将上述结果代入 (7.5.1), 得所求置信区间为

$$(0.131, 0.176).$$

若按公式 (7.5.2) 得置信区间为 (0.130,0.174).

（二）其他分布均值 μ 的区间估计

设总体 X 的均值为 μ, 方差为 $\sigma^2, X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本. 根据中心极限定理, 当样本容量 n 充分大时 (要求 $n > 50$),

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \sim N(0,1)$$

近似成立, 这样可导出 μ 的置信水平为 $1-\alpha$ 的近似的置信区间为

$$\left(\overline{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right). \tag{7.5.3}$$

如果方差 σ^2 未知, 可以用估计量 S^2 代替 σ^2, 由此得到相应的近似置信区间为

$$\left(\overline{X} - \frac{S}{\sqrt{n}} z_{\alpha/2}, \overline{X} + \frac{S}{\sqrt{n}} z_{\alpha/2}\right). \tag{7.5.4}$$

注 当样本容量 $n \leqslant 50$ 时, 根据实际经验, t 分布具有良好的统计稳健性, 即当总体 X 不服从正态分布, 样本数据基本对称时, 枢轴量 $\dfrac{\overline{X} - \mu}{S/\sqrt{n}}$ 仍可以看成近似的 $t(n-1)$ 分布, 从而均值 μ 的置信水平为 $1-\alpha$ 的近似置信区间为

$$\left(\overline{X} \pm \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1)\right). \tag{7.5.5}$$

例 7.5.2 根据实际经验可以认为, 任一路口单位时间内 (如一分钟或一小时或一天等) 的车流量服从泊松分布 $P(\lambda)$. 若以分钟为单位, 对某路口进行 2 小时的记录, 得平均每分钟车流量为 50 辆, 标准差为 10 辆, 试求该路口平均每分钟车流量参数 λ 的置信水平为 0.95 的置信区间.

解 利用 Excel 或查正态分布表得 $Z_{0.025} = 1.96$, 并将样本资料 $n = 120, \overline{x} = 50, s = 10$ 代入 (7.5.4) 得所求置信区间为 $(48.211, 51.789)$.

■ 思考题七

1. 未知参数的估计量与估计值有什么区别?

2. 样本均值 $\overline{X}$ 和样本方差 S^2 分别是总体均值 μ 和总体方差 σ^2 的无偏估计, 问 $\overline{X}=\mu$ 吗? $P\{\overline{X}=\mu\}$ 是多少? $P\{S^2=\sigma^2\}$ 又是多少呢?
3. 设 $\widehat{\theta}$ 是参数 θ 的无偏估计, 且 $D(\widehat{\theta})>0$, 问 $\widehat{\theta}^2$ 是 θ^2 的无偏估计吗?
4. 说明利用矩法和极大似然法求参数点估计量的统计思想.
5. 给出求解参数的矩估计和极大似然估计的主要步骤, 并写出 $0-1$ 分布, 二项分布 $B(n,p)$, 泊松分布 $P(\lambda)$, 均匀分布 $U(a,b)$, 指数分布 $E(\lambda)$ 和正态分布 $N(\mu,\sigma^2)$ 中各参数的矩估计和极大似然估计.
6. 给出估计量的四个评价标准并说明其统计意义.
7. 如何理解置信水平的含义? 置信水平、精确度 (区间平均长度) 和样本容量的关系怎样?
8. 说明枢轴量和统计量的区别.
9. 利用枢轴量法求解参数置信区间的基本步骤, 对正态总体试从有关的统计量出发自行导出几类参数的置信区间.
10. 设总体 $X\sim N(\mu,1)$, μ 是未知参数, $X_1,\cdots,X_n$ 为来自 X 的简单随机样本, $\overline{X}$ 和 S^2 分别是样本均值和样本方差, 则 μ 的置信水平为 $1-\alpha$ 的置信区间应选 $\left(\overline{X}\pm\dfrac{1}{\sqrt{n}}z_{\alpha/2}\right)$, $\left(\overline{X}\pm\dfrac{S}{\sqrt{n}}z_{\alpha/2}\right)$ 还是 $\left(\overline{X}\pm\dfrac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right)$? 为什么?

