参数估计

参数(parameter)是刻画总体某方面概率特性的数量.

参数估计(parameter estimation)：对总体的分布形式已知，但其中的某些参数未知，从总体抽取样本，用某种方法对这个参数进行估计.（机器学习中称为训练，training）

- 点估计 —— 估计未知参数的值
- 区间估计——估计未知参数的取值范围，并使此范围包含未知参数真值的概率为给定的值.

1　参数的点估计

- 矩估计
- 最大似然估计
- Bayes估计
- 点估计的**评价标准**

参数的点估计

根据样本构造一个统计量$\hat{\theta} = \hat{\theta}(X_1, \cdots, X_n)$，用它估计未知参数$\theta$，称为点估计。

$\hat{\theta}(X_1, \cdots, X_n)$称为$\theta$ 的估计量(estimator)

$\hat{\theta}(x_1, \cdots, x_n)$称为$\theta$ 的估计值 (estimate)

点估计的常用方法：

- 矩估计法
- 最大似然估计法
- 最大后验估计

矩估计

例 X—某品牌手机的待机时间，预估其$\mu = E(X)$，抽取$X_1, X_2, \cdots, X_{100}$

$$\hat{\mu} = \frac{1}{100}\sum_{i=1}^{100} X_i = \bar{X}$$

用样本矩去估计相应的总体矩的估计方法称为**矩估计法**（*method of moments,* MoM）:

① 用样本 k 阶原点矩作为总体 k 阶原点矩的估计量，建立含有待估参数的方程,

② 解出待估参数。

理论依据：大数定律

设待估计的参数为$\theta_1, \theta_2, \cdots, \theta_m$

总体的 k 阶原点矩存在，记为 $E(X^k) = \mu_k(\theta_1, \theta_2, \cdots, \theta_m)$

样本$X_1, X_2, \ldots, X_n$ 的k阶原点矩为$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

令 $\boxed{\mu_k(\theta_1, \theta_2, \cdots, \theta_m) = A_k \ , \quad k = 1,2,\cdots,m}$

建立含未知参数 $\theta_1, \theta_2, \ldots, \theta_m$ 的方程组

解方程组，得 m 个统计量:

$$\left.\begin{array}{l}\hat{\theta}_1(X_1, X_2, \cdots, X_n) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{\theta}_m(X_1, X_2, \cdots, X_n)\end{array}\right\}$$

未知参数 $\theta_1, \ldots, \theta_m$ 的矩估计量

代入一组样本值得未知参数 $\theta_1, \ldots, \theta_m$ 的矩估计值

$$\hat{\theta}_1 = \hat{\theta}_1(x_1, x_2, \cdots, x_n)$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\hat{\theta}_m = \hat{\theta}_m(x_1, x_2, \cdots, x_n)$$

例 设总体X有数学期望和方差：$EX=\mu, DX=\sigma^2$

$X_1,\ldots,X_n$是X的一组样本,求 μ,σ^2 的矩估计.

按矩估计法原理
$$\begin{cases} \overline{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i=\mu \\ A_2=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2=E(X^2)=\sigma^2+\mu^2 \end{cases}$$

$$\Longrightarrow \begin{cases} \hat{\mu}=\overline{X} \\ \hat{\sigma}^2=A_2-\hat{\mu}^2 =\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2-\overline{X}^2 \end{cases}$$

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

不论总体服从什么分布, 如果总体期望 μ 与方差σ^2 存在, 则它们的矩估计量分别为$\bar{X}$和S_n^2

例　设总体 X 的概率密度为

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

其中 $\theta>-1$ 是未知参数. $X_1,X_2,\cdots,X_n$ 是总体 X 的一个简单样本，求 θ 的矩估计.

解

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_0^1 x(\theta+1)x^{\theta}dx=\frac{\theta+1}{\theta+2},$$

令 $\underbrace{\dfrac{\theta+1}{\theta+2}}_{\text{总体矩}}=\underbrace{\overline{X}}_{\text{样本矩}}$　　得　$\hat{\theta}=\dfrac{2\overline{X}-1}{1-\overline{X}}$

例　设总体 $X \sim U(0,\theta)$，θ 未知, $X_1,\ldots,X_n$ 是 X 的样本，试求 θ 的矩估计量．

方法1　$\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} = EX = \frac{\theta}{2}$,　　$\hat{\theta} = 2\bar{X}$

方法2　$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = E(X^2) = \int_0^{\theta} \frac{1}{\theta} x^2 dx = \frac{\theta^2}{3}$,　　$\hat{\theta} = \sqrt{\frac{3}{n}\sum_{i=1}^{n} X_i^2}$

使用不同的矩，可能得到不同的矩估计，矩估计不唯一.
一般使用低阶矩进行估计。

矩估计法的优点：

简单易行, 不需要知道总体是什么分布，只需要知道总体 k 阶原点矩.

矩估计法的缺点：

- 当总体分布类型已知时，没有充分利用分布提供的信息 .
- 一般矩估计量不具有唯一性。
- 所需要的总体的某阶矩可能不存在。
- 求解方程组可能很困难。

最大似然估计

点估计的另外一种方法——最大似然估计法（Maximum Likelihood Estimation，MLE），是在总体类型已知条件下使用的一种参数估计方法．

首先由数学家Gauss在1821年提出，Fisher在1922年重新发现了这一方法，并研究了它的一些性质，得到广泛应用.

