另一方面有

$$\frac{\partial \ln L(\theta,\mu)}{\partial \mu} = n\theta = 0 \quad \Rightarrow \quad \theta = 0,$$

此时无法求解 θ 和 μ 的最大似然估计. 回到似然函数的定义

$$L(\theta,\mu) = \begin{cases} \theta^n e^{-\theta \sum_{i=1}^n (X_i-\mu)} & X_1, X_2, \cdots, X_n \geqslant \mu \\ 0 & \text{其它} \end{cases}$$

可以发现 μ 越大似然函数 $L(\theta,\mu)$ 越大, 但须满足 $X_i \geqslant \mu$ $(i \in [n])$. 由此可得最大似然估计

$$\hat{\mu} = \min\{X_1, X_2, \cdots, X_n\},$$

进一步求解可得

$$\hat{\theta} = \frac{1}{\frac{1}{n}\sum_{i=1}^n (X_i - \hat{\mu})}.$$

最大似然估计的思想, 最早始于高斯的误差理论, 到 1912 年由费尔希 (R. A. Fisher) 作为一般的方法正式提出. 很多统计学家对最大似然估计进行了研究, 由于考虑了总体的分布信息, 因而在通常情况下会比矩估计法更好, 但在个别情况下也会出现不理想的结果, 尤其当总体分布信息未知时, 最大似然估计无法使用.

9.2 估计量的评价标准

前面介绍了多种估计方法, 不同的估计方法可能得到不同的估计值, 有必要考虑一个问题: 采用哪一种估计量更好, 或更好的标准是什么. 衡量估计量的常用标准包括: 无偏性, 有效性, 一致性.

9.2.1 无偏性

定义 9.1 设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本, 以及 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ 是 θ 的一个估计量, 若

$$E_{X_1,X_2,\ldots,X_n}\left[\hat{\theta}\right] = E_{X_1,X_2,\ldots,X_n}\left[\hat{\theta}(X_1, X_2, \cdots, X_n)\right] = \theta,$$

则称 $\hat{\theta}$ 为 θ 的无偏估计.

无偏性不要求在任意情况下估计值 $\hat{\theta}$ 都等于 θ, 但在平均期望的情形下有 $E(\hat{\theta}) = \theta$ 成立. 其意义在于无系统性偏差, 无偏性是在有限样本情形下一种常见的估计量标准.

设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本, 则有

- 样本原点矩 $A_k = \sum_{i=1}^n X_i^k/n$ 是 k 阶矩的 $a_k = E[X^k]$ 的无偏估计;
- 样本方差 $S_0^2 = \sum_{i=1}^n (X_i - \bar{X})^2/n$ 是方差 $\mathrm{Var}(X)$ 的有偏估计;
- 修正后的样本方差 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ 是方差 $\mathrm{Var}(X)$ 的无偏估计.

若 $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ 是 θ 的无偏估计, 但并不一定有 $g(\hat{\theta})$ 是 $g(\theta)$ 的无偏估计, 这是因为 $E[\hat{\theta}] = \theta$ 并不能得到 $E[g(\hat{\theta})] = g(\theta)$. 例如根据 $E[\bar{X}] = \mu$ 不能得到 $E[\bar{X}^2] \neq \mu^2$.

例 9.8 设 $X_1, \cdots, X_n$ 是总体 $X \sim e(1/\theta)$ 的一个样本 $(\theta > 0)$, 证明样本均值 $\bar{X} = \sum_{i=1}^n X_i/n$ 和 $n\min\{X_1, X_2, \cdots, X_n\}$ 都是 θ 的无偏估计.

证明 根据指数分布的性质有总体 X 的期望 $E[X] = \theta$. 根据期望的性质有

$$E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n E[X_i] = E[X] = \theta,$$

由此可知样本均值 $\bar{X}$ 是 θ 的无偏估计.

