

Lec-18. 数理统计介绍、随机样本、统计量

主讲教师：吴利苏 (wulisu@sdust.edu.cn)

主 页：wulisu.cn

本节课内容

t 分布

F 分布

正态总体的抽样分布

2. t 分布

定义

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

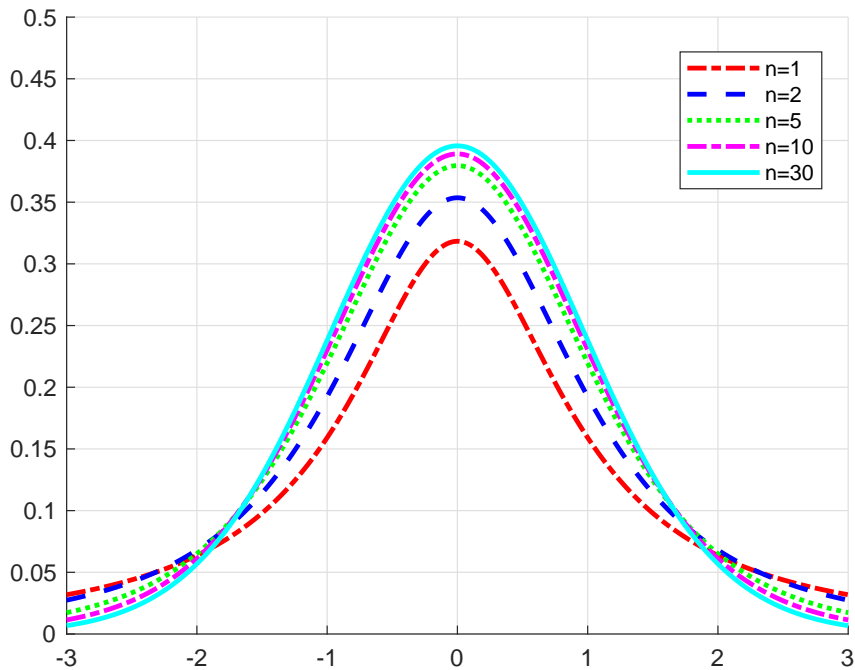
$$t = \frac{X}{\sqrt{Y/n}}$$

为服从自由度为 n 的 t 分布, 记为 $t \sim t(n)$.

概率密度和图像

t 分布又称学生氏分布, 概率密度为

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



性质

1. 关于 $t = 0$ 对称.

2.

$$\lim_{n \rightarrow \infty} h(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

当 n 充分大, t 分布近似 $N(0, 1)$; 当 n 较小时, t 分布与 $N(0, 1)$ 相差很大.

性质

- t 分布的上分位数 对于给定的 α , 满足

$$P\{t > t_{\alpha}(n)\} = \int_{t_{\alpha}(n)}^{\infty} h(t) dt = \alpha$$

的 $t_{\alpha}(n)$ 是 $t(n)$ 分布的上 α 分位数.

性质

- t 分布的上分位数 对于给定的 α , 满足

$$P\{t > t_{\alpha}(n)\} = \int_{t_{\alpha}(n)}^{\infty} h(t) dt = \alpha$$

的 $t_{\alpha}(n)$ 是 $t(n)$ 分布的上 α 分位数.

- $t_{1-\alpha}(n) = -t_{\alpha}(n)$.

分布函数的角度: $1 - F(x) = F(-x)$.

性质

- t 分布的上分位数 对于给定的 α , 满足

$$P\{t > t_{\alpha}(n)\} = \int_{t_{\alpha}(n)}^{\infty} h(t) dt = \alpha$$

的 $t_{\alpha}(n)$ 是 $t(n)$ 分布的上 α 分位数.

- $t_{1-\alpha}(n) = -t_{\alpha}(n)$.

分布函数的角度: $1 - F(x) = F(-x)$.

- 当 $n \leq 45$ 时, 查表 Page-399, 求 $t_{\alpha}(n)$.

性质

- t 分布的上分位数 对于给定的 α , 满足

$$P\{t > t_{\alpha}(n)\} = \int_{t_{\alpha}(n)}^{\infty} h(t) dt = \alpha$$

的 $t_{\alpha}(n)$ 是 $t(n)$ 分布的上 α 分位数.

- $t_{1-\alpha}(n) = -t_{\alpha}(n)$.

分布函数的角度: $1 - F(x) = F(-x)$.

- 当 $n \leq 45$ 时, 查表 Page-399, 求 $t_{\alpha}(n)$.
- 当 $n > 45$ 时, $t_{\alpha}(n) \approx z_{\alpha}$.

