6.6 区间估计

3. $0.50, 1.25, 0.80, 2.00$ 是取自总体 X 的样本，已知 $Y = \ln X$ 服从正态分布 $N(\mu, 1)$。

(1) 求 μ 的置信水平为 95% 的置信区间；

解：

$$\frac{1-0.95}{2} = 0.025, \quad u_{0.025} = -1.96, \quad \frac{u_{0.025}}{\sqrt{n}} = \frac{-1.96}{\sqrt{4}} = -0.98$$
$$\overline{\ln x} = \frac{1}{4}(\ln 0.50 + \ln 1.25 + \ln 0.80 + \ln 2.00) = 0$$

所以 μ 的置信水平为 95% 的置信区间为 $[-0.98, 0.98]$。 □

(2) 求 X 的数学期望的置信水平为 95% 的置信区间。

解：

$$EX = Ee^Y = e^{EY} = e^{\mu}$$

所以 EX 的置信水平为 95% 的置信区间为 $[e^{-0.98}, e^{0.98}]$，即$[0.3753110988514, 2.66445624192942]$。

□

5. 已知某种材料的抗压强度 $X \sim N(\mu, \sigma^2)$，现随机地抽取 10 个试件进行抗压试验，测得数据如下：

482 493 457 471 510 446 435 418 394 469

(1) 求平均抗压强度 μ 的置信水平为 95% 的置信区间；

解：

由于 σ 未知，所以 μ 的置信区间为 $\left[\bar{x} - t_{1-0.025}(10-1)s/\sqrt{10}, \bar{x} + t_{1-0.025}(10-1)s/\sqrt{10}\right]$
之后计算得

$$\bar{x} = 457.5, \quad s \approx 35.21757768$$

所以 μ 的置信水平为 95% 的置信区间为

$$[457.5 - 2.2622 \times 35.21757768/\sqrt{10}, 457.5 - 2.2622 \times 35.21757768/\sqrt{10}]$$

即

$$[432.306385526736, 482.693614473264]$$

□

(2) 若已知 $\sigma = 30$，求平均抗压程度 μ 的置信水平为 95% 的置信区间；

解：

由于 σ 已知，所以 μ 的 95% 置信区间为 $[\bar{x} - u_{0.975}\sigma/\sqrt{n}, \bar{x} + u_{0.975}\sigma/\sqrt{n}]$，代入得

$$\left[457.5 - 1.96 \times 30/\sqrt{10}, 457.5 + 1.96 \times 30/\sqrt{10}\right]$$

即

$$[438.90580735821, 476.09419264179]$$

□

(3) 求 σ 的置信水平为 95% 的置信区间。

解：

μ 未知时 σ 的置信水平为 95% 的置信区间为 $\left[\frac{s\sqrt{n-1}}{\sqrt{\chi^2_{0.975}(n-1)}}, \frac{s\sqrt{n-1}}{\sqrt{\chi^2_{0.025}(n-1)}}\right]$，代入得

$$\left[\frac{35.21757768 \times \sqrt{10-1}}{\sqrt{19.0228}}, \frac{35.21757768 \times \sqrt{10-1}}{\sqrt{2.7004}}\right]$$

即

$$[24.2238693218913, 64.2934434191729]$$

□

6. 在一批货物中随机抽取 80 件，发现有 11 件不合格品，试求这批货物的不合格品率的置信水平为 0.90 的置信区间。

解：

样本的分布为 $b(1,p)$。由于样本量较大，可以使用近似置信区间，即 $\left[\bar{x} - u_{0.95}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}, \bar{x} + u_{0.95}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$，其中 $\bar{x} = \frac{11}{80} = 0.1375$，$n = 80$，$u_{0.95} = 1.645$，代入得

$$\left[0.1375 - 1.645 \times \sqrt{\frac{0.1375 \times (1 - 0.1375)}{80}}, 0.1375 + 1.645 \times \sqrt{\frac{0.1375 \times (1 - 0.1375)}{80}}\right]$$

即

$$[0.0741638282314373, 0.200836171768563]$$

□

9. 设从总体 $X \sim N(\mu_1, \sigma_1^2)$ 和总体 $Y \sim N(\mu_2, \sigma_2^2)$ 中分别抽取容量为 $n_1 = 10, n_2 = 15$ 的独立样本，可计算得 $\bar{x} = 82, s_x^2 = 56.5, \bar{y} = 76, s_y^2 = 52.4$。

