第三讲 偏相关系数

2023.9.22

如何变换/生成两个不相关/高度相关的随机变量？

- 随机向量
- 去相关化
- 决定系数和典则相关系数
- 偏相关系数

随机向量的协方差/相关系数矩阵

多个随机变量之间的相关性如何度量？我们一般通过两两之间的相关性，即协方差矩阵或相关系数矩阵进行刻画。

均值

定义：随机向量$\mathbf{x}=(x_1,\ldots,x_n)^{\mathsf{T}}$的均值定义为

$$E(\mathbf{x})=(Ex_1,\ldots,Ex_n)^{\mathsf{T}}$$

协方差矩阵

定义：随机向量$\mathbf{x}_{n\times 1},\mathbf{y}_{m\times 1}$的协方差矩阵定义为

$$\operatorname{cov}(\mathbf{x},\mathbf{y})=E(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{y}-\boldsymbol{\mu}_y)^{\mathsf{T}}=\left(\operatorname{cov}(x_i,y_j),_{1\le i\le n,1\le j\le m}\right)$$

其中$\boldsymbol{\mu}_x=E(\mathbf{x}),\boldsymbol{\mu}_y=E(\mathbf{y})$，特别地，随机向量$\mathbf{x}$的方差-协方差矩阵

$$\operatorname{var}(\mathbf{x})=\operatorname{cov}(\mathbf{x})=\operatorname{cov}(\mathbf{x},\mathbf{x})=E(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)^{\mathsf{T}}=E\mathbf{x}\mathbf{x}^{\mathsf{T}}-\boldsymbol{\mu}_x\boldsymbol{\mu}_x^{\mathsf{T}}.$$

记号：随机向量 $\mathbf{x}\sim(\boldsymbol{\mu},\Sigma)\Leftrightarrow E(\mathbf{x})=\boldsymbol{\mu},\ \operatorname{var}(\mathbf{x})=\Sigma$

性质

命题1. 设$\mathbf{x},\mathbf{y}$分别为$n\times 1$的随机向量，$\mathbf{z},\mathbf{w}$是$m\times 1$随机向量，则

(1) $E(\mathbf{x}+\mathbf{y})=E(\mathbf{x})+E(\mathbf{y})$,

$E\left(A\mathbf{x}+\mathbf{b}\right)=AE(\mathbf{x})+\mathbf{b}$，任意常数矩阵$A$和向量$\mathbf{b}$.

(2) $\operatorname{cov}(\mathbf{x}+\mathbf{y},\mathbf{z}+\mathbf{w})=\operatorname{cov}(\mathbf{x},\mathbf{z})+\operatorname{cov}(\mathbf{x},\mathbf{w})+\operatorname{cov}(\mathbf{y},\mathbf{z})+\operatorname{cov}(\mathbf{y},\mathbf{w})$

$\operatorname{var}(\mathbf{x}+\mathbf{y})=\operatorname{var}(\mathbf{x})+\operatorname{var}(\mathbf{y})+\operatorname{cov}(\mathbf{x},\mathbf{y})+\operatorname{cov}(\mathbf{y},\mathbf{x})$

(3) $\operatorname{cov}\left(A\mathbf{x}+\mathbf{b},B\mathbf{y}+\mathbf{c}\right)=A\operatorname{cov}\left(\mathbf{x},\mathbf{y}\right)B^{\mathsf{T}}$, 其中$A,B$为常数矩阵,$\mathbf{b},\mathbf{c}$为常数向量. 特别地，$\operatorname{var}(A\mathbf{x}+\mathbf{b})=A\operatorname{var}(\mathbf{x})A^{\mathsf{T}}$

证明:(3) $\operatorname{cov}\left(A\mathbf{x}+\mathbf{b},B\mathbf{y}+\mathbf{c}\right)\overset{\text{定义}}{=}\mathrm{E}(A\mathbf{x}+\mathbf{b}-A\boldsymbol{\mu}_x-\mathbf{b})(B\mathbf{y}+\mathbf{c}-B\boldsymbol{\mu}_y-\mathbf{c})^{\mathsf{T}}$

$=A\left[E(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{y}-\boldsymbol{\mu}_y)^{\mathsf{T}}\right]B^{\mathsf{T}}=A\operatorname{cov}\left(\mathbf{x},\mathbf{y}\right)B^{\mathsf{T}}$

相关系数矩阵

随机向量$\mathbf{x}$的相关系数矩阵$R=\left(\rho_{ij}\right)$，其中ρ_{ij}为x_i, x_j的相关系数。

若$\Sigma=(\sigma_{ij})$为随机向量$\mathbf{x}$的协方差矩阵，则$\rho_{ij}=\dfrac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$，且

$$R=\left(\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}\right)=D^{-1/2}\Sigma D^{-1/2}$$

其中$D=diag(\Sigma)=diag(\sigma_{11},\ldots,\sigma_{nn})$.

