课程主页：http://staff.ustc.edu.cn/~ynyang/2022

第四讲 随机向量之间的相关性

2022.9.23

如何用一个实数（最好是介于0和1之间）度量两个随机向量之间的相关性？

Corr-2. 随机向量之间的相关性

- 典则相关系数、决定系数

多个随机变量的相关性通常以相关系数矩阵表示，它代表了各对(pair-wise)随机变量之间的相关程度，不是一个整体相关程度的度量。

- ❑ 能否不用成对的方式计算多个随机变量的相关性？
- ❑ 如何度量两组随机变量(两个随机向量)整体上的相关性？

多元分析中的“典则相关分析(Canonical correlation analysis)”以及回归分析中的决定系数回答了后一问题， 其基本思想是，把两个随机向量分别组合(pool)成两个随机变量，然后计算两个pooled变量之间的相关系数 (还是成对！)：

寻找两个最优的方向$\mathbf{a} \in R^q, \mathbf{b} \in R^p$，使得投影坐标(随机变量) $\mathbf{a}^{\mathrm{T}}\mathbf{x}$和$\mathbf{b}^{\mathrm{T}}\mathbf{y}$的相关系数最大，该最大相关系数定义为向量$\mathbf{x}$，$\mathbf{y}$ 的相关性度量，称为（第一）典则相关系数（参见命题5）。

但下面我们采用正交分解（即两个向量的去相关化-还是成对！），得到同样的相关性度量。

正交分解/去相关化

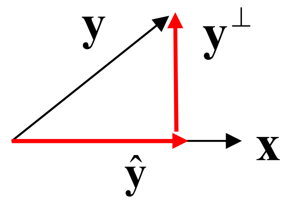
假设随机向量$\mathbf{y}_{q\times 1}$和$\mathbf{x}_{p\times 1}$的方差-协方差矩阵为

$$\Sigma = \mathrm{cov}\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \Sigma_{\mathbf{yy}} & \Sigma_{\mathbf{yx}} \\ \Sigma_{\mathbf{xy}} & \Sigma_{\mathbf{xx}} \end{pmatrix},$$

去相关化：$\mathbf{y}^{\perp} = \mathbf{y} - \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}$（与$\mathbf{x}$不相关，也与$\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}$不相关），

记$\hat{\mathbf{y}} = \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}$，则去相关化$\mathbf{y}^{\perp} = \mathbf{y} - \hat{\mathbf{y}} \Leftrightarrow$ 正交分解：

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{y}^{\perp}，\ \hat{\mathbf{y}}\text{与}\mathbf{y}^{\perp}\text{不相关(正交)}$$



$\mathbf{y}=\mathbf{y}^{\perp}+\hat{\mathbf{y}}$两边同时求方差(理解成求长度平方，勾 股定理）：

方差分解

$$\begin{array}{cccc} \mathrm{var}(\mathbf{y}) = & \mathrm{var}(\hat{\mathbf{y}}) & + & \mathrm{var}(\mathbf{y}^{\perp}) \\ \Sigma_{\mathbf{yy}} = & \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}} & + & \Sigma_{\mathbf{yy\bullet x}} \end{array}$$

(总方差)　(**x**能解释的部分)　(与**x**无关的部分)

我们用**x**所能解释的方差$\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}$在总方差$\Sigma_{\mathbf{yy}}$中的"占比"表示**x**与**y**的关系密切程度：

$$\Phi=\Sigma_{\mathbf{yy}}^{-1/2}\left(\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\right)\Sigma_{\mathbf{yy}}^{-1/2}$$

当$q>1$时，Φ是矩阵，我们可用矩阵的某个数字特征，比如行列式、迹、特征根等概括Φ。

典则相关系数 ($q > 1$)

Φ的最大特征根的平方根$\sqrt{\lambda_{\max}(\Phi)}$称为第一典则相关系数（Canonical correlation coefficient）。Φ的第二大特征根的平方根称为第二典则相关系数，用来描述第一典则相关系数未能刻画的相关性。

性质：（1）典则相关系数介于0和1之间。
（2）典则相关系数关于$\mathbf{x}, \mathbf{y}$对称。

证明：（1）因为方差分解公式中 $\Sigma_{\mathbf{yy}\bullet\mathbf{x}} \geq 0$

$$\Rightarrow \ 0 \leq \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}} \leq \Sigma_{\mathbf{yy}} \quad \Rightarrow \ 0 \leq \Phi = \Sigma_{\mathbf{yy}}^{-1/2}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2} \leq I_q.$$

