南京大学
数理统计期中考试

天影　　　陈相相

2023 年 1 月 9 日

评语: 这张试卷是王立洪老师出的, 主要考察内容为《应用数理统计》的一至二章. 整张试卷没有绝对的难题, 在考前认真记住书上的知识点、弄懂作业题和课堂练习题就足够了. 此外, 不知道为什么王老师今年特别执着于对置信区间的考察, 这对我们来说是一个提醒.

感谢陈韵雯同学对试卷排版的帮助.

一、**(30 分)** 设总体 $X \sim N(\mu,\sigma^2)$, $X_1, X_2, \cdots, X_n$ 为 X 的样本.

(1) 设 $Y = (X_1, X_2, \cdots, X_n)'$, $v = (\mu, \mu, \cdots, \mu)'$, 求 $U = (Y-v)'(Y-v)$ 的分布.

(2) 已知 $F_{0.95}(1,1) = 161$, 求 c, 满足

$$P\left\{\frac{n(X_1 - X_2)^2}{2(X_1 + \cdots + X_n - n\mu)^2} \le c\right\} = 0.05.$$

(3) 求经验密度函数 $F_n(X)$ 的期望、方差和渐近分布. 若 1, 1, 1, 1, 2, 0, 0, 1, 3, 1, 0, 0, 2, 4, 0, 3, 1, 4, 0, 2 为 X 的观察值, 求 $F_n(x)$.

分析: (1) (2) 两题都是对教材 1.1.4 节给出的几个重要分布的考察, 需要记住这些分布应该如何构造以及各自的性质和相互之间的关系; 第 (3) 题则是对教材 1.1.3 节经验分布函数的直接考察, 这不是第一次这样考了, 建议直接背诵并默写.

(1) **解:** 由 $X \sim N(\mu,\sigma^2)$ 知 $\dfrac{X_i - \mu}{\sigma} \sim N(0,1)$. 从而由 χ^2 分布的定义和其与 Γ 分布的关系可得

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n) = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right).$$

因此,

$$U = (Y-v)'(Y-v) = \sigma^2 \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \Gamma\left(\frac{n}{2}, \frac{1}{2\sigma^2}\right).$$

(2) **解:** 由样本之间的独立性, 我们得到

$$\frac{X_1 - X_2}{\sqrt{2}\sigma} \sim N(0,1), \quad \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} \sim N(0,1),$$

于是

$$\frac{(X_1 - X_2)^2}{2\sigma^2} \sim \chi^2(1), \quad \frac{(X_1 + \cdots + X_n - n\mu)^2}{n\sigma^2} \sim \chi^2(1).$$

从而由 F 分布的定义, 有

$$F := \frac{n(X_1 - X_2)^2}{2(X_1 + \cdots + X_n\mu)^2} \sim F(1,1)$$

并且 $\frac{1}{F} \sim F(1,1)$. 那么由 $P\{F \le c\} = 0.05$, 我们有 $P\left\{\frac{1}{F} < \frac{1}{c}\right\} = 1 - P\{F \le c\} = 0.95$. 再根据 $F_{0.95}(1,1) = 161$ 知 $\frac{1}{c} = 161$, 即 $c = \frac{1}{161}$.

(3) **解:** 我们可以将经验分布函数写为 $F_n(x) = \frac{1}{n}\sum_{k=1}^{n}\mu(x - X_k)$, 其中 $\mu(x)$ 是单位阶跃函数:

$$\mu(x) = \begin{cases} 1, & x > 0, \\ 0, & x \le 0. \end{cases}$$

因为 $\mu(x - X_1),\ \mu(x - X_2), \cdots,\ \mu(x - X_n)$ 独立同分布且

$$\begin{aligned} P\{\mu(x - X_k) = 1\} &= P\{X_k \le x\} = F(x), \\ P\{\mu(x - X_k) = 0\} &= 1 - F(x), \end{aligned}$$