设随机变量 $Z = \min\{X_1, X_2, \cdots, X_n\}$, 当 $z > 0$ 时有

$$F_Z(z) = P[Z \leqslant z] = 1 - P[Z > z] = 1 - P[X_1 > z]P[X_2 > z] \cdots P[X_n > z].$$

根据指数分布的密度函数有

$$P[X_i > z] = \int_z^\infty \frac{1}{\theta} e^{-\frac{x}{\theta}} dx = e^{-z/\theta}.$$

由此可得分布函数 $F_Z(z) = 1 - e^{-nz/\theta}$ 以及密度函数

$$f_Z(z) = \begin{cases} \frac{n}{\theta} e^{-\frac{nz}{\theta}} & z \geqslant 0 \\ 0 & z < 0, \end{cases}$$

即随机变量 $Z \sim e(n/\theta)$. 根据指数分布的性质有 $E[Z] = \theta/n$, 由此证明 nZ 是 θ 的无偏估计.

9.2.2 有效性

若 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 是 θ 的无偏估计, 此时可以比较方差

$$\mathrm{Var}(\hat{\theta}_1) = E[(\hat{\theta}_1 - \theta)^2] \qquad \text{和} \qquad \mathrm{Var}(\hat{\theta}_2) = E[(\hat{\theta}_2 - \theta)^2].$$

一般来说: 方差越小估计量越好.

定义 9.2 设 $\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$ 和 $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$ 是 θ 的两个无偏估计, 若

$$\mathrm{Var}(\hat{\theta}_1) \leqslant \mathrm{Var}(\hat{\theta}_2),$$

则称估计量 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效.

例 9.9 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X \sim e(1/\theta)$ 的一个样本 $(\theta > 0)$, 证明样本均值 $\bar{X} = \sum_{i=1}^n X_i/n$ 比 $n\min\{X_1, X_2, \cdots, X_n\}$ 估计参数 θ 更有效.

证明 设 $Z=\min\{X_1,X_2,\cdots,X_n\}$, 从例 9.8 可知样本均值 $\bar{X}$ 和 nZ 是参数 θ 的无偏估计. 对样本均值 $\bar{X}$, 根据方差的性质有

$$\mathrm{Var}(\bar{X})=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i)=\frac{\theta^2}{n}.$$

从例 9.8 可知随机变量 $Z\sim e(n/\theta)$, 根据指数分布的性质有 $\mathrm{Var}(Z)=\theta^2/n^2$, 由此可得

$$\mathrm{Var}(nZ)=n^2\mathrm{Var}(Z)=\theta^2,$$

因此当 $n\geqslant 1$ 时有 $\mathrm{Var}(\bar{X})<\mathrm{Var}(nZ)$ 成立, 故 $\bar{X}$ 比 nZ 估计参数 θ 更有效.

例 9.10 若 $X_1,X_2,\cdots,X_n$ 是总体 X 的一个样本, 设非负常数 $\lambda_1,\cdots,\lambda_n$ 满足 $\sum_{i=1}^n\lambda_i=1$, 证明样本均值 $\bar{X}$ 比 $\sum_{i=1}^n\lambda_iX_i$ 有效.

证明 根据期望的性质很容易发现样本均值 $\bar{X}$ 和 $\sum_{i=1}^n\lambda_iX_i$ 均是期望 $E[X]$ 的无偏估计. 根据样本的独立同分布条件有

$$\mathrm{Var}\left(\bar{X}\right)=\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i)=\frac{\mathrm{Var}(X)}{n}.$$

进一步根据方差的性质和不等式 $\sum_{i=1}^n\lambda_i^2/n\geqslant(\sum_{i=1}^n\lambda_i/n)^2=1/n^2$ 有

$$\mathrm{Var}\left(\sum_{i=1}^{n}\lambda_iX_i\right)=\sum_{i=1}^{n}\lambda_i^2\mathrm{Var}(X_i)=\sigma^2\sum_{i=1}^{n}\lambda_i^2\geqslant\frac{\mathrm{Var}(X)}{n},$$

由此完成证明.

下面给出最有效的统计量, 即 **Rao-Cramer 不等式**.

定理 9.2 设随机变量 X 的概率概率密度为 $f(x;\theta)$ 或分布函数为 $F(x;\theta)$, 以及设

$$\mathrm{Var}_0(\theta)=\left(nE_X\left[\left(\frac{\partial\ln f(X;\theta)}{\partial\theta}\right)^2\right]\right)^{-1}\quad\text{或}\quad\mathrm{Var}_0(\theta)=\left(nE_X\left[\left(\frac{\partial\ln F(X;\theta)}{\partial\theta}\right)^2\right]\right)^{-1},$$

则对任意的无偏估计量 $\hat{\theta}$ 有

$$\mathrm{Var}(\hat{\theta})\geqslant\mathrm{Var}_0(\theta),$$

称 $\mathrm{Var}_0(\theta)$ 为估计量 $\hat{\theta}$ 方差的下界. 当 $\mathrm{Var}(\hat{\theta})=\mathrm{Var}_0(\theta)$ 时称 $\hat{\theta}$ 为最有效估计量、简称有效估计量.

