1 数理统计的基本概念

1.1 总体，样本，统计量

总体: 研究对象全体，可以用一个随机变量 X 及其分布 F 表示

个体: 总体中的每一个成员

抽样: 按照一定法则，从总体中抽取部分个体观测的过程，常用有不放回抽样和放回抽样，在样本容量相对总体过小时这两个方法可以视为等同

样本: 抽样选中的部分个体，用 n 个随机变量来表示

样本容量: 样本中个体的数量

样本观测值: 由样本中的 n 个随机变量观察得到的 n 个数据

简单随机抽样满足:1) 代表性: 总体中每个个体都有相同的被抽到的概率，则样本中 n 个随机变量和总体 X 有相同的分布

2）独立性:n 个随机变量相互独立

简单随机样本: 由简单随机抽样得到的样本

统计量: 设 $X_1, X_2, \cdots, X_n$ 是总体 X 的一个样本，若 $T = T(X_1, X_2, \cdots, X_n)$ 是样本的函数, 且不含任意未知参数，则称 T 是一个统计量

常见统计量: 样本均值:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

样本方差:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

样本标准差:

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

样本 k 阶原点矩:

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

极大次序统计量:

$$X_{(n)} = max\{X_1, X_2, \cdots, X_n\}$$

极小次序统计量:

$$X_{(1)} = min\{X_1, X_2, \cdots, X_n\}$$

1.2 常用统计量的分布

1.2.1 标准正态

正态分布的可加性: 对于相互独立的随机变量 $X_1, X_2, \cdots, X_n$ 且 $X_i \sim N(\mu_i, \sigma_i^2)$

$$\sum_{i=1}^{n} a_i X_i + c \sim N(\sum_{i=1}^{n} a_i\mu_i + c, \sum_{i=1}^{n} a_i^2\sigma_i^2)$$

1.2.2 χ^2 分布

1.χ^2 分布: 设 $X_1, X_2, \cdots, X_n$ 相互独立且服从标准正态分布，则

$$\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$$

服从自由度为 n 的 χ^2 分布，记为 $\chi^2(n)$，其密度函数为:

$$f(x) = \begin{cases} \frac{1}{2(\frac{n}{2})\Gamma(\frac{n}{2})} e^{-\frac{x}{2}} x^{\frac{n}{2}-1} & x > 0, \\ 0 & x \le 0 \end{cases}$$

2. 性质 1）可加性: 若 X$\sim \chi^2(n)$,Y$\sim \chi^2(m)$, 则 $X + Y \sim \chi^2(n+m)$
2)$E[\chi^2(n)] = n, D[\chi^2(n)] = 2n$

1.2.3 t 分布

1. 设 X∼N（0，1），Y$\sim \chi^2(n)$, 且 X 和 Y 相互独立，则

$$t = \frac{X}{\sqrt{Y/n}}$$

服从自由度为 n 的 t 分布，记为 t（n），其密度函数为

$$f(X) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1 + \frac{x^{2^{-\frac{n+1}{2}}}}{n})$$

当 n 充分大时，t 分布可以看作标准正态

1.2.4 F 分布

1.$X \sim \chi^2(n), Y \sim \chi^2(m)$, 且 X，Y 相互独立，则

$$F = \frac{X/n}{Y/m}$$

服从自由度为 n，m 的 F 分布，记为 F(n,m)
2. 性质: 1)$X \sim F(n,m)$, 则 $\frac{1}{X} \sim F(m,n)$
2) 若 $t \sim t(n)$, 则 $t^2 \sim F(1,n)$

1.3 正态总体的抽样分布

1. 目的: 总体为正态分布，进行抽样，求出抽样得到的样本的统计量服从什么分布

1.3.1 单正态总体的抽样分布定理

设总体 $X \sim N(\mu, \sigma^2)$, 简单随机样本 $X_1, X_2, \cdots, X_n$，样本均值 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, 样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$, 则
1)$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, 已知方差
2)$\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$, 且 $\overline{X}, S^2$ 相互独立
3)$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$，未知方差

1.3.2 双正态总体的抽样分布定理

设总体 $X\sim N(\mu_1,\sigma_1^2),Y\sim N(\mu_2,\sigma_2^2)$ 相互独立,简单随机样本 $X_1,X_2,\cdots,X_n;Y_1,Y_2,\cdots,Y_m$,则有

1)$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}\sim N(0,1)$;$\sigma$ 已知

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

3) 若 $\sigma_1^2=\sigma_2^2$，但未知大小

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_\omega\sqrt{\frac{1}{n}+\frac{1}{m}}}\sim t(n+m-2)$$

其中

$$S_\omega=\sqrt{\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}}$$

1.4 抽样分布的上 α 分位点

1.4.1 标准正态

设随机变量 $Z\sim N(0,1)$, 对 $\alpha\in(0,1)$，实数 z_α 满足

$$P\{Z>z_\alpha\}=\alpha$$

则称 z_α 是标准正态上的 α 分位点

note:1）α 是概率，z_α 是对应 x 的取值；2）$z_{1-\alpha}=-z_\alpha$

1.4.2 χ^2 分布

设随机变量 $\chi^2\sim\chi^2(n)$, 对 $\alpha\in(0,1)$，实数 $\chi^2_\alpha(n)$ 满足

$$P\{\chi^2>\chi^2_\alpha(n)\}=\alpha$$

则称 $\chi^2_\alpha(n)$ 是标准正态上的 α 分位点

$$P\{\chi^2(n)\le\chi^2_{1-\alpha}(n)\}=\alpha$$

note:χ^2 分布图像不对称

1.4.3 t 分布

设随机变量 $t\sim t(n)$, 对 $\alpha\in(0,1)$，实数 $t_\alpha(n)$ 满足

$$P\{t>t_\alpha(n)\}=\alpha$$

则称 $t_\alpha(n)$ 是标准正态上的 α 分位点

note:$t_{1-\alpha}=-t_\alpha$

1.4.4 F 分布

设随机变量 $F \sim F(n,m)$, 对 $\alpha \in (0,1)$，实数 $F_\alpha(n,m)$ 满足

$$P\{F > F_\alpha(n,m)\} = \alpha$$

则称 $F_\alpha(n,m)$ 是标准正态上的 α 分位点

note:F 分布图像不对称但是 $F_{1-\alpha}(n,m) = \frac{1}{F_\alpha(m,n)}$, 注意参数位置的调整

1.4.5 一般情况

设连续型随机变量 Y, 对 $\alpha \in (0,1)$，实数 Y_α 满足

$$P\{Y > Y_\alpha\} = \alpha$$

则称 $t_\alpha(n)$ 是标准正态上的 α 分位点

易得: 1)$P\{Y > Y_\alpha\} = \alpha$

2)$P\{Y < Y_{1-\alpha}\} = \alpha$

3)$P\{Y < Y_{1-\alpha/2}或Y > Y_{\alpha/2}\} = \alpha$

4)$P\{Y < Y_\alpha\} = 1-\alpha$

5)$P\{Y > Y_{1-\alpha}\} = 1-\alpha$

6)$P\{Y_{1-\alpha/2} < Y < Y_{\alpha/2}\} = 1-\alpha$