所以 $0 \leq \lambda_{\max}(\Phi) \leq 1.$

（2）因为 $\Phi = \Sigma_{\mathbf{yy}}^{-1/2}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}$ 与 $\Psi = \Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1/2}$
有相同的非0特征根，$\lambda_{\max}(\Phi) = \lambda_{\max}(\Psi)$。

二次型的极值

引理1. 若$A \geq 0$(半正定), 则对任何$\mathbf{v} \neq \mathbf{0}$, $\frac{\mathbf{v}^{\mathrm{T}} A \mathbf{v}}{\mathbf{v}^{\mathrm{T}} \mathbf{v}} \leq \lambda_{max}(A)$最大特征根, 当$\mathbf{v} \propto \mathbf{v}_{\max}$(最大特征根对应的特征向量)时达到极大。

证明：假设A的谱分解为$A = O \Lambda O^{\mathrm{T}}, O^{\mathrm{T}} O = O O^{\mathrm{T}} = I, \Lambda = diag(\lambda_1, \ldots, \lambda_n)$，假设$\lambda_1 \geq \ldots \geq \lambda_n$。记$\mathbf{u} = O^{\mathrm{T}} \mathbf{v}$, 则

$$\frac{\mathbf{v}^{\mathrm{T}} A \mathbf{v}}{\mathbf{v}^{\mathrm{T}} \mathbf{v}} = \frac{\mathbf{v}^{\mathrm{T}} O \Lambda O^{\mathrm{T}} \mathbf{v}}{\mathbf{v}^{\mathrm{T}} \mathbf{v}} = \frac{\mathbf{u}^{\mathrm{T}} \Lambda \mathbf{u}}{\mathbf{u}^{\mathrm{T}} \mathbf{u}} = \frac{\sum \lambda_i u_i^2}{\sum u_i^2} \leq \lambda_1 = \max(\lambda_i),$$

当$u_2 = \ldots = u_n = 0$，即$\mathbf{v}$与O的第2至n列正交时，达到极大，此时$\mathbf{v} \propto \mathbf{v}_1 = \mathbf{v}_{\max}$。

典则相关系数的最优性

命题5. 假设$\mathbf{x}, \mathbf{y}$分别为$p \times 1, q \times 1$随机向量, $\Phi = \Sigma_{\mathbf{yy}}^{-1/2} \Sigma_{\mathbf{yx}} \Sigma_{\mathbf{xx}}^{-1} \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1/2}$，则第一典则相关系数等于$\mathbf{x}$，$\mathbf{y}$线性组合的相关系数的最大值。具体地,

$$\max_{\mathbf{a} \neq \mathbf{0} \in R^p, \mathbf{b} \neq \mathbf{0} \in R^q} \rho(\mathbf{a}^{\mathrm{T}} \mathbf{x}, \mathbf{b}^{\mathrm{T}} \mathbf{y}) = \lambda_{\max}^{1/2}(\Phi)$$

等号在$\mathbf{a} \propto \Sigma_{\mathbf{xx}}^{-1} \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1/2} \mathbf{v}_{\max}$，$\mathbf{b} \propto \Sigma_{\mathbf{yy}}^{-1/2} \mathbf{v}_{\max}$时达到，其中$\mathbf{v}_{\max}$为$\Phi$的特征根$\lambda_{\max}(\Phi)$对应的特征向量。【这是典则相关系数通常的定义方式】

注：q=1时，即y是随机变量时, b不起作用（hw3第6题）

证明: $\left|\rho(\mathbf{a}^\top\mathbf{x},\mathbf{b}^\top\mathbf{y})\right| = \frac{\left|\operatorname{cov}(\mathbf{a}^\top\mathbf{x},\mathbf{b}^\top\mathbf{y})\right|}{\sqrt{\operatorname{var}(\mathbf{a}^\top\mathbf{x})}\sqrt{\operatorname{var}(\mathbf{b}^\top\mathbf{y})}} = \frac{\left|\mathbf{a}^\top\Sigma_{\mathbf{xy}}\mathbf{b}\right|}{\sqrt{\mathbf{a}^\top\Sigma_{\mathbf{xx}}\mathbf{a}}\sqrt{\mathbf{b}^\top\Sigma_{\mathbf{yy}}\mathbf{b}}}$

