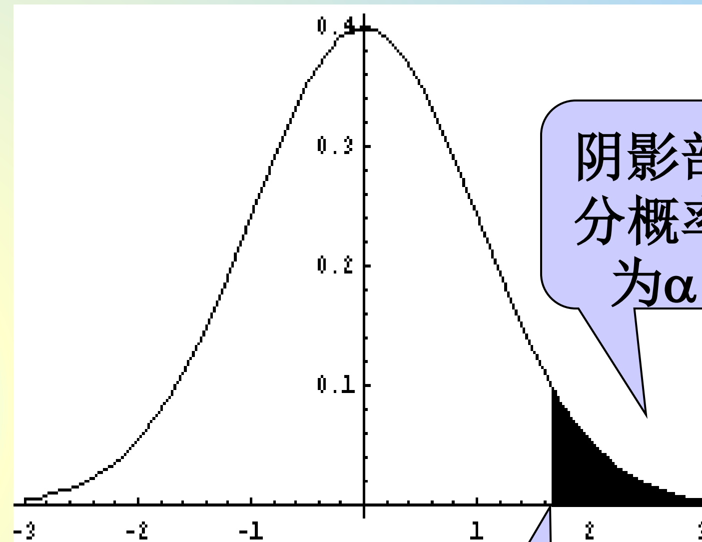
§2 常用的分布

一、四个常用分布

1. 标准正态分布 $X \sim N(0,1)$

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}},x\in R$$



上侧分位数u_α(0< α<1）：

$$P\{X>u_\alpha\}=\int_{u_\alpha}^{+\infty}f(x)dx=\alpha$$

对于正态分布有： $\Phi(u_\alpha)=1-\alpha$

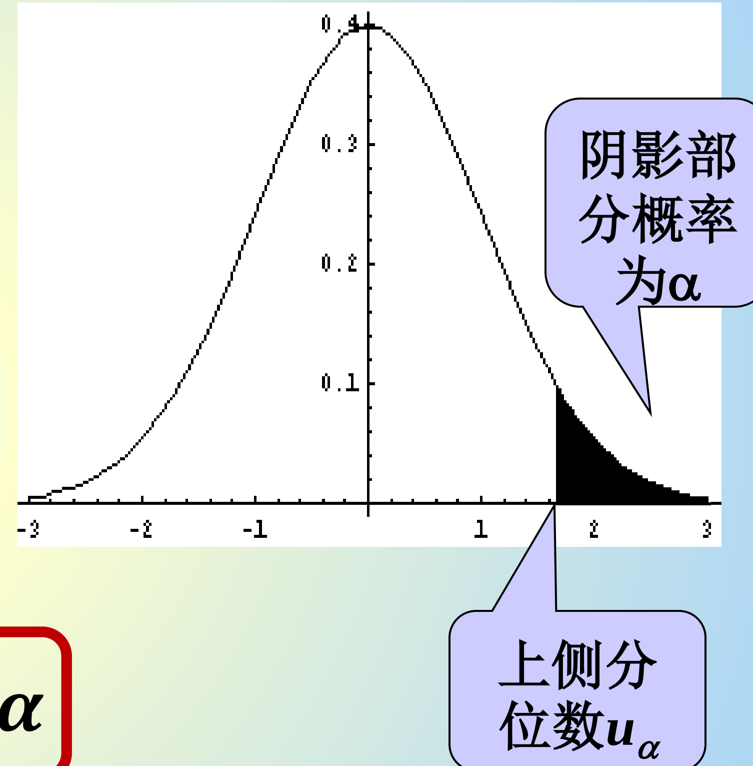
$$\because P\{X>u_\alpha\}==1-P\{X\le u_\alpha\}=1-\Phi(u_\alpha)=\alpha$$

查表： 如 α＝0.025 时， u_α＝？

$$\Phi(u_{0.025})=1-0.025=0.975 \Longrightarrow u_{0.025}=1.96$$

1. 标准正态分布

若标准正态分布的上侧分位数为$u_\alpha (0<\alpha<1)$



一般正态分布$X\sim N\left(\mu,\sigma^2\right)$的

上侧分位数u_α^*的计算方法:

$$P\{X>u_\alpha^*\}=\alpha$$

$$\Longrightarrow \boxed{P\left\{\frac{X-\mu}{\sigma}>\frac{u_\alpha^*-\mu}{\sigma}\right\}=\alpha}$$

$$\frac{X-\mu}{\sigma}\sim N(0,1)\Longrightarrow \boxed{P\left\{\frac{X-\mu}{\sigma}>u_\alpha\right\}=\alpha}$$

$$\therefore \frac{u_\alpha^*-\mu}{\sigma}=u_\alpha \Longrightarrow u_\alpha^*=\mu+\sigma\cdot u_\alpha$$

2. 自由度为n的χ^2 (卡方)分布　$\chi^2 \sim \chi^2(n)$

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

其中,$\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}e^{-x}dx, \alpha>0$ 为Gama 函数.

