2021 年秋统计学习题 05 参考解答

1. 设总体的概率密度函数如下

$$f(x|\theta) = \frac{1}{2}(1+\theta x), \quad -1 < x < 1, -1 < \theta < 1,$$

$X_1, X_2, \cdots, X_n$ 为样本，求 θ 的矩估计，并证明它是弱相容的.

解 1) 总体的一阶矩为

$$\begin{aligned} EX &= \int_{-1}^{1} x \cdot \tfrac{1}{2}(1+\theta x)\,dx \\ &= \int_{-1}^{1} \left(\tfrac{1}{2}x + \tfrac{1}{2}\theta x^2\right) dx \\ &= \tfrac{1}{6}\theta x^3 \Big|_{-1}^{1} \\ &= \tfrac{1}{3}\theta \end{aligned}$$

由 $\bar{X} = \frac{1}{3}\theta$ 可解得 θ 的矩估计为

$$\hat{\theta} = 3\bar{X}.$$

2) 总的二阶矩为

$$EX^2 = \int_{-1}^{1} x^2 \cdot \tfrac{1}{2}(1+\theta x)\,dx$$

$$= \tfrac{1}{6}x^3 \Big|_{-1}^{1} = \tfrac{1}{3}.$$

因此总体的方差为

$$\mathrm{Var}X = EX^2 - (EX)^2 = \frac{1}{3} - \frac{\theta^2}{9}.$$

由 Chebyshev 不等式（注意到 $E3\bar{X} = \theta$），可得

对任意 $\varepsilon > 0$,

$$\mathbb{P}(|3\bar{X} - \theta| > \varepsilon) \leq \frac{Var(3\bar{X})}{\varepsilon^2} = \frac{9}{n\varepsilon^2}\left(\frac{1}{3} - \frac{\theta^2}{9}\right).$$

由此可见，当 $n \to +\infty$ 时，

$$\mathbb{P}(|3\bar{X} - \theta| > \varepsilon) \to 0,$$

即 $3\bar{X}$ 依概率收敛到 θ.

注：令 $Y_i = 3X_i$，则 $3\bar{X} = \bar{Y}$，$E\bar{Y} = \theta$，

$Var\,Y_i = 9\,Var\,X_i = 9\left(\frac{1}{3} - \frac{\theta^2}{9}\right) = 3 - \theta^2$ 有限，可

对 $Y_1, Y_2, \cdots, Y_n$ 应用弱大数定律直接得知

$\bar{Y}$ 依概率收敛到 θ.

2. 设样本 $X_1, X_2, \cdots, X_n$ 来自正态总体 $N(\theta,\theta)$，其中 $\theta>0$ 为参数.

(a). 求 θ 的极大似然估计 $\hat{\theta}$.

(b). 求 $\hat{\theta}$ 的近似方差.

解 a) 似然函数为

$$L(\theta)=\prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi\theta}}\exp\left(-\frac{1}{2\theta}(x_i-\theta)^2\right)\right]$$

$$=(2\pi\theta)^{-\frac{n}{2}}\exp\left[-\frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\theta)^2\right]$$

对数似然函数为

$$\ell(\theta)=-\frac{n}{2}\log(2\pi\theta)-\frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\theta)^2.$$

它关于 θ 的导数为

$$\ell'(\theta)=-\frac{n}{2\theta}+\frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i-\theta)^2+\frac{1}{\theta}\sum_{i=1}^{n}(x_i-\theta)$$

$$=\frac{1}{2\theta^2}\left[-n\theta+\sum_{i=1}^{n}(x_i-\theta)(x_i-\theta+2\theta)\right]$$

$$=\frac{1}{2\theta^2}\left[-n\theta+\sum_{i=1}^{n}(x_i-\theta)(x_i+\theta)\right]$$

$$=\frac{1}{2\theta^2}\left[-n\theta+\sum_{i=1}^{n}(x_i^2-\theta^2)\right]$$

$$=\frac{-n}{2\theta^2}\left[\theta^2+\theta-\frac{1}{n}\sum_{i=1}^{n}x_i^2\right]. \qquad (1)$$

