

第六次作业

题目 1. (1) 令 X 是来自分布 $f(x; \theta) = \theta x^{\theta-1} I_{(0,1)}(x)$, $\theta > 0$.

(a) 求一个枢轴量, 并且使用它求出关于 θ 的一个置信区间估计.

(b) 令 $Y = -1/\log X$, 则 $(Y/2, Y)$ 关于 θ 的置信区间, 求出关于该区间的置信系数, 并且找到一个关于 θ 更优的置信区间.

解答. (a) 由于 X 分布函数 $F(x) = \int_0^x \theta t^{\theta-1} dt = x^\theta$, ($0 < x < 1$). 由于 $F(X) \sim I(0, 1)$, 则 $-\log F(X) \sim \Gamma(1, 1) \Rightarrow -2 \log F(X) \sim \chi^2(2)$. 设 $q_1, q_2 \in (0, \infty)$, $q_1 < q_2$, 满足

$$\gamma = P(q_1 < -2 \log F(x) < q_2) = P(q_1 < -2 \theta \log X < q_2) = P\left(\frac{q_1}{-2 \log X} < \theta < \frac{q_2}{-2 \log X}\right)$$

q_1, q_2 取等尾概率 $q_1 = \chi^2_{\frac{1-\gamma}{2}}(2)$, $q_2 = \chi^2_{\frac{1+\gamma}{2}}(2)$, 则 θ 的一个系数为 γ 的置信区间为

$$\left(\frac{\chi^2_{\frac{1-\gamma}{2}}(2)}{-2 \log X}, \frac{\chi^2_{\frac{1+\gamma}{2}}(2)}{-2 \log X}\right)$$

(b) 由于 $(Y/2, Y)$ 是 θ 的一个置信区间, 则

$$P(Y/2 < \theta < Y) = P(1/2 < \theta/Y < 1) = P(1/2 < -\theta \log X < 1)$$

由于 $-\theta \log X \sim \Gamma(1, 1)$, 于是 $P(Y/2 < \theta < Y) = \int_{\frac{1}{2}}^1 e^{-x} = e^{-\frac{1}{2}} - e^{-1}$, 于是该区间的置信系数 $\gamma = e^{-\frac{1}{2}} - e^{-1}$. 设 $q_1, q_2 \in [0, \infty)$, $q_1 < q_2$ 满足

$$\gamma = P(q_1 < -\theta \log X < q_2) = P\left(\frac{q_1}{-\log X} < \theta < \frac{q_2}{-\log X}\right)$$

则期望区间长度 $l = \mathbb{E}\left[\frac{q_1}{\log X} - \frac{q_2}{\log X}\right] = \mathbb{E}\left[\frac{1}{\log X}\right](q_1 - q_2)$, 由于 $\mathbb{E}\left[\frac{1}{\log X}\right] < 0$, 则极小化 $l \iff$ 极大化 $q_1 - q_2$, 转化为求解极大化问题

$$\begin{aligned} \max \quad & z = q_1 - q_2, \\ \text{s.t.} \quad & \int_{q_1}^{q_2} e^{-x} dx = e^{-q_1} - e^{-q_2} = \gamma, \\ & 0 \leq q_1 < q_2. \end{aligned}$$

于是 $q_1 = -\log(e^{-q_2} + \gamma)$, 由于 $q_1 \geq 0$ 代入得到 $q_2 \geq -\log(1 - \gamma)$ 且

$$z = -\log(e^{-q_2} + \gamma) - q_2, \quad z' = \frac{-\gamma}{e^{-q_2} + \gamma} < 0$$

则 z 关于 q_2 单调递减, 于是 $q_2 = -\log(1 - \gamma)$ 时 z 有极大值, 区间长度有极小值, 对应的 $q_1 = 0$, 则更优的置信区间为

$$\left(0, \frac{\log(1 - \gamma)}{\log X}\right) = \left(0, \frac{\log(1 + e^{-1} - e^{-\frac{1}{2}})}{\log X}\right)$$

题目 2. (2) 令 X_1, \dots, X_n 是来自 $N(\theta, \theta)$, $\theta > 0$ 的随机样本. 求出一个枢轴量, 并用它得到关于 θ 的置信区间估计.

