日期	科目	班级	姓名	学号	序号
2022 年 10 月 16 日	数理统计	强基数学 002	吴天阳	2204210460	50

第三次作业

题目 1. 设 $X_1,\cdots,X_n$ 是来自某一个分布的随机样本，期望为 μ 方差为 σ^2.

1. 证明 $\sum\limits_{i=1}^n a_iX_i$ 是 μ 的无偏估计，其中 $\sum\limits_{i=1}^n a_i=1$.
2. 若 $\sum\limits_{i=1}^n a_i=1$，证明当 $a_i=\dfrac{1}{n}$, $(i=1,2,\cdots,n)$ 时，$\text{Var}\left(\sum\limits_{i=1}^n a_i\right)$ 取到最小值.

证明. 1. $\mathbb{E}\left(\sum\limits_{i=1}^n a_iX_i\right)=\sum\limits_{i=1}^n a_i\mathbb{E}(X_i)=\mu\sum\limits_{i=1}^n a_i=\mu$，于是 $\text{bias}\left(\sum\limits_{i=1}^n a_iX_i\right)=\mathbb{E}\left(\sum\limits_{i=1}^n a_iX_i\right)-\mu=0$，所以 $\sum\limits_{i=1}^n a_iX_i$ 是无偏估计.

2.

$$
\begin{aligned}
\text{Var}\left(\sum_{i=1}^n a_iX_i\right)&=\mathbb{E}\left(\left(\sum_{i=1}^n a_iX_i\right)^2\right)-\mu^2=(\sigma^2+\mu^2)\sum_{i=1}^n a_i^2+2\mu^2\sum_{1\leqslant i<j\leqslant n}a_ia_j-\mu^2\\
&=\sigma^2\sum_{i=1}^n a_i^2+\mu^2\left(\sum_{i=1}^n a_i^2+2\sum_{1\leqslant i<j\leqslant n}a_ia_j-1\right)\\
&=\sigma^2\left(\sum_{i=1}^n\left(a_i-\frac{1}{n}\right)^2+\frac{1}{n}\right)+\mu^2\left(\left(\sum_{i=1}^n a_i\right)^2-1\right)\\
&=\sigma^2\left(\sum_{i=1}^n\left(a_i-\frac{1}{n}\right)^2+\frac{1}{n}\right)
\end{aligned}
$$

所以当 $a_i=\dfrac{1}{n}$ 时，$\text{Var}\left(\sum\limits_{i=1}^n a_i\right)$ 取到最小值. □

题目 2. 设 $X_1,X_2,\cdots,X_n$ 是来自 $f(x;\theta)=\dfrac{\theta}{x^2}I_{[\theta,\infty)}(x)$ 的随机变量，其中 $\theta>0$.

1. 求 θ 的 MLE.
2. $Y_1=\min\{X_1,\cdots,X_n\}$ 是充分统计量么?

解答. 1.

$$
\mathcal{L}(\theta)=\prod_{i=1}^n f(x_i;\theta)=\prod_{i=1}^n\frac{\theta}{x_i^2}I_{[\theta,\infty)}(x_i)=\frac{\theta^n}{\prod_{i=1}^n X_i^2}I_{(-\infty,y_1]}(\theta)
$$

所以 $\hat\theta=Y_1$.

2. 由于 $\prod\limits_{i=1}^n f(x_i;\theta)=\dfrac{\theta^n}{\prod_{i=1}^n x_i^2}I_{(-\infty,\theta]}(y_1)$，令

$$
S=Y_1,\ g(s;\theta)=\theta^2I_{(-\infty,\theta]}(y_1),\ h(x_1,\cdots,x_n)=\left(\prod_{i=1}^n x_i^2\right)^{-1}
$$

则 $\prod\limits_{i=1}^n f(x_i;\theta)=g(s;\theta)h(x_1,\cdots,x_n)$，由因子分解定理可知 Y_1 是充分统计量.

题目 3. 设 $X_1,\cdots,X_n$ 是来自 $f(x;\theta)=\dfrac{1}{\theta}I_{[0,\theta]}(x)$ 的随机变量，其中 $\theta>0$，定义 $Y_n=\max\{X_1,\cdots,X_n\},Y_1=\min\{X_1,\cdots,X_n\}$.

1, 2. 对 θ 分别求解矩估计 T_1，最大似然估计 T_2，并分别计算期望和均方误差.

3. 对于所有形式为 aY_n 的估计量，其中 a 是依赖于 n 的常数，其中具有最小的均方误差记为 T_3，求出 T_3 并计算期望和均方误差.

4. 令 $T_4=Y_1+Y_n$，求出 T_4 的期望和均方误差.

5. 如何确定哪个估计量最优.

6. 求关于总体方差的 MLE.

解答. 1. 矩估计：$\bar{X}=E(X)=\dfrac{\theta}{2}\Rightarrow T_1=2\bar{X}$. 期望：$\mathbb{E}(T_1)=\mathbb{E}(2\bar{X})=\theta$，所以 T_1 是无偏估计. 均方误差：$\mathrm{MSE}(T_1)=\mathrm{Var}(T_1)=4\dfrac{\mathrm{Var}(X)}{n}=\dfrac{\theta^2}{3n}$.

