问题 如何评价同一参数的不同估计量？

- **解决方案** 利用损失函数（Loss function）来研究不同估计量的性能，从中选取最优。
- **基本思想** 如果估计量$\hat{\theta}$接近真实值$\vec{\theta}$，则$\hat{\theta}$ 是合理的且遭受的损失小；否则，如果$\hat{\theta}$远离$\vec{\theta}$，则选择$\hat{\theta}$遭受的损失就大。
- **损失函数的基本性质** 非负，且随着$\|\hat{\theta}-\vec{\theta}\|$的增加而增加（这里$\|\cdot\|$ 是欧氏距离），函数常记为$\mathcal{L}(\vec{\theta},\hat{\theta})$.
- **常用损失函数**
  1. $\mathcal{L}(\vec{\theta},\hat{\theta})=\|\hat{\theta}-\vec{\theta}\|$，绝对误差损失（Absolute Error Loss）；
  2. $\mathcal{L}(\vec{\theta},\hat{\theta})=\|\hat{\theta}-\vec{\theta}\|^2$，平方误差损失（Squared Error Loss）；
  3. $\mathcal{L}(\vec{\theta},\hat{\theta})=\omega(\vec{\theta})\|\hat{\theta}-\vec{\theta}\|^2$，加权平方损失（Weighted Squared Error Loss）.

2.1 均方误差

问题 $\mathcal{L}(\vec{\theta}, \widehat{\theta}) = \mathcal{L}\left(\vec{\theta}, \widehat{\theta}(\vec{X})\right)$ 是随机变量，需要量化！

- 解决方案：考虑如下函数
- 定义2.1 [【0】定义7.4.1] 风险函数（Risk function）

$$R(\vec{\theta}) = \mathbb{E}_{\vec{\theta}}\left[\mathcal{L}\left(\vec{\theta}, \widehat{\theta}(\vec{X})\right)\right].$$

即，估计量$\widehat{\theta}$将遭受的平均损失。

- 定义2.2 [【0】定义3.4.1] 参数$\vec{\theta}$的估计量$\widehat{\theta}$的均方误差（Mean Square Error, 简记MSE）定义为

$$MSE(\vec{\theta}, \widehat{\theta}) = \mathbb{E}_{\vec{\theta}}\|\widehat{\theta} - \vec{\theta}\|^2.$$

- **定义2.3** 参数$\vec{\theta}$的估计量$\widehat{\vec{\theta}}$的**偏倚**（Bias）定义为

$$Bias_{\vec{\theta}}(\widehat{\vec{\theta}}) = \mathbb{E}_{\vec{\theta}}\left(\widehat{\vec{\theta}}\right) - \vec{\theta}$$

对于一个估计量$\widehat{\vec{\theta}}$，如果它的偏倚（关于$\vec{\theta}$）恒等于0，则称其为**无偏**的(Unbiased)，即

$$\mathbb{E}_{\vec{\theta}}\left(\widehat{\vec{\theta}}\right) = \vec{\theta}, \forall \vec{\theta} \in \Theta.$$

注1 $MSE(\widehat{\theta}) = Var_{\theta}(\widehat{\theta}) + |Bias_{\theta}(\widehat{\theta})|^2$，即MSE由两部分组成：一是该估计量$\widehat{\theta}$的变异性（精度），二是它的偏倚（准确度）。一个估计量具有好的MSE性质，就意味着在方差和偏倚两项上综合较小。

注2 上述定义参考【0】定义3.1.2.

举例说明

Example (2.1)

设$X_1, \ldots, X_n$ $i.i.d. \sim Poisson(\lambda)$, $\lambda > 0$未知.

1. 由例1.11注1可知，若λ具有先验分布$Gamma(\alpha, \beta)$，其贝叶斯估计为$\widehat{\lambda}_B = \mathbb{E}(\lambda|\vec{X}) = \frac{\sum_{i=1}^n X_i + \alpha}{n+\beta}$，求$\widehat{\lambda}_B$的均方误差MSE;
2. 由例1.1和1.6(1)可知，λ的矩估计和极大似然估计均为$\widehat{\lambda}_{MoM} = \widehat{\lambda}_{MLE} = \overline{X}$，求这个估计的均方误差MSE.
3. 当$\alpha = \beta = 1$且$\lambda > 5$时，比较$\widehat{\lambda}_B$和$\widehat{\lambda}_{MLE}$在均方误差意义下哪个更优？

注 对于问题（3），若无$\lambda > 5$的信息，则只能判断，存在$0 < \lambda_1 < \lambda_2 < \infty$,

1. 当$\lambda < \lambda_1$或$\lambda > \lambda_2$时，$\widehat{\lambda}_{MLE}$ 优于$\widehat{\lambda}_B$;
2. 当$\lambda_1 < \lambda < \lambda_2$时，$\widehat{\lambda}_B$优于$\widehat{\lambda}_{MLE}$.