例 9.11 设 $X_1,X_2,\cdots,X_n$ 是来自总体 $X\sim e(1/\theta)$ 的一个样本 $(\theta>0)$, 则样本均值 $\bar{X}=\sum_{i=1}^nX_i/n$ 为 θ 的最有效估计量.

解 根据独立同分布假设有

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{\theta^2}{n}.$$

根据指数分布的概率密度函数有

$$\frac{\partial \ln f(X;\theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \ln\left(\frac{1}{\theta}\exp\left(-\frac{X}{\theta}\right)\right) = \frac{\partial}{\partial \theta}\left(-\ln\theta - \frac{X}{\theta}\right) = -\frac{1}{\theta} + \frac{X}{\theta^2}.$$

结合总体 X 的期望 $E[X]=\theta$ 和方差 $\mathrm{Var}(X)=\theta^2$ 有

$$E\left[\frac{\partial \ln f(X;\theta)}{\partial \theta}\right]^2 = E\left[\left(-\frac{1}{\theta} + \frac{X}{\theta^2}\right)^2\right] = \frac{E[(X-\theta)^2]}{\theta^4} = \frac{E[(X-E[X])^2]}{\theta^4} = \frac{\mathrm{Var}(X)}{\theta^4} = \frac{1}{\theta^2}.$$

根据定理 9.2 可知

$$\mathrm{Var}_0(X) = \frac{1}{n}\left(E_X\left[\left(\frac{\partial \ln f(X;\theta)}{\partial \theta}\right)^2\right]\right)^{-1} = \frac{\theta^2}{n} = \mathrm{Var}(\bar{X}),$$

由此可知最大似然估计 $\bar{X}$ 是 θ 的有效估计量.

9.2.3 一致性

定义 9.3 设 $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$ 是 θ 的一个估计量, 若当 $n \to \infty$ 时有 $\hat{\theta}_n \xrightarrow{P} \theta$ 成立, 即对任意 $\epsilon > 0$ 有

$$\lim_{n\to 0} P[|\hat{\theta}_n - \theta| > \epsilon] = 0,$$

则称 $\hat{\theta}_n$ 为 θ 的一致估计量.

估计量的一致性刻画了在足够多样本情形下估计量 $\hat{\theta}$ 能有效逼近真实值 θ, 一致性是对估计的基本要求, 不满足一致性的估计量一般不予考虑. 下面给出满足一致性的充分条件:

定理 9.3 设 $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$ 是 θ 的一个估计量, 若满足以下两个条件:

$$\lim_{n\to\infty} E\left[\hat{\theta}_n\right] = \theta \quad \text{和} \quad \lim_{n\to\infty} \mathrm{Var}\left(\hat{\theta}_n\right) = 0,$$

则 $\hat{\theta}_n$ 为 θ 的一致估计量.

证明 根据概率的性质有

$$P\left[|\hat{\theta}_n - \theta| > \epsilon\right] \leqslant P\left[\left|\hat{\theta}_n - E[\hat{\theta}_n]\right| > \epsilon/2\right] + P\left[\left|E[\hat{\theta}_n] - \theta\right| > \epsilon/2\right].$$

根据 $\lim_{n\to\infty} E[\hat{\theta}_n]=\theta$ 可知对任意 $\epsilon>0$, 存在一个N_0, 当 $n\geqslant N_0$ 有 $|E[\hat{\theta}_n]-\theta|\leqslant\theta/2$, 于是有

$$\lim_{n\to\infty} P\left[\left|E[\hat{\theta}_n]-\theta\right|>\epsilon/2\right]=0;$$

再根据 Chebyshev 不等式有

$$\lim_{n\to\infty} P\left[\left|\hat{\theta}_n-E[\hat{\theta}_n]\right|>\epsilon/2\right]\leqslant\lim_{n\to\infty}\frac{4}{\epsilon^2}\mathrm{Var}(\hat{\theta}_n)=0.$$

由此完成证明.