其中 $F(x)$ 为总体 X 的分布函数, 所以

$$\mu(x - X_k) \sim B(1, F(x)),$$

从而

$$nF_n(x) = \sum_{k=1}^{n}\mu(x - X_k) \sim B(n, F(x)).$$

因此

$$\begin{aligned} E[F_n(x)] &= \frac{1}{n}E[nF_n(x)] = \frac{1}{n}\cdot nF(x) = F(x), \\ D[F_n(x)] &= \frac{1}{n^2}D[nF_n(x)] = \frac{F(x)[1 - F(x)]}{n}. \end{aligned}$$

· 由棣莫弗-拉普拉斯中心极限定理知, 随机变量序列 $\{\mu(x - X_k)\}$ 服从中心极限定理, 即

$$\frac{\sqrt{n}[F_n(x) - F(x)]}{\sqrt{F(x)[1 - F(x)]}} \xrightarrow{L} \zeta \sim N(0,1)\ (n \to \infty),$$

所以当 n 充分大时有

$$F_n(x) \approx \frac{\sqrt{F(x)[1 - F(x)]}}{\sqrt{n}}\zeta + F(x).$$

这说明有渐近分布

$$F_n(x) \sim N\left(F(x), \frac{F(x)[1-F(x)]}{n}\right).$$

最后我们给出题干所给观察值对应的经验分布函数

$$F_n(x) = \begin{cases} 0, & x \le 0, \\ \dfrac{3}{10}, & 0 < x \le 1, \\ \dfrac{13}{20}, & 1 < x \le 2, \\ \dfrac{4}{5}, & 2 < x \le 3, \\ \dfrac{9}{10}, & 3 < x \le 4, \\ 1, & x > 4. \end{cases}$$

二、(25 分)

(1) 设 ξ 服从对数正态分布

$$f(x, \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\ln x - \mu)^2\right) \quad x > 0.$$

求 μ, σ^2 的极大似然估计.

(2) 设 $\xi \sim U(0, \theta)$.

(i) 求 θ 的矩估计, 极大似然估计及 $1-\alpha$ 置信区间.

(ii) 验证 $T = \left(\prod_{i=1}^{n} \xi_i\right)^{\frac{1}{n}}$ 是 $g(\theta) = \theta \mathrm{e}^{-1}$ 的相合估计.

分析: 第一问求矩估计和极大似然估计均属于基本技能, 按部就班计算即可. 第二问中, 对于均匀分布的置信区间我们往往考虑 $\dfrac{\xi_{(n)}}{\theta}$ 与 1 的比较. 相合估计无非是对弱大数定律的运用, 难点仅仅是要对估计量取对数.

(1) **解:** 构造似然函数 $L(\mu, \sigma) = \prod_{i=1}^{n}\left[\dfrac{1}{\xi_i \sigma\sqrt{2\pi}} \exp\left(-\dfrac{1}{2\sigma^2}(\ln \xi_i - \mu)^2\right)\right]$, 则

$$\ln L(\mu, \sigma) = -\sum_{i=1}^{n} \ln \sqrt{2\pi}\xi_i - n \ln \sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln \xi_i - \mu)^2.$$

分别对 μ, σ 求偏导, 令

$$\begin{cases} \dfrac{\partial \ln L}{\partial \mu} = -\dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(\mu - \ln \xi_i) = 0, \\ \dfrac{\partial \ln L}{\partial \sigma} = -\dfrac{n}{\sigma} + \dfrac{1}{\sigma^3}\displaystyle\sum_{i=1}^{n}(\ln \xi_i - \mu)^2 = 0, \end{cases}$$

解得 $\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}\ln\xi_i,\ \hat{\sigma}=\left[\frac{1}{n}\sum_{i=1}^{n}(\ln\xi_i-\hat{\mu})^2\right]^{\frac{1}{2}}$. 由极大似然估计的不变性知

$$\begin{cases}\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}\ln\xi_i,\\ \widehat{\sigma^2}=\frac{1}{n}\sum_{i=1}^{n}\left(\ln\xi_i-\frac{1}{n}\sum_{i=1}^{n}\ln\xi_i\right)^2.\end{cases}$$