▶ 习题七

1. 设总体 X 的密度函数为

$$f(x;\theta)=\begin{cases}\dfrac{6x(\theta-x)}{\theta^3}, & 0<x<\theta,\\ 0, & \text{其他},\end{cases}$$

$\theta>0$ 未知, $X_1,\cdots,X_n$ 是来自总体 X 的简单随机样本, 求 θ 的矩估计量 $\widehat{\theta}$, 并计算 $E(\widehat{\theta})$ 和 $D(\widehat{\theta})$.
2. 设湖中有 N 条鱼 (N 未知), 现钓出 r 条, 做上记号后放回湖中, 一段时间后, 再钓出 S 条 ($S\geqslant r$), 结果发现其中 t 条有记号, 试用极大似然法估计湖中鱼的数量 N.
3. 设总体 X 的取值为 0, 1, 2, 从总体取得容量是 n 的样本 $X_1,\cdots,X_n$, 记其中取 0, 1, 2 的个数分别为 $n_0,n_1,n_2,n_0+n_1+n_2=n$, 求下列总体概率分布律中未知参数的矩估计量和极大似然估计量. 特别地, 当样本值是 0, 2, 2, 2, 1, 0, 0, 0, 1, 1 时, 求相应参数的矩估计值和极大似然估计值.

   (1) $0<\theta<1$ 未知

X	0	1	2
p	θ	$\dfrac{1-\theta}{2}$	$\dfrac{1-\theta}{2}$

(2) $0<\theta<1$ 未知

X	0	1	2
p	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

(3) $0<\theta<1, 0<\lambda<1, 0<\theta+\lambda<1$ 未知

X	0	1	2
p	θ	λ	$1-\theta-\lambda$

4. 设 $X_1,\cdots,X_n$ 是来自总体 X 的简单随机样本, 求下列总体 X 的密度函数中未知参数 θ 的矩估计量和极大似然估计量, 并对总体所获得的样本值, 求相应分布参数 θ 的矩估计值和极大似然估计值.

(1) $f(x;\theta)=\begin{cases} 2^{-\theta}\theta x^{\theta-1}, & 0<x<2, \\ 0, & \text{其他}, \end{cases} \quad \theta>0;$

样本值: 0.45　0.2　0.5　0.47　0.35　1.63　0.14　0.06　0.89　0.34

(2) $f(x;\theta)=\begin{cases} \dfrac{1}{2-\theta}, & \theta\leqslant x<2, \\ 0, & \text{其他}, \end{cases} \quad \theta<2;$

样本值: 0.95　0.63　1.69　1.97　0.84　1.81　0.53　0.35　1.34　0.82

(3) $f(x;\theta)=\dfrac{1}{2\theta}\mathrm{e}^{-\frac{|x|}{\theta}}, -\infty<x<+\infty;$

样本值: -0.05　-0.47　0.01　-0.03　-0.18　1.65　-0.64　-1.05　0.41　-0.19

5. 设总体 X 的密度函数为

$$f(x;\theta)=\begin{cases} \dfrac{x}{\theta}\mathrm{e}^{-\frac{x^2}{2\theta}}, & x>0, \\ 0, & \text{其他}, \end{cases}$$

$\theta>0$ 未知, $X_1,\cdots,X_n$ 是来自总体 X 的简单随机样本, 求:

(1) θ 的极大似然估计量;

(2) $E(X^2)$ 的极大似然估计量;

(3) $P\{X>1\}$ 的极大似然估计量.

6. 设总体 X 具有均值 μ, 方差 $\sigma^2, X_1,\cdots,X_{10}$ 为 X 的简单随机样本.