若随机向量$\mathbf{x}_{n\times 1}\sim(\boldsymbol{\mu},\Sigma)$，即$E(\mathbf{x})=\boldsymbol{\mu}, \mathrm{var}(\boldsymbol{\mu})=\Sigma$，设$D=diag(\Sigma)$，$R$为$\mathbf{x}$的相关系数矩阵。

(1) 中心化：$\mathbf{y}=\mathbf{x}-\boldsymbol{\mu}\sim(\mathbf{0},\Sigma)$

(2) 标准化：$\mathbf{y}=\Sigma^{-1/2}(\mathbf{x}-\boldsymbol{\mu})\sim(\mathbf{0},I_n)$

$$\mathbf{y}=D^{-1/2}(\mathbf{x}-\boldsymbol{\mu})\sim(\mathbf{0},R)$$

去相关化

假设任意$q\times1$随机向量$\mathbf{y}$，$p\times1$随机向量$\mathbf{x}$, 方差-协方差矩阵：

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

$\mathbf{y}$与$\mathbf{x}$相关，协方差为$q\times p$矩阵$\Sigma_{\mathbf{yx}}$。我们希望从$\mathbf{y}$中消去与$\mathbf{x}$有关的线性成分：　求$A_{q\times p}$, 使得$\mathbf{y}-A\mathbf{x}$与$\mathbf{x}$不相关.

解：$0=\operatorname{cov}(\mathbf{y}-A\mathbf{x},\mathbf{x})=\operatorname{cov}(\mathbf{y},\mathbf{x})-A\operatorname{cov}(\mathbf{x},\mathbf{x})=\Sigma_{\mathbf{yx}}-A\Sigma_{\mathbf{xx}}$

$\Rightarrow A=\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Rightarrow \mathbf{y}-A\mathbf{x}=\mathbf{y}-\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}$与$\mathbf{x}$不相关。

去相关化

我们称$\mathbf{y}^{\perp}=\mathbf{y}-\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}$为$\mathbf{y}$的去相关化（与$\mathbf{x}$不相关）。
$\mathbf{y}^{\perp}$为$\mathbf{y}$消除了与$\mathbf{x}$有关的成分之后的剩余。

容易验证：

$$\begin{aligned}\mathrm{var}(\mathbf{y}^{\perp}) &= \mathrm{cov}(\mathbf{y}-\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}, \mathbf{y}-\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}) \\ &= \mathrm{cov}(\mathbf{y},\mathbf{y})-\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathrm{cov}(\mathbf{x},\mathbf{y})-\mathrm{cov}(\mathbf{y},\mathbf{x})\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}+\mathrm{cov}(\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}, \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}) \\ &= \Sigma_{\mathbf{yy}}-\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}} \overset{\Delta}{=} \Sigma_{\mathbf{yy}\bullet\mathbf{x}},\end{aligned}$$

$\Rightarrow$

$$\mathrm{cov}\begin{pmatrix}\mathbf{y}^{\perp}\\ \mathbf{x}\end{pmatrix}=\begin{pmatrix}\Sigma_{\mathbf{yy}\bullet\mathbf{x}} & 0\\ 0 & \Sigma_{\mathbf{xx}}\end{pmatrix}$$

对角化

去相关化：$\begin{pmatrix}\mathbf{y}\\ \mathbf{x}\end{pmatrix}\to\begin{pmatrix}\mathbf{y}^{\perp}\\ \mathbf{x}\end{pmatrix}=\begin{pmatrix}I_q & -\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\\ 0 & I_p\end{pmatrix}\begin{pmatrix}\mathbf{y}\\ \mathbf{x}\end{pmatrix}$