举例说明

练习 回忆本章例1.1注，自行比较在均方误差意义下, $\widehat{\lambda}_{MoM}$ 和$\widehat{\lambda}^*_{MoM} = \widehat{\mu}_2$哪个更优？

作业1 设$X_1, \ldots, X_n$ $i.i.d. \sim Bernoulli(\theta)$, $\theta \in (0,1)$未知.

1. 回忆例1.9，在先验分布$\vartheta \sim Beta(a,b)$下，θ的贝叶斯估计为$\widehat{\theta}_B = \frac{\sum_{i=1}^n X_i + a}{n+a+b}$. 求$\widehat{\theta}_B$的均方误差$MSE(\widehat{\theta}_B)$;
2. 分别求θ的矩估计$\widehat{\theta}_{MoM}$和极大似然估计$\widehat{\theta}_{MLE}$，以及它们各自的均方误差。

作业2 设$X_1, \ldots, X_n$ $i.i.d. \sim U(0,\theta)$, $\theta > 0$未知.

1. 分别求θ的矩估计$\widehat{\theta}_{MoM}$和极大似然估计$\widehat{\theta}_{MLE}$，以及它们各自的均方误差。
2. 设θ的先验分布$\vartheta \sim U(0,1)$，求θ的贝叶斯估计及其均方误差;
3. 在均方误差意义下，上述三个估计量哪个更优？

2.2 最佳无偏估计

问题 由例2.1（3）可知，基于MSE，对不同估计量进行比较，未必能找到一个“最佳MSE”的估计量！

建议 在一个更小的估计量类型范围内考虑问题— 无偏估计类。具体而言，对于参数$\tau(\theta)$，考虑估计类

$$\mathcal{C}_\tau = \{W : \mathbb{E}_\theta(W) = \tau(\theta)\}.$$

注 $\forall W_1, W_2 \in \mathcal{C}_\tau$，均有$Bias_\theta W_1 = Bias_\theta W_2$，从而

$$MSE(W_1) - MSE(W_2) = Var_\theta(W_1) - Var_\theta(W_2).$$

即在$\mathcal{C}_\tau$中，对不同估计量的MSE进行比较，可以仅基于方差大小。

- **定义2.4**（【0】定义3.4.2） 估计量W^*称为$\tau(\theta)$ 的**最佳无偏估计量**（Best Unbiased Estimator），如果它满足$\mathbb{E}_\theta(W^*) = \tau(\theta)$, $\forall\theta \in \Theta$，并且对任何一个其它的满足$\mathbb{E}_\theta W = \tau(\theta)$的估计量$W$，都有$Var_\theta(W^*) \le Var_\theta(W), \forall\theta \in \Theta$。$W^*$也称为$\tau(\theta)$的**一致最小方差无偏估计量**（Uniform Minimum Variance Unbiased Estimator），简记UMVUE.

问题 如何寻找UMVUE？

方法1 寻找方差达到如下Cramér-Rao 不等式下界的无偏估计。

Theorem (2.1，Cramér-Rao 不等式)

[【0】定理3.5.1，【1】定理7.3.9] 设$X_1, \dots, X_n$具有联合密度函数$f(\vec{x})$，$W(\vec{X})$ 是任意一个统计量。若f 和W 满足如下条件：

1. 任意$\vec{x} \in \mathcal{X}, \theta \in \Theta$，$\frac{\partial f(\vec{x}|\theta)}{\partial \theta}$存在，且在$f$支撑集上不恒等于*0;*
2. $\frac{\partial}{\partial \theta} \int f(\vec{x}|\theta) d\vec{x} = \int \frac{\partial}{\partial \theta} f(\vec{x}|\theta) d\vec{x}$;
3. $\frac{d}{d\theta} \mathbb{E}_\theta[W(\vec{X})] = \int \frac{\partial}{\partial \theta}[W(\vec{x}) f(\vec{x}|\theta)] d\vec{x}$.

