4.3 HW 3

4.3.1 习题

作业 3 链接

记号: 对于任何随机变量或向量 a, b, c, 我们记 $\Sigma_{ab} = \text{cov}(a, b), \Sigma_{ab\bullet c} = \Sigma_{ab} - \Sigma_{ac}\Sigma_{cc}^{-1}\Sigma_{cb}$

练习 4.1 假设 $n \times 1$ 随机向量 $\mathbf{x} = (x_1, ..., x_n)^\top \sim (\boldsymbol{\mu}, \Sigma)$，即 $E(\mathbf{x}) = \boldsymbol{\mu}, \text{cov}(\mathbf{x}) = \Sigma$，设 $\mathbf{a}, \mathbf{b} \in R^n$ 为任意两个常数向量

1. 求随机变量 $\mathbf{a}^\top\mathbf{x}$ 的均值和方差;
2. 求 $(x_1 + ... + x_n)/n$ 的均值和方差;
3. 求 $\mathbf{a}^\top\mathbf{x}$ 和 $\mathbf{b}^\top\mathbf{x}$ 的协方差和相关系数 $\rho(\mathbf{a}^\top\mathbf{x}, \mathbf{b}^\top\mathbf{x})$。
4. 求 $E||\mathbf{x}||^2$.

解

1. $E\boldsymbol{a}^T\boldsymbol{x} = \boldsymbol{a}^T\mu$, $Var(\boldsymbol{a}^T\boldsymbol{x}) = cov(\boldsymbol{a}^T\boldsymbol{x}, \boldsymbol{a}^T\boldsymbol{x}) = \boldsymbol{a}^T\Sigma\boldsymbol{a}$.
2. $E(\frac{x_1+...+x_n}{n}) = \mu$, $var(\frac{x_1+...+x_n}{n}) = \boldsymbol{a}^T\Sigma\boldsymbol{a}$, where $\boldsymbol{a} = \frac{1}{n}\mathbf{1}$.
3. $cov(\boldsymbol{a}^Tx, \boldsymbol{b}^Tx) = \boldsymbol{a}^T\Sigma\boldsymbol{b}$, $\rho(\boldsymbol{a}^Tx, \boldsymbol{b}^Tx) = \frac{cov(\boldsymbol{a}^Tx, \boldsymbol{b}^Tx)}{\sqrt{Var(\boldsymbol{a}^T\boldsymbol{x})Var(\boldsymbol{b}^T\boldsymbol{x})}} = \frac{\boldsymbol{a}^T\Sigma\boldsymbol{b}}{\sqrt{\boldsymbol{a}^T\Sigma\boldsymbol{a}\boldsymbol{b}^T\Sigma\boldsymbol{b}}}$.
4. $E\|x\|^2 = \Sigma_{i=1}^n Ex_i^2 = \Sigma_{i=1}^n(Var(x_i) + (Ex_i)^2) = tr\Sigma + \|\mu\|^2$.

练习 4.2 假设 $\mathbf{y}, \mathbf{x}, \mathbf{z}$ 为三个随机向量，记它们的方差-协方差矩阵为

$$\Sigma = \text{var}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \end{pmatrix} = \begin{pmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} & \Sigma_{\mathbf{xz}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} & \Sigma_{\mathbf{yz}} \\ \Sigma_{\mathbf{zx}} & \Sigma_{\mathbf{zy}} & \Sigma_{\mathbf{zz}} \end{pmatrix}$$

令 $\mathbf{x}, \mathbf{y}$ 关于 $\mathbf{y}$ 的去相关化分别为

$$\mathbf{x}^\perp = \mathbf{x} - \Sigma_{\mathbf{xz}}\Sigma_{\mathbf{zz}}^{-1}\mathbf{z},\ \mathbf{y}^\perp = \mathbf{y} - \Sigma_{\mathbf{yz}}\Sigma_{\mathbf{zz}}^{-1}\mathbf{z}$$

验证

$$\text{var}\begin{pmatrix} \mathbf{x}^\perp \\ \mathbf{y}^\perp \\ \mathbf{z} \end{pmatrix} = \begin{pmatrix} \Sigma_{\mathbf{xx\bullet z}} & \Sigma_{\mathbf{xy\bullet z}} & 0 \\ \Sigma_{\mathbf{yx\bullet z}} & \Sigma_{\mathbf{yy\bullet z}} & 0 \\ 0 & 0 & \Sigma_{\mathbf{zz}} \end{pmatrix}$$

证明 只要用 $\Sigma_{\boldsymbol{xy\bullet z}}$ 的定义验证

$$var(\boldsymbol{x}^\perp) = \Sigma_{\boldsymbol{xx\bullet z}}$$
$$var(\boldsymbol{y}^\perp) = \Sigma_{\boldsymbol{yy\bullet z}}$$
$$cov(\boldsymbol{x}^{\perp T}\boldsymbol{y}^\perp) = \Sigma_{\boldsymbol{xy\bullet z}}.$$