(1) a 取什么值时, $a\sum\limits_{i=1}^{9}(X_{i+1}-X_i)^2$ 是 σ^2 的无偏估计量?

(2) b 取什么值时, $b\sum_{i=1}^{5}(X_i-X_{i+5})^2$ 是 σ^2 的无偏估计量?

7. 设总体 $X\sim N(\mu,\sigma^2),\mu,\sigma^2$ 未知, X_1,X_2,X_3 是 X 的简单随机样本, 用

$$\widehat{\mu}_1=\frac{1}{2}X_1+\frac{1}{4}X_2+\frac{1}{4}X_3,\widehat{\mu}_2=2X_1-2X_2+X_3,\widehat{\mu}_3=\frac{1}{3}X_1+\frac{1}{3}X_2+\frac{1}{3}X_3$$

估计参数 μ, 它们都是无偏估计量吗? 如果是, 哪个更有效? 如果不是, 在均方误差准则下, 哪个更优?

8. 设总体 $X\sim N(\mu,\sigma^2),\mu,\sigma^2$ 未知, 抽取三个独立样本 $(X_1,X_2),(Y_1,Y_2,Y_3),(Z_1,Z_2,Z_3,Z_4)$, S_1^2,S_2^2,S_3^2 分别是对应的样本方差. 设 $T=aS_1^2+bS_2^2+cS_3^2$, 其中 a,b,c 是实数.

(1) 求 T 是 σ^2 的无偏估计量的充要条件;

(2) 问 a,b,c 取何值才是最有效估计量?

9. 设总体 X 的密度函数为

$$f(x;\theta)=\begin{cases}\dfrac{2x}{\theta^2}, & 0<x<\theta,\\ 0, & \text{其他},\end{cases}$$

$\theta>0$ 未知, $X_1,\cdots,X_n(n\geqslant 4)$ 是来自总体 X 的简单随机样本.

(1) 求 θ 的矩估计量 $\widehat{\theta}_1$ 和极大似然估计量 $\widehat{\theta}_2$;

(2) 在均方误差准则下, 判断哪个估计量更优?

(3) 判断两个估计量是否为 θ 的相合估计量?

10. 设总体 X 的密度函数为

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}\mathrm{e}^{-\frac{x}{\theta}}, & x>0,\\ 0, & \text{其他},\end{cases}$$

$\theta>0$ 未知, $X_1,\cdots,X_n$ 是来自总体 X 的简单随机样本.

(1) 证明: 样本均值是 θ 的矩估计量, 也是极大似然估计量;

(2) 在形如 $c\sum_{i=1}^{n}X_i$ 的估计中求 c, 使其在均方误差准则下最优;

(3) 判断由 (2) 得到的估计量是否为 θ 的相合估计量?

11. 设总体 X 的密度函数为

$$f(x;\theta)=\begin{cases}\mathrm{e}^{-(x-\theta)}, & x\geqslant\theta,\\ 0, & \text{其他},\end{cases}$$

θ 未知, $X_1,\cdots,X_n$ 是来自总体 X 的简单随机样本.

(1) 求 θ 的极大似然估计量 $\widehat{\theta}$;

(2) 求 $\widehat{\theta}-\theta$ 的概率密度;

(3) 判断 $\widehat{\theta}-\theta$ 是否可以取为关于 θ 的区间估计问题的枢轴量;

(4) 求 θ 的置信水平为 $1-\alpha$ 的单侧置信下限.

12. 某机器生产的螺杆直径 X (单位: mm) 服从正态分布 $N(\mu, 0.3^2)$. 现随机抽取 5 只, 测得直径为: 22.3, 21.5, 21.8, 21.4, 22.1. 试以 95% 的置信水平计算该机器所生产螺杆平均直径 μ 的置信区间.