根据上述定理可以研究一些估计量的一致性.

例 9.12 设 $X_1,X_2,\cdots,X_n$ 是来自总体 $X\sim e(1/\theta)$ 的一个样本, 则样本均值 $\bar{X}=\sum_{i=1}^n X_i/n$ 为 θ 的一致估计量.

证明 根据指数分布和期望的性质有

$$E[\bar{X}]=\frac{1}{n}\sum_{i=1}^n E[X_i]=\theta.$$

再根据方差的性质

$$\lim_{n\to\infty}\mathrm{Var}(\bar{X})=\lim_{n\to\infty}\frac{\theta^2}{n}=0,$$

由此完成证明.

例 9.13 设 $X_1,X_2,\cdots,X_n$ 是来自总体 $X\sim U(0,\theta)$ 的一个样本, 证明 θ 的最大似然估计量是一致估计量.

证明 根据前面的例题得到 θ 的最大似然估计为 $\hat{\theta}_n=\max(X_1,X_2,\cdots,X_n)$. 设随机变量 $Z=\max(X_1,X_2,\cdots,X_n)$, 则由 Z 的分布函数

$$F_Z(z)=P[Z\leqslant z]=P[\max(X_1,X_2,\cdots,X_n)\leqslant z]=\prod_{i=1}^n P[X_i\leqslant z]=\begin{cases}1 & z>\theta\\ (z/\theta)^n & z\in[0,\theta]\\ 0 & z<0.\end{cases}$$

由此得到当 $z\in[0,\theta]$ 时随机变量 Z 的概率密度函数 $f_Z(z)=nz^{n-1}/\theta^n$, 进一步有

$$E\left[\hat{\theta}_n\right]=E[Z]=\int_0^\theta\frac{nz^n}{\theta^n}dz=\frac{n}{n+1}\theta,$$

因此 $\hat{\theta}$ 是 θ 的有偏估计. 另一方面有

$$E\left[Z^2\right]=\int_0^\theta\frac{nz^{n+1}}{\theta^n}dz=\frac{n}{n+2}\theta^2,$$

从而得到

$$\mathrm{Var}\left(\hat{\theta}_n\right) = \mathrm{Var}(Z) = E[Z^2] - (E[Z])^2 = \frac{n}{n+2}\theta^2 - (\frac{n\theta}{n+1})^2 = \frac{n}{(n+1)^2(n+2)}\theta^2,$$

于是有

$$\lim_{n\to\infty} E[\hat{\theta}_n] = \theta \quad 和 \quad \lim_{n\to\infty} \mathrm{Var}(\hat{\theta}_n) = 0,$$

由此可得 $\hat{\theta}_n$ 是 θ 的有偏但一致估计量.

9.3 区间估计

区间估计问题: 设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本, θ 为总体 X 的分布函数 $F(x,\theta)$ 的未知参数, 根据样本估计 θ 的范围 $(\hat{\theta}_1, \hat{\theta}_2)$, 其中 $\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$ 和 $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$, 使得以较大的概率保证有 $\theta \in (\hat{\theta}_1, \hat{\theta}_2)$ 成立. 具体而言, 对任意给定 $\alpha \in (0,1)$, 有

$$P\left[\hat{\theta}_1(X_1, X_2, \cdots, X_n) < \theta < \hat{\theta}_2(X_1, X_2, \cdots, X_n)\right] \geqslant 1 - \alpha.$$

定义 9.4 (置信区间与置信度) 设 $X_1, X_2 \cdots, X_n$ 是来自总体 X 的样本, 总体 X 的分布函数含未知参数 θ, 找出统计量 $\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$ 和 $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$ $(\hat{\theta}_1 < \hat{\theta}_2)$, 使得

$$P\left[\hat{\theta}_1 < \theta < \hat{\theta}_2\right] \geqslant 1 - \alpha$$

成立, 则称 $1-\alpha$ 为置信度, $[\hat{\theta}_1, \hat{\theta}_2]$ 为 θ 的置信度为 $1-\alpha$ 的置信区间.