例：袋子中的黑球白球，不知哪种多？
每次取一个球，有放回取三次，结果第一次和第三次是白球，第二次是黑球。哪种球多？

白球多！

假设参数是未知的不变量，选择参数值使实验结果具有最大可能性，这就是最大似然法的基本思想 .

设θ是盒中白球所占比例。

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

最大

求最大似然估计的一般步骤

(1) 构造似然函数$L(\theta)$

设$X_1,\cdots,X_n$是来自X的样本,

$x_1,\cdots,x_n$是$X_1,\cdots,X_n$的一个样本值;

若总体X属离散型，其分布律

$$P(X=x)=p(x;\theta),\quad \theta\in\Theta$$

似然函数

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

若总体X属连续型，其概率密度$f(x;\theta),\theta\in\Theta$

似然函数

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

(2) 求似然函数$L(\theta)$ 的最大值点

选择使$L(\theta)$达到最大的参数$\hat{\theta}$，作为θ的估计，

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

$\hat{\theta}(x_1,\cdots,x_n)$称为参数$\theta$的<u>最大似然估计值</u>.

$\hat{\theta}(X_1,\cdots,X_n)$称为参数$\theta$的<u>最大似然估计量</u>.

θ 可由下式求得：

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

因$L(\theta)$与$\ln L(\theta)$在同一θ 处取到极值，
θ 的最大似然估计$\hat{\theta}$ 也可从下述方程解得：

$$\frac{d}{d\theta}\ln L(\theta) = 0.$$

注1 如果有多个未知参数, 如$\theta_1,\ldots,\theta_k$ 时

设X 的密度(或分布)为 $f(x;\theta_1,\cdots,\theta_k)$

则似然函数为 $L(x_1,\cdots,x_n;\theta_1,\cdots,\theta_k)$

可令 $$\frac{\partial L}{\partial \theta_i}=0 \text{或} \frac{\partial \ln L}{\partial \theta_i}=0, i=1,\cdots,k.$$ 似然方程组

解方程组求得$\theta_1,\cdots,\theta_k$的最大似然估计.

注2 用上述方法求参数的最大似然估计值有时行不通，例如$L(\theta)$不是θ的连续可导函数，或参数空间是有界区域，这时一般利用定义进行判断分析求解。

注3 **最大似然估计不变性**

若$\hat{\theta}$是θ的最大似然估计, 则$g(\hat{\theta})$是 $g(\theta)$ 的最大似然估计.

例 考虑一系列独立的硬币投掷试验，θ是每次正面向上的概率，固定$\boldsymbol{n}$， $\boldsymbol{k}$是$\boldsymbol{n}$次投掷中正面向上的次数．试找出基于$\boldsymbol{k}$的θ的最大似然估计.

$$L(\theta) = C_n^k \theta^k (1-\theta)^{n-k}, \qquad 0 \le k \le n.$$

$$\ln L(\theta) = \ln C_n^k + k\ln\theta + (n-k)\ln(1-\theta)$$

$$\hat{\theta}_{\mathrm{ML}} = \frac{k}{n}$$

例 设总体 X 的概率密度为

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

其中 $\theta>-1$ 是未知参数. $X_1,X_2,\cdots,X_n$ 是总体 X 的一个简单样本，求θ 的最大似然估计.

解：似然函数为 $L(\theta)=\prod_{i=1}^{n}[(\theta+1)x_i^{\theta}]=(\theta+1)^n(\prod_{i=1}^{n}x_i)^{\theta}$

对数似然函数为 $\ln L(\theta)=n\ln(\theta+1)+\theta\sum_{i=1}^{n}\ln x_i$

求导并令其为0 $(0<\boldsymbol{x_i}<1)$

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

解得

$$\hat{\theta}=-\frac{n}{\sum_{i=1}^{n}\ln x_i}-1$$

可以验证二阶导数<0，
所以是θ**的最大似然估计**

例 设总体 $X \sim N(\mu,\sigma^2)$, $x_1, x_2, \ldots, x_n$ 是 X 的样本值, 求 μ, σ^2 的最大似然估计.

解 $L(x_1, x_2, \cdots, x_n; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}} e^{-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}}$

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

似然方程组

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

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

例 设$X \sim G(p), x_1, \cdots, x_n$是来自$X$的一个样本值，试求参数$p$与$E(X)$的最大似然估计.

解： X的分布律为：$P(X=k)=p(1-p)^{k-1}$， $k=1,2,\cdots$

似然函数为

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

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

解得p的最大似然估计值 $\hat{p}=\dfrac{\mathrm{n}}{\sum_{i=1}^{n} x_i}=\dfrac{1}{\bar{x}}$

$E(X)$的最大似然估计为$\widehat{\mathrm{E}(X)}=\dfrac{1}{\hat{p}}=\bar{x}$

例　设某种元件使用寿命 X 的概率密度为

$$f(x)=\begin{cases}2e^{-2(x-\theta)}, & x\geq\theta\\ 0, & 其他\end{cases}$$

其中 $\theta>0$ 是未知参数. 设 $x_1,\cdots,x_n$ 是样本观测值，求θ 的极大似然估计.