则有不等式

$$Var_\theta[W(\vec{X}] \geq \frac{|\frac{d}{d\theta}\mathbb{E}_\theta[W(\vec{X})]|^2}{\mathbb{E}_\theta\{[\frac{\partial}{\partial \theta} \log f(\vec{X}|\theta)]^2\}}. \tag{1}$$

Corollary (2.1)

在定理*2.1* 条件下，若$X_1, \ldots, X_n$ *i.i.d.* $\sim X$, 总体X具有*p.d.f./p.m.f.* $f(x|\theta)$，令

$$\mathrm{I}(\theta) = \mathbb{E}_\theta\{[\frac{\partial}{\partial\theta}\log f(X|\theta)]^2\},$$

则在$\mathrm{I}(\theta) > 0$条件下，

$$Var_\theta[W(\vec{X}] \geq \frac{\{\frac{d}{d\theta}\mathbb{E}_\theta[W(\vec{X})]\}^2}{n\mathrm{I}(\theta)}. \tag{2}$$

注1 $\mathrm{I}(\theta)$称为X关于θ的Fisher信息量（参考【0】3.5.2(3)）。再令

$$\mathrm{I}_n(\theta) = \mathbb{E}_\theta\{[\frac{\partial}{\partial\theta}\log f(\vec{X}|\theta)]^2\},$$

称$\mathrm{I}_n(\theta)$为样本$X_1, \ldots, X_n$关于θ 的Fisher信息量。固定样本容量n，若估计的方差达到C-R不等式的下界，则$\mathrm{I}_n(\theta)$ 越大，表明估计的精度越高，这意味着该模型本身提供的信息量越大。

注2 定理2.1中条件（2），（3）分别等价于

$$\begin{aligned}\int \frac{\partial}{\partial\theta} f(x|\theta)dx &= \mathbb{E}_\theta[\frac{\partial}{\partial\theta}\log f(X|\theta)] = 0 \\ \frac{d}{d\theta}\mathbb{E}_\theta[W(\vec{X})] &= \mathbb{E}_\theta[W(\vec{X})\frac{\partial}{\partial\theta}\log f(\vec{X}|\theta)].\end{aligned}$$

注3 在样本$i.i.d.$情形下，若

1. $\int \frac{\partial}{\partial\theta} f(x|\theta)dx = 0$;
2. $Var(W) < \infty$;
3. $0 < \mathrm{I}(\theta) < \infty$.

则定理2.1中条件（2），（3）自动满足。

注4 若$f(x|\theta)$进一步满足

$$\frac{\partial}{\partial\theta}\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(X|\theta)\right] = \int \frac{\partial}{\partial\theta}\left\{[\frac{\partial}{\partial\theta}\log f(x|\theta)]f(x|\theta)\right\}dx,$$

则

$$\mathrm{I}(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right].$$

Corollary (2.2)

在定理*2.1* 条件下，若$W(\vec{X})$ 是$\tau(\theta)$的一个无偏估计，且*C-R*不等式*(2)*等号成立，则W是$\tau(\theta)$的*UMVUE*。

注1 定理2.1和推论2.1, 2.2对离散型随机样本同样适用，只要满足条件（1）—（3），此时积分变求和。

注2 如果$\frac{\partial}{\partial\theta}[f(\vec{x}|\theta)]W(\vec{x})$和$\frac{\partial}{\partial\theta}[f(\vec{x}|\theta)]$关于$\vec{x}$积分/求和一致收敛，则条件（1）—（3）自动成立。

注3 不等式(2)是单参数情形，对于多参数情形的C-R不等式参考【0】3.5.3小节。

举例说明

Example (2.2)

（例2.1续）设$X_1,\ldots,X_n$ $i.i.d. \sim Poisson(\lambda)$, $\lambda>0$未知，$n\geq 2$.
(1)验证$\widehat{\lambda}_{MoM}=\overline{X}$和$\widehat{\lambda}^*_{MoM}=S^2$均是$\lambda$的无偏估计;
(2)分别计算$Var(\widehat{\lambda}_{MoM})$和$Var(\widehat{\lambda}^*_{MoM})$，进而说明在均方误差MSE意义下，哪个估计更优？
(3)$\widehat{\lambda}_{MoM}$是否是λ的UMVUE?

问题 C-R不等式成立需要关键性假设条件（1)—（3），即积分和求导次序可互换。若此条件不满足，则不等式(2)可能不成立。

Example (2.3)

[【1】例7.3.13]设$X_1,\ldots,X_n$ $i.i.d. \sim U(0,\theta)$, $\theta>0$未知。记$X_1,\ldots,X_n$的联合密度函数为f，证明：$Y=\frac{n+1}{n}X_{(n)}$是θ的一个无偏估计量，但

$$Var_\theta(Y)<\frac{\{\frac{d}{d\theta}\mathbb{E}_\theta(Y)\}^2}{\mathbb{E}_\theta\{[\frac{\partial}{\partial\theta}\log f(\vec{X}|\theta)]^2\}}.$$