例

设 $T \sim t(n)$, $t(n)$ 的上 α 分位数满足

$$P\{T > t_{\alpha}(n)\} = \int_{t_{\alpha}(n)}^{\infty} h(y) dy = \alpha,$$

求 $t_{0.05}(10)$, $t_{0.025}(15)$ 的值.

例

设 $T \sim t(n)$, $t(n)$ 的上 α 分位数满足

$$P\{T > t_{\alpha}(n)\} = \int_{t_{\alpha}(n)}^{\infty} h(y) dy = \alpha,$$

求 $t_{0.05}(10)$, $t_{0.025}(15)$ 的值.

解: $t_{0.05}(10) = 1.8125$, $t_{0.025}(15) = 2.1315$. □

3. F 分布

定义

设 $U \sim \chi^2(n_1)$, $V \sim \chi^2(n_2)$ 且 U, V 相互独立, 则称

$$F = \frac{U/n_1}{V/n_2}$$

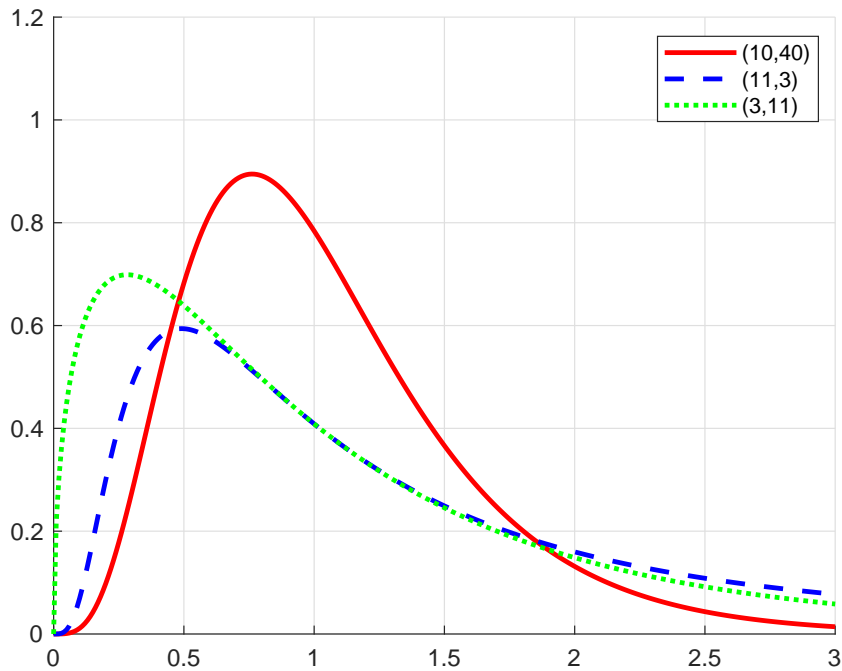
服从自由度为 (n_1, n_2) 的 F 分布, 记为
 $F \sim F(n_1, n_2)$.

3. F 分布的概率密度和图像

概率密度

$$\begin{aligned}\phi(y) &= \begin{cases} \frac{\Gamma(\frac{n_1+n_2}{2}) (\frac{n_1}{n_2})^{\frac{n_1}{2}} y^{\frac{n_1}{2}-1}}{\Gamma(\frac{n_1}{2}) \Gamma(\frac{n_2}{2}) (1+\frac{n_1}{n_2}y)^{\frac{n_1+n_2}{2}}} & y > 0; \\ 0 & \text{其他.} \end{cases} \\ &= \begin{cases} \frac{n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}} y^{\frac{n_1}{2}-1}}{B(\frac{n_1}{2}, \frac{n_2}{2}) (n_2+n_1y)^{\frac{n_1+n_2}{2}}} & y > 0; \\ 0 & \text{其他.} \end{cases}\end{aligned}$$

其中 Beta 函数 $B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.



3. F 分布的性质

- 若 $F \sim F(n_1, n_2)$, 则 $\frac{1}{F} \sim F(n_2, n_1)$.
- 上分位数 对于给定的 α , $0 < \alpha < 1$, 满足条件

$$P\{F > F_\alpha(n_1, n_2)\} = \int_{F_\alpha(n_1, n_2)}^{\infty} \phi(y) dy = \alpha$$

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

求 F 分布的上 α 分位数

- 查表 Page401-404

$$F_{0.025}(8, 7) = 4.90, F_{0.05}(30, 14) = 2.31.$$

求 F 分布的上 α 分位数

- 查表 Page401-404

$$F_{0.025}(8, 7) = 4.90, F_{0.05}(30, 14) = 2.31.$$

- F 分布的上 α 分位数满足

$$F_{1-\alpha}(n_1, n_2) = \frac{1}{F_{\alpha}(n_2, n_1)}. \quad (1)$$

求 F 分布的上 α 分位数

- 查表 Page401-404

$$F_{0.025}(8, 7) = 4.90, F_{0.05}(30, 14) = 2.31.$$

- F 分布的上 α 分位数满足

$$F_{1-\alpha}(n_1, n_2) = \frac{1}{F_{\alpha}(n_2, n_1)}. \quad (1)$$

- 利用上式，可以求分布表中未列出的常用的上 α 分位数.