对应于协方差矩阵相合（或合同）分块对角化：

$$\begin{pmatrix}I_q & -\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\\ 0 & I_p\end{pmatrix}\begin{pmatrix}\Sigma_{\mathbf{yy}} & \Sigma_{\mathbf{yx}}\\ \Sigma_{\mathbf{xy}} & \Sigma_{\mathbf{xx}}\end{pmatrix}\begin{pmatrix}I_q & 0\\ -\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}} & I_p\end{pmatrix}=\begin{pmatrix}\Sigma_{\mathbf{yy}\bullet\mathbf{x}} & 0\\ 0 & \Sigma_{\mathbf{xx}}\end{pmatrix}$$

典则相关系数和决定系数

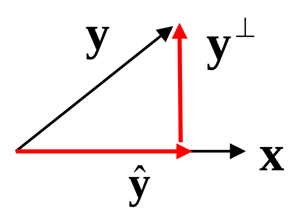
$\mathbf{y}$和$\mathbf{x}$的协方差矩阵$\Sigma_{\mathbf{yx}} = \mathrm{cov}(\mathbf{y}, \mathbf{x})$或相关系数矩阵度量了$\mathbf{y}$和$\mathbf{x}$各个分量之间的相关程度，这是一个$q \times p$矩阵，我们希望汇总矩阵所有元素，得到向量$\mathbf{y}$和$\mathbf{x}$的一个相关性的数字度量。

利用前述去相关化技术，我们可以定义典则相关系数或决定系数用于度量两个随机向量之间的相关性。

正交分解

$\mathbf{y}^{\perp} = \mathbf{y} - \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}$与$\mathbf{x}$不相关，也与$\hat{\mathbf{y}} = \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}$不相关，所以有正交分解：

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{y}^{\perp},\quad \hat{\mathbf{y}} \perp \mathbf{y}^{\perp}$$（这里$\perp$表示不相关）



正交分解 $\mathbf{y}=\hat{\mathbf{y}}+\mathbf{y}^{\perp}$两边同时求方差（理解成长度）：

方差分解
勾股定理

$$\mathrm{var}(\mathbf{y}) = \mathrm{var}(\hat{\mathbf{y}}) + \mathrm{var}(\mathbf{y}^{\perp})$$

$$\Sigma_{\mathbf{yy}} = \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}} + \Sigma_{\mathbf{yy\bullet x}}$$

(总方差)　(与**x**有关的部分)　(与**x**无关的部分)

即**y**的方差$\Sigma_{\mathbf{yy}}$中，与**x**所有关的部分$\hat{\mathbf{y}}$的方差为$\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}$，它在总方差$\Sigma_{\mathbf{yy}}$中的"比例"表示**x**与**y**的关系密切程度：

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

当然也可以定义比例为：$\Psi=\left(\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\right)\Sigma_{\mathbf{yy}}^{-1}$，但该矩阵不对称。

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

在多元分析中，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，等等。

定义若A, B是$n\times n$对称矩阵，定义$A\le B$,若$A-B\le 0$(半负定).

引理1:(1) 若$A\le B$,则对任何$m\times n$矩阵C, $CAC^{\top}\le CBC^{\top}$

(2)若$A\le I$,则$\lambda(A)\le 1$

证:(1) 对任何$\mathbf{x}\in \mathrm{R}^m$，$\mathbf{x}^{\top}(CAC^{\top}-CBC^{\top})\mathbf{x}=\mathbf{y}^{\top}(A-B)\mathbf{y}\le 0$，其中$\mathbf{y}=C^{\top}\mathbf{x}$

(2)假设正交矩阵O使得$OAO^{\top}=\Lambda$对角阵，$A\le I\Rightarrow OAO^{\top}\le OO^{\top}=I$,

即$\Lambda\le I$,即对角阵$\Lambda-I\le 0$，对角元≤ 0。

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

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

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

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，$\lambda_{\max}(\Phi) = \lambda_{\max}(\Psi)$。

回归分析中$q=1$，即y是随机变量，此时$0 \le \Phi \le 1$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，$\phi = \dfrac{\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}}{\Sigma_{yy}}$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) ϕ的最优性（命题3）：$\phi = \max\limits_{\boldsymbol{\beta} \in R^p}\left(\rho(y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\right)^2$.