2. MLE：$\mathcal{L}(\theta)=\displaystyle\prod_{i=1}^n f(x_i;\theta)=\frac{1}{\theta^n}\prod_{i=1}^n I_{[0,\theta]}(x_i)=\frac{1}{\theta^n}I_{[y_n,\infty)}(\theta)I_{[0,y_n]}(y_1)$，则 $T_2=Y_n$.

由于 $f_{Y_n}(x)=nF_X(x)^{n-1}f_X(x)=n\dfrac{x^{n-1}}{\theta^n}I_{[0,\theta]}(x)$，于是 T_2 的期望为

$$\mathbb{E}(T_2)=\mathbb{E}(Y_n)=\int_0^\theta n\frac{x^n}{\theta^n}\,\mathrm{d}x=\frac{n}{n+1}\theta,$$
$$\mathbb{E}(T_2^2)=\mathbb{E}(Y_n^2)=\int_0^\theta n\frac{x^{n+1}}{\theta^n}\,\mathrm{d}x=\frac{n}{n+2}\theta^2.$$

均方误差：$\mathrm{MSE}(T_2)=\mathbb{E}((T_2-\theta)^2)=\mathbb{E}(T_2^2)-2\theta\mathbb{E}(T_2)+\theta^2=\dfrac{2}{(n+1)(n+2)}\theta^2$.

3. 设 $T_3=aY_n$，由第二问可知 $\mathbb{E}(T_3)=\dfrac{an}{n+1}\theta$, $\mathbb{E}(T_3^2)=\dfrac{a^2n}{n+2}\theta^2$，则

$$\mathrm{MSE}(T_3)=\frac{a^2n}{n+2}\theta^2-\frac{2an}{n+1}\theta^2+\theta^2=\frac{a^2n(n+1)-2an(n+2)+(n+1)(n+2)}{(n+1)(n+2)}\theta^2.$$

令 $\varphi(a)=a^2n(n+1)-2an(n+2)+(n+1)(n+2)$，则 $\varphi'(a_0)=0\Rightarrow a_0=\dfrac{n+2}{n+1}$，所以 $T_3=\dfrac{n+2}{n+1}Y_n$，对应的期望和均方误差分别为

$$\mathbb{E}(T_3)=\frac{n+2}{n+1}\mathbb{E}(T_2)=\frac{n(n+2)}{(n+1)^2}\theta,\quad \mathrm{MLE}(T_3)=\frac{\theta^2}{(n+1)^2}.$$

4. 由次序统计量的性质可知 $f_{Y_1,Y_n}=\dfrac{n(n-1)(y_n-y_1)^{n-2}}{\theta^n}$, $(0\leqslant y_1\leqslant y_n\leqslant\theta)$，于是

$$
\begin{aligned}
\mathbb{E}(Y_1Y_n)&=\int_0^\theta\int_{y_1}^\theta y_1y_n\frac{n(n-1)(y_n-y_1)^{n-2}}{\theta^n}\,\mathrm{d}y_n\,\mathrm{d}y_1\\
&\xlongequal{x=y_n-y_1}\frac{n(n-1)}{\theta^n}\int_0^\theta y_1\int_0^{\theta-y_1}(x+y_1)x^{n-2}\,\mathrm{d}x\,\mathrm{d}y_1\\
&=\frac{n(n-1)}{\theta^n}\int_0^\theta y_1\left(\frac{(\theta-y_1)^n}{n}+\frac{y_1(\theta-y_1)^{n-1}}{n-1}\right)\mathrm{d}y_1\\
&\xlongequal{y=\theta-y_1}\frac{n(n-1)}{\theta^n}\int_0^\theta\left((\theta-y)\frac{y^n}{n}+(\theta-y)^2\frac{y^{n-1}}{n-1}\right)\mathrm{d}y\\
&=\frac{n(n-1)}{\theta^n}\left(\frac{1}{n(n+1)}-\frac{1}{n(n+2)}+\frac{1}{n(n-1)}-\frac{2}{(n+1)(n-1)}+\frac{1}{(n+2)(n-1)}\right)\theta^{n+2}\\
&=\frac{\theta^2}{n+2},
\end{aligned}
$$

由于 $f_{Y_1}=\dfrac{n}{\theta}(1-\dfrac{y_1}{\theta})^{n-1}$，则

$$
\begin{aligned}
\mathbb{E}(Y_1)&=\int_0^\theta\frac{n}{\theta}y_1\left(1-\frac{y_1}{\theta}\right)^{n-1}\mathrm{d}y_1\xlongequal{\theta(1-x)=y_1}\theta n\int_0^1(x^{n-1}-x^n)\,\mathrm{d}x=\frac{\theta}{n+1},\\
\mathbb{E}(Y_n)&=\int_0^\theta\frac{n}{\theta}y_1^2\left(1-\frac{y_1}{\theta}\right)^{n-1}\mathrm{d}y_1\xlongequal{\theta(1-x)=y_1}\theta^2 n\int_0^1(1-x)^2x^{n-1}\,\mathrm{d}x=\frac{2}{(n+1)(n+2)}\theta^2,
\end{aligned}
$$