例如： $F_{0.95}(12, 9) = 0.357.$

(1) 式的证明: $F \sim F(n_1, n_2)$

$$\begin{aligned} 1 - \alpha &= P\{F > F_{1-\alpha}(n_1, n_2)\} \\ &= P\left\{\frac{1}{F} < \frac{1}{F_{1-\alpha}(n_1, n_2)}\right\} \\ &= 1 - P\left\{\frac{1}{F} \geq \frac{1}{F_{1-\alpha}(n_1, n_2)}\right\} \\ &= 1 - P\left\{\frac{1}{F} > \frac{1}{F_{1-\alpha}(n_1, n_2)}\right\} \end{aligned}$$

$\frac{1}{F} \sim F(n_2, n_1)$, 则

$$P\left\{\frac{1}{F} > \frac{1}{F_{1-\alpha}(n_1, n_2)}\right\} = \alpha = P\left\{\frac{1}{F} > \frac{1}{F_{1-\alpha}(n_1, n_2)}\right\},$$

故 $F_{1-\alpha}(n_1, n_2) = \frac{1}{F_{\alpha}(n_2, n_1)}$.

例

X, Y, Z 相互独立, 服从 $N(0, 1)$, 则

(1) $X^2 + Y^2 + Z^2 \sim \chi^2(3),$

(2) $\frac{X}{\sqrt{(Y^2+Z^2)/2}} \sim t(2),$

(3) $\frac{2X^2}{Y^2+Z^2} \sim F(1, 2).$

例

若 $t \sim t(n)$, 则 $t^2 \sim F(1, n).$

4. 正态总体的抽样分布

- 设总体 X 的期望、方差存在,

$$E(X) = \mu, \quad D(X) = \sigma^2.$$

- 设 X_1, \dots, X_n 是来自总体 X 的样本,

$$\bar{X} = \frac{1}{n} \sum X_i, \quad S^2 = \frac{1}{n-1} \left(\sum X_i^2 - n\bar{X}^2 \right)$$

分别是样本均值和样本方差, 则有

$$E(\bar{X}) = \mu, \quad D(\bar{X}) = \frac{\sigma^2}{n}.$$

4. 正态总体的抽样分布

$$\begin{aligned} E(S^2) &= E\left(\frac{1}{n-1} \left(\sum X_i^2 - n\bar{X}^2\right)\right) \\ &= \frac{1}{n-1} \left(\sum E(X_i^2) - nE(\bar{X}^2)\right) \\ &= \frac{1}{n-1} \left(\sum (\sigma^2 + \mu^2) - n(\sigma^2/n + \mu^2)\right) \\ &= \sigma^2. \end{aligned}$$

$$D(S^2) = D\left(\frac{1}{n-1} \left(\sum X_i^2 - n\bar{X}^2\right)\right) = ?.$$

4. 正态总体的抽样分布

$$\begin{aligned} E(S^2) &= E\left(\frac{1}{n-1} \left(\sum X_i^2 - n\bar{X}^2\right)\right) \\ &= \frac{1}{n-1} \left(\sum E(X_i^2) - nE(\bar{X}^2)\right) \\ &= \frac{1}{n-1} \left(\sum (\sigma^2 + \mu^2) - n(\sigma^2/n + \mu^2)\right) \\ &= \sigma^2. \end{aligned}$$

$$D(S^2) = D\left(\frac{1}{n-1} \left(\sum X_i^2 - n\bar{X}^2\right)\right) = ?.$$

一般不能由 μ, σ^2 表示.

定理 (正态总体的抽样分布)

设总体 $X \sim N(\mu, \sigma^2)$, X_1, \dots, X_n 是样本, 样本均值和样本方差分别为

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

则

(1) $\bar{X} \sim N(\mu, \sigma^2/n).$

(2) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, 且 \bar{X} 与 S^2 相互独立.

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

(2) 的证明见本章末二维码.