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

$$\phi = \frac{\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}}{\Sigma_{yy}} = \frac{(\Sigma_{yx})^2}{\Sigma_{yy}\Sigma_{xx}} = (\rho_{xy})^2$$

所以决定/典则系数是两个随机变量相关系数的推广。

典则相关系数的最优性

命题3. 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】

引理1. 若$A\geq 0$(半正定), 则对任何$\mathbf{v}\neq\mathbf{0}$, $\dfrac{\mathbf{v}^{\mathrm{T}}A\mathbf{v}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}}\leq\lambda_{max}(A)$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$A=O\Lambda O^{\mathrm{T}}, O^{\mathrm{T}}O=OO^{\mathrm{T}}=I, \Lambda=diag(\lambda_1,\ldots,\lambda_n)$，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$\mathbf{v}\propto\mathbf{v}_1=\mathbf{v}_{\max}$。

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

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

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

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

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

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

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

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$\mathbf{v} \propto \mathbf{v}_{\max} = \mathbf{v}_{\max}(B)$.

偏相关系数 (partial correlation)

设y, x 为1维随机变量，干扰因素$\mathbf{z}$为随机向量或变量，
我们希望在控制$\mathbf{z}$的条件下计算x, y之间的相关性，即偏相关系数。

理解为控制$\mathbf{z}$的条件下，x, y不相关

偏相关系数

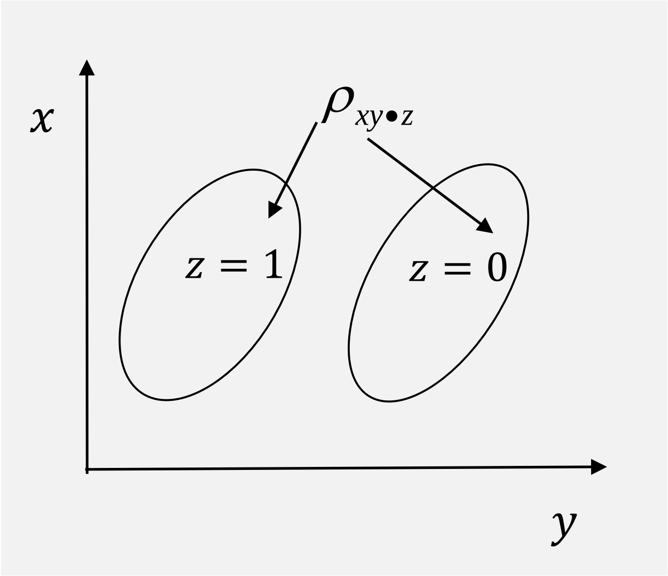
定义：令 $y^{\perp} = y - \Sigma_{y\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\mathbf{z}$， $x^{\perp} = x - \Sigma_{x\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\mathbf{z}$（它们都与$\mathbf{z}$不相关，即消除了$\mathbf{z}$的干扰)，偏相关系数定义为$y^{\perp}$ 与$x^{\perp}$的相关系数：

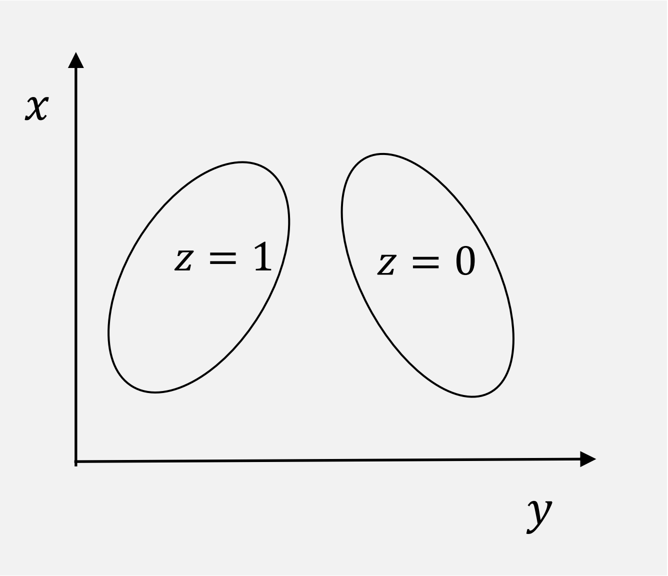
$$\rho_{yx\bullet\mathbf{z}} = \rho_{y^{\perp}x^{\perp}} = \frac{\operatorname{cov}(y^{\perp}, x^{\perp})}{\sqrt{\operatorname{var}(x^{\perp})\operatorname{var}(y^{\perp})}}。$$