利用C-R不等式判定UMVUE的一个简单方法

Corollary (2.3)

在定理*2.1* 条件下，如果$W(\vec{X})$ 是$\tau(\theta)$的一个无偏估计，则W是$\tau(\theta)$的*UMVUE* 如果$\exists$ 函数$a(\theta)$，*s.t.*

$$a(\theta)(W(\vec{X})-\tau(\theta))=\frac{\partial}{\partial\theta}log\mathcal{L}(\theta|\vec{X}).$$

这里，$\mathcal{L}(\theta|\vec{X})=f(\vec{X}|\theta)$ 为样本$X_1,\ldots,X_n$的似然函数。

注 若样本分布族为指数族，其联合$p.d.f.$为

$$f(\vec{x}|\theta)=C(\theta)\exp\{Q(\theta)T(\vec{x})\}h(\vec{x}),$$

则其必满足定理2.1 条件(1)—(2)。因此由上述推论2.3可得，对于$\tau(\theta)$的方差有限的无偏估计$W(\vec{X})$，如果$\exists$函数$a(\theta)$, $\theta\in\Theta$, s.t.

$$a(\theta)(W(\vec{X})-\tau(\theta))=Q'(\theta)T(\vec{X})+C'(\theta)/C(\theta) \qquad (*)$$

则W必是$\tau(\theta)$的UMVUE。

练习 【0】例3.5.1 ∼ 例3.5.5。证明步骤：(1) 指明样本分布族为指数族; (2) 证明给出的估计量W无偏; (3) 找出$a(\theta)$使得上式$(*)$成立。

有效性

- 例2.2中的UMVUE的方差达到C-R不等式的下界，这样的估计称为有效估计。
- **定义2.5 有效性**（【0】定义3.1.3）设有样本$\vec{X} = (X_1, \ldots, X_n)$，$W_1(\vec{X})$和$W_2(\vec{X})$为$\tau(\theta)$的两个不同无偏估计量，若$Var_\theta(W_1) \le Var_\theta(W_2), \forall \theta \in \Theta$，且至少存在一个$\theta \in \Theta$，使得严格不等式成立，则称估计量$W_1(\vec{X})$比$W_2(\vec{X})$有效。

注 比较例2.2中λ的两个无偏估计量$\overline{X}$和S^2哪个更有效？

- **定义2.6 估计效率和有效估计**（【0】定义3.5.2）
  设$W(\vec{X})$为$\tau(\theta)$的无偏估计量，
  1. 比值$e_W(\theta) = \dfrac{|\tau'(\theta)|^2}{\mathrm{I}_n(\theta) Var_\theta(W)}$称为$W$的效率（efficiency）；
  2. 当$e_W(\theta) = 1$时，称$W(\vec{X})$为$\tau(\theta)$的有效估计（effective estimator）；
  3. 若$W(\vec{X})$不是$\tau(\theta)$的有效估计，但$\lim_{n\to\infty} e_W(\theta) = 1$，则称$W(\vec{X})$为$\tau(\theta)$的渐近有效估计（asymptotically effective estimator）．

注1 对于满足C-R不等式条件(1)–(3)的分布族和无偏估计W，必有$0 < e_W(\theta) \leq 1$; 且为有效估计时，W必为$UMVUE$.

注2 若分布族和无偏估计W不满足C-R不等式条件(1)—(3)，则可能出现$e_W(\theta) > 1$，见例2.3。此时讨论有效性失去意义。

问题 (1) $UMVUE$是否一定是有效估计？
(2) 在满足C-R不等式条件(1)—(3)的前提下，$UMVUE$是否一定是有效估计？

作业 习题3 Ex. 22, 41, 42, 43.

Lemma (2.1, Rao-Blackwell Theorem, 简记R-B引理)

（【0】引理3.4.1，【1】定理7.3.17）设W是$\tau(\theta)$的任意一个无偏估计量，而T是关于θ的一个充分统计量，定义$\phi(T)=\mathbb{E}(W|T)$，则$\mathbb{E}_\theta\phi(T)=\mathbb{E}_\theta W=\tau(\theta)$，而且

$$Var_\theta[\phi(T)]\leq Var_\theta(W),\forall\theta\in\Theta.$$

即$\phi(T)$是比W更有效的$\tau(\theta)$的无偏估计量。

- **证明** 见【0】.