(2) (i) **解:** 由方程 $E(\xi)=\frac{\theta}{2}=\overline{\xi}$ 知, θ 的矩法估计量为

$$\hat{\theta}=2\overline{\xi}.$$

由 $L(\theta)=\frac{1}{\theta^n}$ 以及 $\theta\geq\xi_{(n)}$ 知, 当 $\theta=\xi_{(n)}$ 时, $L(\theta)$ 取得最大值. 因此, θ 的极大似然估计为

$$\hat{\theta}=\xi_{(n)}.$$

由上述过程知 $\frac{\xi_{(n)}}{\theta}$ 的取值应该在 1 附近并小于等于 1, 所以设 $k_1<\frac{\xi_{(n)}}{\theta}<k_2$. 由顺序统计量的知识可得 $\xi_{(n)}$ 的分布函数为

$$F(x)=P\{\xi_{(n)}\leq x\}=(P\{\xi\leq x\})^n=\left(\frac{x}{\theta}\right)^n,$$

所以

$$P\left\{k_1<\frac{\xi_{(n)}}{\theta}<k_2\right\}=F(k_2\theta)-F(k_1\theta)=k_2^n-k_1^n=1-\alpha.$$

取 $k_1=\left(\frac{\alpha}{2}\right)^{\frac{1}{n}},\ k_2=\left(1-\frac{\alpha}{2}\right)^{\frac{1}{n}}$, 则 θ 的置信区间为 $\left(\xi_{(n)}\left(1-\frac{\alpha}{2}\right)^{-\frac{1}{n}},\ \xi_{(n)}\left(\frac{\alpha}{2}\right)^{-\frac{1}{n}}\right)$.

注: k_1, k_2 的选取并非唯一, 只需要满足 $k_1<k_2$ 和 $k_2^n-k_1^n=1-\alpha$ 即可.

(ii) **证明:** 将两边取对数并整理为 $\frac{1}{n}\sum_{i=1}^{n}\left(-\ln\frac{\xi_i}{\theta}\right)$. 令 $Y=-\ln\frac{\xi}{\theta}$, 则它的分布函数为

$$G(y)=P\{Y<y\}=1-P\{\xi\leq\theta\mathrm{e}^{-y}\}=1-\mathrm{e}^{-y}.$$

这说明 $Y\sim\Gamma(1,1)$, 从而 $E(Y_i)=D(Y_i)=1$ 均有界. 由切比雪夫大数定律知, 对任意的 $\varepsilon>0$ 有

$$\lim_{n\to\infty}P\left\{\left|-\frac{1}{n}\sum_{i=1}^{n}\ln\frac{\xi_i}{\theta}-1\right|>\varepsilon\right\}=0,$$

从而

$$\frac{1}{n}\ln\left(\prod_{i=1}^{n}\xi_i\right)\xrightarrow{P}\ln\frac{\theta}{\mathrm{e}},$$

即有

$$T = \left(\prod_{i=1}^{n} \xi_i\right)^{\frac{1}{n}} \xrightarrow{P} \theta \mathrm{e}^{-1} = g(\theta).$$

故 T 是 $g(\theta)$ 的相合估计. $\square$

三、**(15 分)** 设 $\xi_1, \xi_2, \cdots, \xi_{10}$ 是总体 $\xi \sim N(a, \sigma^2)$ 的样本, 且 $\sum_{i=1}^{10} \xi_i = 10$, $\sum_{i=1}^{10} \xi_i^2 = 100$. 求 a, σ^2 的 0.95 置信区间.

$(t_{0.975}(9) = 2.26,\ \chi^2_{0.975}(9) = 19.02,\ \chi^2_{0.025}(9) = 2.7.)$

分析: 本题就是考察一个正态总体的两个参数的置信区间, 完全是书上的内容. 唯一要注意的是, 置信区间的推导是不可少的, 直接默写结论是会被王老师扣分的.

解: 由于 $\overline{\xi}$ 是 a 的无偏估计, 我们认为 a 的置信区间为 $(\overline{\xi} - c,\ \overline{\xi} + c)$, 其中, c 为某个常数. 由于 $T := \dfrac{\overline{\xi} - a}{S/\sqrt{n-1}} \sim t(n-1)$, 我们得到

$$1 - \alpha = P\{|\overline{\xi} - a| < c\} = 2P\left\{T < \frac{c}{S/\sqrt{n-1}}\right\} - 1,$$