- 由定义， $|\rho_{yx\bullet\mathbf{z}}| \le 1$。
- 后面将会证明，多元正态假设下，偏相关系数$\rho_{yx\bullet\mathbf{z}}$实际上是给定$\mathbf{z}$的条件下$x, y$ 的条件相关系数。
- $\rho_{yx\bullet\mathbf{z}}=0$： 控制$\mathbf{z}$的条件下，$x$和$y$不相关（正态假设下条件独立）
- $\rho_{yx\bullet\mathbf{z}}$不依赖于$\mathbf{z}$的具体值，所以只有在各个组内$x, y$分布形状类似的情况下，偏相关系数才有意义（参见下页）。

只有在各个组内x, y分布形状类似的情况下，偏相关系数才有意义（下图）。

下图计算偏相关系数$\rho_{xy\bullet z}$没有意义





记号：$\Sigma_{ab\bullet c} = \Sigma_{ab} - \Sigma_{ac}\Sigma_{cc}^{-1}\Sigma_{cb}$,
其中$\Sigma_{ab} = \mathrm{cov(a,b)}$等等。

$$\text{命题2.}\quad \rho_{yx\bullet\mathbf{z}} = \frac{\Sigma_{xy\bullet\mathbf{z}}}{\sqrt{\Sigma_{xx\bullet\mathbf{z}}}\sqrt{\Sigma_{yy\bullet\mathbf{z}}}}.$$

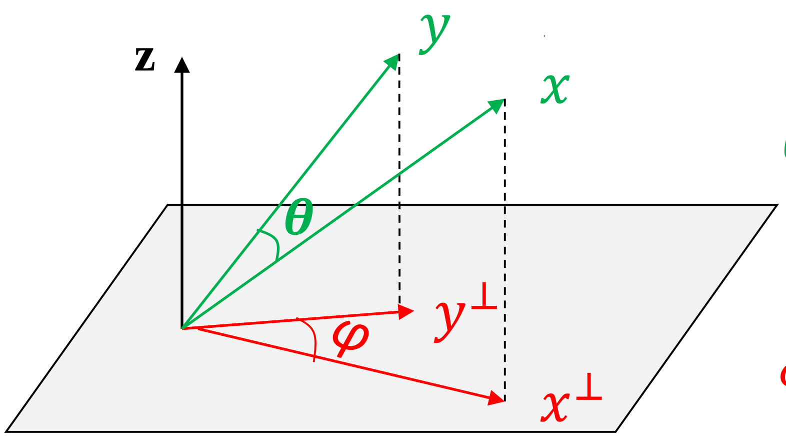
证：$\mathrm{cov}(x^{\perp}, y^{\perp}) = \mathrm{cov}(y - \Sigma_{y\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\mathbf{z},\quad x - \Sigma_{x\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\mathbf{z})$

$= \mathrm{cov}(y, x) - \Sigma_{y\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\mathrm{cov}(\mathbf{z},\ x) - \mathrm{cov}(y, \mathbf{z})\Sigma_{\mathbf{zz}}^{-1}\Sigma_{\mathbf{z}x} + \Sigma_{y\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\mathrm{cov}(\mathbf{z}, \mathbf{z})\Sigma_{\mathbf{zz}}^{-1}\Sigma_{\mathbf{z}x}$

$= \Sigma_{yx} - \Sigma_{y\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\Sigma_{\mathbf{z}x} = \Sigma_{xy\bullet\mathbf{z}}$，类似地，$\mathrm{var}(x^{\perp}) = \Sigma_{xx\bullet\mathbf{z}}, \mathrm{var}(y^{\perp}) = \Sigma_{yy\bullet\mathbf{z}}$,

所以$\rho_{yx\bullet\mathbf{z}} = \dfrac{\mathrm{cov}(y^{\perp}, x^{\perp})}{\sqrt{\mathrm{var}(x^{\perp})\,\mathrm{var}(y^{\perp})}} = \dfrac{\Sigma_{xy\bullet\mathbf{z}}}{\sqrt{\Sigma_{xx\bullet\mathbf{z}}}\sqrt{\Sigma_{yy\bullet\mathbf{z}}}}$

图示：相关系数与偏相关系数



θ代表了相关系数ρ_{xy}

φ 代表了偏相关系数$\rho_{xy\bullet z}$

$$\rho_{yx\bullet\mathbf{z}} = \frac{\operatorname{cov}(y^{\perp}, x^{\perp})}{\sqrt{\operatorname{var}(x^{\perp})\operatorname{var}(y^{\perp})}} = \frac{\Sigma_{xy\bullet\mathbf{z}}}{\sqrt{\Sigma_{xx\bullet\mathbf{z}}}\sqrt{\Sigma_{yy\bullet\mathbf{z}}}},$$