解答. 由于 $\frac{\bar{X} - \theta}{\sqrt{\theta/n}} \sim N(0, 1)$, $\frac{X_i - \theta}{\sqrt{\theta}} \sim N(0, 1)$, 于是 $\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sqrt{\theta}} \right)^2 = \frac{(n-1)S^2}{\theta} \sim \chi^2(n-1)$, 于是

$$\frac{\bar{X} - \theta}{\sqrt{\theta/n}} \cdot \frac{\sqrt{\theta(n-1)}}{\sqrt{(n-1)S^2}} = \sqrt{n} \frac{\bar{X} - \theta}{S} \sim t(n-1)$$

设 $q_1, q_2 \in \mathbb{R}$, $q_1 < q_2$, 满足

$$\gamma = P(q_1 < \sqrt{n} \frac{\bar{X} - \theta}{S} < q_2) = P\left(\bar{X} - \frac{S}{\sqrt{n}} q_2 < \theta < \bar{X} - \frac{S}{\sqrt{n}} q_1\right)$$

由于 t 分布为对称分布, 则 $q_2 = -q_1 = t_{\frac{1+\gamma}{2}}(n-1)$ 是使得置信区间最小的上下限, 则 θ 的系数为 γ 的置信区间为

$$\left(\bar{X} - \frac{S}{\sqrt{n}} t_{\frac{1+\gamma}{2}}(n-1), \bar{X} + \frac{S}{\sqrt{n}} t_{\frac{1+\gamma}{2}}(n-1) \right)$$

题目 3. (5) 令 X_1, \dots, X_n 是来自 $f(x; \theta) = \theta e^{-\theta x} I_{(0, \infty)}(x)$ 的随机样本.

- (a) 求解关于总体均值的 100γ 置信度区间.
- (b) 求解关于总体方差的 100γ 置信度区间.
- (c) 求解上述两个区间同时覆盖真实均值和方差的概率.
- (d) 求解关于 $e^{-\theta} = P(X > 1)$ 的置信区间估计.
- (e) 求出仅与 Y_1 相关的枢轴量, 并且使用它求出关于 θ 的置信区间估计.

解答. (a) 总体均值为 $\mathbb{E}(\bar{X}) = \mathbb{E}(X) = 1/\theta$. 由于 θ 为尺度参数, $\theta X \sim \gamma(1, 1)$, 则 $\theta \sum_{i=1}^n X_i \sim$

$\Gamma(n, 1) \Rightarrow 2\theta \sum_{i=1}^n X_i \sim \chi^2(2n)$. 设 $q_1, q_2 \in [0, \infty)$, $q_1 < q_2$, 满足

$$\gamma = P(q_1 < 2\theta \sum_{i=1}^n X_i < q_2) = P\left(\frac{2n\bar{X}}{q_2} < \frac{1}{\theta} < \frac{2n\bar{X}}{q_1}\right)$$

对 q_1, q_2 取等尾概率得 $q_1 = \chi_{\frac{1-\gamma}{2}}^2(2n)$, $q_2 = \chi_{\frac{1+\gamma}{2}}^2(2n)$, 于是总体均值 $1/\theta$ 的置信度为 100γ 的区间为

$$\left(\frac{2n\bar{X}}{\chi_{\frac{1+\gamma}{2}}^2(2n)}, \frac{2n\bar{X}}{\chi_{\frac{1-\gamma}{2}}^2(2n)} \right)$$

(b) 由于总体方差为 $\text{Var}(X) = 1/\theta^2$, 且 $P(q'_1 < 1/\theta < q'_2) = P(q_1'^2 < 1/\theta < q_2'^2)$, 由 (a) 可得, 总体方差 $1/\theta^2$ 的置信度为 100γ 的区间为

$$\left(\left[\frac{2n\bar{X}}{\chi_{\frac{1+\gamma}{2}}^2(2n)} \right]^2, \left[\frac{2n\bar{X}}{\chi_{\frac{1-\gamma}{2}}^2(2n)} \right]^2 \right)$$

(c) 由于 (b) 是由 (a) 直接推出的, 所以同时包含住真实均值和方差的概率为 γ .