解： 当有一个$x_i<\theta$,则$L(\theta)=0$

当所有$x_i\geq\theta$时　$L(\theta)=\prod_{i=1}^{n}[2e^{-2(x_i-\theta)}]=2^n e^{-2\sum(x_i-\theta)}$

$$\ln L(\theta)=n\ln 2-2\sum_{i=1}^{n}(x_i-\theta)$$

$\dfrac{d\ln L(\theta)}{d\theta}=2n>0$　$L(\theta)$单调增加　且要满足$\theta\leq x_i$

θ的最大似然估计　$\hat{\theta}=\min\{x_1,x_2,\cdots,x_n\}$

例(2)　设$X \sim U[a,b]$; a, b未知，$x_1,\cdots,x_n$是一个样本值，求a,b的最大似然估计值。

解：　X的概率密度为$f(x;a,b)=\begin{cases}\frac{1}{b-a}, a \le x \le b;\\ 0 \quad\quad, 其它\end{cases}$

似然函数为$L(a,b)=\begin{cases}\frac{1}{(b-a)^n}, a \le x_1,\cdots,x_n \le b;\\ 0 \quad\quad, 其它\end{cases}$

似然函数只有当$a \le x_i \le b, i = 1,2,\ldots,n$ 时才能获得最大值，

且$b-a$越小（即a越大, b越小），L越大.

令　$x_{\min} = \min\{x_1, x_2,\ldots, x_n\}, x_{\max} = \max\{x_1, x_2,\ldots, x_n\}$

取　$\hat{a} = x_{\min}$，$\hat{b} = x_{\max}$　这就是a, b的最大似然估计值。

填空题 1分

设置

设某产品合格率p可能的取值为0<p<1，为估计p，现从大批的该产品中随机抽查了10件，发现恰有8件产品合格. 则该产品合格率p的极大似然估计值为 [填空1]

作答

Bayes估计

① 把未知参数 θ 看作随机变量，可用一个概率分布去描述，这个分布称为**先验分布**；

② 在获得样本之后，总体分布、样本与先验分布通过贝叶斯公式结合起来，得到一个关于 θ 的新分布，称为**后验分布**；

③ 基于 θ 的后验分布进行估计.

$X = (x_1, x_2, \dots, x_n)$ 是i.i.d.的一组抽样,

$$P(\theta \mid X) = \frac{P(X \mid \theta) \times P(\theta)}{P(X)}$$

由后验分布 $P(\theta \mid X)$ 得到θ的点估计有三种常用的方法:

- 使用后验分布的密度函数最大值点作为 θ 的点估计的**最大后验估计(maximum a posteriori, MAP)**.

$$\hat{\theta}_{MAP} = \arg\max_{\theta} p(\theta|X).$$

- 使用后验分布的中位数作为 θ 的点估计的**后验中位数估计**.

$$p(\theta \leq \hat{\theta}_{MED}|X) = 0.5.$$

- 使用后验分布的均值作为 θ 的点估计的**后验期望估计**.

$$\hat{\theta}_B = \mathbb{E}[\theta|X] = \int \theta p(\theta|X) d\theta.$$

用得最多的是后验期望估计, 一般也称为贝叶斯估计.

最大后验估计(Maximum A Posteriori, MAP)

假设$X = (x_1, x_2, \dots, x_n)$ 是i.i.d.的一组抽样，

$$P(\theta \mid X) = \frac{P(X \mid \theta) \times P(\theta)}{P(X)}$$

$$\hat{\theta}_{MAP} = \operatorname{argmax} P(\theta \mid X)$$

$$= \arg\min - \log P(\theta \mid X)$$

$$= \arg\min - \log P(X \mid \theta) - \log P(\theta) + \log P(X)$$ （贝叶斯定理)

$$= \arg\min - \log P(X \mid \theta) - \log P(\theta)$$ （$P(X)$因为与 θ 无关)

MLE和**MAP**的不同在于是否有先验项。

例 设某事件 A 在一次试验中发生的概率为 θ, 为估计 θ, 对试验进行了n次独立观测, 其中事件A发生了 X 次,

显然 $X \mid \theta \sim B(n, \theta)$, 即

$$P(X = x \mid \theta) = \mathrm{C}_n^x \theta^x (1-\theta)^{n-x},\ x = 0,1,\cdots,n.$$

假若我们在试验前对事件 A 没有什么了解, 从而对其发生的概率 θ 也没有任何信息. 在这种场合, 贝叶斯本人建议采用 "同等无知"的原则使用区间 (0,1) 上的均匀分布 $U(0,1)$ 作为 θ 的先验分布, 因为它取 (0,1) 上的每一点的机会均等. 贝叶斯的这个建议称为贝叶斯假设.

利用贝叶斯公式求出 θ 的后验分布:

X 和 θ 的联合分布

$$h(x,\theta)=\mathrm{C}_n^x\theta^x(1-\theta)^{n-x},\ x=0,1,\cdots,n,\ 0<\theta<1,$$

X 的边缘分布

$$m(x)=\mathrm{C}_n^x\int_0^1\theta^x(1-\theta)^{n-x}\,\mathrm{d}\theta=\mathrm{C}_n^x\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)},$$

θ 的后验分布

$$\pi(\theta\mid x)=\frac{h(x,\theta)}{m(x)}$$

$$=\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^{(x+1)-1}(1-\theta)^{(n-x+1)-1},0<\theta<1.$$

结果说明 $\theta\mid x\sim Beta(x+1,n-x+1)$

其最大后验估计$\hat{\theta}_{MAP}=\frac{x}{n}$

其后验期望估计 $\hat{\theta}_B=\frac{x+1}{n+2}$

假如不用先验信息, 只用总体信息与样本信息, 那么事件 A 发生的概率的最大似然估计 $\hat{\theta}_{MLE}=\frac{x}{n}$,

某些场合, 贝叶斯估计要比最大似然估计更合理一点.

