区间估计

设总体X的分布函数为$F(x;\theta)$,未知参数$\theta \in \boldsymbol{\Theta} \subset \mathbb{R}$, $X_1, X_2, \dots, X_n$是来自总体X的样本，

找两个统计量$\underline{\theta} = \underline{\theta}(X_1, \dots, X_n)$和$\overline{\theta} = \overline{\theta}(X_1, \dots, X_n)$满足

- 可靠度要强： $P\{\underline{\theta}(X_1, \dots, X_n) < \theta < \overline{\theta}(X_1, \dots, X_n)\}$ 要大
- 精度要高： 区间长度$\overline{\theta}(X_1, \dots, X_n) - \underline{\theta}(X_1, \dots, X_n)$要小

精度 ⟸ 矛盾 ⟹ $\boxed{\text{可靠度}}$ → 优先保证

区间估计

设总体X的分布函数为$F(x;\theta)$, 未知参数$\theta \in \Theta \subset \mathbb{R}$,

$X_1, X_2, \dots, X_n$是来自总体X的样本，$0 < \alpha < 1$ 是给定值.

若两个统计量$\underline{\theta} = \underline{\theta}(X_1, \dots, X_n)$和$\overline{\theta} = \overline{\theta}(X_1, \dots, X_n)$满足

$$P\{\underline{\theta} < \theta < \overline{\theta}\} = 1 - \alpha \quad 对一切\theta \in \Theta,$$

则称随机区间$(\underline{\theta}, \overline{\theta})$是$\theta$的置信度为$1 - \alpha$的置信区间,

分别称$\underline{\theta}$和$\overline{\theta}$为置信度为$1 - \alpha$的置信下限和置信上限.

$$P\{\underline{\theta}(X_1,\ldots,X_n) < \theta < \overline{\theta}(X_1,\ldots,X_n)\} = 1-\alpha$$

给定 α，固定样本容量n, 然后进行多次抽样，每次抽样都可得到一个区间,

这些区间并非每个都包含未知参数的真值，包含θ的真值的区间

大约占$100(1-\alpha)\%$，即只能以$100(1-\alpha)\%$ 的可靠度保证，

将样本观察值带入$(\underline{\theta},\overline{\theta})$ 所得的区间包含θ的真值.

$$P\{\underline{\theta}(X_1,\ldots,X_n) < \theta < \overline{\theta}(X_1,\ldots,X_n)\} = 1-\alpha$$

置信区间$(\underline{\theta},\overline{\theta})$的长度反映了区间估计的精度

置信度$1-\alpha$反映了置信区间包含未知参数真值的可靠度.

$$P\{\underline{\theta}(X_1,\dots,X_n) < \theta < \overline{\theta}(X_1,\dots,X_n)\} = 1-\alpha$$

引例 已知幼儿身高服从正态分布，现从5~6岁的幼儿中随机地抽查了9人，

其高度(单位 cm)分别为：

$n = 9$

$\boxed{115, 120, 131, 115, 109, 115, 115, 105, 110;}$ $\dashrightarrow$ $\bar{x} = 115$

假设标准差$\sigma = 7$，置信度为95%，试求总体均值μ的置信区间.

$\alpha = 0.05$

$$P\left\{\underline{\mu}(X_1,\dots,X_n) < \mu < \overline{\mu}(X_1,\dots,X_n)\right\} = 1-\alpha$$

此例是方差已知，要求总体均值μ的区间估计，故考虑μ的无偏估计$\overline{X}$.

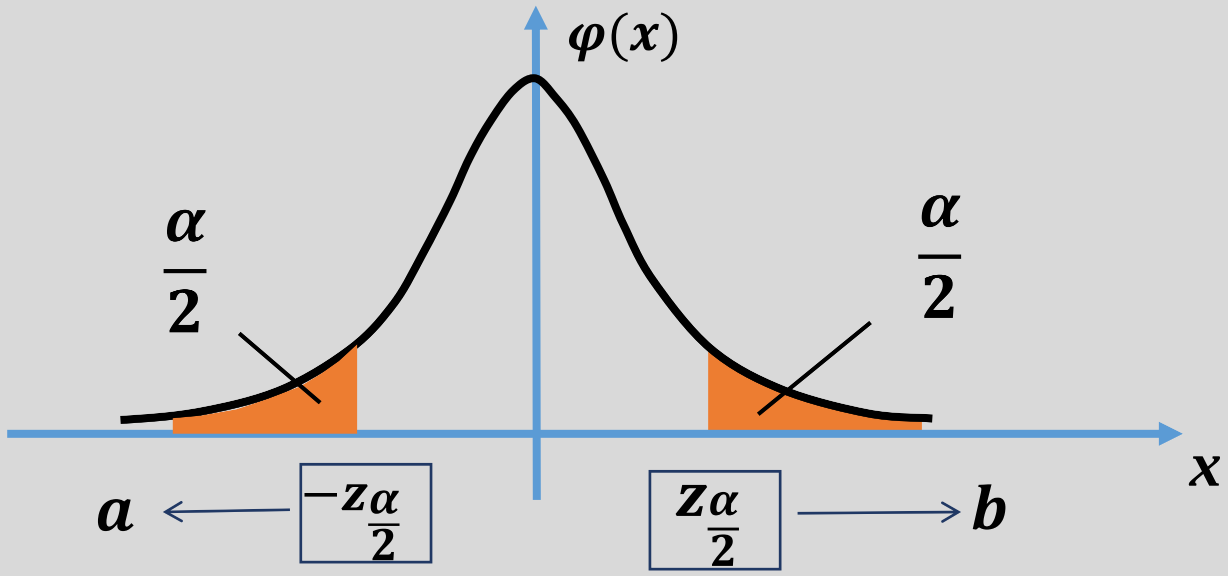
由于已知 $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ 枢轴量

故对给定的α，要求常数a, b,使得

$$P\left\{a < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < b\right\} = 1-\alpha$$

$$P\left\{\underline{\mu}(X_1, \dots, X_n) < \mu < \overline{\mu}(X_1, \dots, X_n)\right\} = 1 - \alpha$$

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



即得 $P\left\{\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| < z_{\frac{\alpha}{2}}\right\} = 1 - \alpha$

$$P\left\{\underline{\mu}(X_1,\dots,X_n) < \mu < \overline{\mu}(X_1,\dots,X_n)\right\} = 1-\alpha$$

整理得 $$P\left\{\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}\right\} = 1-\alpha$$

故μ的置信度为$1-\alpha$的置信区间为 $$\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}\right)$$

查表得 $z_{0.025} = 1.96$ 并带入已知值得

$$\left(115 - 1.96 \times \frac{7}{\sqrt{9}}, 115 + 1.96 \times \frac{7}{\sqrt{9}}\right) = (110.43, 119.57)$$

$n = 9$ $\bar{x} = 115$ 标准差$\sigma = 7$，置信度$1-\alpha = 95\%$

$$P\{\underline{\theta}(X_1,\dots,X_n)<\theta<\overline{\theta}(X_1,\dots,X_n)\}=1-\alpha$$

区间估计的步骤：

枢轴量

- 选取一个即含样本又含未知参数θ且分布已知的函数$Z(X_1,\dots,X_n;\theta)$

- 对给定的置信度$1-\alpha$,根据枢轴量的分布的分位点求出待定常数a,b

  使得对一切$\theta\in\Theta$, $P\{a<Z(X_1,\dots,X_n;\theta)<b\}=1-\alpha$

- 利用不等式变形，求出未知参数θ的置信区间.