日期	科目	班级	姓名	学号
2022 年 10 月 1 日	数理统计	强基数学 002	吴天阳	2204210460

第二次作业

题目 1. 设 $X_1,\cdots,X_n$ 来自 $N(0,1)$ 的随机样本，令

$$\bar{X}_k=\frac{1}{k}\sum_{i=1}^{k}X_i,\quad \bar{X}_{n-k}=\frac{1}{n-k}\sum_{i=k+1}^{n}X_i,$$

求下列分布：(1) $\frac{1}{2}(\bar{X}_k+\bar{X}_{n-k})$. (2) $k\bar{X}_k^2+(n-k)\bar{X}_{n-k}^2$. (3) X_1^2/X_2^2. (4) X_1/X_n.

解答. (1) $\bar{X}_k\sim N(0,\frac{1}{k})$, $\bar{X}_{n-k}\sim N(0,\frac{1}{n-k})$，所以 $\frac{\bar{X}_k+\bar{X}_{n-k}}{2}\sim(0,\frac{1}{4}(\frac{1}{k}+\frac{1}{n+k}))$.

(2) $\sqrt{k}\bar{X}_k\sim N(0,1)$, $\sqrt{n-k}\bar{X}_{n-k}\sim N(0,1)$，所以 $k\bar{X}_k^2+(n-k)\bar{X}_{n-k}^2\sim\chi(2)$.

(3) X_1^2, $X_2^2\overset{iid}{\sim}\chi(1)$，所以 $\frac{X_1^2}{X_2^2}\sim F(1,1)$.

(4) 设 $U=X_1/X_2, V=X_2$，于是 $x_1=uv, x_2=v$，$\det(J)=\begin{vmatrix}v&u\\0&1\end{vmatrix}=v$，且 $p_{X_1,X_2}(x_1,x_2)=\frac{1}{2\pi}e^{-\frac{x_1^2+x_2^2}{2}}$，则

$$p_{U,V}(u,v)=P_{X_1,X_2}(uv,v)|v|=\frac{1}{2\pi}e^{-\frac{u^2v^2+v^2}{2}}|v|=\frac{|v|}{2\pi}e^{-\frac{v^2(u^2+1)}{2}},$$

则

$$P_{X_1/X_2}(u)=P_U(u)=2\int_0^\infty\frac{v}{2\pi}e^{-\frac{v^2(u^2+1)}{2}}\,dv=\frac{1}{\pi(u^2+1)},$$

所以 $U\sim\text{Cauchy}(0,1)$ 为标准 Cauchy 分布.

题目 2. 设 $X_1,\cdots,X_n$ 是来自双参数指数分布

$$f(x,\mu,\sigma)=\frac{1}{\sigma}\exp\left\{-\frac{x-\mu}{\sigma}\right\},\quad x\geqslant\mu$$

的简单随机样本，其中 $u\in\mathbb{R},\sigma>0$，求 μ,σ 和 $P(X_1\geqslant t)$, $(t>\mu)$ 的矩估计和 MLE.

解答. 矩估计：由于

$$\begin{aligned}E(X)&=\int_\mu^\infty\frac{x}{\sigma}\exp\left\{-\frac{x-\mu}{\sigma}\right\}\,dx=\frac{1}{\sigma}\left(\int_0^\infty xe^{\frac{x}{\sigma}}+\mu e^{-\frac{x}{\sigma}}\right)\,dx\xlongequal{\text{凑 Gamma 分布}}\sigma+\mu\\E(X^2)&=\int_\mu^\infty\frac{x^2}{\sigma}\exp\left\{-\frac{x-\mu}{\sigma}\right\}\,dx=\frac{1}{\sigma}\left(\int_0^\infty x^2e^{-\frac{x}{\sigma}}+2\mu\int_0^\infty xe^{-\frac{x}{\sigma}}+\mu^2\int_0^\infty e^{-\frac{x}{\sigma}}\right)\,dx\\&\xlongequal{\text{凑 Gamma 分布}}2\sigma^2+2\mu\sigma+\mu^2\end{aligned}$$

则 μ,σ 的矩估计为

$$\begin{cases}\bar{X}=\sigma+\mu,\\M_2'=2\sigma^2+2\mu\sigma+\mu^2.\end{cases}\Rightarrow\begin{cases}\hat{\mu}=\bar{X}-\sqrt{M_2'-\bar{X}^2},\\\hat{\sigma}=\sqrt{M_2'-\bar{X}^2}.\end{cases}$$