例

设总体 $X \sim N(\mu, \sigma^2)$, X_1, \dots, X_n 是样本, 则

$$(1) \quad \frac{(n-1)S^2}{\sigma^2} = \frac{\sum (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1).$$

$$(2) \quad \frac{\sum (X_i - \mu)^2}{\sigma^2} = \sum \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n).$$

其原因是在 (1) 式中有一个约束条件

$$\sum_{i=1}^n (X_i - \bar{X}) = 0.$$

(3) $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ 的证明:

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1), \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

相互独立, 由 t 分布的定义

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

□

例

设总体 $X \sim N(\mu, \sigma^2)$, X_1, \dots, X_n 是样本, 样本方差为 $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$. 求 $D(S^2)$.

解: $\because \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$

$$\therefore D\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1).$$

$$\therefore D(S^2) = \frac{2\sigma^4}{n-1}.$$

□

- 随 n 增大, $D(S^2)$ 减小, 样本方差的方差变小.
所以, 可用样本方差 S^2 推断总体方差.

定理 (两个正态总体的抽样分布)

设 X_1, \dots, X_{n_1} 与 Y_1, \dots, Y_{n_2} 分别是正态总体 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$ 的样本, 且相互独立. 样本均值分别为 \bar{X}, \bar{Y} ; 样本方差分别为 S_1^2, S_2^2 . 则

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

证明 (1):

$$\chi_1^2 = \frac{(n_1 - 1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1),$$

$$\chi_2^2 = \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1)$$

S_1^2, S_2^2 独立.

\therefore

$$F = \frac{S_1^2}{S_2^2} \bigg/ \frac{\sigma_1^2}{\sigma_2^2} = \frac{\chi_1^2/(n_1 - 1)}{\chi_2^2/(n_2 - 1)} \sim F(n_1 - 1, n_2 - 1).$$

证明 (2): 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 时,

$$\begin{aligned}\bar{X} - \bar{Y} &\sim N\left(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_1}\right) \\ U &= \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)\end{aligned}$$

又

$$\frac{(n_1 - 1)S_1^2}{\sigma^2} \sim \chi^2(n_1 - 1), \frac{(n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2(n_2 - 1)$$

且相互独立, 故有 χ^2 分布的可加性,

$$V = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2).$$

U, V 相互独立. 由 t 分布的定义,

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

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

U, V 相互独立的证明见本章末二维码.

思考: 若 σ^2 未知, 为什么用 S_w^2 来估计 σ^2 , 而不用 S_1^2 或 S_2^2 来估计 σ^2 呢?

思考: 若 σ^2 未知, 为什么用 S_w^2 来估计 σ^2 , 而不用 S_1^2 或 S_2^2 来估计 σ^2 呢?

- $E(S_1^2) = E(S_2^2) = E(S_w^2) = \sigma^2$;
- $D(S_1^2) = \frac{2\sigma^4}{n_1-1}, D(S_2^2) = \frac{2\sigma^4}{n_2-1},$

$$D(S_w^2) = \frac{2\sigma^4}{n_1 + n_2 - 2}$$

S_w^2 包含更多信息, 具有更小的方差.

- 对于单个正态总体 $N(\mu, \sigma^2)$, 得到了 \bar{X}, S^2 的分布, 用于对 μ, σ^2 进行推断 (区间估计, 假设检验).
- 对于两个独立正态总体 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$, 得到了 $\bar{X} - \bar{Y}, S_1^2/S_2^2$ 的分布, 用于对 $\mu_1 - \mu_2, \frac{\sigma_1^2}{\sigma_2^2}$ 进行推断.

定理 3 的证明

令 $Z_i = \frac{X_i - \mu}{\sigma}$, $i = 1, 2, \dots, n$, 则由定理 3 的假设知,
 Z_1, Z_2, \dots, Z_n 相互独立, 且都服从 $N(0,1)$ 分布, 而

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i = \frac{\bar{X} - \mu}{\sigma};$$

$$\begin{aligned} \frac{(n-1)S^2}{\sigma^2} &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} = \sum_{i=1}^n \left[\frac{(X_i - \mu) - (\bar{X} - \mu)}{\sigma} \right]^2 \\ &= \sum_i^n (Z_i - \bar{Z})^2 = \sum_i^n Z_i^2 - n\bar{Z}^2. \end{aligned}$$

取一 n 阶正交矩阵 $A = (a_{ij})$ ，其中第一行的元素均为 $1/\sqrt{n}$.
作正交变换

$$\mathbf{Y} = \mathbf{AZ},$$

其中

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix}.$$

由于 $Y_i = \sum_{j=1}^n a_{ij}Z_j, i = 1, 2, \dots, n$. 故 Y_1, Y_2, \dots, Y_n 仍为正态变量, 由 $Z_i \sim N(0, 1), i=1, 2, \dots, n$ 知

$$E(Y_i) = E\left(\sum_{j=1}^n a_{ij}Z_j\right) = \sum_{j=1}^n a_{ij}E(Z_j) = 0.$$