(d) 由 (a) 可知, $P\left(\frac{q_1}{2n\bar{X}} < \theta < \frac{q_2}{2n\bar{X}}\right) = P\left(\exp\left\{-\frac{q_2}{2n\bar{X}}\right\} < e^{-\theta} < \exp\left\{-\frac{q_1}{2n\bar{X}}\right\}\right) = \lambda$, 于是 $e^{-\theta}$ 的系数为 λ 的置信区间为

$$\left(\exp\left\{-\frac{q_2}{2n\bar{X}}\right\}, \exp\left\{-\frac{q_1}{2n\bar{X}}\right\}\right)$$

(e) 由于 $f_{Y_1}(y) = nf_X(y)(1 - F_X(y))^{n-1} = n\theta e^{-n\theta y} \sim \Gamma(1, n\theta)$, 令 $Z = \theta Y_1$, 则 $f_Z(z) = f_{Y_1}(z/\theta)/\theta = ne^{-nz} \sim \Gamma(1, n)$. 设 $q_1, q_2 \in [0, \infty)$, $q_1 < q_2$, 满足

$$1 - \alpha = P(q_1 < \theta Y_1 < q_2) = P\left(\frac{q_1}{Y_1} < \theta < \frac{q_2}{Y_1}\right)$$

于是期望区间长度为 $l = \mathbb{E}\left[\frac{q_2}{Y_1} - \frac{q_1}{Y_1}\right] = \mathbb{E}\left[\frac{1}{Y_1}\right](q_2 - q_1)$, 由于 $\mathbb{E}\left[\frac{1}{Y_1}\right] > 0$, 于是最小化 $l \iff$ 最小化 $q_2 - q_1$, 转化为求解极小化问题

$$\begin{aligned} \min \quad & z = q_2 - q_1, \\ \text{s.t.} \quad & \int_{q_1}^{q_2} ne^{-nz} dz = e^{-nq_1} - e^{-nq_2} = 1 - \alpha \\ & 0 \leq q_1 < q_2. \end{aligned}$$

则 $q_1 = -\frac{1}{n} \log(1 - \alpha + e^{-nq_2})$, 由于 $q_1 \geq 0$ 代入可得 $q_2 \geq -\frac{1}{n} \log \alpha$ 且

$$z = q_1 + \frac{1}{n} \log(1 - \alpha + e^{-nq_2}), f'(q_2) = 1 - \frac{e^{-nq_2}}{1 - \alpha + e^{-nq_2}} > 0$$

则 z 关于 q_2 单调递增, 当 $q_2 = -\frac{1}{n} \log \alpha$ 时 z 有极小值, 对应 $q_1 = 0$, 于是 θ 的系数为 $1 - \alpha$ 的置信区间为

$$\left(0, -\frac{\log \alpha}{nY_1}\right)$$

题目 4. (18) 为了在正常耕作条件下测试两个新型杂交玉米品种, 一家种子公司在 Iowa 选择了八个农场, 并在每个农场的试验田里终止了两个新品种. 八个地点的产量为 (换算为蒲式耳/英尺)

Line A:	86	87	56	93	84	93	75	79
Line B:	80	79	58	91	77	82	74	66

假设假设两个收益率满足联合正态分布, 求解两个品种平均收益率之差满足系数为 95 的置信区间.

解答. 设品种 A, B 收益率对应的样本为 $X_1, \dots, X_8; Y_1, \dots, Y_8$, 则 $(X_1, Y_1), \dots, (X_8, Y_8) \stackrel{iid}{\sim} N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right)$, 其中 $\rho = \frac{\text{Cov}(X, Y)}{\sigma_1\sigma_2}$. 令 $D_i = X_i - Y_i$ 则 $D_i \sim N(\mu_2 - \mu_1, \sigma_D^2)$,

其中 $\sigma_D^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$, 则

$$\frac{\bar{D} - (\mu_2 - \mu_1)}{\sigma_D/\sqrt{n}} \sim N(0, 1), \frac{\sum_{i=1}^n (D - \bar{D})^2}{\sigma_D^2} = \frac{(n-1)S_D^2}{\sigma_D^2} \sim \chi^2(7)$$

于是 $\frac{\bar{D} - (\mu_2 - \mu_1)}{S_D/\sqrt{n}} \sim t(7)$, 则置信度为 95 的区间为

$$\left(\bar{D} - t_{0.975}(7) \frac{S_D}{\sqrt{8}}, \bar{D} + t_{0.975}(7) \frac{S_D}{\sqrt{8}} \right)$$

由于 $D_i = 6, 8, -2, 2, 7, 11, 1, 13$, 则 $\bar{D} = 5.75$, $S_D \approx 5.12$, 查表可得 $t_{0.975}(7) \approx 2.365$, 于是品种 A 与品种 B 收益率均值差的置信度为 95 的区间为 (1.469, 10.031).