自由度的含义

Γ函数的主要性质： $\Gamma(1)=1, \Gamma(\frac{1}{2})=\sqrt{\pi}, \Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1)$

当$n=2$时，卡方分布就是指数分布$E\left(\frac{1}{2}\right)$

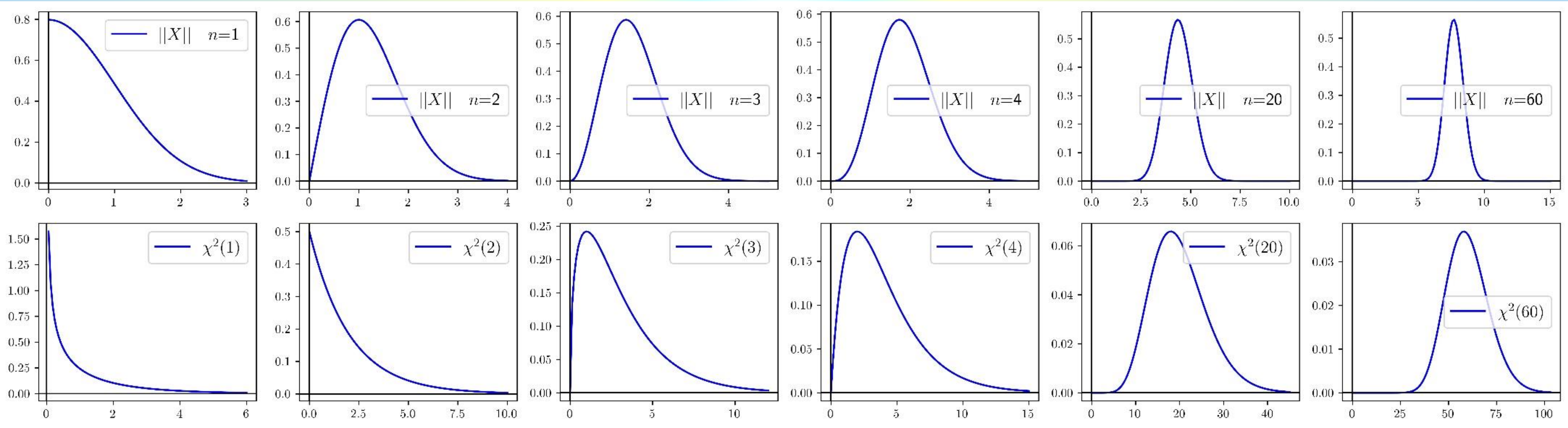
2. 自由度为n的χ^2 (卡方)分布 $\chi^2 \sim \chi^2(n)$

定理6.2.1 设 $X_1, X_2, \dots, X_n$相互独立且都服从标准正态分布，则

$$\chi^2 = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$$

即随机变量 χ^2 服从自由度为 n 的卡方分布.

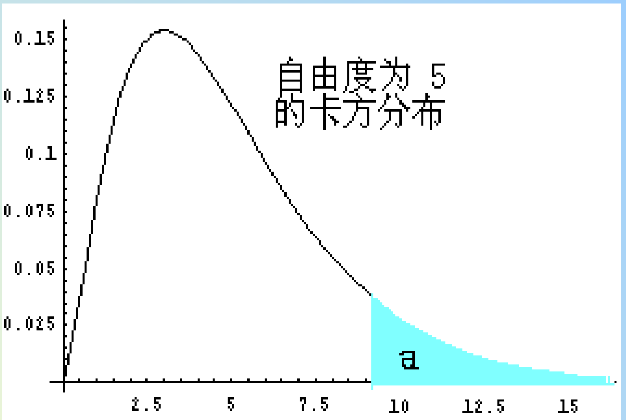
例 统计量的分布 (之一)



当$n=2$时，卡方分布就是指数分布

• $\chi^2(n)$ 的上侧分位数 $\chi^2_\alpha(n)$($0<\alpha<1$):

$$P\{\chi^2>\chi^2_\alpha(n)\}=\int_{\chi^2_\alpha(n)}^{+\infty} f_{\chi^2}(x)dx=\alpha$$



TIPS

例 查表计算

• χ^2分布的三条性质：

性质1.（数字特征）设$\chi^2\sim\chi^2(n)$，则有

$$E(\chi^2)=n, \qquad D(\chi^2)=2n$$

性质2.（可加性）设Y_1,Y_2相互独立且$Y_1\sim\chi^2(n_1)$，$Y_2\sim\chi^2(n_2)$，则 $Y_1+Y_2\sim\chi^2(n_1+n_2)$

性质3.（大样本分位数）当n足够大（如$n>45$）时，有 $\chi^2_\alpha(n)\approx n+u_\alpha\sqrt{2n}$，其中$u_\alpha$满足$\Phi(u_\alpha)=1-\alpha$.

