4.2 HW 2

作业 2 链接

练习 4.1 随机向量 $\mathbf{x}=(x_1,\ldots,x_n)^\top$, 假设随机变量 $x_1,\ldots,x_n$ iid $\sim N(0,1)$, 我们称随机向量 $\mathbf{x}$ 服从 n 元标准正态分布, 记作 $\mathbf{x}\sim N_n(0,I_n)$ 。假设 A 是 $n\times n$ 正交矩阵, 证明 $\mathbf{y}=A\mathbf{x}\sim N_n(0,I_n)$.

证明 设 $A=(a_{ij})_{n\times n}$

$\because A$ 是正交阵 $\quad\therefore \sum_{j=1}^n a_{ij}^2=1\quad\forall i,\quad \sum_{j=1}^n a_{kj}a_{ij}=0\quad\forall k\neq l$

$\because Y=\begin{pmatrix} y_1\\ \vdots\\ y_n\end{pmatrix}=Ax\quad y_i=\sum_{j=1}^n a_{ij}x_j$. 而 $x_1,\cdots x_n\overset{\text{iid}}{\sim}N(0,1)$

$\therefore y_i\sim N(0,1)$

又对 $\forall k\neq l\quad \operatorname{Cov}(y_k,y_l)=\sum_{j=1}^n a_{kj}a_{lj}=0$, 即 y_k 与 y_l 独立

$$\therefore Y=Ax\sim N_n(0,I_n)$$

练习 4.2 设 $\mathbf{a},\mathbf{b}\in R^n$ 为常数向量, 假设 $\|\mathbf{a}\|=\|\mathbf{b}\|=1$, 且 $\mathbf{a}^\top\mathbf{b}=0$ 。假设 $n\times 1$ 随机向量 $\mathbf{x}\sim N_n(0,I_n)$, 试证明

$$\sqrt{n-2}\times\frac{\mathbf{a}^\top\mathbf{x}}{\sqrt{\|\mathbf{x}\|^2-(\mathbf{a}^\top\mathbf{x})^2-(\mathbf{b}^\top\mathbf{x})^2}}\sim t_{n-2}.$$

证明 构造 $n\times n$ 的正交矩阵 $A=\begin{pmatrix}a^\top\\ b^\top\\ *\end{pmatrix}$, 令 $y=Ax=\begin{pmatrix}a^\top x\\ b^\top x\\ *\end{pmatrix}=\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}$ 由题 1 可知, $y_1,\cdots y_n\overset{\text{iid}}{\sim}N(0,1)$

即 $y\sim N_n(0,I_n)$

$$\because \|y\|^2=y^\top y=x^\top A^\top Ax=x^\top I_n x=x^\top x=\|x\|^2$$

$$\therefore \|x\|^2-(a^\top x)^2-(b^\top x)^2=\|y\|^2-y_1^2-y_2^2=\sum_{i=3}^n y_i^2\sim\chi^2_{n-2}$$

$$\therefore \sqrt{n-2}\frac{a^\top x}{\sqrt{\|x\|^2-(a^\top x)^2-(b^\top x)^2}}=\frac{y_1}{\sqrt{\sum_{i=3}^n y_i^2/(n-2)}}\sim t_{n-2}$$

练习 4.3 假设两组样本:

$$y_1,y_2,\ldots,y_{n_0}\text{ iid }\sim N(\mu_0,\sigma^2),$$

$$y_{n_0+1},y_{n_0+2},\ldots,y_{n_0+n_1}\text{ iid }\sim N(\mu_1,\sigma^2).$$

零假设 $H_0:\mu_1=\mu_0$ 的两样本 t-检验统计量为

$$t_1=\frac{\bar{y}_1-\bar{y}_0}{\sqrt{\left(\frac{1}{n_0}+\frac{1}{n_1}\right)s^2}},$$

