数理统计笔记

目录

1 样本与抽样分布

1.1 基本概念

定义 1.1.1（样本）：设随机变量 X 服从分布 F，若随机变量序列 $X_1, X_2, ..., X_n$ 具有同一分布 F 且相互独立，则称这一随机变量序列为从总体 F 或总体 X 得到的容量为 n 的样本，$x_1, x_2, ..., x_n$ 为 X 的 n 个独立观测值。

反之，若一随机变量序列是总体 F 的一个样本，则序列中的随机变量同分布为 F，且相互独立。

定义 1.1.2（经验分布函数）：有 样本 $x_1, x_2, ..., x_n$，用$S(x)$，$-\infty < z < \infty$ 表示 $x_1, x_2, ..., x_n$ 中不大于 x 的随机变量的个数，定义经验分布函数 $F(z)$ 为

$$F_{n(x)} = \frac{1}{n}S(x), \quad -\infty < x < \infty$$

1.1.1 统计量

定义 1.1.1.1（统计量与统计量的观测值）：若有一随机变量序列 $X_1, X_2, ..., X_n$ 是总体 F 的一个容量为 n 的样本，则称不含有位置参数的函数函数 $g(X_1, X_2, ..., X_n)$ 为统计量。

由定义可知，$g(X_1, X_2, ..., X_n)$ 也是一个随机变量，若有 $x_1, x_2, ..., x_n$ 是样本的观测值，则 $g(x_1, x_2, ..., x_n)$ 是随机变量 $g(X_1, X_2, ..., X_n)$ 的观测值。

有总体 $X, E(X) = \mu, D(X) = \sigma^2$，下方为常见的统计量:

定义 1.1.1.2（样本平均值）：$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

根据定义可得 $E\left(\overline{X}\right) = n\mu, D\left(\overline{X}\right) = \frac{\sigma^2}{n}$

定义 1.1.1.3（样本方差）：$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$

根据定义可得，$E(S^2) = D(X) = \sigma^2$

定义 1.1.1.4（样本标准差）：$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$

定义 1.1.1.5（样本 k 阶原点矩）：$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$，$\quad k = 1, 2, 3, ...$

定义 1.1.1.6 （样本 k 阶中心矩）：$B_k = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^k, \quad k = 2, 3, ...$

1.2 抽样分布

抽样分布即为统计量为 $g(X_1, X_2, ..., X_n)$ 的分布，如 χ^2 分布。

1.2.1 χ^2 分布

定义 1.2.1.1 （ χ^2 分布）：设样本 $X_1, X_2, ..., X_n$ 相互独立，且均服从 $N(0,1)$ 分布，则有 $X = X_1^2 + X_2^2 + ... + X_n^2$ 服从自由度为 n 的 χ^2 分布，即 $X \sim \chi^2(n)$。

χ^2 分布有如下几条性质:

性质:

1. 可加性

若 $X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$ 则
$X + Y \sim \chi^2(n_1 + n_2)$.

2. 均值与方差

若 $X \sim \chi^2(n)$，则 $E(X) = n, D(X) = 2n$.

3. 上 α 分位点

在 χ^2 分布的密度图形中，当 $x = x_\alpha$ 时，$x > x_\alpha$ 的面积为 α，称此点为上 α 分位点。
此时有 $P\{X > x_\alpha\} = \alpha$.

1.2.2 t 分布

定义 1.2.2.1 （t 分布）：若有 $X \sim N(0,1), Y \sim \chi^2(n)$ 且相互独立，则

$$\frac{X}{\sqrt{\frac{Y}{n}}} = t \sim t(n)$$

1.2.3 F 分布

定义 1.2.3.1 （F 分布）：若有 $X_1 \sim \chi^2(n_1), X_2 \sim \chi^2(n_2)$ 且相互独立，则

$$\frac{\frac{X_1}{n_1}}{\frac{X_2}{n_2}} = F \sim F(n_1, n_2)$$

5. 正态总体的样本均值与样本方差的分布.

(1) 设总体 $X \sim N(\mu, \sigma^2)$，$X_1, X_2, \cdots, X_n$ 为总体 X 的一个样本，则:

① $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ ； ② $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ ； ③ $\bar{X}$ 与 S^2 独立.

定理四 设 $X_1, X_2, \cdots, X_{n_1}$ 与 $Y_1, Y_2, \cdots, Y_{n_2}$ 分别是来自正态总体 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$ 的样本，且这两个样本相互独立[①]. 设 $\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$，$\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$ 分别是这两个样本的样本均值；$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2$，$S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$ 分别是这两个样本的样本方差，则有

1° $\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$；

2° 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 时，

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2),$$

其中

$$S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}, \quad S_w = \sqrt{S_w^2}.$$