- χ^2分布的三条性质：

性质3.（大样本分位数） 当n足够大(如$n > 45$)时，有

$$\chi^2_\alpha(n) \approx n + u_\alpha\sqrt{2n}$$，其中u_α满足$\Phi(u_\alpha) = 1 - \alpha$.

性质3的分析：

(1) $X_i \sim N(0,1),\ X_i^2 \sim \chi^2(1), i = 1, \cdots, n$ 相互独立

$$\chi^2 = \sum_{i=1}^{n} {X_i}^2 \sim \chi^2(n)$$

根据独立同分布中心极限定理，当n足够大时， χ^2近似为正态分布 —— 参数如何定？

(2) 根据卡方分布性质(1) $E(\chi^2) = n$， $D(\chi^2) = 2n$

(3) χ^2近似为一般正态分布， $\chi^2 \sim N(n, 2n)$

其分位数为 $\chi^2_\alpha(n) \approx n + u_\alpha\sqrt{2n}$

$$X \sim N(\mu, \sigma^2) \rightarrow u^*_\alpha = \mu + \sigma \cdot u_\alpha$$

3. 自由度为 n 的 t 分布 $T \sim t(n)$

(又称学生氏分布----第一个研究者以Student作笔名发表文章)

$$f(x)=\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},\quad x\in R$$

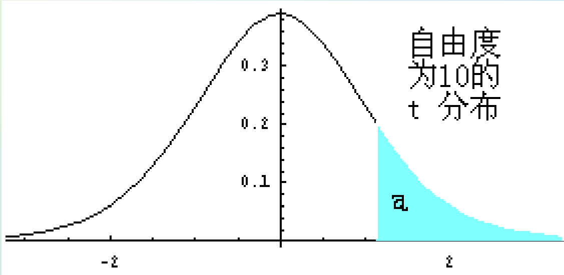
定理6.2.2 设 随机变量 X, Y 相互独立， $X \sim N(0,1)$，$Y \sim \chi^2(n)$，则

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

即随机变量 T 服从自由度为 n 的 t 分布.

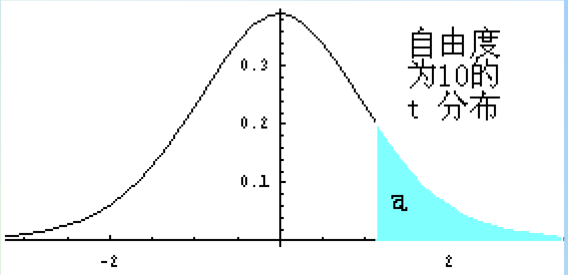
- $t(n)$ 的上侧分位数 $t_\alpha(n)$ $(0<\alpha<1)$