例如: 在产品抽样检验中只区分合格品和不合格品, θ 表示不合格品率。"抽检3个全是不合格品"与 "抽检10个全是不合格品" 是有差别的两个事件, 前者质量很差, 后者则不可救药.

这种差别用 $\hat{\theta}_{MLE}$ 反映不出(两者都是1)

而 $\hat{\theta}_B$ 分别是 $(3+1)/(3+2)=0.80$ 和 $(10+1)/(10+2)=0.917$.

点估计的评价标准

点估计具有不唯一性，即对于同一个未知参数，不同的方法得到的估计量可能不同，**怎么评价一个估计量的好坏?**

(1) 无系统偏差, 即 $\hat{\theta}$ 平均起来应与 θ 值相同, 也就是说 $E\left[\hat{\theta}(X_1,X_2,\cdots,X_n)-\theta\right]=0$. 这就是**无偏性**的要求.

(2) 波动性小, 即 $|\hat{\theta}-\theta|$ 平均起来应尽可能小, 为数学上便于处理, 可用 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 的均方误差 $E\left[\left(\hat{\theta}(X_1,X_2,\cdots,X_n)-\theta\right)^2\right]$ 表示. 这就是**有效性**的要求.

(3) 随着样本容量 n 的增大, $\hat{\theta}$ 应越来越接近于 θ, 亦即 $|\hat{\theta}-\theta|$ 按某种概率意义收敛于0 . 这就是**相合性**的要求.

均方误差准则(Mean Squared Error, MSE)

定义： 设$\hat{\theta}$是参数θ的点估计，方差存在，则称

$E(\hat{\theta}-\theta)^2$是估计量$\hat{\theta}$的均方误差，记为$Mse(\hat{\theta})$

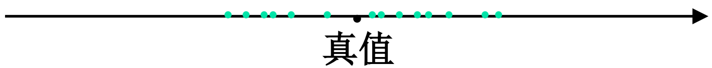
- 估计量的均方误差由方差(variance)和偏差(bias)两部分组成。

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

- 均方误差越小的估计量越好。
- 在所有估计量中，均方误差最小的称为最优估计量。

(1) 无偏性(unbiased)

估计量是随机变量，对于不同的样本值会得到不同的估计值．我们希望估计值在未知参数真值附近摆动，它的期望值等于未知参数的真值。这就是无偏性标准。



设$\hat{\theta}(X_1,\cdots,X_n)$是未知参数$\theta$的估计量，若$E(\hat{\theta})=\theta$，则称$\hat{\theta}$为$\theta$的无偏估计．

无偏估计的实际意义：无系统误差.

若$\lim\limits_{n\to\infty} E(\hat{\theta})=\theta$，则称$\hat{\theta}$是$\theta$渐近无偏估计(asymptotically unbiased).

实际应用中应尽量选择无偏估计。如果无偏估计不存在或很难计算，则选择渐进无偏估计。

例 设总体 X 的 k 阶矩 $\alpha_k = E(X^k)$存在,

$(X_1, X_2, \cdots, X_n)$ 是总体X 的样本,

证明: 不论 X 服从什么分布(但k 阶矩存在),

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$ 是 α_k 的无偏估计量.

证

由于 $E(X_i^k) = \alpha_k \quad i = 1, 2, \cdots, n$

$$E(A_k) = E(\frac{1}{n}\sum_{i=1}^{n} X_i^k) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^k) = \frac{1}{n}\cdot n\cdot\alpha_k = \alpha_k$$

- 样本均值 $\bar{X}$ 是总体期望 $E(X)$ 的无偏估计量。
- 样本二阶原点矩$A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$是总体二阶原点矩$E(X^2)$的无偏估计量。

例 设总体 X 的期望μ与方差σ^2存在，X 的样本为$(X_1, X_2, \cdots, X_n)$

(1)原样本方差(Raw sample variance)

$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 不是 $D(X)$的无偏估计,是 $D(X)$的渐进无偏估计

$$E(S_n^2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2;$$

$$\lim_{n\to\infty} E\left(S_n^2\right) = \lim_{n\to\infty} \frac{n-1}{n}\sigma^2 = \sigma^2$$

(2)样本修正方差(Bessel’s Correction)

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

$$E\left(S^2\right) = \sigma^2$$

$$\boxed{E(S_n^2)=\frac{n-1}{n}\sigma^2, \qquad E(S^2)=\sigma^2}$$

推导

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

$$E(\overline{X})=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\mu \qquad\qquad D(\overline{X})=\frac{1}{n}\sigma^2$$

$$E(S_n^2)=E\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}X_i^2\right)-E\left(\overline{X}^2\right)=E\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}X_i^2\right)-\left[D(\overline{X})+E^2(\overline{X})\right]$$

$$=\sigma^2+\mu^2-\left(\frac{1}{n}\sigma^2+\mu^2\right)=\frac{n-1}{n}\sigma^2$$

$$E(S^2)=E\left[\frac{n}{n-1}S_n^2\right]=\frac{n}{n-1}E(S_n^2)=\sigma^2$$

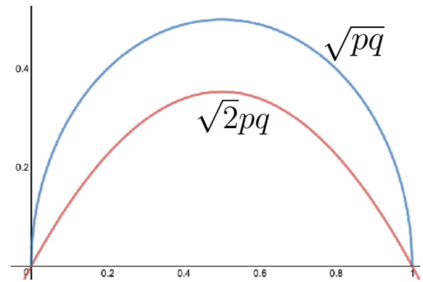
样本标准差是无偏的吗？

$$S \stackrel{\text{def}}{=} +\sqrt{S^2} = +\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2} \qquad \sigma = \sqrt{\sigma^2}$$

E(S) = σ ?