除了记号"•z"外,

偏相关系数与相关系数表达式相同

三元情形

当z也是一维随机变量时,相关系数矩阵

$$R=\begin{pmatrix}1 & \rho_{xy} & \rho_{xz}\\ \rho_{xy} & 1 & \rho_{yz}\\ \rho_{xz} & \rho_{yz} & 1\end{pmatrix},$$ 则偏相关系数$\rho_{xy\bullet z}$具有如下形式

$$\rho_{xy\bullet z}=\frac{\rho_{xy}-\rho_{xz}\rho_{yz}}{\sqrt{1-\rho_{xz}^2}\sqrt{1-\rho_{yz}^2}}$$

其中ρ_{ab}代表随机变量a,b的Pearson相关系数.

验证：x,y,z都是随机变量，Σ_{xx},Σ_{xy}等等都是实数/标量,

$$\Sigma_{xy\bullet z}=\Sigma_{xy}-\Sigma_{xz}\Sigma_{zy}/\Sigma_{zz}=\sqrt{\Sigma_{xx}}\sqrt{\Sigma_{yy}}\left(\rho_{xy}-\rho_{xz}\rho_{yz}\right),$$

$$\Sigma_{xx\bullet z}=\Sigma_{xx}-\Sigma_{xz}^2/\Sigma_{zz}=\Sigma_{xx}(1-\rho_{xz}^2),\Sigma_{yy\bullet z}=\Sigma_{yy}(1-\rho_{yz}^2).$$

注1: $|\rho_{xy\bullet z}|\le 1\Leftrightarrow 1-\rho_{xy}^2-\rho_{yz}^2-\rho_{zx}^2+2\rho_{xy}\rho_{xz}\rho_{yz}\ge 0\Leftrightarrow \det(R)\ge 0$

注2: $\rho_{xy\bullet z}=0\Leftrightarrow \rho_{xy}=\rho_{xz}\rho_{yz}\Leftrightarrow \sigma_{xy}\sigma_{zz}=\sigma_{xz}\sigma_{yz}$的意义?

例1.　假设三个变量的相关系数矩阵如下

$$R = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

即两两相关系数都是ρ, 则必定$\rho \geq -\frac{1}{2}$ (正定性约束)，且 $\rho_{xy\bullet z} = \frac{\rho}{1+\rho}$，这表明偏相关系数比相关系数更靠近0:

- 当$\rho \geq 0$时，且 $0 \leq \rho_{xy\bullet z} \leq \rho$
- 当$-1/2 \leq \rho \leq 0$时，$\rho \leq \rho_{xy\bullet z} \leq 0$

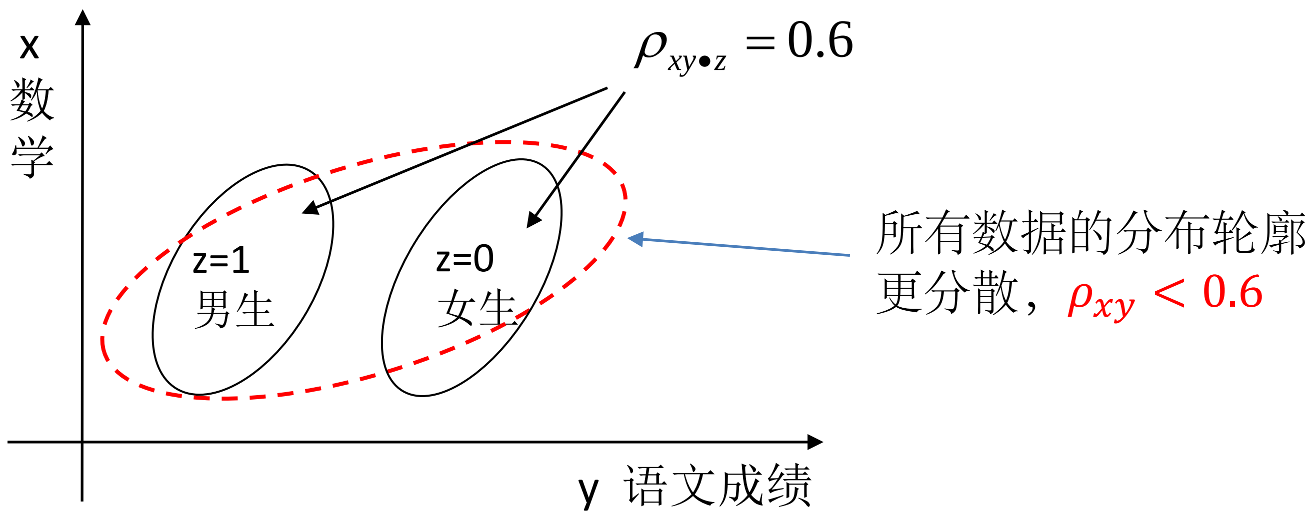
思考：为什么有约束$\rho \geq -\frac{1}{2}$，稳定性的要求？

