

第六章 数理统计的基本概念

随机样本与统计量

随机样本: 从总体中随机地取 n 个个体, 称为一个随机样本

简单随机样本: 满足以下两个条件的随机样本 $(X_1, X_2, \dots, X_n)$

- 1) 每个 X_i 与 X 同分布
- 2) $X_1, X_2, \dots, X_n$ 是相互独立的随机变量

若总体 X 具有概率密度 $f(x)$, 则样本 $(X_1, X_2, \dots, X_n)$ 具有联合密度函数:

$$f_n(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i)$$

统计量: 样本的不含任何未知参数的函数

常用统计量: 设 $(X_1, X_2, \dots, X_n)$ 为取自总体 X 的样本, 以下为常用统计量

- 1) 样本均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, 则有 $E(\bar{X}) = E(X)$, $D(\bar{X}) = \frac{1}{n} D(X)$
- 2) 样本方差 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, $E(S^2) = D(X)$
- 3) 样本矩: $\begin{cases} k \text{ 阶矩: } A_k = \frac{1}{n} \sum_{i=1}^n X_i^k \quad (k=1, 2, \dots) \\ k \text{ 阶中心矩: } B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \quad (k=1, 2, \dots) \end{cases}$

当总体数字特征未知时

- 一般, 用样本均值 $\bar{X}$ 作为总体均值 $\mu = E(X)$ 的估计;
- 用样本方差 S^2 作为总体方差 $\sigma^2 = E(X - \mu)^2$ 的估计;
- 用样本的原点矩 A_k 作为总体原点矩 $\mu_k = E(X^k)$ 的估计;
- 用样本的中心矩 B_k 作为总体中心矩 $\nu_k = E(X - \mu)^k$ 的估计.
(假设总体各阶矩存在)
- 总体方差的估计可以用 S^2 也可以用 B_2 , 主要的区别是 S^2 作为总体方差估计是无偏估计, 但 B_2 作为总体方差的估计是有偏的

χ^2 -分布; t -分布; F -分布

统计量的分布称为抽样分布

χ^2 分布: 设随机变量 $X_1, X_2, \dots, X_n$ 相互独立, $X_i \sim N(0, 1)$ ($i=1, 2, \dots, n$)

则称 $\chi_n^2 = \sum_{i=1}^n X_i^2$

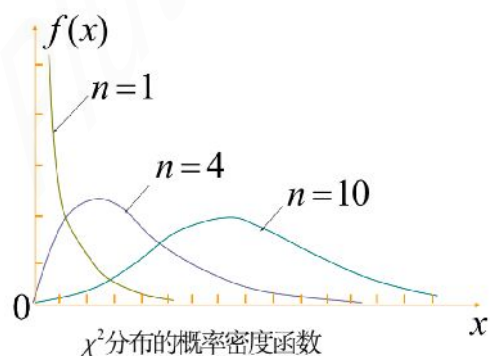
服从自由度为 n 的 χ^2 分布, 记为 $\chi^2 \sim \chi^2(n)$

$\chi^2(n)$ 分布的概率密度为 $f_n(y) = \begin{cases} \frac{1}{2\Gamma(\frac{n}{2})} (\frac{y}{2})^{\frac{n}{2}-1} e^{-\frac{y}{2}}, & y > 0 \\ 0, & y \leq 0 \end{cases}$ 其中 $\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx$

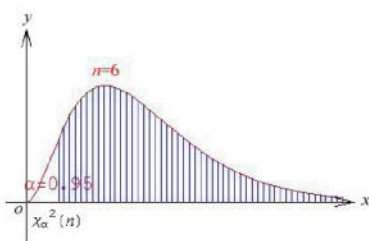
χ^2 分布的性质

- 1) 设 $\chi^2 \sim \chi^2(n)$, 则有 $E(\chi^2) = n$, $D(\chi^2) = 2n$
- 2) 设 $Y_1 \sim \chi^2(n_1)$, $Y_2 \sim \chi^2(n_2)$, 且 Y_1, Y_2 相互独立, 则有 $Y_1 + Y_2 \sim \chi^2(n_1 + n_2)$

$$\begin{cases} \Gamma(a+1) = a\Gamma(a) \\ \Gamma(n+1) = n!, \Gamma(\frac{1}{2}) = \sqrt{\pi} \end{cases}$$



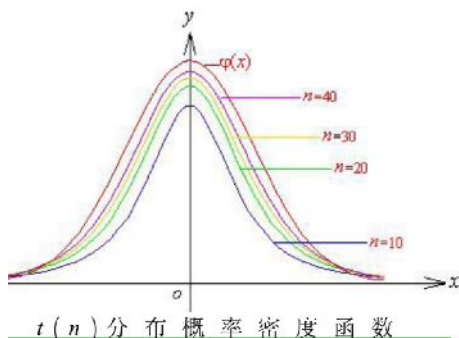
对给定的概率 $\alpha, 0 < \alpha < 1$, 称满足条件 $\int_{\chi_{\alpha}^2(n)}^{\infty} f_n(y) dy = \alpha$ 的点 $\chi_{\alpha}^2(n)$ 为 $\chi^2(n)$ 分布的上 α 分位数, 上 α 分位数 $\chi_{\alpha}^2(n)$ 的值可查 χ^2 分布表



