习题6.4 第14 题第（1）小题的判断无偏性两种方法：

14．设 $X_1, X_2, \cdots, X_n$ 为独立同分布变量，$0<\theta<1$，

$$P\{X_1=-1\}=\frac{1-\theta}{2},\quad P\{X_1=0\}=\frac{1}{2},\quad P\{X_1=1\}=\frac{\theta}{2}。$$

（1）求 θ 的MLE $\hat{\theta}_1$，并问 $\hat{\theta}_1$ 是否无偏。

解：总体 X 概率函数为

$$p(x;\theta)=\left(\frac{1-\theta}{2}\right)^{\frac{x(x-1)}{2}}\left(\frac{1}{2}\right)^{-(x+1)(x-1)}\left(\frac{\theta}{2}\right)^{\frac{x(x+1)}{2}}=\frac{1}{2}(1-\theta)^{\frac{x^2-x}{2}}\theta^{\frac{x^2+x}{2}}，\ x=-1,0,1，$$

则似然函数 $L(\theta)=\prod_{i=1}^{n}\frac{1}{2}(1-\theta)^{\frac{x_i^2-x_i}{2}}\theta^{\frac{x_i^2+x_i}{2}}=\frac{1}{2^n}(1-\theta)^{\frac{1}{2}\left(\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}x_i\right)}\theta^{\frac{1}{2}\left(\sum_{i=1}^{n}x_i^2+\sum_{i=1}^{n}x_i\right)}$，

有 $\ln L(\theta)=\frac{1}{2}\left(\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}x_i\right)\ln(1-\theta)+\frac{1}{2}\left(\sum_{i=1}^{n}x_i^2+\sum_{i=1}^{n}x_i\right)\ln\theta-n\ln 2$，

令 $\frac{d\ln L(\theta)}{d\theta}=\frac{1}{2}\left(\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}x_i\right)\cdot\frac{-1}{1-\theta}+\frac{1}{2}\left(\sum_{i=1}^{n}x_i^2+\sum_{i=1}^{n}x_i\right)\cdot\frac{1}{\theta}=0$，

得 $\theta=\frac{\sum_{i=1}^{n}x_i^2+\sum_{i=1}^{n}x_i}{2\sum_{i=1}^{n}x_i^2}=\frac{1}{2}+\frac{\sum_{i=1}^{n}x_i}{2\sum_{i=1}^{n}x_i^2}$，

故 θ 的MLE $\hat{\theta}_1=\frac{1}{2}+\frac{\sum_{i=1}^{n}X_i}{2\sum_{i=1}^{n}X_i^2}$；

方法一：设 $X_1, X_2, \cdots, X_n$ 中取值 $-1, 0, 1$ 分别有 n_{-1}, n_0, n_1 次，有 $n_{-1}+n_0+n_1=n$，

则 $\sum_{i=1}^{n}X_i^2=n_1+n_{-1}$，$\sum_{i=1}^{n}X_i=n_1-n_{-1}$，

即 θ 的MLE $\hat{\theta}_1=\frac{\sum_{i=1}^{n}X_i^2+\sum_{i=1}^{n}X_i}{2\sum_{i=1}^{n}X_i^2}=\frac{n_1}{n_1+n_{-1}}$；

因 $E(\hat{\theta}_1)=E\left(\frac{n_1}{n_{-1}+n_1}\right)=E\left[E\left(\frac{n_1}{n_{-1}+n_1}\middle|n_{-1}+n_1\right)\right]=E\left[\frac{1}{n_{-1}+n_1}E(n_1|n_{-1}+n_1)\right]$，

且 $P\{X=1\,|\,X=-1或X=1\}=\frac{P\{X=1\}}{P\{X=-1或X=1\}}=\frac{\frac{\theta}{2}}{\frac{1-\theta}{2}+\frac{\theta}{2}}=\theta$，

则在 $n_{-1}+n_1=m$ 的条件下，n_1 服从二项分布 $b(m,\theta)$，$E(n_1|n_{-1}+n_1=m)=m\theta$，

可得 $E(n_1|n_{-1}+n_1)=(n_{-1}+n_1)\theta$，

故 $E(\hat{\theta}_1)=E\left[\frac{1}{n_{-1}+n_1}E(n_1|n_{-1}+n_1)\right]=E(\theta)=\theta$，$\hat{\theta}_1$ 是 θ 的无偏估计；