又由 $\text{Cov}(Z_i, Z_j) = \delta_{ij}$, $i, j = 1, 2, \dots, n$, 知

$$\begin{aligned}\text{Cov}(Y_i, Y_k) &= \text{Cov}\left(\sum_{j=1}^n a_{ij}Z_j, \sum_{l=1}^n a_{kl}Z_l\right) \\ &= \sum_{j=1}^n \sum_{l=1}^n a_{ij}a_{kl} \text{Cov}(Z_j, Z_l) = \sum_{j=1}^n a_{ij}a_{kj} = \delta_{ik}\end{aligned}$$

(由正交矩阵的性质), 故 Y_1, Y_2, \dots, Y_n 两两不相关. 又由于 n 维随机变量 (Y_1, Y_2, \dots, Y_n) 是由 n 维正态随机变量 (X_1, X_2, \dots, X_n) 经由线性变换而得到的, 因此 (Y_1, Y_2, \dots, Y_n) 也是 n 维正态随机变量 (参见第 4 章 §4).

于是由 Y_1, Y_2, \dots, Y_n 两两不相关可推得 Y_1, Y_2, \dots, Y_n 相互独立 (参见第 4 章 § 4), 且有 $Y_i \sim N(0, 1), i=1, 2, \dots, n$. 而

$$Y_1 = \sum_{j=1}^n a_{1j} Z_j = \sum_{j=1}^n \frac{1}{\sqrt{n}} Z_j = \sqrt{n} \bar{Z};$$

$$\begin{aligned} \sum_{i=1}^n Y_i^2 &= Y^T Y = (A\mathbf{Z})^T (A\mathbf{Z}) = \mathbf{Z}^T (\mathbf{A}^T \mathbf{A}) \mathbf{Z} \\ &= \mathbf{Z}^T \mathbf{I} \mathbf{Z} = \mathbf{Z}^T \mathbf{Z} = \sum_{i=1}^n Z_i^2, \end{aligned}$$

于是

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n Z_i^2 - n\bar{Z}^2 = \sum_{i=1}^n Y_i^2 - Y_1^2 = \sum_{i=2}^n Y_i^2.$$

由于 Y_2, Y_3, \dots, Y_n 相互独立,

且 $Y_i \sim N(0, 1), i = 2, 3, \dots, n$, 知 $\sum_{i=2}^n Y_i^2 \sim \chi^2(n-1)$. 从而证得

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

再者, $\bar{X} = \sigma \bar{Z} + \mu = \frac{\sigma Y_1}{\sqrt{n}} + \mu$ 仅依赖于 Y_1 ,

而 $S^2 = \frac{\sigma^2}{n-1} \sum_{i=2}^n Y_i^2$ 仅依赖于 Y_2, Y_3, \dots, Y_n .

再由 Y_1, Y_2, \dots, Y_n 的独立性, 推知 \bar{X} 与 S^2 相互独立.

定理 3 的推广

定理 3 中 \bar{X} 与 S^2 相互独立这一结论, 还能推广到多个同方差正态总体的情形.

例如, 对于两个同方差正态总体的情形. 设 $\bar{X}, \bar{Y}, S_1^2, S_2^2$ 是定理 5 的 2° 中所说的正态总体 $N(\mu_1, \sigma^2), N(\mu_2, \sigma^2)$ 的样本均值和样本方差. 只要引入正交矩阵

$$T = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{pmatrix},$$

其中 \mathbf{A}_i 为 n_i 阶正交矩阵, 其第一行元素都是 $1/\sqrt{n_i}$ ($i = 1, 2$).

与上面同样的做法，考察向量 $Z = TV$ 各分量的独立性，其中

$$V^T = (V_1, V_2, \dots, V_n),$$

$$V_i = (X_i - \mu_1)/\sigma, \quad i = 1, 2, \dots, n_1,$$

$$V_{n_1+j} = (Y_j - \mu_2)/\sigma, \quad j = 1, 2, \dots, n_2, \quad n_1 + n_2 = n,$$

就可证得 $\bar{X}, \bar{Y}, S_1^2, S_2^2$ 相互独立.

对于 $m(m \geq 2)$ 个同方差的正态总体的情形，设 \bar{X}_i, S_i^2 分别是总体 $N(\mu_i, \sigma^2), i = 1, 2, \dots, m$ 的样本均值和样本方差，且设各样本相互独立，则 $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_m, S_1^2, S_2^2, \dots, S_m^2$ 相互独立.