由第二小问可知 $\mathbb{E}(Y_n)=\dfrac{n}{n+1}\theta$, $\mathbb{E}(Y_n^2)=\dfrac{n}{n+2}\theta^2$，于是

$$
\begin{aligned}
\mathbb{E}(T_4)&=\mathbb{E}(Y_1+Y_n)=\mathbb{E}(Y_1)+\mathbb{E}(Y_n)=\frac{\theta}{n+1}+\frac{n}{n+1}\theta=\theta,\text{（所以}T_4\text{是无偏估计）}\\
\mathbb{E}(T_4^2)&=\mathbb{E}(Y_1^2)+2\mathbb{E}(Y_1Y_n)+\mathbb{E}(Y_n^2)=\frac{n^2+3n+4}{(n+1)(n+2)}\theta^2,\\
\mathrm{MSE}(T_4)&=\mathrm{Var}(T_4)=\mathbb{E}(T_4^2)-\mathbb{E}(T_4)^2=\frac{n^2+3n+4}{(n+1)(n+2)}\theta^2-\theta^2=\frac{2}{(n+1)(n+2)}\theta^2.
\end{aligned}
$$

5. T_3 具有最小的 MSE，所以 T_3 最优.

6. 由于总体方差是关于 θ 的函数，则其 MLE 为 $\dfrac{\hat{\theta}^2}{12}$，其中 $\hat{\theta}$ 为 θ 的估计量.

题目 4. 设 $X_1,\cdots,X_n$ 是来自均匀分布 $U(\theta,2\theta)$, $\theta>0$ 的样本，试给出充分统计量.

解答. 由于 $f(x)=\dfrac{1}{\theta}I_{[\theta,2\theta]}(x)$，于是

$$
\prod_{i=1}^n f(x_i)=\frac{1}{\theta^n}\prod_{i=1}^n I_{[\theta,2\theta]}(x_i)=\frac{1}{\theta^n}I_{[\theta,y_n]}(y_1)I_{[y_1,2\theta]}(y_n),
$$

所以 θ 的充分统计量可以为 Y_1,Y_n.

题目 5. 设 $X_1,\cdots,X_n$ 是来自 Gamma 分布 $\{\Gamma(a,\lambda):\alpha>0,\lambda>0\}$ 的样本，寻求 α,λ 的充分统计量.

解答. 由于 $f(x)=\dfrac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\lambda x}$，于是

$$\prod_{i=1}^n f(x_i)=\frac{\lambda^{n\alpha}}{\Gamma(\alpha)^n}\left(\prod_{i=1}^n x_i\right)^{\alpha-1}\mathrm{e}^{-\lambda\sum\limits_{i=1}^n x_i}$$

所以 α,λ 的充分统计量可以是 $\sum\limits_{i=1}^n X_i,\prod\limits_{i=1}^n X_i$.

题目 6. 设 $X_1,\cdots,X_n$ 是 $N(\mu,\sigma_1^2)$ 的样本，$Y_1,\cdots,Y_m$ 是来自 $N(\mu,\sigma_2^2)$ 的样本，这两个样本相互独立，试给出 $\mu,\sigma_1^2,\sigma_2^2$ 的充分统计量.

解答. 由于 $f_X(x)=\dfrac{1}{\sqrt{2\pi\sigma_1^2}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma_1^2}}$，$f_Y(x)=\dfrac{1}{\sqrt{2\pi\sigma_2^2}}\mathrm{e}^{-\frac{(y-\mu)^2}{2\sigma_1^2}}$，于是

$$\begin{aligned}\prod_{i=1}^n f_X(x_i)\prod_{j=1}^m f_Y(y_j)&=\frac{1}{(2\pi\sigma_1^2)^{\frac{n}{2}}(2\pi\sigma_2^2)^{\frac{m}{2}}}\exp\left\{-\frac{\sum_{i=1}^n(x_i-\mu)^2}{2\sigma_1^2}-\frac{\sum_{j=1}^n(y_j-\mu)^2}{2\sigma_2^2}\right\}\\&=\frac{1}{(2\pi\sigma_1^2)^{\frac{n}{2}}(2\pi\sigma_2^2)^{\frac{m}{2}}}\exp\left\{-\frac{\sum_{i=1}^n x_i^2-2\mu\sum_{i=1}^n x_i+\mu^2}{2\sigma_1^2}-\frac{\sum_{j=1}^n y_j-2\mu\sum_{j=1}^m y_j+\mu^2}{2\sigma_2^2}\right\}\end{aligned}$$

所以 $\mu,\sigma_1^2,\sigma_2^2$ 的充分统计量可以为 $\sum\limits_{i=1}^n X_i^2,\sum\limits_{i=1}^n X_i,\sum\limits_{j=1}^m Y_i^2,\sum\limits_{j=1}^m Y_i$.