其中 $\bar{y}_0=\frac{1}{n_0}\sum_{i=1}^{n_0}y_i,\bar{y}_1=\frac{1}{n_1}\sum_{i=n_0+1}^{n_0+n_1}y_i,s^2=\frac{1}{n_0+n_1-2}\left(\sum_{i=1}^{n_0}(y_i-\bar{y}_0)^2+\sum_{i=n_0+1}^{n_0+n_1}(y_i-\bar{y}_1)^2\right)$ 。另外一方面, 我们记 $x_i=0,1\leq i\leq n_0$ 为第一组的标号, $x_i=1,n_0+1\leq i\leq n_0+n_1$ 为第二组的标号。记 r 为 $(x_i,y_i),i=1,\ldots,n=n_0+n_1$ 的样本相关系数, x,y 的相关性检验统计量

$$t_2=\sqrt{n-2}\frac{r}{\sqrt{1-r^2}},$$

试验证 $t_1=t_2$ 。

证明

$$
\begin{aligned}
r &= \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}, \quad \bar{x} = \frac{n_1}{n_0+n_1}, \quad \bar{y} = \frac{n_0\bar{y}_0 + n_1\bar{y}_1}{n_0+n_1} \\
S_{xx} &= \sum_{i=1}^{n}(x_i-\bar{x})^2 = n_0\left(\frac{n_1}{n_0+n_1}\right)^2 + n_1\left(\frac{n_0}{n_0+n_1}\right)^2 = \frac{n_1 n_0}{n_1+n_0} \\
S_{xy} &= \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) \\
&= n_0\left(-\frac{n_1}{n_0+n_1}\right)(\bar{y}_0-\bar{y}) + n_1\left(\frac{n_0}{n_0+n_1}\right)(\bar{y}_1-\bar{y}) \\
&= \frac{n_0 n_1^2(\bar{y}_1-\bar{y}_0)}{(n_0+n_1)^2} + \frac{n_0^2 n_1(\bar{y}_1-\bar{y}_0)}{(n_0+n_1)^2} \\
&= \frac{n_0 n_1(\bar{y}_1-\bar{y}_0)}{n_0+n_1} \\
S_{yy} &= \sum_{i=1}^{n}(y_i-\bar{y}) = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = \left(\sum_{i=1}^{n} y_i^2 - n_0\bar{y}_0^2 - n_1\bar{y}_1^2\right) + \left(n_0\bar{y}_0^2 + n_1\bar{y}_1^2 - n\bar{y}^2\right) \\
&= (n_0+n_1-2)s^2 + \frac{n_0 n_1}{n_0+n_1}(\bar{y}_0-\bar{y}_1)^2 \\
r &= \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \sqrt{\frac{n_0 n_1}{n_0+n_1}}\frac{\bar{y}_1-\bar{y}_0}{\sqrt{S_{yy}}}, 1-r^2 = 1-\frac{n_0 n_1}{n_0+n_1}\frac{(\bar{y}_1-\bar{y}_0)^2}{S_{yy}} = \frac{(n_0+n_1-2)s^2}{S_{yy}} \\
\therefore \frac{r}{\sqrt{1-r^2}} &= \sqrt{\frac{n_0 n_1}{n_0+n_1}}\frac{\bar{y}_1-\bar{y}_0}{\sqrt{(n_0+n_1-2)s^2}} = \frac{\bar{y}_1-\bar{y}_0}{\sqrt{\left(\frac{1}{n_0}+\frac{1}{n_1}\right)s^2}}\cdot\frac{1}{\sqrt{n-2}} \\
\therefore \sqrt{n-2}\cdot\frac{r}{\sqrt{1-r^2}} &= t_1
\end{aligned}
$$

✍ **练习 4.4** 假设有二元数据 $(x_i, y_i), i = 1, \ldots, n$iid, 其中 x_i, y_i 各分别仅取 0,1 值, 其样本相关系数记为 r 。另一方面, 这种二元属性 (离散) 数据常常以列联表的形式表示: 其中 a 为 $x_i = y_i = 1$ 的样本个数 (即 $a = \sum x_i y_i, c = \sum(1-x_i)y_i$ 等等), $n_1 = a+b, n_0 = c+d, m_1 = a+c, m_0 = b+d$ 。该列联表的独立性检验的 Pearson 卡方统计量为

$$X^2 = \frac{n(ad-bc)^2}{n_0 n_1 m_0 m_1}.$$

试验证等于 $X^2 = nr^2$ 。

证明

$$r^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} = a - \frac{n_1 m_1}{n}$$

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = n_1 - \frac{n_1^2}{n}$$

$$S_{yy} = m_1 - \frac{m_1^2}{n}$$

$$\therefore r^2 = \frac{\left(a - \frac{n_1 m_1}{n}\right)^2}{\left(n_1 - \frac{n_1^2}{n}\right)\left(m_1 - \frac{m_1^2}{n}\right)} = \frac{(na - n_1 m_1)^2}{n_0 n_1 m_1 m_0}$$

$$= \frac{(ad - bc)^2}{n_0 n_1 m_0 m_1}$$

$$\therefore X^2 = nr^2$$