注 给定任一统计量T，条件期望$\mathbb{E}_\theta(W|T)$从方差意义上都是对W的一个改进，但如果T不是一个充分统计量，则$\mathbb{E}_\theta(W|T)$表达式可能会依赖于θ，即$\mathbb{E}_\theta(W|T)$可能不是一个统计量。

练习1 设$X_1,\ldots,X_n\ i.i.d.\sim U(0,\theta)$，$Y=\mathbb{E}(\overline{X}|X_1)$是否是一个统计量？

练习2 设$X_1,\ldots,X_n\ i.i.d.\sim Bernoulli(p)$，$S=\sum_{i=1}^n X_i$，验证$\phi(S)=\mathbb{E}(X_1|S)$ 的方差比X_1小。（【0】例3.4.2）

寻找UMVUE的方法2：零无偏估计法

问题1 经过充分统计量改进后的无偏估计$\phi(T)$是否就是UMVUE?

问题2 参数θ的UMVUE是否唯一？

Lemma (2.2)

对于一个参数统计模型，参数$\tau(\theta)$的*UMVUE* 唯一。

Theorem (2.2)

（【0】定理*3.4.1*，【1】定理*7.3.20*）
设$\mathbb{E}_\theta(W)=\tau(\theta)$，$Var_\theta W<\infty$，则$W$是$\tau(\theta)$的*UMVUE*当且仅当$W$与0的所有无偏估计量不相关。

注 上述定理中也可以限定为0的所有方差有限的无偏估计量。

寻找UMVUE的方法2'

问题 定理2.2对于具体问题，操作上有一定难度。

Corollary (2.4)

（【0】推论3.4.1） 设$T = T(\vec{X})$是θ的充分统计量，$h(T)$是$\tau(\theta)$的一个无偏估计，$Var(h(T)) < \infty, \forall\theta \in \Theta$。则$h(T)$ 是$\tau(\theta)$的$UMVUE$，当且仅当$h(T)$与任一零无偏估计$U(T)$都不相关。

注 这里零无偏估计$U(T)$是充分统计量T的函数。

Example (2.4)

（【0】例3.4.6） 设$X_1, \ldots, X_n$ $i.i.d. \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}, \sigma^2 > 0$ 均未知，求μ的UMVUE.

注1 例2.4中，S^2是σ^2的UMVUE，见习题3，Ex. 30 (见下页作业)。

注2 分别验证$\overline{X}$和S^2是否各为μ, σ^2的有效估计。

注3 由注1，2的结果可知，在C-R不等式条件（1）–（3）成立前提下，有效估计$\Rightarrow$UMVUE，但UMVUE$\not\Rightarrow$有效估计。

寻找UMVUE的方法2'

Example (2.5)

（例2.3续，【0】例3.4.5） 设$X_1,\dots,X_n$ $i.i.d. \sim U(0,\theta)$, $\theta>0$ 未知，求θ的UMVUE.

注1 观察例2.4，样本均值$\overline{X}$是总体均值的UMVUE；而在例2.5中，$\frac{n+1}{2n}X_{(n)}$是总体均值的UMVUE。这说明在同一评价标准下，某一统计量的优劣不仅取决于统计量本身，而且与统计模型有关。

注2 练习【0】例3.4.3, 3.4.4.

注3 回忆第一章例7.2，可知$T=X_{(n)}$同时也是分布族$U(0,\theta)$的完全统计量，因此不存在非平凡的可测函数h，使得$\mathbb{E}[h(T)]=0$；即，例2.5中基于T的零无偏估计量存在唯一$U(T)=0$.

问题 是否基于充分完全统计量T的无偏估计$h(T)$必为UMVUE?

Theorem (2.3, Lehmann-Scheffè 定理，简记L-S定理)

（【0】定理3.4.2，【1】定理7.3.23）设T是参数θ的充分完全统计量，而$\phi(T)$是任意一个仅基于T的估计量，且$\mathbb{E}_\theta\phi(T)=\tau(\theta)$，则$\phi(T)$是$\tau(\theta)$的$UMVUE$.

- 证明 见【0】.

注 由定理证明过程可知，若W是$\tau(\theta)$的任一无偏估计量，则$\phi(T)=\mathbb{E}(W|T)$就是$\tau(\theta)$的UMVUE。

Example (2.6)

（【0】例3.4.9）设$X_1,\ldots,X_n$ $i.i.d. \sim Poisson(\lambda)$, $\lambda>0$ 未知，求

1. $g_1(\lambda)=\lambda^r$的UMVUE, 这里$r\in\mathbb{N}^+$;
2. $g_2(\lambda)=\mathbb{P}_\lambda(X_1=x)$的UMVUE.

练习 【0】例3.4.7 $\sim$ 3.4.12, 3.5.9.

作业 习题3 Ex. 30, 31, 32, 33, 36, 38, 40, 44.