于是 $P(X_1 \geqslant t)$ 的矩估计为

$$P(X_1 \geqslant t) = \int_t^{\infty} \frac{1}{\sigma} \exp\left\{-\frac{x-\mu}{\sigma}\right\} \mathrm{d}x = \sqrt{\frac{\pi}{\sigma}} \int_{t-\mu}^{\infty} \frac{1}{\sqrt{\pi\sigma}} \exp\left\{-\frac{x}{\sigma}\right\} \mathrm{d}x = \sqrt{\pi}\hat{\sigma}(1-\Phi(t-\hat{\mu})).$$

MLE：似然函数 $L(\mu,\sigma) = \prod_{i=1}^{n} \frac{1}{\sigma} \exp\left\{-\frac{x_i-\mu}{\sigma}\right\} = \frac{1}{\sigma^n} \exp\left\{-\frac{\sum_{i=1}^n x_i - n\mu}{\sigma}\right\}$，对数似然为 $\log L(\mu,\sigma) = -n\log\sigma - \frac{1}{\sigma}\left(\sum_{i=1}^n x_i - n\mu\right)$，但方程 $\frac{\partial \log L}{\partial \mu} = \frac{n}{\sigma} = 0$ 无解，所以该分布不存在 MLE.

题目 3. 设 $X_1,\cdots,X_n$ 和 $Y_1,\cdots,Y_n$ 分别来自正态总体 $N(\mu_1,\sigma^2)$ 和 $N(\mu_2,\sigma^2)$ 的随机样本，求 μ_1,μ_2,σ^2 的 MLE.

解答. 似然函数为 $L(\mu_1,\mu_2,\sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x_i-\mu_1)^2}{2\sigma^2}} \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(y_j-\mu_2)^2}{2\sigma^2}}$,对数似然为 $\log L(\mu_1,\mu_2,\sigma^2) = -2n\log(\sqrt{2n\sigma^2}) - \frac{1}{2\sigma^2}\left(\sum_{i=1}^n (x_i-\mu_1)^2 + \sum_{j=1}^n (y_j-\mu_2)^2\right)$，于是

$$\begin{cases} \frac{\partial \log L}{\partial \mu_1} = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i-\mu_1) = 0, \\ \frac{\partial \log L}{\partial \mu_2} = \frac{1}{\sigma^2}\sum_{j=1}^n (y_i-\mu_2) = 0, \\ \frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma^2} + \frac{1}{2\sigma^4}\left(\sum_{i=1}^n (x_i-\mu_1)^2 + \sum_{j=1}^n (y_j-\mu_2)^2\right) = 0. \end{cases} \Rightarrow \begin{cases} \hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^n x_i = \bar{X}, \\ \hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^n y_i = \bar{Y}, \\ \hat{\sigma}^2 = \frac{1}{2n}\left(\sum_{i=1}^n (x_i-\hat{\mu}_1)^2 + \sum_{j=1}^n (y_j-\mu_2)^2\right) \end{cases}$$

题目 4. 设 $X_1,\cdots,X_n$ 来自均匀分布 $U(\theta,2\theta)$ 的随机样本，其中 $\theta>0$，求：(1) θ 的 MLE $\hat{\theta}$；(2) 判断 $\hat{\theta}$ 是否是无偏估计，如果不是无偏估计，基于 $\hat{\theta}$ 构造一个无偏估计.

解答. (1) 似然函数 $L(\theta) = \frac{1}{\theta}\prod_{i=1}^n I_{[\theta,2\theta]}(x_i)$，令 $Y_1,\cdots,Y_n$ 为 $X_1,\cdots,X_n$ 的次序统计量，则 $L(\theta)=1$ 时，有 $y_1 \geqslant \theta, y_n \leqslant 2\theta \Rightarrow \frac{y_n}{2} \leqslant \theta \leqslant y_1$. 则 $L(\theta) = \frac{1}{\theta} I_{[\frac{y_n}{2},y_1]}(\theta)$，因此 $\hat{\theta} = \frac{y_n}{2}$.

(2) 由于 $F_{Y_n}(y) = P(X_1,\cdots,X_n \leqslant y) = (F_X(y))^n = (\frac{y-\theta}{\theta})^n$, $y \in [\theta,2\theta]$，于是

$$f_{Y_n}(y) = \frac{n}{\theta}(\frac{y-\theta}{\theta})^{n-1} I_{[\theta,2\theta]}(y)$$

则

$$E(\hat{\theta}) = E(\frac{y_n}{2}) = \frac{1}{2}\int_\theta^{2\theta} y\frac{n}{\theta}(\frac{y-\theta}{\theta})^{n-1}\,\mathrm{d}y = \frac{n}{2\theta^n}\int_0^\theta (y+\theta)y^{n-1}\,\mathrm{d}y = \frac{2n+1}{2n+2}\theta.$$

所以 $\hat{\theta}$ 为 θ 的有偏估计，构造 $\hat{\theta}' = \frac{2n+2}{2n+1}\hat{\theta} = \frac{n+1}{2n+1}y_n$ 是 θ 的无偏估计.

