

当 $n \rightarrow \infty$ 时, 随机变量 $T \sim t(n)$ 的概率密度

$$f(x) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

因此当 n 足够大时, $f(x)$ 可被近似为 $\mathcal{N}(0, 1)$ 的密度函数.

8.4.3 F 分布

定义 8.11 设随机变量 $X \sim \chi^2(m)$ 和 $Y \sim \chi^2(n)$ 相互独立, 称随机变量

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

服从自由度为 (m, n) 的 F -分布, 记 $F \sim F(m, n)$.

随机变量 $F \sim F(m, n)$ 的概率密度为

$$f(x) = \begin{cases} \frac{\Gamma(\frac{m+n}{2}) (\frac{m}{n})^{\frac{m}{2}} x^{\frac{m}{2}-1}}{\Gamma(\frac{m}{2}) \Gamma(\frac{n}{2}) (1 + \frac{mx}{n})^{\frac{m+n}{2}}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

若随机变量 $F \sim F(m, n)$, 则 $1/F \sim F(n, m)$.

课题练习:

- 独立同分布随机变量 X_1, X_2, \dots, X_n 满足 $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, 求 $\sum_{i=1}^n (X_i - \mu_i)^2 / \sigma_i^2$ 的分布.
- 设 X_1, X_2, \dots, X_9 和 Y_1, Y_2, \dots, Y_9 是分别来自总体 $\mathcal{N}(0, 9)$ 的两个独立样本, 求 $(X_1 + X_2 + \dots + X_9) / \sqrt{Y_1^2 + Y_1^2 + \dots + Y_9^2}$ 的分布.
- 设 X_1, X_2, \dots, X_{2n} 来自总体 $\mathcal{N}(0, \sigma_2)$ 的样本, 求 $(X_1^2 + X_3^2 + \dots + X_{2n-1}^2) / (X_2^2 + X_4^2 + \dots + X_{2n}^2)$ 的分布.

8.5 正态分布的抽样定理

定理 8.5 若 X_1, X_2, \dots, X_n 是来自总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的一个样本, 则有

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}) \quad \text{和} \quad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

定理 8.6 若 X_1, X_2, \dots, X_n 是来自总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的一个样本, 其样本均值和修正样本方差分别为

$$\bar{X} = \sum_{i=1}^n X_i / n \quad \text{和} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

则有 \bar{X} 和 S^2 相互独立, 且

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

该定理的证明超出了本书的范畴. 该定理是其他抽样定理的基础.

定理 8.7 若 X_1, X_2, \dots, X_n 是来自总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的一个样本, 其样本均值为 \bar{X} , 修正样本方差为 S^2 , 则有

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

证明 根据正态分布的性质有 $(\bar{X} - \mu)/\sigma\sqrt{n} \sim \mathcal{N}(0, 1)$, 再根据定理 8.6 可知 $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$, 于是有

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \bigg/ \sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}} \sim t(n-1).$$

定理 8.8 若 X_1, X_2, \dots, X_m 和 Y_1, Y_2, \dots, Y_n 是分别来自总体 $X \sim \mathcal{N}(\mu_X, \sigma^2)$ 和 $Y \sim \mathcal{N}(\mu_Y, \sigma^2)$ 的两个样本, 其样本均值分别 \bar{X} 和 \bar{Y} , 修正样本方差分别为 S_X^2 和 S_Y^2 , 则有

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}} \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2).$$

证明 根据正态分布的性质有 $\bar{X} \sim \mathcal{N}(\mu_X, \sigma^2/m)$ 和 $\bar{Y} \sim \mathcal{N}(\mu_Y, \sigma^2/n)$, 以及

$$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_X - \mu_Y, \left(\frac{1}{m} + \frac{1}{n}\right)\sigma^2\right),$$

进一步有

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim \mathcal{N}(0, 1).$$

根据定理 8.6 有 $\frac{(m-1)S_X^2}{\sigma^2} \sim \chi^2(m-1)$ 和 $\frac{(n-1)S_Y^2}{\sigma^2} \sim \chi^2(n-1)$, 由此得到

$$\frac{(m-1)S_X^2 + (n-1)S_Y^2}{\sigma^2} \sim \chi^2(m+n-2).$$

从而完成证明.