$$P\{T>t_\alpha(n)\}=\int_{t_\alpha(n)}^{+\infty} f_T(x)dx=\alpha$$



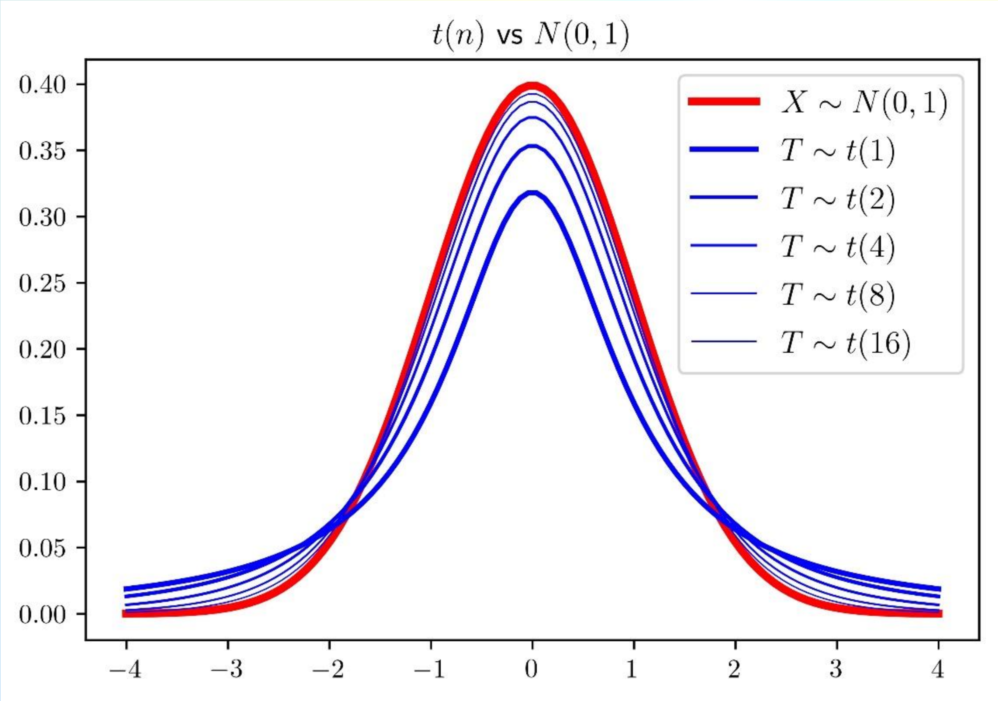
T 分布的特点：

- 关于纵轴对称 $\boxed{t_{1-\alpha}(n) = -t_\alpha(n)}$
- n 较大时(n>30)， $t_\alpha(n) \approx u_\alpha$



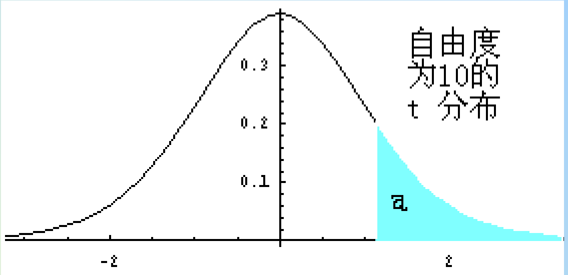
$\because P\{T \le -t_\alpha\} = P\{T > t_\alpha\} = \alpha$

$\therefore P\{T > -t_\alpha\} = 1-\alpha$ 即 $t_{1-\alpha} = -t_\alpha$



T 分布的特点：

- 关于纵轴对称 $\boxed{t_{1-\alpha}(n) = -t_{\alpha}(n)}$
- n 较大时(n>30)， $t_{\alpha}(n) \approx u_{\alpha}$