例　总体X服从B(1,p),样本容量n=2　$\sigma = \sqrt{p(1-p)} = \sqrt{pq}$

x_1,x_2	$P(x_1,x_2)$	$\overline{x}$	s^2	s
0,0	q^2	0	0	0
0,1	qp	½	½	$\frac{1}{\sqrt{2}}$
1,0	pq	½	½	$\frac{1}{\sqrt{2}}$
1,1	p^2	1	0	0



$$E(S) = q^2\cdot 0 + qp\cdot\frac{1}{\sqrt{2}} + pq\cdot\frac{1}{\sqrt{2}} + p^2\cdot 0 = \sqrt{2}\cdot pq < \sqrt{pq}$$

一般的σ无偏估计量不存在。

例 设总体 X 的密度函数为 $f(x;\theta)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}} & x>0,\\ 0 & x\le 0\end{cases}$

$\theta>0$为参数，$X_1,X_2,\cdots,X_n$为 X 的一个样本.

求θ的最大似然估计量, 并判断它是否无偏估计量.

解 似然函数 $L(\theta)=\dfrac{1}{\theta^n}e^{-\frac{1}{\theta}\sum\limits_{i=1}^{n}x_i}$, $x_i>0$ $\quad\ln L(\theta)=-n\ln\theta-\dfrac{1}{\theta}\sum\limits_{i=1}^{n}x_i$

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L(\theta)=-\frac{n}{\theta}+\frac{\sum\limits_{i=1}^{n}x_i}{\theta^2}\overset{令}{=}0 \longrightarrow \hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$$

θ的最大似然估计量为 $\hat{\theta}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i=\bar{X}$

$E(\hat{\theta})=E(\bar{X})=E(X)=\theta$ 是无偏估计量.

设λ=$\dfrac{1}{\theta}$，λ的最大似然估计量是什么？是无偏估计吗？

$$\hat{\lambda}=\frac{1}{\hat{\theta}}=\frac{1}{\bar{X}} \qquad E(\hat{\lambda})=E\left(\frac{1}{\bar{X}}\right)\neq\frac{1}{E(\bar{X})}=\lambda$$

例 设 $X_1, X_2, \cdots, X_n$ 是取自$P(\lambda)$的样本，证明$\bar{X}$,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \text{和} a\bar{X} + (1-a)S^2 \ (0 \le a \le 1)$$

都是λ的无偏估计.

解

$$X \sim P(\lambda) \Rightarrow E(X) = D(X) = \lambda$$

$$E(\overline{X}) = E(X) = \lambda$$

$$E(S^2) = D(X) = \lambda$$

$$E[a\overline{X} + (1-a)S^2] = aE(\overline{X}) + (1-a)E(S^2)$$

$$= [a + (1-a)]\lambda = \lambda$$

(2) 有效性

一个参数往往有不止一个无偏估计, 若$\hat{\theta}_1$和$\hat{\theta}_2$都是参数$\boldsymbol{\theta}$的无偏估计量，二者谁更优?

由无偏估计的性质得到 $E(\hat{\theta}_1)=\theta \qquad E(\hat{\theta}_2)=\theta$

$$MSE(\hat{\theta}_1) = E(\hat{\theta}_1-\theta)^2 = D(\hat{\theta}_1)$$

$$MSE(\hat{\theta}_2) = E(\hat{\theta}_2-\theta)^2 = D(\hat{\theta}_2)$$

无偏估计中方差小的更好, 这就是“**有效性**”.

- 设 $\hat{\theta}_1$、$\hat{\theta}_2$ 是末知参数 θ 的两个无偏估计量, 若$D(\hat{\theta}_1) < D(\hat{\theta}_2)$, 则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ **有效(efficient)**.
- 设 $\hat{\theta}^*$ 是末知参数 θ 的无偏估计量, 若对于 θ 的任意一个无偏估计量 $\hat{\theta}$, 都有 $D(\hat{\theta}^*) \leqslant D(\hat{\theta})$, 则称 $\hat{\theta}^*$ 为 θ 的**最小方差无偏估计**.

例 设总体 X 的均值 $E(X)=\mu$ 未知, $-\infty<\mu<+\infty$, $(X_1,X_2,\cdots,X_n)$ 是取自总体 X 的样本，在形如$\hat{\mu}=\sum_{i=1}^{n}a_iX_i$ 且 $\sum_{i=1}^{n}a_i=1$的无偏估计量中，求出最小方差无偏估计量$\hat{\mu}^*$。

解 记 $D(X)=\sigma^2$, 则

$$D(\hat{\mu})=D\left(\sum_{i=1}^{n}a_iX_i\right)=\sum_{i=1}^{n}a_i^2D(X_i)=\left(\sum_{i=1}^{n}a_i^2\right)\sigma^2.$$

问题归结为在条件 $\sum_{i=1}^{n}a_i=1$ 下, 求 $a_1,a_2,\cdots,a_n$ 使 $\sum_{i=1}^{n}a_i^2$ 取得最小值.