令 $\mathbf{u} = \Sigma_{\mathbf{xx}}^{1/2}\mathbf{a}$,

$\mathbf{v} = \Sigma_{\mathbf{yy}}^{1/2}\mathbf{b}$

$$= \frac{\left|\mathbf{u}^\top\left(\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{v}\right)\right|}{\sqrt{\mathbf{u}^\top\mathbf{u}}\sqrt{\mathbf{v}^\top\mathbf{v}}}$$

$$\leq \frac{\sqrt{\mathbf{u}^\top\mathbf{u}}\sqrt{\mathbf{v}^\top\Sigma_{\mathbf{yy}}^{-1/2}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{v}}}{\sqrt{\mathbf{u}^\top\mathbf{u}}\sqrt{\mathbf{v}^\top\mathbf{v}}} \quad (\text{Cauchy-Schwartz不等式})$$

$$= \sqrt{\frac{\mathbf{v}^\top\Sigma_{\mathbf{yy}}^{-1/2}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{v}}{\mathbf{v}^\top\mathbf{v}}}$$

$$\leq \lambda_{\max}{}^{1/2}\left(\Sigma_{\mathbf{yy}}^{-1/2}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}\right) \quad (\text{引理1})$$

第一个不等式的等号成立当且仅当 $\mathbf{u} \propto \Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{v}$，第二个不等式等号成立当且仅当 $\mathbf{v} \propto \mathbf{v}_{\max} = \mathbf{v}_{\max}(B)$.

回归分析关心的是$q=1$，即y是随机变量的情形，此时Φ是正数，我们称之为（回归方程的）决定系数。

决定系数 ($q=1$)

y是一元随机变量时，$R^2=\Phi=\dfrac{\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}}{\Sigma_{yy}}$为$\mathbf{x}$能解释的部分在$y$的方差中所占比例，称为回归方程的决定系数。它度量了随机变量y与随机向量$\mathbf{x}$之间的相关程度。

注：

(1) R^2的最优性参见命题5.

(2) 如果$\mathbf{x}$，$\mathbf{y}$都是一维随机变量(记为x, y)，则$\Sigma_{yx}=\Sigma_{xy}$，

$$R^2=\frac{\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}}{\Sigma_{yy}}=\frac{(\Sigma_{yx})^2}{\Sigma_{yy}\Sigma_{xx}}=(\rho_{xy})^2，\text{典则相关系数}=\rho_{xy}$$

所以典则相关系数和R可看作是两个随机变量相关系数的推广。

至此，我们介绍了

Corr：　两个随机变量的Pearson相关系数

Corr-0：两个随机向量之间的去相关化

Corr-1：两个随机向量$\mathbf{x}, \mathbf{y}$之间的相关性度量-典则相关系数/决定系数：

将$\mathbf{y}$分解为：$\mathbf{y}=$与$\mathbf{x}$有关的部分+与$\mathbf{x}$无关的部分，考虑第一部分的占比。

Corr-2：两个随机变量x, y的偏相关系数（控制$\boldsymbol{z}$)：

将x, y分别对$\boldsymbol{z}$去相关化，得$x^{\perp}, y^{\perp}$，其Pearson相关系数定义为偏相关系数。

下面，考虑两个随机向量$\mathbf{x}, \mathbf{y}$在控制$\boldsymbol{z}$条件下的相关性（Corr−1 + corr−2）

Corr-3：两个随机向量$\mathbf{x}, \mathbf{y}$在控制$\boldsymbol{z}$条件下的相关性度量：

- 首先将$\mathbf{x}, \mathbf{y}$关于$\boldsymbol{z}$去相关化，得到与$\boldsymbol{z}$不相关的 $\mathbf{x}^{\perp}, \mathbf{y}^{\perp}$
- 计算随机向量$\mathbf{x}^{\perp}, \mathbf{y}^{\perp}$之间的典则相关系数，应称为偏典则相关系数。
  特别地，当y是随机变量时，称为偏决定系数。

Corr-3：向量之间的偏（部分）相关性

如何度量控制随机向量**z**的条件下，随机向量**y**和**x**相关性？
首先在**y**，**x**中消除**z**的影响（即去相关化），得到$\mathbf{y}^{\perp}, \mathbf{x}^{\perp}$，
计算它们的典则相关系数，得到数值度量。