题目 5. 设 X 服从对数正态分布，即 $\ln X \sim N(\mu,\sigma^2)$，其中 $\mu \in \mathbb{R}, \sigma > 0$. 设 $X_1,\cdots,X_n$ 来自总体 X 的随机样本，求：(1) μ,σ^2 的 MLE；(2) 求 $E(X)$ 的 MLE.

解答. (1) 令 $Y=\log X$，则 $f_X(x)=f_Y(\log x)\dfrac{1}{x}=\dfrac{1}{x\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(\log x-\mu)^2}{2\sigma^2}}$，则似然函数

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

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

于是

$$\begin{cases}\dfrac{\partial\log L}{\partial\mu}=\dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^n(\log x_i-\mu)=0,\\ \dfrac{\partial\log L}{\partial\sigma^2}=-\dfrac{n}{2\sigma^2}+\dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^n(\log x_i-\mu)^2=0.\end{cases}\quad\Rightarrow\quad\begin{cases}\hat\mu=\dfrac{1}{n}\displaystyle\sum_{i=1}^n\log x_i,\\ \hat\sigma^2=\dfrac{1}{n}\displaystyle\sum_{i=1}^n(\log x_i-\hat\mu)^2.\end{cases}$$

(2) $E(X)$ 的 MLE 为

$$E(x)=\int_0^\infty\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(\log x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x=\int_R\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2-2\sigma^2x}{2\sigma^2}}\,\mathrm{d}x$$
$$\overset{\text{凑 Gauss 分布}}{=\!=\!=\!=}\frac{1}{\sqrt{2\pi\hat\sigma^2}}\mathrm{e}^{2\hat\mu+\hat\sigma^2}.$$

编程作业 1. 假设 $X_1,\cdots,X_n$ 是来自总体 X 的随机样本，$X\sim\chi^2(k)$.

(1) 求样本均值 $\bar X$ 的密度函数.

(2) 求样本均值的渐近分布.

(3) 通过编程比较，在不同样本量下，样本均值的密度函数和其渐近分布的密度函数图像.

解答. (1) $n\bar X=\displaystyle\sum_{i=1}^n X_i$，由于 $X_i\sim\chi(k)=\Gamma(\dfrac{k}{2},\dfrac{1}{2})$，且 Gamma 函数有可加性，则 $N\bar X\sim\Gamma(\dfrac{nk}{2},\dfrac{1}{2})$

$$p_{\bar X}(x)=p_{n\bar X}(nx)n=n^{\frac{nk}{2}}\frac{(\frac{1}{2})^{\frac{nk}{2}}}{\Gamma(\frac{nk}{2})}x^{\frac{nk}{2}-1}\mathrm{e}^{-\frac{nx}{2}}.$$

(2) $\chi^2(k)$ 的 $\mu=k,\sigma^2=2k$，则 $\bar X$ 的渐近分布为 $N(k,\dfrac{2k}{n})$.

(3) 设定每次计算密度函数时使用 10^7 个样本，k 表示卡方分布的自由度，N 表示每个样本的采样数，结果如图 1 所示.

编程作业 2. 在一个图上画出标准正态分布的密度曲线和 $t(1),t(3),t(30),t(100)$ 的密度曲线.

解答. 直接绘图，结果如图 2 所示.

编程作业 3. 令 $X_1,\cdots,X_n$ 是来自均匀分布 $U[\mu-\sqrt3\sigma,\mu+\sqrt3\sigma]$ 的随机样本，其中 $\mu\in\mathbb{R},\sigma>0$. 编程比较 μ 的矩估计和 MLE 的偏，方差和均方误差.

解答. 由于 $E(x)=\displaystyle\int_{\mu-\sqrt3\sigma}^{\mu+\sqrt3\sigma}\frac{1}{2\sqrt3\sigma}x\,\mathrm{d}x=\mu$，则 μ 的矩估计为 $\hat\mu=\bar X$.

由课上习题可知，μ 的 MLE 估计为 $\hat\mu=\dfrac{y_1+y_n}{2}$，其中 $Y_1,\cdots,Y_n$ 为 $X_1,\cdots,X_n$ 的次序统计量.