注意: 置信区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 是随机区间, $1-\alpha$ 为该区间包含 θ 的概率/可靠程度. 若 $\alpha = 0.05$, 则置信度为 95%. 通常采用 95% 的置信度, 有时也可 99% 或 90%等.

i) $\hat{\theta}_2 - \hat{\theta}_1$ 反映了估计精度, 长度越小精度越大.

ii) α 反映了估计的可靠度, α 越小可靠度越高.

iii) 给定 α, 区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 的选取并不唯一确定, 通常选长度最小的一个区间.

置信区间的求解方法: **枢轴变量法**.

1) 先找一样本函数 $W(X_1, X_2, \cdots, X_n; \theta)$ 包含待估参数 θ, 但不含其它参数, 函数 W 的分布已知, 称 W 为枢轴变量.

2) 给定置信度 $1-\alpha$, 根据 W 的分布找出临界值 a 和 b, 使得 $P[a < W < b] = 1-\alpha$ 成立.

3) 根据 $a < W < b$ 解出 $\hat{\theta}_1 < \theta < \hat{\theta}_2$, 则 $(\hat{\theta}_1, \hat{\theta}_2)$ 为 θ 的置信度为 $1-\alpha$ 的置信区间.

9.3.1 正态总体, 方差已知, 求期望的区间估计

设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的样本, 若方差 σ^2 已知. 给定 $\alpha \in (0,1)$, 确定置信度为 $1-\alpha$ 下 μ 的置信区间 $[\hat{\theta}_1, \hat{\theta}_2]$. 令样本均值为 $\bar{X} = \sum_{i=1}^n X_i/n$, 根据正太分布的性质找出枢轴变量:

$$W = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1),$$

给定置信度 $1-\alpha$, 找出临界值 a 和 b 使得

$$P[a < W < b] = 1-\alpha.$$

根据正态分布的性质、对称性和上分位点可知

$$P[W \geqslant \mu_{\alpha/2}] = \alpha/2 \quad \text{和} \quad P[W \leqslant -\mu_{\alpha/2}] = \alpha/2.$$

求解可得 $a = -\mu_{\alpha/2}$ 和 $b = \mu_{\alpha/2}$. 于是有

$$P[-\mu_{\alpha/2} < W < \mu_{\alpha/2}] = 1-\alpha.$$

根据 $W = (\bar{X}-\mu)/(\sigma/\sqrt{n})$ 可得

$$P\left[\bar{X} - \frac{\sigma}{\sqrt{n}}\mu_{\alpha/2} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}}\mu_{\alpha/2}\right] = 1-\alpha.$$

例 9.14 某地区儿童身高服从正态分布, 现随机抽查 9 人, 高度分别为 115, 120, 131, $115, 109, 115, 115$, $105, 110$, 已知 $\sigma^2 = 7$ 和置信度为 95%, 求期望 μ 的置信区间 ($\mu_{0.025} = 1.96$).

9.3.2 正态总体, 方差未知, 求期望的区间估计

设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的样本, 若方差 σ^2 未知, 考虑期望 μ 的置信度为 $1-\alpha$ 的置信区间. 设 $\bar{X} = \sum_{i=1}^n X_i/n$ 和 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$, 根据正态总体抽样定理可知:

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1).$$

由此设枢轴变量

$$W = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1).$$

给定置信度 $1-\alpha$, 设临界值 a 和 b 满足

$$P[a \leqslant W \leqslant b] = 1-\alpha \quad \Rightarrow \quad b = t_{\alpha/2}(n-1),\ a = -t_{\alpha/2}(n-1).$$

整理可得

$$P\left[\bar{X} - \frac{S}{\sqrt{n}}t_{\alpha/2}(n-1) < \mu < \bar{X} + \frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right] = 1-\alpha.$$

9.3.3 正态总体, 求方差 σ^2 的置信区间

设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的样本, 考虑方差 σ^2 的置信度为 $1-\alpha$ 的置信区间. 设修正样本方差 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$, 根据正态总体抽样定理有

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

由此设枢轴变量 $W = (n-1)S^2/\sigma^2$, 设临界值 a 和 b 满足

$$P[a \leqslant W \leqslant b] = 1-\alpha.$$

根据 χ^2 分布的不对称性, 采用概率对称的区间

$$P[W \leqslant a] = P[W \geqslant b] = \alpha/2 \quad \Rightarrow \quad b = \chi^2_{\alpha/2}(n-1), \quad a = \chi^2_{1-\alpha/2}(n-1).$$