定理 8.9 若 X_1, X_2, \dots, X_m 和 Y_1, Y_2, \dots, Y_n 是分别来自总体 $X \sim \mathcal{N}(\mu_X, \sigma^2)$ 和 $Y \sim \mathcal{N}(\mu_Y, \sigma^2)$ 的两个样本, 其修正样本方差分别为 S_X^2 和 S_Y^2 , 则有

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F(m-1, n-1).$$

证明 根据定理 8.6 有 $\frac{(m-1)S_X^2}{\sigma_X^2} \sim \chi^2(m-1)$ 和 $\frac{(n-1)S_Y^2}{\sigma_Y^2} \sim \chi^2(n-1)$, 由此得到

$$\frac{\frac{(m-1)S_X^2}{\sigma_X^2}/(m-1)}{\frac{(n-1)S_Y^2}{\sigma_Y^2}/(n-1)} \sim F(m-1, n-1).$$

8.5.1 分位数(点)

定义 8.12 对给定的 $\alpha \in (0, 1)$ 和随机变量 X , 满足

$$P(X > \lambda_\alpha) = \alpha$$

的实数 λ_α 称为上侧 α 分位数(点).

对正态分布 $X \sim \mathcal{N}(0, 1)$, 给定 $\alpha \in (0, 1)$, 满足 $P(X > \mu_\alpha) = \int_{\mu_\alpha}^{\infty} f(x)dx = \alpha$ 的点 μ_α 称为正态分布上侧 α 分位点, 由对称性可知 $\mu_{1-\alpha} = -\mu_\alpha$.

对 $\chi^2(n)$ 分布 $X \sim \chi^2(n)$, 给定 $\alpha \in (0, 1)$, 满足 $P(X \geq \chi_\alpha^2(n)) = \alpha$ 的点 $\chi_\alpha^2(n)$ 称为 $\chi^2(n)$ 分布上侧 α 分位点. 当 $n \rightarrow \infty$ 时有 $\chi_\alpha^2(n) \approx \frac{1}{2}(\mu_\alpha + \sqrt{2n-1})^2$, 其中 μ_α 表示正态分布上侧 α 分位点.

对 t -分布 $X \sim t(n)$, 给定 $\alpha \in (0, 1)$, 满足 $P(X > t_\alpha(n)) = \alpha$ 的点 $t_\alpha(n)$ 称为 $t(n)$ -分布上侧 α 分位点. 由对称性可知 $t_{(1-\alpha)}(n) = -t_\alpha(n)$.

对 F -分布 $X \sim F(m, n)$, 给定 $\alpha \in (0, 1)$, 满足 $P[X > F_\alpha(m, n)] = \alpha$ 的点 $F_\alpha(m, n)$ 称为 $F(m, n)$ 分布上侧 α 分位点.

对于 F -分布, 有如下性质:

引理 8.1 对 F 分布的分位点有

$$F_{(1-\alpha)}(m, n) = \frac{1}{F_\alpha(n, m)}.$$

证明 设 $X \sim F(m, n)$, 根据定义有

$$1 - \alpha = P(X > F_{1-\alpha}(m, n)) = P\left(\frac{1}{X} < \frac{1}{F_{1-\alpha}(m, n)}\right) = 1 - P\left(\frac{1}{X} \geq \frac{1}{F_{1-\alpha}(m, n)}\right).$$

再根据 $1/X \sim F(n, m)$, 结合上式有

$$\alpha = P\left(\frac{1}{X} \geq \frac{1}{F_{1-\alpha}(m, n)}\right) = P\left(\frac{1}{X} > \frac{1}{F_{1-\alpha}(m, n)}\right)$$

于是有 $F_\alpha(n, m) = 1/F_{1-\alpha}(m, n)$.

课堂习题:

- 若随机变量 $X \sim t(n)$, 求 $Y = X^2$ 的分布.
- 设 X_1, X_2, \dots, X_5 是来自总体 $\mathcal{N}(0, 1)$ 的样本, 令 $Y = c_1(X_1 + X_3)^2 + c_2(X_2 + X_4 + X_5)^2$. 求常数 c_1, c_2 使 Y 服从 χ^2 分布.
- 设 X_1, X_2 是来自总体 $\mathcal{N}(0, \sigma^2)$ 的样本, 求 $\frac{(X_1 + X_2)^2}{(X_1 - X_2)^2}$ 的分布.
- 设 X_1, X_2, \dots, X_{10} 是总体 $\mathcal{N}(\mu, 1/4)$ 的样本, i) 若 $\mu = 0$, 求 $P(\sum_{i=1}^{10} X_i^2 \geq 4)$; ii) 若 μ 未知, 求 $P(\sum_{i=1}^{10} (X_i - \bar{X})^2 \geq 2.85)$.
- 设 X_1, X_2, \dots, X_{25} 是总体 $\mathcal{N}(12, \sigma^2)$ 的样本, i) 若 $\sigma = 2$, 求 $P(\sum_{i=1}^{25} X_i/25 \geq 12.5)$; ii) 若 σ 未知但知道修正样本方差为 $S^2 = 5.57$, 求 $P(\sum_{i=1}^{25} X_i/25 \geq 12.5)$.