故 $\dfrac{c}{S/\sqrt{n-1}} = t_{1-\alpha/2}(n-1)$, 从而 a 的 $1-\alpha$ 置信区间为

$$\left(\overline{\xi} - \frac{S}{\sqrt{n-1}} t_{1-\alpha/2}(n-1),\ \overline{\xi} + \frac{S}{\sqrt{n-1}} t_{1-\alpha/2}(n-1)\right).$$

考虑到 $\widetilde{S}^2$ 是 σ^2 的最优无偏估计量, 所以 $\dfrac{\widetilde{S}^2}{\sigma^2}$ 应当接近 1, 故可设 $k_1 < \dfrac{\widetilde{S}^2}{\sigma^2} < k_2$. 所以 σ^2 的置信区间为

$$\left(\frac{\widetilde{S}^2}{k_2},\ \frac{\widetilde{S}^2}{k_1}\right).$$

因为 $\chi^2 := \dfrac{(n-1)\widetilde{S}^2}{\sigma^2} = \dfrac{nS^2}{\sigma^2} \sim \chi^2(n-1)$, 所以

$$1 - \alpha = P\left\{\frac{\widetilde{S}^2}{k_2} < \sigma^2 < \frac{\widetilde{S}^2}{k_1}\right\} = P\{(n-1)k_1 < \chi^2 < (n-1)k_2\},$$

即

$$\alpha = P\{\chi^2 \geq (n-1)k_2\} + P\{\chi^2 \leq (n-1)k_1\}.$$

令

$$\frac{\alpha}{2} = P\{\chi^2 \leq (n-1)k_1\}$$

与

$$\frac{\alpha}{2} = P\{\chi^2 \geq (n-1)k_2\} = 1 - P\{\chi^2 < (n-1)k_2\},$$

则

$$k_1=\frac{1}{n-1}\chi^2_{\alpha/2}(n-1),\ k_2=\frac{1}{n-1}\chi^2_{1-\alpha/2}(n-1).$$

故 σ^2 的 $1-\alpha$ 置信区间为

$$\left(\frac{nS^2}{\chi^2_{1-\alpha/2}(n-1)},\ \frac{nS^2}{\chi^2_{\alpha/2}(n-1)}\right).$$

由题干内容可知, $n=10,\ \overline{\xi}=1,\ S=3,\ \alpha=0.05$. 因此 $a,\ \sigma^2$ 的 0.95 置信区间分别为 $(-1.26,3.26)$ 和 $(4.7319,33.3333)$.

四、(30 分) 设 $\xi\sim\Gamma(k,\lambda)$, k 已知, $k\in\mathbb{Z}_+$.

(1) 求 λ^{-1} 的有效估计量, 该估计量是否是 λ^{-1} 的一致最小方差无偏估计量和相合估计量?

(2) 求 λ^{-1} 的 $1-\alpha$ 置信区间.

(3) 设 λ^{-1} 的先验分布 $I\Gamma(1,\alpha_0)$, α_0 已知. 在平方损失下求 λ 的贝叶斯估计量与 λ 的 $1-\alpha$ 贝叶斯置信区间.

分析: 有效估计量、置信区间和贝叶斯估计量的求法是基本能力, 相合估计量的验证在上题中已做过分析. 要注意的是, 虽然贝叶斯置信区间的结果往往长得很丑, 但并不会因此不做考察, 笔者就吃了这个亏; 但事实上这个概念在理解和使用上没有难度.

(1) **解:** 构造似然函数 $L(\lambda)=\prod_{i=1}^{n}\frac{\lambda^k}{\Gamma(k)}\xi_i^{k-1}e^{-\lambda\xi_i}$, 可得

$$\ln L(\lambda)=-n\ln\left[\Gamma(k)\right]+nk\ln\lambda+(k-1)\sum_{i=1}^{n}\ln\xi_i-n\lambda\overline{\xi},$$

从而

$$\frac{\partial\ln L(\lambda)}{\partial\lambda}=\frac{nk}{\lambda}-n\overline{\xi}=-nk\left(\frac{\overline{\xi}}{k}-\frac{1}{\lambda}\right).$$

又因为 $E\left(\frac{\overline{\xi}}{k}\right)=\frac{1}{nk}\sum_{i=1}^{n}E(\xi_i)=\frac{1}{nk}\cdot n\frac{k}{\lambda}=\frac{1}{\lambda}$, 即 $\frac{\overline{\xi}}{k}$ 是 $\frac{1}{\lambda}$ 的无偏估计量, 所以 λ^{-1} 的有效估计量为 $\frac{\overline{\xi}}{k}$.