根据枢轴变量 $W = (n-1)S^2/\sigma^2$ 可得

$$P\left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}\right] = 1-\alpha.$$

9.3.4 双正态总体情形

设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ 的样本, 设 $Y_1, Y_2, \cdots, Y_m$ 是总体 $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$的样本, 令

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i, \quad \bar{Y} = \frac{1}{m}\sum_{i=1}^m Y_i, \quad S_1^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2, \quad S_2^2 = \frac{1}{m-1}\sum_{i=1}^m (Y_i - \bar{Y})^2.$$

考虑 $\mu_1 - \mu_2$ 和 σ_1^2/σ_2^2 的置信度为 $1-\alpha$ 的区间估计.

1) 已知方差 σ_1^2 和 σ_2^2, 求 $\mu_1 - \mu_2$ 的置信区间. 根据正态分布的性质有

$$\bar{X} \sim \mathcal{N}(\mu_1, \frac{\sigma_1^2}{n}), \quad \bar{Y} \sim \mathcal{N}(\mu_2, \frac{\sigma_2^2}{m}) \quad \bar{X} - \bar{Y} \sim \mathcal{N}(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}),$$

进一步有

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim \mathcal{N}(0, 1).$$

于是设枢轴变量

$$W = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim \mathcal{N}(0, 1),$$

求解置信区间

$$P\left[\bar{X}-\bar{Y}-\mu_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}<\mu_1-\mu_2<\bar{X}-\bar{Y}+\mu_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right]=1-\alpha.$$

2) 若 σ_1^2 和 σ_2^2未知, 但已知 $\sigma_1^2=\sigma_2^2$, 设

$$S_W=\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2},$$

则考虑枢轴变量

$$W=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\frac{1}{n}+\frac{1}{m}}}\sim t(n+m-2).$$

于是有

$$P\left[-t_{\alpha/2}(n+m-2)<\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\frac{1}{n}+\frac{1}{m}}}<t_{\alpha/2}(n+m-2)\right]=1-\alpha.$$

3) 求方差比 σ_1^2/σ_2^2 的置信度为 $1-\alpha$ 的置信区间. 设枢轴变量

$$W=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n-1,m-1),$$

根据 F 分布的不对称性, 采用概率对称的区间

$$P[W\leqslant a]=P[W\geqslant b]=\alpha/2\quad\Rightarrow\quad b=F_{\alpha/2}(n-1,m-1),\quad a=F_{1-\alpha/2}(n-1,m-1).$$

由此可得置信区间

$$P\left[\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n-1,m-1)}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}(n-1,m-1)}\right]=1-\alpha.$$

9.3.5 单侧置信区间

对某些实际问题, 我们往往只关心置信区间的上限或下限, 例如, 次品率只关心上限, 产品的寿命只关心下限, 由此引入单侧置信区间及其估计.

定义 9.5 (单侧置信区间) 给定 $\alpha\in(0,1)$, 若样本 $X_1,\cdots,X_n$ 的统计量 $\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\cdots,X_n)$ 满足

$$P\left[\theta>\hat{\theta}_1\right]\geqslant 1-\alpha,$$

则称 $(\hat{\theta}_1,+\infty)$ 为 θ 的置信度为 $1-\alpha$ 的单侧置信区间, $\hat{\theta}_1$ 称为单侧置信下限.

同理定义单侧置信上限. 对正态总体, 可以将相关置信区间的估计都扩展到单侧置信估计, 枢轴变量的定理类似, 我们将不再重复讨论, 下面仅举两个实例:

例 9.15 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的样本, 若方差 σ^2 已知, 求 μ 的置信度为 $1-\alpha$ 的单侧置信下限和上限.

解 设样本均值 $\bar{X} = \sum_{i=1}^n X_i/n$, 根据正态分布的性质考虑枢轴变量

$$W = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1),$$

于是有

$$P\left[\frac{\bar{X}-\mu}{S/\sqrt{n}} < \mu_\alpha\right] = 1-\alpha, \quad P\left[\frac{\bar{X}-\mu}{S/\sqrt{n}} > -\mu_\alpha\right] = 1-\alpha,$$

整理计算完成估计.