13. 某厂生产的灯泡寿命 X (单位: h) 服从 $N(\mu, \sigma^2), \mu, \sigma^2$ 未知, 从已生产的一批灯泡中随机抽取 15 只, 测得其寿命如下:

4 040	2 990	2 964	3 245	3 026	3 633	3 387	4 136
3 595	3 194	3 714	2 831	3 845	3 410	3 004	

(1) 求 μ 的置信水平为 95% 的置信区间;

(2) 求 μ 的置信水平为 95% 的单侧置信下限.

14. 为比较甲、乙两种肥料对产量的影响, 研究者选择了 10 块田地, 将每块田地分成大小相同的两块, 随机选择一块用甲肥料, 另一块用乙肥料, 其他条件保持相同, 得到的产量 (单位: 公斤) 数据如下:

甲肥料 x_i: 109　98　97　100　104　102　94　99　103　108

乙肥料 y_i: 107　105　110　118　109　113　111　95　112　101

假设成对数据差服从正态分布, 求均值差的置信水平为 95% 的置信区间.

15. 求 13 题中总体方差 σ^2 的置信水平为 95% 的置信区间和单侧置信上限.

16. 下面 16 个数字来自计算机的正态随机数生成器:

8.801	3.817	8.223	6.374	9.252	7.352	13.781	7.599
13.134	4.465	6.533	7.021	9.015	7.325	7.041	9.560

(1) 你认为生成的正态分布的均值和方差是多少?

(2) 求均值的置信水平为 95% 的置信区间;

(3) 求方差的置信水平为 95% 的置信区间.

17. 已知某种电子管使用寿命 (单位: h) 服从正态分布 $N(\mu, \sigma^2), \mu, \sigma^2$ 未知, 从一批电子管中随机抽取 16 只, 检测结果得样本标准差为 300 h. 试求:

(1) σ 的置信水平为 95% 的置信区间;

(2) σ 的置信水平为 95% 的单侧置信上限.

18. 为了解某市两所高校学生的消费情况, 在两所高校各随机调查 100 人, 调查结果为: 甲校学生月平均消费 803 元, 标准差 75 元; 乙校学生月平均消费 938 元, 标准差 102 元. 假设甲校学生月平均消费额 (单位: 元) $X \sim N(\mu_1, \sigma^2)$, 乙校学生月平均消费额 (单位: 元) $Y \sim N(\mu_2, \sigma^2), \mu_1, \mu_2, \sigma^2$ 未知, 两样本相互独立. 求两校学生月平均消费额差值 $\mu_1 - \mu_2$ 的置信水平为 95% 的置信区间和单侧置信上限.

19. 某厂的一台瓶装灌装机, 每瓶的净重量 X 服从正态分布 $N(\mu_1, \sigma_1^2)$, 从中随机抽出 16 瓶, 称得其净重的平均值为 456.64 g, 标准差为 12.8 g; 现引进了一台新灌装机, 其每瓶的净重量 Y 服从正态分布 $N(\mu_2, \sigma_2^2)$, 抽取产品 12 件, 称得其净重的平均值为 451.34 g, 标准差为 11.3 g.

    (1) 假设 $\sigma_1 = 13, \sigma_2 = 12$, 求 $\mu_1 - \mu_2$ 的置信水平为 95% 的置信区间;

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

    (3) 求 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的置信水平为 95% 的置信区间.

20. 某超市负责人需要比较郊区 A 和郊区 B 居民的平均收入来确定合适的分店地址. 假设两郊区居民的收入均服从正态分布, 对两个郊区的居民分别进行抽样调查, 各抽取 64 户家庭, 计算得郊区 A 居民的年人均收入为 3.276 万元, 标准差为 0.203 万元; 郊区 B 居民的年人均收入为 3.736 万元, 标准差为 0.421 万元. 假设两个正态总体的方差不相等, 求两郊区居民年人均收入的置信水平为 95% 的近似置信区间.

21. 某餐厅为了解顾客对餐厅新开发的菜品的满意程度, 随机调查来餐厅就餐的顾客 80 人, 结果发现有 55 人满意, 求满意比例 p 的置信水平为 95% 的置信区间.