$1=(\sum_{i=1}^{n}c_i)^2=\sum_{i=1}^{n}c_i^2+2\sum_{1\le i<j\le n}c_ic_j\le\sum_{i=1}^{n}c_i^2+\sum_{1\le i<j\le n}(c_i^2+c_j^2)=n\sum_{i=1}^{n}c_i^2$

即 $\sum_{i=1}^{n}c_i^2\ge\frac{1}{n}$

当 $a_1=a_2=\cdots=a_n=\frac{1}{n}$, 即 $\hat{\mu}=\hat{\mu}^*=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}$ 时, $D(\hat{\mu})$ 取得最小值

$D(\hat{\mu}^*)=\frac{1}{n}\sigma^2$.

样本均值 $\bar{X}$ 是形如 $\sum_{i=1}^{n}a_iX_i$ 且 $\sum_{i=1}^{n}a_i=1$ 的无偏估计量中方差最小的.

(3) 一致性(相合性，consistence)

无偏性和有效性都是在样本容量 n 固定的前提下提出的，随着试验次数 n 的不断增加，样本所包含的信息量也逐步增加．因此，一个 “好” 的估计在样本容量 n 增加时，必须越来越接近于参数的真值．

定义 设$\hat{\theta}=\hat{\theta}(X_1, X_2, \cdots, X_n)$是总体参数$\theta$的估计量.

若$n\to\infty$时, $\hat{\theta}$ 依概率收敛于θ, 即

$$\forall\varepsilon > 0,\ \lim_{n\to\infty} P\left(\left|\hat{\theta}-\theta\right| \geq \varepsilon\right) = 0$$

则称 $\hat{\theta}$ 是参数θ的一致(或相合)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.
- 样本 k 阶矩是总体 k 阶矩的一致性估计量
- 矩估计量一般为一致估计量
- 在一定条件下，极大似然估计具有一致性

定理（马尔可夫不等式） 设 $g(X)$ 是随机变量 X 的非负连续函数, 如果 $E[g(X)]$ 存在, 则对于任一正常数 c, 均有

$$P\{g(X) \geqslant c\} \leqslant \frac{E[g(X)]}{c}.$$

证 由于 g 是非负连续函数, 所以

$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x \geqslant \int_{\{x|g(x)\geqslant c)} g(x)f(x)\mathrm{d}x \\ &\geqslant c\int_{\{x|g(x)\geqslant c)} f(x)\mathrm{d}x = cP\{g(X) \geqslant c\}, \end{aligned}$$

由 $c > 0$ 知, $P\{g(X) \geqslant c\} \leqslant \frac{E[g(X)]}{c}$成立. $\square$

定理 设 $\hat{\theta}_n$ 是 θ 的一个估计量, 若

$$\lim_{n\to\infty} E\left(\hat{\theta}_n\right) = \theta, \lim_{n\to\infty} D\left(\hat{\theta}_n\right) = 0,$$

则 $\hat{\theta}_n$ 是 θ 的相合估计.

证 对任给的 $\varepsilon > 0$, 有

$$P\left(\left|\hat{\theta}_n - \theta\right| \geqslant \varepsilon\right) = P\left(\left(\hat{\theta}_n - \theta\right)^2 \geqslant \varepsilon^2\right) \leqslant \frac{E\left(\hat{\theta}_n - \theta\right)^2}{\varepsilon^2} \text{ (马尔可夫不等式)}.$$

又 $D\left(\hat{\theta}_n\right) = D\left(\hat{\theta}_n - \theta\right) = E\left(\hat{\theta}_n - \theta\right)^2 - \left[E\left(\hat{\theta}_n - \theta\right)\right]^2$, 得

$$P\left(\left|\hat{\theta}_n - \theta\right| \geqslant \varepsilon\right) \leqslant \frac{D\left(\hat{\theta}_n\right) + \left[E\left(\hat{\theta}_n - \theta\right)\right]^2}{\varepsilon^2} = \frac{D\left(\hat{\theta}_n\right) + \left[E\left(\hat{\theta}_n\right) - \theta\right]^2}{\varepsilon^2},$$

根据已知条件，当 $n \to \infty$ 时, $E\left(\hat{\theta}_n\right) \to \theta, D\left(\hat{\theta}_n\right) \to 0$,

故 $P\left(\left|\hat{\theta}_n - \theta\right| \geqslant \varepsilon\right) \to 0$, 因而 $\hat{\theta}_n$ 是 θ 的相合估计.

例 若总体X的 $E(X)=\mu$ 和 $D(X)=\sigma^2$ 存在, 则样本均值 $\boldsymbol{\bar{X}}$ 是总体均值 μ 的相合估计.

证明 由于
$$E(\bar{X})=E(X)=\mu,$$

$$\lim_{n\to\infty} D(\bar{X})=\lim_{n\to\infty}\frac{D(X)}{n}=\lim_{n\to\infty}\frac{\sigma^2}{n}=0$$

故 $\bar{X}$ 是 μ 的相合估计.

例 设总体 $X \sim N(\mu,\sigma^2)$, $\boldsymbol{X_1,\ldots,X_n}$是$X$的一个样本，

则 $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 是σ^2的相合估计量.