记 $t=\frac{1}{n}\sum_{i=1}^{n}x_i^2$，则 $\ell'(\theta)=0$ 的解为

方程

$$\theta^2+\theta-t=0$$

的解

$$\theta=\frac{1}{2}\left[-1\pm\sqrt{1+4t}\right].$$

注意到 $\theta>0$，所以 θ 的极大似然估计为

$$\hat{\theta}=\frac{1}{2}\left(\sqrt{1+4T}+1\right),$$

其中

$$T=\frac{1}{n}\sum_{i=1}^{n}X_i^2.$$

2)

由 (1) 可知

$$l'(\theta)=-\frac{n}{2}-\frac{n}{2}\cdot\frac{1}{\theta}+\frac{n}{2\theta^2}t.$$

因此

$$l''(\theta)=\frac{n}{2}\cdot\frac{1}{\theta^2}-\frac{n}{\theta^3}\cdot t$$

$$=\frac{n}{2\theta^3}\left[\theta-2t\right].$$

所以，θ 的 Fisher 信息为

$$I(\theta)=-E\left[l''(\theta)\mid X\right]$$

$$=\frac{-n}{2\theta^3}\left[\theta-\frac{2}{n}\sum_{i=1}^{n}EX_i^2\right]$$

由于

$$\mathbb{E}X_i^2 = \mathrm{Var}X_i + (\mathbb{E}X_i)^2 = \theta + \theta^2,$$

故

$$I(\theta) = \frac{-n}{2\theta^3}\left[\theta - 2(\theta+\theta^2)\right]$$
$$= \frac{n}{2\theta^3}\left[2\theta^2+\theta\right]$$
$$= n(2\theta+1)/2\theta^2.$$

故 θ 的极大似然估计的近似方差为 $1/I_\theta$，即

$$2\theta^2/n(2\theta+1).$$

3. 设 $X_1, X_2, \cdots, X_n$ 为来自几何分布的样本，用因子分解定理证明 $T = \sum_{i=1}^n X_i$ 是充分统计量.

解 设几何分布总体 X 的分布列如下

$$\mathbb{P}(X=x) = p(1-p)^{x-1}, \quad x=1,2,\cdots,$$

其中 p 为参数，且 $p\in[0,1]$.

则样本的联合分布列为

$$\begin{aligned} f(x|p) &= \prod_{i=1}^n \mathbb{P}(X_i = x_i) \\ &= \prod_{i=1}^n p(1-p)^{x_i-1} \\ &= p^n (1-p)^{\sum_{i=1}^n (x_i-1)} \\ &= p^n (1-p)^{\sum_{i=1}^n x_i - n}. \end{aligned}$$

显然可以将 $f(x|p)$ 写成

$$f(x|p) = g(T(x)|p)\cdot h(x),$$

其中

$$T(x) = \sum_{i=1}^n x_i,$$

$$g(T(x)|p) = p^n(1-p)^{T(x)-n},$$

$$h(x) = 1,$$

所以，由因子分解定理 $T=\sum_{i=1}^n X_i$ 是 p 的充分统计量.

4. 设 X 为来自正态总体 $N(0,\sigma^2)$ 的样本. $|X|$ 是否为充分统计量?

解　因为 X 的分布密度为

$$f(x|\sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-0)^2}{2\sigma^2}\right)$$

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

由因子分解定理可知 $|X|$ 是充分统计量.

5. 设 $X_1, X_2, \cdots, X_n$ 为来自总体概率密度函数如下

$$f(x|\mu,\sigma) = \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right), \quad \mu < x < \infty, 0 < \sigma < \infty$$

的样本. 求 (μ,σ) 的二维充分统计量.