由去相关化定义,

$$cov(\boldsymbol{x}^{\perp T}\boldsymbol{z}) = 0,$$

整合成矩阵形式可知原命题成立。

练习 4.3 假设 y 是随机变量, $\mathbf{x}$ 是 $p \times 1$ 随机向量, 对任何常数向量 $\mathbf{a} \in R^p$，证明 $\mathbf{a}^\top\mathbf{x}$ 和 y 的相关系数 $\rho(\mathbf{a}^\top\mathbf{x}, y) = \frac{\mathbf{a}^\top\Sigma_{\mathbf{x}y}}{\sqrt{\mathbf{a}^\top\Sigma_{\mathbf{xx}}\mathbf{a}}\sqrt{\Sigma_{yy}}}$, 且 $|\rho(\mathbf{a}^\top\mathbf{x}, y)| \leq \sqrt{\frac{\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}}{\Sigma_{yy}}}$.

注 希望同学们通过本题，熟悉矩阵形式的 Cauchy-Schwarz 不等式，并熟练运用配凑的技术。即下面证明的第三个等号和不等式。

证明

$$\begin{aligned}
\rho(\boldsymbol{a}^T\boldsymbol{x},y) &= \frac{cov(\boldsymbol{a}^T\boldsymbol{x},y)}{\sqrt{Var(\boldsymbol{a}^T\boldsymbol{x})Var(y)}} \\
&= \frac{\boldsymbol{a}^T\Sigma_{\boldsymbol{x}y}}{\sqrt{\boldsymbol{a}^T\Sigma_{\boldsymbol{x}\boldsymbol{x}}\boldsymbol{a}\Sigma_{yy}}} \\
&= \frac{(\Sigma_{\boldsymbol{x}\boldsymbol{x}}^{1/2}\boldsymbol{a})^T(\Sigma_{\boldsymbol{x}\boldsymbol{x}}^{-1/2}\Sigma_{\boldsymbol{x}y})}{\sqrt{\boldsymbol{a}^T\Sigma_{\boldsymbol{x}\boldsymbol{x}}\boldsymbol{a}\Sigma_{yy}}} \\
&\overset{C-S}{\leq} \frac{\sqrt{\boldsymbol{a}^T\Sigma_{\boldsymbol{x}\boldsymbol{x}}\boldsymbol{a}}\sqrt{\Sigma_{\boldsymbol{x}y}\Sigma_{\boldsymbol{x}\boldsymbol{x}}^{-1}\Sigma_{y\boldsymbol{x}}}}{\sqrt{\boldsymbol{a}^T\Sigma_{\boldsymbol{x}\boldsymbol{x}}\boldsymbol{a}\Sigma_{yy}}} \\
&= \sqrt{\frac{\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}}{\Sigma yy}}.
\end{aligned}$$

✍ **练习 4.4** 假设 x,y,z 是三个随机变量, 其 Pearson 相关系数矩阵为 3×3 半正定矩阵

$$R = \begin{pmatrix} 1 & \rho_{xy} & \rho_{xz} \\ \rho_{yx} & 1 & \rho_{yz} \\ \rho_{zx} & \rho_{zy} & 1 \end{pmatrix}.$$

1. 证明
$$1-\rho_{xy}^2-\rho_{xz}^2-\rho_{zy}^2+2\rho_{xy}\rho_{yz}\rho_{zx}\geq 0.$$
2. 假设 $\rho_{xy}=\rho_{xz}=\rho_{yz}=\rho$, 说明必定 $\rho\geq -1/2$。

解

1. 等价于原矩阵行列式大于等于 0.
2. 直接求解上问所得方程。

✍ **练习 4.5** 假设 $n\times 1$ 随机向量 $\mathbf{x}=(x_1,...,x_n)^\top$ 的诸分量之间的 Pearson 相关系数都是 ρ, 试证明 $1\geq\rho\geq -1/(n-1)$。并求偏相关系数 $\rho_{x_1x_2\bullet x_3\cdots x_n}$。

解

1. 首先，随机变量之间相关系数是一个归一化的量，与他们的方差无关，所以不妨假设所有分量的方差都是 1，否则对每个随机变量除以它的标准差进行研究。由此可以得到协方差矩阵 Σ 等于相关系数矩阵：
$$R_n = \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix}.$$
注意到 R 的对角元为 1, 非对角元均为 ρ, 因此可以写成
$$R=(1-\rho)I_n+\rho\mathbf{1}\mathbf{1}^T.$$
利用 **Matrix determinant lemma**, 或者用一些初等的线性代数技巧，可以求得
$$\det(R)=(1+\rho\mathbf{1}^T(1-\rho)^{-1}I_n\mathbf{1})\det\left((1-\rho)I_n\right)=\left(1+\frac{n\rho}{1-\rho}\right)(1-\rho)^n.$$
再利用 R 半正定，行列式非负，易得 $-\frac{1}{n-1}\leq\rho\leq 1$.
以上两种求行列式的方法都不容易，还可以从半正定本身的定义出发，由
$$\mathbf{1}^TR\mathbf{1}\geq 0, \text{where } \mathbf{1}=(1,1,1,\cdots,1)^T$$
直接推出 $-\frac{1}{n-1}\leq\rho\leq 1$ 这个结果。

