第 9 讲*

区间估计和置信区间

麻省理工学院 14.30 2006 年春季

Herman Bennett

22 区间估计

区间估计是估计参数θ的另外一种方法。区间估计是指根据给定概率 $(1-\alpha)$，构造包含参数θ真值的一个随机区间。这个区间就被称作置信区间，而概率$(1-\alpha)$则被称为置信水平。

$$P(A(X_1,...,X_n) \le \theta \le B(X_1,...,X_n)) = 1-\alpha \quad \textbf{(71)}$$

例 22.1. 假设一个来自正态分布$N(\mu,\sigma^2)$总体的随机样本，但两个参数均未知。计算σ^2的置信水平为 90%（对称的）的置信区间。

注意：这些讲义不一定是自封的。它们只是对讲座的一种补充而不是替代。

例 22.2. 假设一个随机样本，来自服从正态分布 $N(\mu,\sigma^2)$ 的总体，参数 μ 未知，σ =2。计算 μ 的置信水平为 95%（对称）的置信区间。如果 σ 未知，结果会有怎样的改变？

- 随机区间 $(\bar{X}-\frac{1.96\sigma}{\sqrt{n}},\bar{X}+\frac{1.96\sigma}{\sqrt{n}})$ 包含参数 θ 真值的概率为 95%。但这并不表明 θ 有 95%的概率落在这个区间内，因为 θ 不是一个随机变量！

23 有用的结论

23.1 t 分布

若一个随机变量 X 服从 t 分布，其中参数 ν >0（自由度）。X 的概率密度函数为：

$$X \sim t(\nu): f(x|\nu)=\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\frac{1}{\sqrt{\nu\pi}}\frac{1}{(1+(\frac{x^2}{\nu}))(\nu/2+0.5)} \tag{72}$$

其中 $-\infty < x < \infty$ 且 ν 是正整数。

设 $X \sim N(0,1)$ 和 $Z \sim X_n^{\ 2}$ 是相互独立的随机变量。那么，随机变量 H 服从自由度为 n 的 t 分布。

$$H = \frac{X}{\sqrt{Z/n}} \sim t(n) \tag{73}$$

- 关于 0 点呈对称分布，可得 $t_{a/2,n} = -t_{1-a/2,n}$。

- 当 $n \to \infty$, $t(n) \to N(0,1)$。（见附图和附表）

例 23.1. 重要结论：若一个随机样本服从正态分布，那么，$\frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t(n-1)$。证明这个结论。

23.2 F 分布

随机变量 X 服从 F 分布，其中，参数 $v_1 > 0$ 且 $v_2 > 0$，X 的概率密度函数为：

$$X \sim F_{(v_1,v_2)} : f(x|v_1,v_2) = \frac{\Gamma(\frac{v_1+v_2}{2})}{\Gamma(\frac{v_1}{2})\Gamma(\frac{v_2}{2})}(\frac{v_1}{v_2})^{(v_1/2)}\frac{x^{(v_1/2-1)}}{(1+(\frac{v_1}{v_2})x)^{(v_1/2+v_2/2)}}, \tag{74}$$

其中，$0 < x < \infty$ 且 v_i 是正整数。

设 X和Z 是两个独立的随机变量，其中，$X \sim \chi_n^{\ 2}$，$Z \sim \chi_m^{\ 2}$。那么，随机变量G 服从自由度为 n 和 m 的 F分布。

$$G = \frac{X/n}{Z/m} \sim F_{(n,m)} \tag{75}$$

24 构造θ的置信区间

下面分析参数 θ 在给定有限信息的五种情形。针对每一种情形，研究如何构造置信区间。

24.1 情形 1：$\hat{\theta} \sim N(\theta, Var(\hat{\theta}))$且$\mathrm{Var}(\hat{\theta})$ 已知

我们刚讲解的例 22.2 就是这种情形的一个例子。注意：在例子中$\theta = \mu, \hat{\theta} = \hat{X}$，并且因为$\sigma$ 已知，$\mathrm{Var}(\hat{\theta})$ 也已知。

24.2 情形 2：$\hat{\theta} \sim N(\theta, Var(\hat{\theta}))$且 $\mathrm{Var}(\hat{\theta})$ 未知

例 24.1. 假定如例 22.2 一样中的一个正态随机样本，但现在μ和σ_2都未知。构造μ的置信水平为 95%的置信区间。

24.3 情形 3：$\hat{\theta}$不服从正态分布，但概率密度函数已知

我们刚讲解的例 22.1 就是这种情形的一个例子。注意：在例子中$\theta=\sigma^2,\hat{\theta}=S^2$，$\hat{\theta}$的概率密度函数已知（密度函数为$\hat{\theta}$的函数），且仅取决于一个参数，$\sigma^2$。

24.4　情形 4：$\hat{\theta}$不服从正态分布，概率质量/密度函数未知，且n>30

例 24.2. 设一个容量为 n 的随机样本来自一个总体$f(x)$，且$f(x)$未知。构造$\mu=E(X_i)$置信水平为 99%的一个置信区间。

24.5 情形 5：$\hat{\theta}$不服从正态分布，概率质量/密度函数未知，且n<30