解 总体的分布密度可写为

$$f(x|\mu,\sigma) = \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right)\mathbb{1}_{(\mu,+\infty)}(x),$$

故样本分布为

$$f(x|\mu,\sigma) = \prod_{i=1}^{n} f(x_i|\mu,\sigma)$$

$$= \prod_{i=1}^{n} \frac{1}{\sigma}\exp\left(-\frac{x_i-\mu}{\sigma}\right)\mathbb{1}_{(\mu,+\infty)}(x_i),$$

$$= \frac{1}{\sigma^n}\exp\left(-\frac{1}{\sigma}\sum_{i=1}^{n}(x_i-\mu)\right)\mathbb{1}_{(\mu,+\infty)}(x_{(1)})$$

$$= \frac{1}{\sigma^n}\exp\left(-\frac{1}{\sigma}\sum_{i=1}^{n}x_i + \frac{n\mu}{\sigma}\right)\mathbb{1}_{(\mu,+\infty)}(x_{(1)}).$$

由因子分解定理(取 $h(x)=1$), 可知

$$\left(X_{(1)}, \sum_{i=1}^{n} X_i\right)$$

是 (μ,σ) 的充分统计量.

6. 设 X, Y 为期望有限的随机变量，证明

$$\min_{g(x)} \mathbb{E}(Y - g(X))^2 = \mathbb{E}(Y - \mathbb{E}(Y|X))^2,$$

其中 $g(x)$ 取遍所有的可测函数. $E(Y|X)$ 有时被称为 Y 在 X 上的回归，为给定条件 X 下，Y 的最好的预测.

解 $\mathbb{E}(Y-g(X))^2$

$$= \mathbb{E}\Big([Y - \mathbb{E}(Y|X)] + [\mathbb{E}(Y|X) - g(X)] \Big)^2$$

$$= \mathbb{E}\big(Y-\mathbb{E}(Y|X)\big)^2 + \mathbb{E}\big(\mathbb{E}(Y|X)-g(X)\big)^2 + 2\mathbb{E}\Big([Y-\mathbb{E}(Y|X)]\cdot[\mathbb{E}(Y|X)-g(X)]\Big) \quad (1)$$

因为 $\mathbb{E}(Y|X)$ 是 X 的可测函数，有

$$\mathbb{E}\big(Y - \mathbb{E}(Y|X) \,\big|\, X\big) = 0. \quad (2)$$

又因 $\mathbb{E}(Y|X) - g(X)$ 是 X 的可测函数，由条件期望的性质及(2)可得

$$\mathbb{E}\Big([Y-\mathbb{E}(Y|X)]\cdot[\mathbb{E}(Y|X)-g(X)]\Big)$$

$$= \mathbb{E}\Big(\mathbb{E}\big([Y-\mathbb{E}(Y|X)]\cdot[\mathbb{E}(Y|X)-g(X)] \,\big|\, X\big)\Big)$$

$$= \mathbb{E}\Big([\mathbb{E}(Y|X)-g(X)]\cdot \underbrace{\mathbb{E}\big(Y-\mathbb{E}(Y|X)\,\big|\,X\big)}_{=0}\Big)$$

$$= 0.$$

即(1)中的交叉项为0，从而有

$$\mathbb{E}(Y-g(X))^2 \geq \mathbb{E}(Y-\mathbb{E}(Y|X))^2$$

且等号成立，即 $\mathbb{E}(Y-g(X))^2$ 取到最小值当且仅当 $g(X)=\mathbb{E}(Y|X)$.

7. 证明二项分布属于指数分布族.

解 要将二项分布的分布列（视 $p\in[0,1]$ 为参数）

$$f(x|p)=\mathbb{P}(X=x)=\binom{n}{x}p^x(1-p)^{n-x}$$

写成如下形式

$$f(x|p)=\exp\left(c(p)T(x)+d(p)+S(x)\right).$$

因

$$\begin{aligned}\log f(x|p)&=\log\binom{n}{x}+x\log p+(n-x)\log(1-p)\\&=\log\binom{n}{x}+x\left(\log p-\log(1-p)\right)+n\log(1-p)\\&=\log\binom{n}{x}+x\log\frac{p}{1-p}+n\log(1-p).\end{aligned}$$

故

$$f(x|p)=\exp\left(\left(\log\frac{p}{1-p}\right)\cdot x+n\log(1-p)+\log\binom{n}{x}\right),$$

即取

$$c(p)=\log\frac{p}{1-p},$$

$$T(x)=x,$$

$$d(p)=n\log(1-p)$$

$$S(x)=\log\binom{n}{x}.$$

所以，二项分布是属于指数分布族的.