2. 首先给出求形如 R_n 的矩阵的逆的简便方法，即直接利用 **Sherman–Morrison formula**, 可得

$$\begin{aligned} R_n^{-1} &= ((1-\rho)I_n + \rho\mathbf{1}\mathbf{1}^T)^{-1} \\ &= \frac{1}{1-\rho}I_n - \frac{1}{(1-\rho)^2}\frac{\rho\mathbf{1}\mathbf{1}^T}{1+\rho/(1-\rho)\mathbf{1}^T\mathbf{1}} \\ &= \frac{1}{(1-\rho)(1+(n-1)\rho)}\begin{pmatrix} 1+(n-2)\rho & -\rho & \cdots & -\rho \\ -\rho & 1+(n-2)\rho & \cdots & -\rho \\ \vdots & \vdots & \ddots & \vdots \\ -\rho & -\rho & \cdots & 1+(n-2)\rho \end{pmatrix} \end{aligned}$$

关于求偏相关系数，由于课上只给出了 $\rho_{x_1,x_2\bullet x_3}$ 的显示表达式，没有给更多元，不建议使用递推的方法，而要将多元的情况看作向量形式，如直接将 $\boldsymbol{v} = (x_3, \cdots, x_n)$ 看成向量，用矩阵方法求向量版本的偏相关系数即可。令 $\boldsymbol{u} = (x_1, x_2)$, 则 $\boldsymbol{x} = (\boldsymbol{u}, \boldsymbol{v})$.
此时，

$$\begin{aligned} \Sigma_{\boldsymbol{uu}\bullet\boldsymbol{v}} &= \Sigma_{\boldsymbol{uu}} - \Sigma_{\boldsymbol{uv}}\Sigma_{\boldsymbol{vv}}^{-1}\Sigma_{\boldsymbol{vu}} \\ &= \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} - \begin{pmatrix} \rho & \cdots & \rho \\ \rho & \cdots & \rho \end{pmatrix} R_{n-2}^{-1} \begin{pmatrix} \rho & \rho \\ \vdots & \vdots \\ \rho & \rho \end{pmatrix} \quad = \begin{pmatrix} Var(x_1^\perp) & cov(x_1^\perp, x_2^\perp) \\ cov(x_2^\perp, x_1^\perp) & Var(x_2^\perp) \end{pmatrix}. \end{aligned}$$

再利用偏相关系数的定义即可得到结果

$$\rho_{x_1x_2\bullet x_3\cdots x_n} = \frac{cov(x_1^\perp, x_2^\perp)}{\sqrt{Var(x_1^\perp)Var(x_2^\perp)}} = \frac{\rho}{1+(n-2)\rho}.$$

最后，在已知 R^{-1} 的情况下，一种利用课上结论更简单的方法是，利用课件 *week04 P14* 命题 2 直接得到

$$\rho_{12\bullet\boldsymbol{v}} = -\omega_{ij}/\sqrt{\omega_{11}\omega_{22}} = \frac{\rho}{1+(n-2)\rho},$$

其中, $\Omega = (\omega_{ij}) = R^{-1}$.

4.3.2 补充内容

命题 4.1 (Matrix determinant lemma)

Suppose A is an invertible square matrix and u, v are column vectors. Then the matrix determinant lemma states that

$$\det\left(\mathbf{A} + \mathbf{u}\mathbf{v}^\top\right) = \left(1 + \mathbf{v}^\top\mathbf{A}^{-1}\mathbf{u}\right)\det(\mathbf{A}).$$

命题 4.2 (Sherman–Morrison formula)

Suppose $A \in \mathbb{R}^{n\times n}$ is an invertible square matrix and $u, v \in \mathbb{R}^n$ are column vectors. Then $A + uv^T$ is invertible iff $1 + v^T A^{-1} u \neq 0$. In this case,

$$\left(A + uv^T\right)^{-1} = A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1+v^TA^{-1}u}.$$

Here, uv^T is the outer product of two vectors u and v.

以上两个公式更大的意义在于设计高效的算法。例如当计算样本的 $d\times d$ 维协方差矩阵 $\Sigma_n = \Sigma_{i=1}^n x_i x_i^T$ 的行列式或逆矩阵时，直接计算的复杂度相当高，但可以将原矩阵写为 $\Sigma_n = \Sigma_{n-1} + x_n x_n^T$, 这样在第 n 步便可以直接利用 $n-1$ 步的结果，再进行简单的矩阵乘法，而不必直接处理一个 $d\times d$ 维高阶矩阵。