计。

解

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

似然
方程
组为

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

$$\longrightarrow \begin{cases} \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \\ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \end{cases}$$

μ, σ^2 的极大似然估计分别为

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

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

显然 $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$ 是 μ 的无偏估计

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

$$E(\hat{\sigma}^2) = \frac{n-1}{n} E(S^2) = \frac{n-1}{n} \sigma^2$$

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

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

例3 设总体 X 服从参数为 $1/\theta$ 的指数分布, 概率密度为

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

其中, 参数 $\theta > 0$ 为未知, X_1, \dots, X_n 为来自总体的样本. 试证, \bar{X} 和 $nZ = n\{\min(X_1, \dots, X_n)\}$ 都是 θ 的无偏估计.

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

证

$$\text{故 } E(\bar{X}) = E(X) = \theta$$

\bar{X} 是 θ 的无偏估计量.

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

令 $Z = \min \{ X_1, X_2, \dots, X_n \}$

$$F_Z(z) = 1 - (1 - F_X(z))^n = \begin{cases} 0 & z < 0 \\ 1 - e^{-\frac{nz}{\theta}} & z \geq 0 \end{cases}$$

$$f_Z(z) = \begin{cases} 0 & z < 0 \\ \frac{n}{\theta} e^{-\frac{nz}{\theta}} & z \geq 0 \end{cases}$$

即 $E(Z) = \theta / n \quad E(nZ) = \theta$

故 nZ 是 θ 的无偏估计量.

back

一个参数往往有不只一个无偏估计, 若 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是参数 θ 的无偏估计量, 我们可以比较 $D(\hat{\theta}_1)$ 和 $D(\hat{\theta}_2)$ 的大小来决定二者谁更优.

无偏估计以方差小者为好, 这就引进了有效性这一概念.

2. 有效性

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

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

且存在 $D(\hat{\theta}_1) < D(\hat{\theta}_2)$ 的情形

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

例4 设 (X_1, X_2, \dots, X_n) 为 X 的一个样本, 密度函数为

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x > 0, \\ 0 & x \leq 0 \end{cases} \quad \theta > 0 \text{ 为常数}$$

由前面例2 可知, \bar{X} 与 $n \min\{X_1, X_2, \dots, X_n\}$ 都是 θ 的无偏估计量, 问哪个估计量更有效?

解

$$D(\bar{X}) = \frac{\theta^2}{n}, \quad Z = \min(X_1, X_2, \dots, X_n) \sim E\left(\frac{n}{\theta}\right)$$
$$D(Z) = \frac{\theta^2}{n^2} \quad D(nZ) = \theta^2$$

所以, \bar{X} 比 $n \min\{X_1, X_2, \dots, X_n\}$ 更有效. **back**

3. 一致性 (相合性)

设 $\hat{\theta}$ 为未知参数 θ 的估计量, 若对任意给定的 $\varepsilon > 0$ 都有

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

则称 $\hat{\theta}$ 为参数 θ 的一致(相合)估计量

一致性估计量仅在样本容量 n 足够大时, 才显示其优越性.

关于一致性的常用结论

设总体的 k 阶矩存在，则样本的 k 阶矩是总体 k 阶矩的一致估计量

练习:

设总体 X 的概率密度为

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

其中 $\alpha > -1$ 为未知参数。设 x_1, \dots, x_n 为来自该总体的样本，试求

- (1) 参数 α 的极大似然估计;
- (2) 参数 α 的矩估计。

作业1

1, 2, 4, 5, 7, 9, 10, 11



请休息片刻

马远 踏歌图