(1) 若已知 $\sigma_1^2=64,\sigma_2^2=49$，求 $\mu_1-\mu_2$ 的置信水平为 95% 的置信区间；

解：

σ_1^2 和 σ_2^2 均已知，则 $\mu_1-\mu_2$ 的置信水平为 95% 的置信区间为 $\left[\bar{x}-\bar{y}-u_{0.975}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\bar{x}-\bar{y}+u_{0.975}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right]$，代入得

$$\left[82-76-1.96\times\sqrt{\frac{64}{10}+\frac{49}{15}},\ 82-76+1.96\times\sqrt{\frac{64}{10}+\frac{49}{15}}\right]$$

即

$$[-0.0938876480180258,\ 12.093887648018]$$

□

(2) 若已知 $\sigma_1^2=\sigma_2^2$，求 $\mu_1-\mu_2$ 的置信水平为 95% 的置信区间；

解：

$\sigma_1^2=\sigma_2^2$ 未知，则 $\mu_1-\mu_2$ 的置信水平为 95% 的置信区间为

$$\left[\bar{x}-\bar{y}-\sqrt{\frac{n_1+n_2}{n_1n_2}}s_wt_{0.975}(n_1+n_2-2),\ \bar{x}-\bar{y}+\sqrt{\frac{n_1+n_2}{n_1n_2}}s_wt_{0.975}(n_1+n_2-2)\right]$$

其中 $s_w^2=\dfrac{(n_1-1)s_x^2+(n_2-1)s_y^2}{n_1+n_2-2}$，$t_{0.975}(23)=2.0687$，代入得

$$s_w^2=\frac{(10-1)\times56.5+(15-1)\times52.4}{10+15-2}=\frac{12421}{230}$$

置信区间为

$$\left[82-76-\sqrt{\frac{10+15}{10\times15}}\sqrt{\frac{12421}{230}}\times2.0687,\ 82-76+\sqrt{\frac{10+15}{10\times15}}\sqrt{\frac{12421}{230}}\times2.0687\right]$$

即

$$[-0.206349837966326,\ 12.2063498379663]$$

□

(3) 若对 σ_1^2,σ_2^2 一无所知，求 $\mu_1-\mu_2$ 的置信水平为 95% 的置信区间；

解：

此时为一般场合下，$\mu_1-\mu_2$ 的置信水平为 95% 的近似置信区间为 $[\bar{x}-\bar{y}-s_0t_{0.975}(l),\ \bar{x}-\bar{y}+s_0t_{0.975}(l)]$，其中 $s_0^2=\frac{s_x^2}{n_1}+\frac{s_y^2}{n_2}=\frac{56.5}{10}+\frac{52.4}{15}=\frac{2743}{300}$，$l=\dfrac{s_0^4}{\frac{s_x^4}{n_1^2(n_1-1)}+\frac{s_y^4}{n_2^2(n_2-1)}}=\dfrac{\left(\frac{2743}{300}\right)^2}{\frac{56.5^2}{10^2(10-1)}+\frac{52.4^2}{15^2(15-1)}}=\dfrac{52668343}{2783727}\approx18.92008\approx19$，$t_{0.975}(19)=2.0930$。

所以置信区间为

$$\left[82-76-\sqrt{\frac{2743}{300}}\times 2.0930,\ 82-76+\sqrt{\frac{2743}{300}}\times 2.0930\right]$$

即

$$[-0.328801942179367,\ 12.3288019421794]$$

□

(4) 求 σ_1^2/σ_2^2 的置信水平为 95% 的置信区间。

解：

置信区间为

$$\left[\frac{s_x^2}{s_y^2}\cdot\frac{1}{F_{0.975}(9,14)},\quad \frac{s_x^2}{s_y^2}\cdot\frac{1}{F_{0.025}(9,14)}\right]$$

由于 $F_{\frac{\alpha}{2}}(n_1,n_2)=1/F_{1-\frac{\alpha}{2}}(n_2,n_1)$，所以可以代入得

$$\left[\frac{56.5}{52.4}\cdot\frac{1}{3.21},\ \frac{56.5}{52.4}\cdot 3.80\right]$$

即

$$[0.335901643242729,\ 4.09732824427481]$$

□

12. 设某电子产品的寿命服从指数分布，其密度函数为 $\lambda e^{-\lambda x}I_{\{x>0\}}$，现从此批产品中抽取容量为 9 的样本，测得寿命为（单位：千小时）

$$15\quad 45\quad 50\quad 53\quad 60\quad 65\quad 70\quad 83\quad 90$$

求平均寿命 $1/\lambda$ 的置信水平为 0.9 的置信区间和置信上、下限。

解：

首先尝试构造枢轴量，设样本为 $x_1,x_2,\cdots,x_9$，则 $x_1,x_2,\cdots x_9\overset{\text{i.i.d.}}{\sim}\mathrm{Exp}(\lambda)$，则 $\sum_{i=1}^{9}x_i\sim\mathrm{Ga}(9,\lambda)$，所以 $2\lambda\sum_{i=1}^{9}x_i\sim\mathrm{Ga}(9,\frac{1}{2})=\chi^2(18)$，分布不依赖于 λ，所以 $G=2\lambda\sum_{i=1}^{9}x_i$ 为枢轴量，所以