方法二：$E(\hat{\theta}_1)=\dfrac{1}{2}+\dfrac{1}{2}E\left(\dfrac{\sum\limits_{j=1}^{n}X_j}{\sum\limits_{i=1}^{n}X_i^2}\right)=\dfrac{1}{2}+\dfrac{1}{2}\sum\limits_{j=1}^{n}E\left(\dfrac{X_j}{\sum\limits_{i=1}^{n}X_i^2}\right)=\dfrac{1}{2}+\dfrac{n}{2}E\left(\dfrac{X_1}{X_1^2+\sum\limits_{i=2}^{n}X_i^2}\right)$

$$=\frac{1}{2}+\frac{n}{2}E\left[E\left(\frac{X_1}{X_1^2+\sum\limits_{i=2}^{n}X_i^2}\middle|\sum_{i=2}^{n}X_i^2\right)\right]$$

$$=\frac{1}{2}+\frac{n}{2}E\left(\frac{-1}{1+\sum\limits_{i=2}^{n}X_i^2}\cdot\frac{1-\theta}{2}+\frac{0}{0+\sum\limits_{i=2}^{n}X_i^2}\cdot\frac{1}{2}+\frac{1}{1+\sum\limits_{i=2}^{n}X_i^2}\cdot\frac{\theta}{2}\right)$$

$$=\frac{1}{2}+\frac{n}{2}\left(\theta-\frac{1}{2}\right)E\left(\frac{1}{1+\sum\limits_{i=2}^{n}X_i^2}\right),$$

因$\sum\limits_{i=2}^{n}X_i^2$是$X_2,\cdots,X_n$中取值$-1$或1的次数，则$\sum\limits_{i=2}^{n}X_i^2$服从二项分布$b\left(n-1,\dfrac{1}{2}\right)$，

即$P\left\{\sum\limits_{i=2}^{n}X_i^2=k\right\}=C_{n-1}^{k}\cdot\left(\dfrac{1}{2}\right)^{k}\left(\dfrac{1}{2}\right)^{n-1-k}=\dfrac{(n-1)!}{k!(n-1-k)!}\cdot\left(\dfrac{1}{2}\right)^{n-1}$，$k=0,1,\cdots,n-1$，

则$E(\hat{\theta}_1)=\dfrac{1}{2}+\dfrac{n}{2}\left(\theta-\dfrac{1}{2}\right)\cdot\sum\limits_{k=0}^{n-1}\dfrac{1}{1+k}\cdot\dfrac{(n-1)!}{k!(n-1-k)!}\cdot\left(\dfrac{1}{2}\right)^{n-1}$

$$=\frac{1}{2}+\frac{n}{2}\left(\theta-\frac{1}{2}\right)\cdot\left(\frac{1}{2}\right)^{n-1}\cdot\sum_{k=0}^{n-1}\frac{(n-1)!}{(k+1)!(n-1-k)!}=\frac{1}{2}+\frac{n}{2^n}\left(\theta-\frac{1}{2}\right)\cdot\sum_{k=0}^{n-1}\frac{1}{n}\cdot C_n^{k+1}$$

$$=\frac{1}{2}+\frac{n}{2^n}\left(\theta-\frac{1}{2}\right)\cdot\frac{1}{n}(2^n-1)=\frac{1}{2}+\left(\theta-\frac{1}{2}\right)\cdot\left(1-\frac{1}{2^n}\right)=\frac{1}{2^{n+1}}+\theta\left(1-\frac{1}{2^n}\right),$$

故$\hat{\theta}_1$不是θ的无偏估计；

两种方法结论不同，问题在于当所有样品全部取值0时，有$\sum\limits_{i=1}^{n}X_i^2=n_{-1}+n_1=0$，此时$\hat{\theta}_1=\dfrac{\sum\limits_{i=1}^{n}X_i^2+\sum\limits_{i=1}^{n}X_i}{2\sum\limits_{i=1}^{n}X_i^2}=\dfrac{n_1}{n_1+n_{-1}}$无定义，正确的做法应该是在$\sum\limits_{i=1}^{n}X_i^2=n_{-1}+n_1\neq 0$的条件下，计算$\hat{\theta}_1$的数学期望。事实上在$\sum\limits_{i=1}^{n}X_i^2=n_{-1}+n_1\neq 0$的条件下，$\hat{\theta}_1$的数学期望就是$\theta$。

方法一中相当于当$\sum\limits_{i=1}^{n}X_i^2=n_{-1}+n_1=0$时将$\hat{\theta}_1$取为$\theta$，所以最后的答案就是$\theta$；在方法二中相当于当$\sum\limits_{i=1}^{n}X_i^2=n_{-1}+n_1=0$时将$\hat{\theta}_1$取为$\dfrac{1}{2}$，且$\sum\limits_{i=1}^{n}X_i^2=n_{-1}+n_1=0$即所有样品全部取值0的

概率为$\frac{1}{2^n}$，所以最后的答案就是$\frac{1}{2}\cdot\frac{1}{2^n}+\theta\cdot\left(1-\frac{1}{2^n}\right)=\frac{1}{2^{n+1}}+\theta\left(1-\frac{1}{2^n}\right)$。可见方法一的结论是正确的。