例 9.16 从一批出厂的灯泡中随机抽取 10 盏灯泡, 测试其寿命分别为: 1000, 1500, 1250, 1050, 950, 1000, 1150, 1050, 950, 1000, (单位: 小时). 假设这批灯泡的寿命服从正态分布, 求这批灯泡平均寿命的置信度为 95% 的单侧置信下限.

解 首先计算样本均值和样本修正方差分别为

$$\bar{X} = \sum_{i=1}^{10} X_i/10 = 1090 \quad \text{和} \quad S^2 = \sum_{i=1}^{10}(X_i-\bar{X})^2/9 = 8800/3.$$

根据正态分布的性质考虑枢轴变量

$$W = \frac{\bar{X}-\mu}{S/3} \sim t(9),$$

于是有

$$P\left[\frac{\bar{X}-\mu}{S/3} < t_{0.05}(9)\right] = 0.95,$$

查表 $t_{0.05}(9) = 1.833$ 可得

$$\mu > \bar{X} - t_{0.05}(9)S/3 = 1090 - \sqrt{8800/3} \times 1.833/3 > 1056.$$

9.3.6 非正态分布的区间估计

设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本, 若总体 X 的分布未知或非正态分布, 我们可以给出总体期望 $\mu = E[X]$ 的区间估计, 方法分为两种: 利用 Concentration 不等式和中心极限定理.

(1) 首先考虑 Concentration 不等式, 若总体 $X \in [a,b]$, 设 $\bar{X} = \sum_{i=1}^n X_i/n$, 根据 Concentration 不等式有

$$P\left[|\mu - \bar{X}| \geqslant \epsilon\right] \leqslant 2\exp(-2n\epsilon^2/(b-a)^2).$$

令 $\alpha = 2\exp(-2n\epsilon^2/(b-a)^2)$ 求解 $\epsilon = \sqrt{(b-a)^2\ln(2/\alpha)/n}$, 于是有

$$P\left[\bar{X} - \sqrt{(b-a)^2\ln(2/\alpha)/n} < \mu < \bar{X} + \sqrt{(b-a)^2\ln(2/\alpha)/n}\right] > 1-\alpha.$$

可基于其它 Concentration 不等式给出类似的置信区间估计, 以及其它 sub-Gaussian 型随机变量的期望的置信区间估计.

(2) 利用中心极限定理, 求枢轴变量的近似分布, 再给出置信区间估计. 设总体 X 的期望 $E(X)=\mu$, 方差 $\mathrm{Var}(X)=\sigma^2$, 利用中心极限定理设枢轴变量

$$W = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}.$$

枢轴变量 W 的分布近似于标准正态分布 $\mathcal{N}(0,1)$. 当方差 σ^2 已知时有

$$P\left[-\mu_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \mu_{\alpha/2}\right] \approx 1-\alpha.$$

当方差 σ^2 未时, 用修正样本方差 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ 代替方差 σ^2, 于是有

$$P\left[-\mu_{\alpha/2} < \frac{\bar{X}-\mu}{S^2/\sqrt{n}} < \mu_{\alpha/2}\right] \approx 1-\alpha.$$

例 9.17 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X \sim \mathrm{Ber}(p)$ 的样本, 求 p 的置信度为 $1-\alpha$ 的区间估计.

解 根据 Bernoulli 分布的性质有 $X_i \in \{0,1\}$ 以及 $p = E[X]$, 根据 Chernoff 不等式有

$$P[|\bar{X}-p| > \epsilon p] \leqslant 2\exp(-np\epsilon^2/3),$$

设 $\alpha = 2\exp(-np\epsilon^2/3)$, 于是有

$$P\left[\bar{X} - \sqrt{3p\ln(2/\alpha)/n} < p < \bar{X} + \sqrt{3p\ln(2/\alpha)/n}\right] \geqslant 1-\alpha,$$

最后求解 p 的置信区间.

方法二: 根据 Bernoulli 分布的性质有 $E[X]=p$ 和 $\mathrm{Var}(X)=p(1-p)$, 设枢轴变量

$$W = \frac{n\bar{X}-np}{\sqrt{np(1-p)}}$$

根据中心极限定理可知 W 近似于标准正态分布 $\mathcal{N}(0,1)$. 于是有

$$P\left[-\mu_{\alpha/2} < \frac{n\bar{X}-np}{\sqrt{np(1-p)}} < \mu_{\alpha/2}\right] \approx 1-\alpha.$$