通过程序计算，取 $\mu=0,\sigma=1$，每个样本大小 $n=10^5$，总共取 10^5 个样本，计算得到

$$\text{矩估计：}\quad E(\hat{\mu})\approx 1.96\times10^{-6},\ Var(\hat{\mu})\approx 0.003,\ \text{MSE}(\hat{\mu})\approx 0.003$$
$$\text{MLE：}\quad E(\hat{\mu})\approx 6.19\times10^{-8},\ Var(\hat{\mu})\approx 2.46\times10^{-5},\ \text{MSE}(\hat{\mu})\approx 2.46\times10^{-5}$$

由此看出 MLE 的估计效果优于矩估计方法.

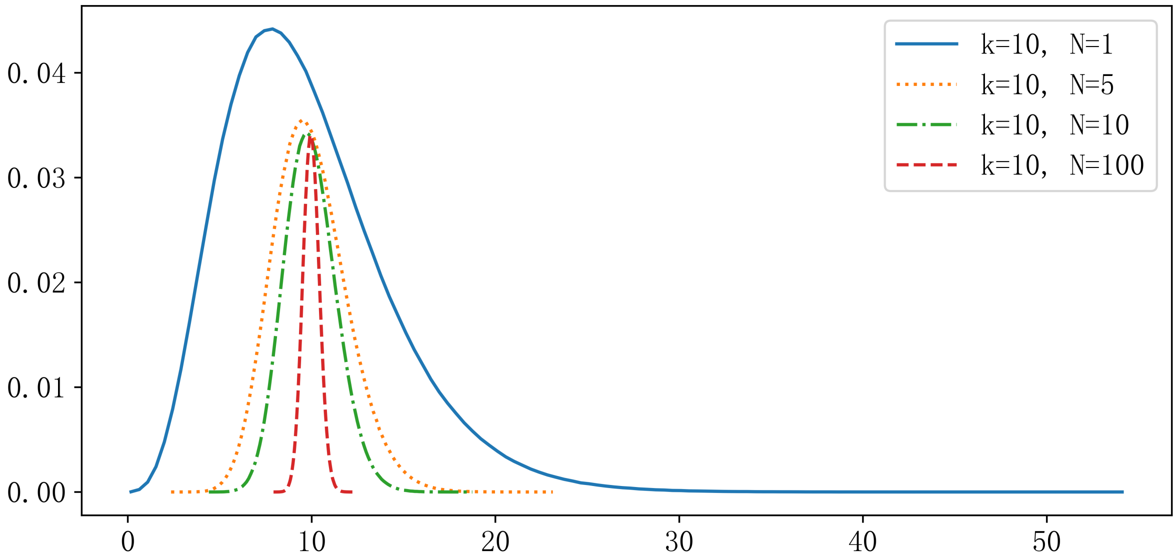

图 1: 第一题

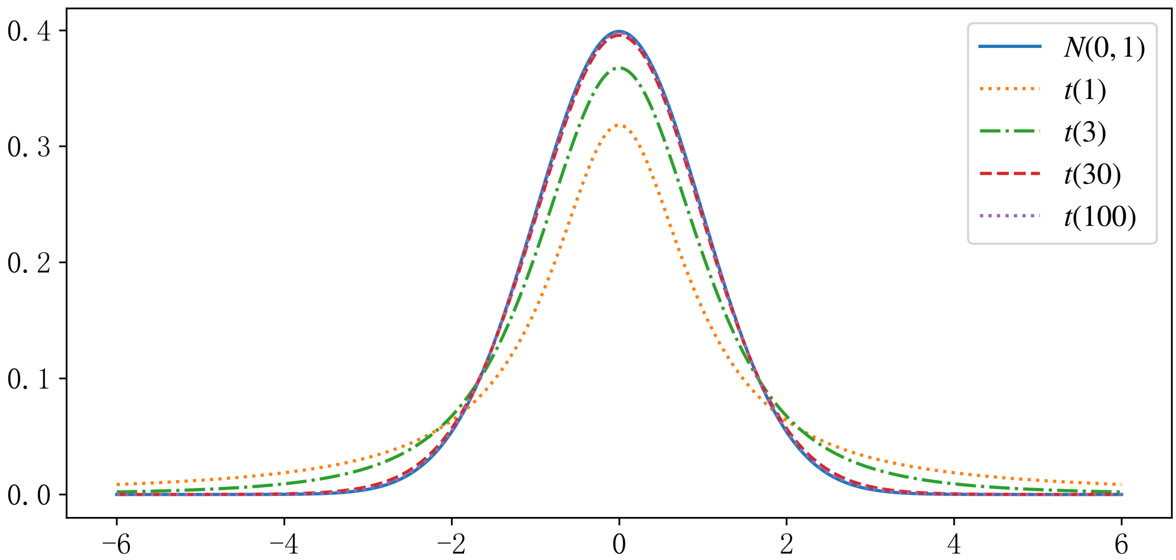

图 2: 第二题