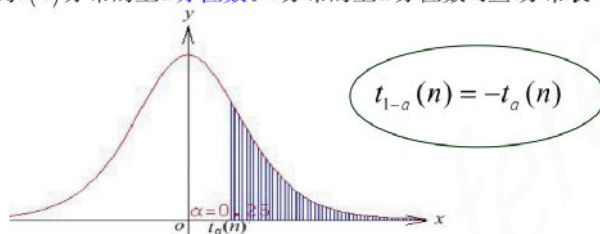
t-分布: 设 $X \sim N(0,1)$, $Y \sim \chi^2(n)$, 且 X, Y 相互独立, 则称随机变量 $T = \frac{X}{\sqrt{\frac{Y}{n}}}$ 服从自由度为 n 的 t 分布, 记作 $T \sim t(n)$.

$t(n)$ 分布的概率密度为:

$$f(t, n) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < t < +\infty$$



对给定的 α , $0 < \alpha < 1$, 称满足条件 $\int_{t_\alpha(n)}^{\infty} f(t, n) dt = \alpha$ 的点 $t_\alpha(n)$ 为 $t(n)$ 分布的上 α 分位数. t 分布的上 α 分位数可查 t 分布表



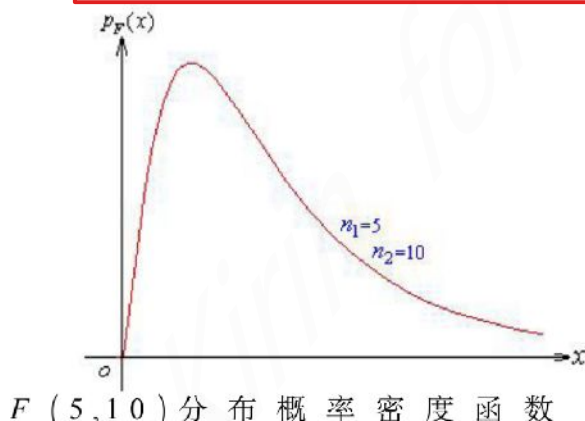
F分布: 设 $X \sim \chi^2(n_1)$, $Y \sim \chi^2(n_2)$, 且 X, Y 独立, 则称随机变量 $F = \frac{X/n_1}{Y/n_2}$ 服从自由度 (n_1, n_2) 的 F 分布, 记作 $F \sim F(n_1, n_2)$

若 $F \sim F(n_1, n_2)$, 则 $F^{-1} \sim F(n_2, n_1)$

$F(n_1, n_2)$ 分布的概率密度为

$$f(x, n_1, n_2) = \begin{cases} \frac{1}{B(\frac{n_1}{2}, \frac{n_2}{2})} n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}} x^{\frac{n_1}{2}-1} (n_2 + n_1 x)^{-\frac{n_1+n_2}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

其中 $B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a+b)}$

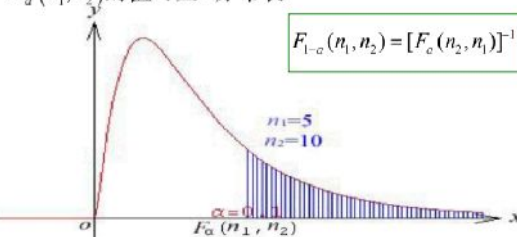


对于给定的 α , $0 < \alpha < 1$, 称满足条件

$$\int_{F_\alpha(n_1, n_2)}^{\infty} f(x; n_1, n_2) dx = \alpha$$

的点 $F_\alpha(n_1, n_2)$ 为 $F(n_1, n_2)$ 分布的上 α 分位数.

$F_\alpha(n_1, n_2)$ 的值可查 F 分布表



正态总体下的抽样分布

设 $x_1, x_2, \dots, x_n$ 为来自正态总体 $N(\mu, \sigma^2)$ 的简单随机样本, $\bar{x}$ 是样本均值, S^2 是样本方差,

则有

- 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 相互独立
- 4) $\frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$

设样本 $(X_1, \dots, X_{n_1})$ 和 $(Y_1, \dots, Y_{n_2})$ 分别来自总体 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$ 且其相互独立，其样本方差分别为 S_1^2, S_2^2 ，则有

$$(1) \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} = \frac{\left(\frac{S_1}{\sigma_1}\right)^2}{\left(\frac{S_2}{\sigma_2}\right)^2} = F(n_1-1, n_2-1)$$

$$(2) \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{\sigma_1}{n_1}\right)^2 + \left(\frac{\sigma_2}{n_2}\right)^2}} \sim N(0, 1)$$

(3) 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 时

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

$$\text{其中 } S_W^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$