$$P\left(\chi^2_{0.05}(18)\leqslant G\leqslant\chi^2_{0.95}(18)\right)=0.9$$

$$P\left(\frac{\chi^2_{0.05}(18)}{2\sum_{i=1}^{9}x_i}\leqslant\lambda\leqslant\frac{\chi^2_{0.95}(18)}{2\sum_{i=1}^{9}x_i}\right)=0.9$$

所以 λ 的置信水平为 0.9 的双侧置信区间为

$$\left[\frac{\chi^2_{0.05}(18)}{2\sum_{i=1}^{9}x_i},\ \frac{\chi^2_{0.95}(18)}{2\sum_{i=1}^{9}x_i}\right]=\left[\frac{9.3905}{2\times 531},\ \frac{28.8693}{2\times 531}\right]=[0.00884227871939736,\ 0.0271838983050847]$$

同理，单侧置信上限为

$$\frac{\chi^2_{0.9}(18)}{2\sum_{i=1}^{9}x_i}=\frac{25.9894}{2\times 531}=0.0244721280602637$$

单侧置信下限为

$$\frac{\chi^2_{0.1}(18)}{2\sum_{i=1}^{9} x_i} = \frac{10.8649}{2\times 531} = 0.0102306026365348$$

所以 $\frac{1}{\lambda}$ 的置信水平为 0.9 的置信区间为

$$\left[\frac{2\times 531}{28.8693}, \ \frac{2\times 531}{9.3905}\right] = [\mathbf{36.7864825264208}, \ \mathbf{113.093019541025}]$$

单侧置信上限为

$$\frac{2\times 531}{10.8649} = \mathbf{97.74595256284}$$

单侧置信下限为

$$\frac{2\times 531}{25.9894} = \mathbf{40.8628133008073}$$

□

13. 设总体 X 的密度函数为

$$p(x;\theta) = \frac{1}{\pi\left[1+(x-\theta)^2\right]}, \quad -\infty < x < \infty, \quad -\infty < \theta < \infty$$

$x_1, x_2, \cdots, x_n$ 为抽自此总体的简单随机样本，求位置参数 θ 的置信水平近似为 $1-\alpha$ 的置信区间。

解：

设 $m_{0.5}$ 表示样本中位数，则根据例 5.3.10，当然 n 较大时，$m_{0.5} \sim N\left(\theta, \frac{\pi^2}{4n}\right)$，将 $m_{0.5}$ 看作样本容量为 1 的样本，则 $m_{0.5}$ 服从方差已知，期望未知的正态分布，所以 θ 的置信水平近似为 $1-\alpha$ 的置信区间为

$$\left[m_{0.5} - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\pi^2}{4n}}/\sqrt{1}, \ m_{0.5} + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\pi^2}{4n}}/\sqrt{1}\right]$$

即

$$\left[m_{0.5} - u_{1-\frac{\alpha}{2}}\frac{\pi}{2\sqrt{n}}, \ m_{0.5} + u_{1-\frac{\alpha}{2}}\frac{\pi}{2\sqrt{n}}\right]$$

□

14. 设 $x_1, x_2, \cdots, x_n$ 为抽自正态总体 $N(\mu, 16)$ 的简单随机样本，为使得 μ 的置信水平为 $1-\alpha$ 的置信区间的长度不大于给定的 L，试问样本容量 n 至少要多少?

解：

$\sigma^2 = 16$ 已知，则置信区间为 $\left[\bar{x} - u_{1-\frac{\alpha}{2}}\sigma/\sqrt{n}, \bar{x} + u_{1-\frac{\alpha}{2}}\sigma/\sqrt{n}\right]$，区间长度为 $2u_{1-\frac{\alpha}{2}} \times 4/\sqrt{n} \leqslant L$，则 $n \geqslant \left(\frac{8u_{1-\frac{\alpha}{2}}}{L}\right)^2 = \frac{64u^2_{1-\frac{\alpha}{2}}}{L^2}$。 □

16. 设 $x_1, x_2, \cdots, x_n$ 是来自 $U(\theta-\frac{1}{2}, \theta+\frac{1}{2})$ 的样本，求 θ 的置信水平为 $1-\alpha$ 的置信区间（提示：证明 $\frac{x_{(n)}+x_{(1)}}{2} - \theta$ 为枢轴量，并求出对应的密度函数）。