证明 $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$ 由卡方分布性质知

$$E\left(\frac{(n-1)S^2}{\sigma^2}\right)=n-1\ ,\ D\left(\frac{(n-1)S^2}{\sigma^2}\right)=2(n-1)$$

$$\therefore E\left(S^2\right)=\sigma^2\ ,\ D\left(S^2\right)=\frac{2\sigma^4}{n-1}$$

$$\lim_{n\to\infty}E(S^2)=\sigma^2\ ,\ \lim_{n\to\infty}D\left(S^2\right)=0$$

S^2 是σ^2的相合估计量.

无偏估计量方差的下界

- **问题**：在样本容量一定的条件下，待估参数的无偏估计量的方差是否可以任意小呢?
- Rao-Cramér 定理

Rao-Cramér 定理

设总体 X 的概率密度函数为 $f(x;\theta)$，$\theta\in\Theta$，这里Θ是$(-\infty,+\infty)$ 上的一个开区间, $(X_1,X_2,\cdots,X_n)$ 是取自总体 X 的样本, 待估函数是$g(\theta)$, 且 $\hat{g}(X_1,X_2,\cdots,X_n)$ 是 $g(\theta)$ 的任意一个无偏估计量, 如果

(1) $E\left[\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2\right]>0$,

(2) $\frac{\partial}{\partial\theta}f(x;\theta)$ 存在, 且 $\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty}f(x;\theta)\mathrm{d}x=\int_{-\infty}^{+\infty}\frac{\partial}{\partial\theta}f(x;\theta)\mathrm{d}x$,

(3) $\frac{\partial}{\partial\theta}g(\theta)$ 存在, 且

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\hat{g}(x_1,\cdots,x_n)\prod_{i=1}^{n}f(x_i;\theta)\mathrm{d}x_1\cdots\mathrm{d}x_n$$
$$=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\hat{g}(x_1,\cdots,x_n)\frac{\partial}{\partial\theta}\left(\prod_{i=1}^{n}f(x_i;\theta)\right)\mathrm{d}x_1\cdots\mathrm{d}x_n,$$

则有 $D[\hat{g}(X_1,X_2,\cdots,X_n)] \geqslant \frac{[g'(\theta)]^2}{nI(\theta)}$,

其中 $I(\theta)=E\left[\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2\right]$ 称为 **Fisher 信息量，**

右端的 $\frac{[g'(\theta)]^2}{nI(\theta)}$ 称为 **Rao-Cramér 下界**.

- 特别地, 当 $g(\theta)=\theta$ 时, $D\left[\hat{\theta}(X_1,X_2,\cdots,X_n)\right] \geqslant \frac{1}{nI(\theta)}$

  表明, 如果未知参数 θ 的某个无偏估计量 $\hat{\theta}$, 其方差 $D(\hat{\theta})$ 等于 Rao-Cramér 下界 $\frac{1}{nI(\theta)}$, 那么这个无偏估计量 $\hat{\theta}$ 必是**最小方差无偏估计量**.
- 设 $\hat{g}(X_1,X_2,\cdots,X_n)$ 是待估函数 $g(\theta)$ 的一个无偏估计量,

$$e(\hat{g})=\frac{[g'(\theta)]^2}{nI(\theta)}/D(\hat{g})$$

称为无偏估计量 $\hat{g}$ 的**效率**. 显然 $e(\hat{g})\leqslant 1$, 如果 $e(\hat{g})=1$, 则称 $\hat{g}$ 为 g 的**优效估计**; 如果 $\hat{g}$ 满 足 $\lim_{n\to+\infty}e(\hat{g})=1$, 则称 $\hat{g}$ 为 g 的**渐近优效估计**.

Fisher 信息量

$$
\begin{aligned}
\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta) &= \frac{\partial}{\partial\theta}\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right] = \frac{\partial}{\partial\theta}\left[\frac{1}{f(X;\theta)}\frac{\partial}{\partial\theta}f(X;\theta)\right] \\
&= -\left[\frac{1}{f(X;\theta)}\frac{\partial}{\partial\theta}f(X;\theta)\right]^2 + \frac{1}{f(X;\theta)}\frac{\partial^2}{\partial\theta^2}f(X;\theta) \\
&= -\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2 + \frac{1}{f(X;\theta)}\frac{\partial^2}{\partial\theta^2}f(X;\theta),
\end{aligned}
$$

$$
E\left[\frac{1}{f(X;\theta)}\frac{\partial^2}{\partial\theta^2}f(X;\theta)\right] = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\frac{\partial^2}{\partial\theta^2}f(x_1,\cdots,x_n;\theta)\mathrm{d}x_1\cdots\mathrm{d}x_n
$$

$$
= \frac{\partial^2}{\partial\theta^2}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}f(x_1,\cdots,x_n;\theta)\mathrm{d}x_1\cdots\mathrm{d}x_n = 0,
$$

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

例 设 $(X_1, X_2, \cdots, X_n)$ 是取自总体 $B(1;p)$ 的一个样本, 其中 $p \in [0,1]$ 但末知, 求 p 的无偏估计量方差的下界.