因为 $\frac{\overline{\xi}}{k}$ 是 λ^{-1} 的无偏有效估计量, 所以它是 λ^{-1} 的一致最小方差无偏估计量.

因为 $E(\xi_i)=\frac{k}{\lambda},\ D(\xi_i)=\frac{k}{\lambda^2}$ 均有界, 由切比雪夫大数定律知, 对任意的 $k\varepsilon>0$,

$$\lim_{n\to\infty}P\left\{\left|\overline{\xi}-\frac{k}{\lambda}\right|>k\varepsilon\right\}=0,$$

从而

$$\lim_{n\to\infty}P\left\{\left|\frac{\overline{\xi}}{k}-\frac{1}{\lambda}\right|>\varepsilon\right\}=0.$$

这说明它是相合估计量.

(2) **解:** 由上述讨论知 $\dfrac{\lambda\overline{\xi}}{k}$ 接近 1, 即有 $k_1 < \dfrac{\lambda\overline{\xi}}{k} < k_2$. 另一方面, $n\overline{\xi} \sim \Gamma(nk, \lambda)$, 从而 $\chi^2 := 2\lambda n\overline{\xi} \sim \Gamma\left(nk, \dfrac{1}{2}\right) = \chi^2(2nk)$. 所以, 我们有

$$1-\alpha = P\left\{k_1 < \frac{\lambda\overline{\xi}}{k} < k_2\right\} = P\{2nkk_1 < \chi^2 < 2nkk_2\}.$$

令

$$\frac{\alpha}{2} = P\{\chi^2 \le 2nkk_1\}$$

与

$$\frac{\alpha}{2} = 1 - P\{\chi^2 < 2nkk_2\},$$

则 $2nkk_1 = \chi^2_{\alpha/2}(2nk)$, $2nkk_2 = \chi^2_{1-\alpha/2}(2nk)$. 得到 λ^{-1} 的 $1-\alpha$ 置信区间为

$$\left(\frac{2n\overline{\xi}}{\chi^2_{1-\alpha/2}(2nk)}, \frac{2n\overline{\xi}}{\chi^2_{\alpha/2}(2nk)}\right).$$

(3) **解:** 因为 λ^{-1} 的先验分布为 $I\Gamma(1, \alpha_0)$, 所以 λ 的先验分布为 $\Gamma(1, \alpha_0)$. 我们由先验分布给出核

$$\begin{aligned} h(y|\mathbf{X}) &\propto \pi(y) f(\mathbf{X}|y) \\ &\propto e^{-\alpha_0 y} \prod_{i=1}^{n} y^k e^{-y x_i} \\ &\propto y^{nk} e^{-(\alpha_0 + n\overline{x})y}. \end{aligned}$$

所以我们得到 $\lambda|\mathbf{X} \sim \Gamma(nk+1, \alpha_0 + n\overline{x})$. 这样, λ 在平方损失下的贝叶斯估计量为

$$\widetilde{\lambda} = E[\lambda|\boldsymbol{\xi}] = \frac{nk+1}{\alpha_0 + n\overline{\xi}}.$$

我们由上述讨论知 $2(n\overline{\xi} + \alpha_0)\lambda|\mathbf{X} \sim \Gamma\left(nk+1, \dfrac{1}{2}\right) = \chi^2(2nk+2)$. 类似地，我们有

$$1-\alpha = P\left\{\chi^2_{\alpha/2}(2nk+2) < 2(n\overline{\xi} + \alpha_0)\lambda < \chi^2_{1-\alpha/2}(2nk+2)\right\},$$

即 λ 的 $1-\alpha$ 贝叶斯置信区间为

$$\left(\frac{\chi^2_{\alpha/2}(2nk+2)}{2(n\overline{\xi} + \alpha_0)}, \frac{\chi^2_{1-\alpha/2}(2nk+2)}{2(n\overline{\xi} + \alpha_0)}\right).$$