解：

设总体为 X，则 $X \sim U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$，则 $X - \theta + \frac{1}{2} \sim U(0,1)$，则根据例 5.3.9，$(Y,Z) = \left(x_{(1)} - \theta + \frac{1}{2}, x_{(n)} - \theta + \frac{1}{2}\right)$ 的联合密度函数为

$$p(y,z) = n(n-1)(z-y)^{n-2}$$

再根据卷积公式，$y + z = x_{(1)} + x_{(n)} - 2\theta + 1$ 的概率密度函数为

$$p(x) = \int_0^1 n(n-1)(x-2t)^{n-2}\mathrm{d}t = \frac{n}{2}x^{n-1} - \frac{n}{2}(x-2)^{n-1}$$

所以 $\frac{(y+z)-1}{2} = \frac{x_{(1)}+x_{(n)}}{2} - \theta$ 的概率密度函数为

$$p'(x) = \frac{n}{2}(2x+1)^{n-1} - \frac{n}{2}(2x-1)^{n-1}$$

显然与 θ 无关，所以令 $G = \frac{x_{(1)}+x_{(n)}}{2} - \theta$ 即为枢轴量，则可知

$$\int_{-\frac{1-\alpha^{\frac{1}{n}}}{2}}^{\frac{1-\alpha^{\frac{1}{n}}}{2}} \frac{n}{2}(2x+1)^{n-1} - \frac{n}{2}(2x-1)^{n-1}\mathrm{d}x = 1-\alpha$$

所以 $-\frac{1-\alpha^{\frac{1}{n}}}{2} \leqslant G = \frac{x_{(1)}+x_{(n)}}{2} - \theta \leqslant \frac{1-\alpha^{\frac{1}{n}}}{2}$ ，所以 $\frac{x_{(1)}+x_{(n)}}{2} - \frac{1-\alpha^{\frac{1}{n}}}{2} \leqslant \theta \leqslant \frac{x_{(1)}+x_{(n)}}{2} + \frac{1-\alpha^{\frac{1}{n}}}{2}$ 所以 θ 的置信水平为 $1-\alpha$ 的置信区间为

$$\left[\frac{x_{(1)}+x_{(n)}}{2} - \frac{1-\alpha^{\frac{1}{n}}}{2},\ \frac{x_{(1)}+x_{(n)}}{2} + \frac{1-\alpha^{\frac{1}{n}}}{2}\right]$$

□

19. 设总体 X 的密度函数为

$$p(x,\theta) = e^{-(x-\theta)}I_{\{x>\theta\}}, \quad -\infty < \theta < \infty,$$

$x_1, x_2, \cdots, x_n$ 为抽自此总体的简单随机样本。

(1) 证明：$x_{(1)} - \theta$ 的分布与 θ 无关，并求出此分布；

证明：

令 $y = x - \theta$，则 $p(y) = e^{-y}I_{\{y>0\}}$。由于 $y = x - \theta$ 是单调增函数，所以 $y_{(1)} = x_{(1)} - \theta$。$F(y) = \int_0^y e^{-t}\mathrm{d}t = 1 - e^{-y}$，从而次序统计量 $y_{(1)} = x_{(1)} - \theta$ 的概率密度函数为

$$p_{(1)}(y) = \frac{n!}{0!(n-1)!}[F(y)]^0\,[1-F(y)]^{n-1}\,p(y) = n\left(e^{-y}\right)^{n-1}e^{-y}I_{\{y>0\}} = ne^{-ny}I_{\{y>0\}}$$

所以 $x_{(1)} - \theta \sim \mathrm{Exp}(n)$，与 θ 无关。 □

(2) 求 θ 的置信水平为 $1-\alpha$ 的置信区间。

解：

$$P(c \leqslant x_{(1)} - \theta \leqslant d) = \int_c^d n e^{-ny} \mathrm{d}y$$

因为被积函数在 $[0, +\infty)$ 上单调递减，所以区间长度最短则 $c = 0$，所以

$$\int_0^d n e^{-ny} \mathrm{d}y = -e^{-ny}\Big|_0^d = 1 - e^{-nd} = 1 - \alpha$$

所以 $d = \dfrac{-\ln\alpha}{n}$。

而 $c \leqslant x_{(1)} - \theta \leqslant d \implies x_{(1)} - d \leqslant \theta \leqslant x_{(1)} - c \implies x_{(1)} + \frac{\ln\alpha}{n} \leqslant \theta \leqslant x_{(1)}$，所以 θ 的置信水平为 $1 - \alpha$ 的置信区间为

$$\left[x_{(1)} + \frac{\ln\alpha}{n},\ x_{(1)}\right]$$

□