解: X的概率密度函数为 $f(x;p) = p^x(1-p)^{1-x}, x = 0,1.$

$$\frac{\partial}{\partial p}\ln f(x;p) = \frac{\partial}{\partial p}\{x\ln p + (1-x)\ln(1-p)\} = \frac{x}{p} - \frac{1-x}{1-p} = \frac{x-p}{p(1-p)},$$

$$\begin{aligned} I(p) &= E\left[\left(\frac{X-p}{p(1-p)}\right)^2\right] = \frac{1}{p^2(1-p)^2}E[(X-p)^2] \\ &= \frac{1}{p^2(1-p)^2}D(X) = \frac{1}{p^2(1-p)^2}p(1-p) = \frac{1}{p(1-p)}. \end{aligned}$$

因此, p 的无偏估计量的方差下界是 $\frac{p(1-p)}{n}$.

如果取样本均值 $\bar{X}$ 为 p 的估计量, 它显然是无偏估计量, 由 $D(\bar{X}) = \frac{1}{n}D(X) = p(1-p)/n$ 知, $\bar{X}$ 的方差等于 Rao-Cramér 下界 $p(1-p)/n$, 故 $\bar{X}$ 是 p 的最小方差无偏估计量.

例 设 $(X_1, X_2, \cdots, X_n)$ 是取自正态总体 $N(\mu; \sigma^2)$ 的样本, 讨论 μ、σ^2 的无偏估计量的方差下界.

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

$$\ln f(x; \mu, \sigma^2) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2.$$

(1) μ的无偏估计的方差下界

$$\frac{\partial}{\partial\mu}\ln f(x; \mu, \sigma^2) = \frac{x-\mu}{\sigma^2}, \ \frac{\partial^2}{\partial\mu^2}\ln f(x; \mu, \sigma^2) = -\frac{1}{\sigma^2}, I(\mu) = \frac{1}{\sigma^2}.$$

于是, μ 的无偏估计的方差下界是 σ^2/n. 因为样本均值 $\bar{X}$ 的方差 $D(\bar{X}) = \sigma^2/n$ 达到这个下界, 从而 $\bar{X}$ 是 μ 的最小方差无偏估计量.

(2) σ^2 的无偏估计的方差下界.

$$\frac{\partial}{\partial\sigma^2}\ln f\left(x;\mu,\sigma^2\right)=-\frac{1}{2\sigma^2}+\frac{1}{2\sigma^4}(x-\mu)^2,$$

$\frac{\partial^2}{\partial(\sigma^2)^2}\ln f\left(x;\mu,\sigma^2\right)=\frac{1}{2\sigma^4}-\frac{(x-\mu)^2}{\sigma^6}$，得 $I\left(\sigma^2\right)=-\left(\frac{1}{2\sigma^4}-\frac{1}{\sigma^4}\right)=\frac{1}{2\sigma^4}$.

因此, σ^2 的无偏估计的方差下界是 $\frac{2}{n}\sigma^4$.

- 当 μ 已知时, σ^2 的极大似然估计量 $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ 是无偏的. 由于 $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n)$, $D\left[\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right]=2n$, $D[\sum_{i=1}^{n}(X_i-\mu)^2]=2n\sigma^4$.

于是, $D\left(\hat{\sigma}^2\right)=\frac{1}{n^2}D[\sum_{i=1}^{n}(X_i-\mu)^2]=\frac{2}{n}\sigma^4$, 即 $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ 的方差达到 Rao-Cramér下界, 因此它是 σ^2 的最小方差无偏估计量.

- 当 μ 未知时, 样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ 是 σ^2 的无偏估计. 由于 $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2\sim\chi^2(n-1)$,

$$D\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right]=2(n-1),\ D\left[\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right]=2(n-1)\sigma^4.$$

$$\begin{aligned}D\left(S^2\right)\quad&=D\left(\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right)=\frac{1}{(n-1)^2}D\left(\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right)\\&=\frac{1}{(n-1)^2}\cdot 2(n-1)\sigma^4=\frac{2}{n-1}\sigma^4>\frac{2}{n}\sigma^4,\end{aligned}$$

即 S^2 的方差没有达到 Rao-Cramér 下界.

- 对于正态总体 $N\left(\mu;\sigma^2\right)$, 样本均值 $\bar{X}$ 是 μ 的优效估计, 样本方差 S^2 是 σ^2 的渐近优效估计.

填空题 1分

设置

方法一:

$\ln P = x\ln a - a - \ln x!$

求导: $\frac{x}{a} - 1$

再求导: $-\frac{x}{a^2}$

$E(\frac{x}{a^2}) = \frac{E(X)}{a^2} = \frac{1}{a}$　　故 $\frac{1}{nI(\theta)} = \frac{a}{n}$

设 $(X_1, X_2, \cdots, X_n)$ 是取自Poisson分布的样本, 概率分布是

$$P\{X = x\} = \frac{a^x}{x!}e^{-a} \quad (x = 0,1,2,\cdots,),$$ 其中$a > 0$ 未知,

a的无偏估计量的方差下界是 [填空1] .

方法二: $P\{X=x\} = \frac{a^x}{x!}e^{-a}$.

$\ln P = x\ln a - \ln x! - a$.

$\frac{\partial \ln P}{\partial a} = \frac{x}{a} - 1$.

$\therefore E(\frac{X}{a} - 1)^2 = E(\frac{X^2}{a^2} - \frac{2X}{a} + 1)$

$= \frac{1}{a^2}E(X^2) - \frac{2}{a}EX + 1$

$= \frac{1}{a^2}(a + a^2) - \frac{2}{a}\cdot a + 1$

$= \frac{1}{a}$

则为 $\frac{a}{n}$.

作答