假设有随机向量 $\mathbf{y}_{q\times 1}$和$\mathbf{x}_{p\times 1}$，干扰因素 **z**，令

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

其方差 - 协方差矩阵

$$\mathrm{var}\begin{pmatrix} \mathbf{y}^{\perp} \\ \mathbf{x}^{\perp} \end{pmatrix} = \begin{pmatrix} \Sigma_{\mathbf{yy\bullet z}} & \Sigma_{\mathbf{yx\bullet z}} \\ \Sigma_{\mathbf{xy\bullet z}} & \Sigma_{\mathbf{xx\bullet z}} \end{pmatrix} = \begin{pmatrix} \Sigma_{\mathbf{yy}} & \Sigma_{\mathbf{yx}} \\ \Sigma_{\mathbf{xy}} & \Sigma_{\mathbf{xx}} \end{pmatrix} - \begin{pmatrix} \Sigma_{\mathbf{yz}} \\ \Sigma_{\mathbf{xz}} \end{pmatrix} \Sigma_{\mathbf{zz}}^{-1} \left(\Sigma_{\mathbf{zy}},\ \Sigma_{\mathbf{zx}} \right)$$

类似于典则相关系数 / 决定系数，考虑 $\mathbf{y}^{\perp}$中能用$\mathbf{x}^{\perp}$解释的“比例”：

$$\Phi^{\perp} = \Sigma_{\mathbf{y}^{\perp}\mathbf{y}^{\perp}}^{-1/2} \left(\Sigma_{\mathbf{y}^{\perp}\mathbf{x}^{\perp}} \Sigma_{\mathbf{x}^{\perp}\mathbf{x}^{\perp}}^{-1} \Sigma_{\mathbf{x}^{\perp}\mathbf{y}^{\perp}} \right) \Sigma_{\mathbf{y}^{\perp}\mathbf{y}^{\perp}}^{-1/2} = \Sigma_{\mathbf{yy\bullet z}}^{-1/2} \left(\Sigma_{\mathbf{yx\bullet z}} \Sigma_{\mathbf{xx\bullet z}}^{-1} \Sigma_{\mathbf{xy\bullet z}} \right) \Sigma_{\mathbf{yy\bullet z}}^{-1/2}$$

$q>1$时$\Phi^{\perp}$是一个矩阵，我们取其 最大特征根$\lambda_{\max}(\Phi^{\perp})$的平方根，即$\mathbf{y}^{\perp}$和$\mathbf{x}^{\perp}$的典则相关系数，度量“控制$\mathbf{z}$条件下，随机向量$\mathbf{x}$和随机向量$\mathbf{y}$之间的相关程度”。

偏典则相关系数($q>1$)

我们称$\sqrt{\lambda_{\max}(\Phi^{\perp})}$为随机向量$\mathbf{x}$和随机向量$\mathbf{y}$之间的第一偏典则相关系数。

$q>1$情形属于多元分析，下 面不再赘述。

本课程主要关心 $q=1$的情形（即 y是随机变量）

偏决定系数 ($q=1$)

$q=1$时(y是随机变量)，$\Phi^{\perp}$是介于0,1之间的实数，记作 $R^2_{y\mathbf{x}\bullet\mathbf{z}}$

$$R^2_{y\mathbf{x}\bullet\mathbf{z}} = \frac{\Sigma_{y^{\perp}\mathbf{x}^{\perp}}\Sigma^{-1}_{\mathbf{x}^{\perp}\mathbf{x}^{\perp}}\Sigma_{\mathbf{x}^{\perp}y^{\perp}}}{\Sigma_{y^{\perp}y^{\perp}}} = \frac{\Sigma_{y\mathbf{x}\bullet\mathbf{z}}\Sigma^{-1}_{\mathbf{xx}\bullet\mathbf{z}}\Sigma_{\mathbf{x}y\bullet\mathbf{z}}}{\Sigma_{yy\bullet\mathbf{z}}}$$

我们称之为控制随机向量 $\mathbf{z}$条件下，响应变量 y和随机向量 $\mathbf{x}$之间的决定系数。

至此，对于y是一维随机变量的情形(q=1)，$\mathbf{x}=\begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix}$, 我们定义了

- ✓ 决定系数：描述了随机变量y和向量$\mathbf{x}$的整体相关性.
- ✓ 偏决定系数： 描述了y 与 $\mathbf{x}$的局部相关性，即控制$\mathbf{x}_2$时， y与$\mathbf{x}_1$的偏相关性。