例 查表计算： $t_{0.95}(20) = ?$ $t_{0.95}(80) = ?$

解： $t_{0.95}(20) = t_{1-0.05}(20) = -t_{0.05}(20) = -1.7247$

$$t_{0.95}(80) = -t_{0.05}(80) \approx -u_{0.05} = -1.645$$

$$\boxed{\Phi(u_{\alpha}) = 1-\alpha}$$

4. F 分布　　　$F \sim F(n_1, n_2)$

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

其中，n_1 为 F 分布的第一自由度，n_2 为 F 分布的第二自由度

定理6.2.3 设随机变量X，Y 相互独立， $X \sim \chi^2(n_1)$，$Y \sim \chi^2(n_2)$，则

$$F=\frac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$$

即随机变量 F 服从第一自由度为n_1，第二自由度为n_2的F分布。

推论: 若　$F \sim F(n_1, n_2) \Rightarrow \dfrac{1}{F} \sim F(n_2, n_1)$

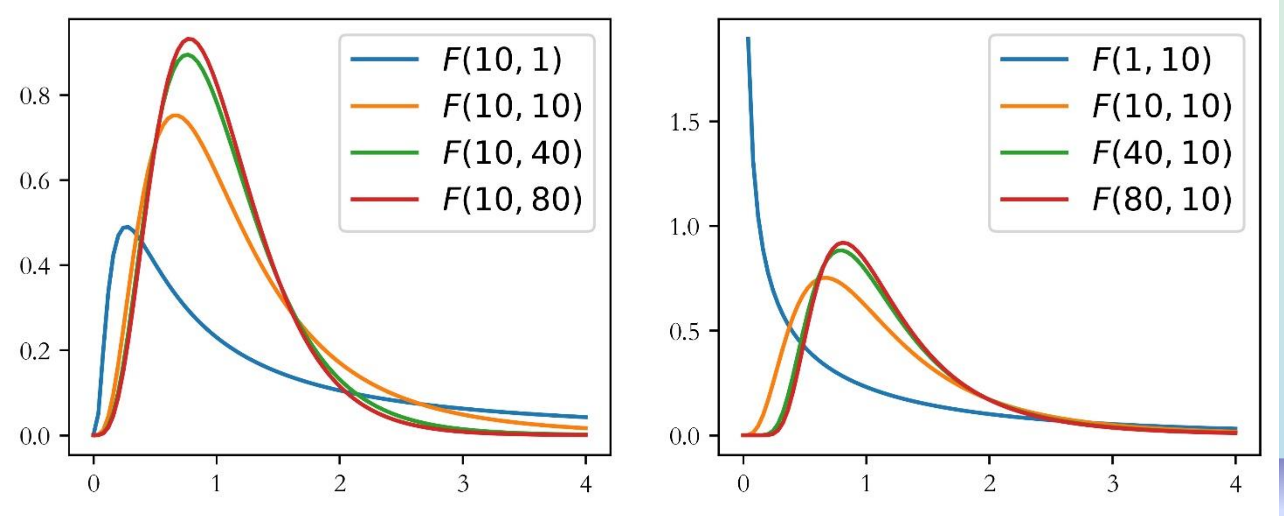
4. F 分布　　　$F \sim F(n_1, n_2)$

定理6.2.3 设 随机变量X，Y 相互独立， $X \sim \chi^2(n_1)$ ，$Y \sim \chi^2(n_2)$，则

$$F = \frac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$$

即随机变量 F 服从第一自由度为n_1，第二自由度为n_2的 F分布。

推论: 若 $F \sim F(n_1, n_2) \Rightarrow \frac{1}{F} \sim F(n_2, n_1)$



• $F(n_1, n_2)$的上侧分位数$F_\alpha(n_1, n_2)$ $(0<\alpha<1)$:

$$P\{F > F_\alpha(n_1,n_2)\} = \int_{F_\alpha(n_1,n_2)}^{+\infty} f_F(x)dx = \alpha$$

推论: 若 $F \sim F(n_1,n_2) \Rightarrow F_{1-\alpha}(n_1,n_2) = \dfrac{1}{F_\alpha(n_2,n_1)}$

证:
$$1-\alpha = P\{F > F_{1-\alpha}(n_1,n_2)\} = P\{\frac{1}{F} < \frac{1}{F_{1-\alpha}(n_1,n_2)}\}$$

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

$$又 \because \frac{1}{F} \sim F(n_2,n_1) \Rightarrow P\left\{\boxed{\frac{1}{F} > F_\alpha(n_2,n_1)}\right\} = \alpha$$

由分位数定义得到

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

TIPS

例 统计量的分布 (之二)

二、抽样分布定理

设$X_1, X_2, \cdots, X_n$是正态总体$X \sim N(\mu, \sigma^2)$的样本，$\bar{X}$，S^2分别是样本均值和样本方差，则

定理6.2.4 (1) $\bar{X}$，S^2相互独立 (2) $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

(3) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ (4) $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

证明： (4) 由(2)得 $U = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

由(3)得 $V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

由(1)可知：U和V相互独立，再由t分布构造定理可得

$$\boxed{\frac{U}{\sqrt{V/(n-1)}}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \Big/ \sqrt{\frac{(n-1)S^2}{\sigma^2} * \frac{1}{n-1}} = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

定理6.2.5 设$X_1, X_2, \cdots, X_{n_1}$和$Y_1, Y_2, \cdots, Y_{n_2}$分别是来自正态总体$X \sim N(\mu_1, \sigma_1^2)$和$Y \sim N(\mu_2, \sigma_2^2)$的样本，且相互独立，$\overline{X}, S_1^2$和$\overline{Y}, S_2^2$为各自的样本均值和样本方差，则

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

(2) 当$\sigma_1^2 = \sigma_2^2$时，

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

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

[(2)分析] $\overline{X}-\overline{Y}$服从正态分布，$S^2$可化为$\chi^2$分布，二者组合而成的统计量应服从 t分布；再确定其系数与参数.

证明： (2) 由于$\overline{X}\sim N\left(\mu_1,\dfrac{\sigma_1^2}{n_1}\right),\overline{Y}\sim N\left(\mu_2,\dfrac{\sigma_2^2}{n_2}\right)$，从而

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

当$\sigma_1^2=\sigma_2^2=\sigma^2$时，可化为

$$U=\frac{\left(\overline{X}-\overline{Y}\right)-(\mu_1-\mu_2)}{\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim N(0,1)$$

【注意：比较该统计量与目标统计量的相似性】

根据抽样分布定理6.2.4 —(3) 可知

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

当$\sigma_1^2=\sigma_2^2=\sigma^2$时，根据卡方分布可加性，有

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

$$U=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1)$$

由于$\overline{X},S_1^2$和$\overline{Y},S_2^2$相互独立，从而U与V也相互独立，故

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