例2. 高中男、女生的数学成绩 x与语文成绩 y 的相关系数都为0.6, 假设男女生数学成绩没有差异，但女生的语文成绩要好于男生。如果男女生混合在一起，得到的相关系数是大于、等于还是小于0.6？

以$z=1$、0分别代表男、女。已知$\rho_{xz}=0$(数学与性别不相关),

已知$\rho_{xy\bullet z}=0.6$（在各个z组内x,y的分布/相关性相同）。

$$故0.6=\rho_{xy\bullet z}=\frac{\rho_{xy}}{\sqrt{1-\rho_{yz}^2}}\geq\rho_{xy}$$



偏相关系数的检验

类似于相关系数的检验，等价于多重回归模型中回归系数的检验。不要求证明。

精确检验

数据$(x_i, y_i, \mathbf{z}_i), i=1,\ldots,n.\quad H_0: \rho_{xy\bullet\mathbf{z}}=0$

假设$y_i \mid x_i, \mathbf{z}_i, i=1,\ldots,n$ iid服从正态分布，记样本偏相关系数为$r_{xy\bullet z}$，设变量总个数为p。检验统计量在原假设下

$$T=\sqrt{n-p}\,\frac{r_{xy\bullet\mathbf{z}}}{\sqrt{1-r_{xy\bullet\mathbf{z}}^2}} \sim t_{n-p}$$

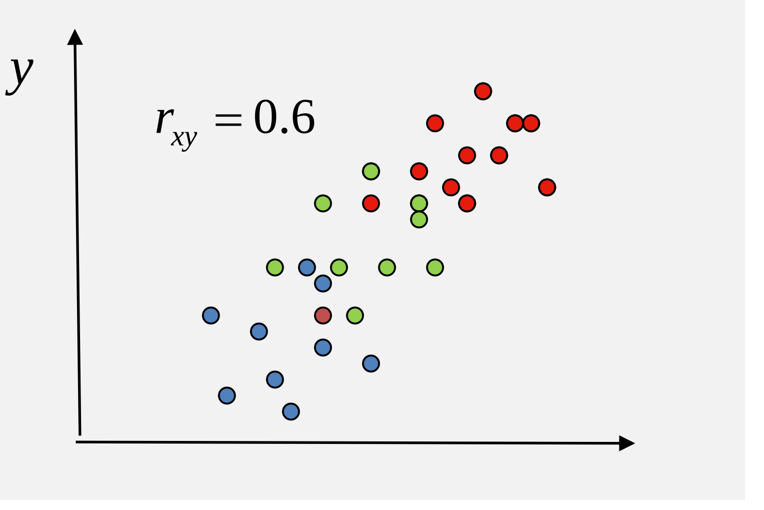
大样本检验

对一般总体，H_0下，$z=\sqrt{n-p}\times r_{xy\bullet\mathbf{z}} \xrightarrow{d} N(0,1)$，$n\to\infty$

注： p=2时（没有z），即为相关系数的检验

例3. 阅读能力(y)、身高(x)和年龄段(z,三种颜色)数据如下图，相关系数矩阵如下，求偏相关系数 $r_{xy\bullet z}$，检验阅读与身高是否相关（n=100）。

$$
\begin{array}{c|c}
 & \begin{array}{ccc} \text{x} & \text{y} & \text{z} \end{array} \\
\begin{array}{c} \text{x} \\ \text{y} \\ \text{z} \end{array} & \begin{pmatrix} 1 & 0.6 & 0.8 \\ 0.6 & 1 & 0.7 \\ 0.8 & 0.7 & 1 \end{pmatrix}
\end{array}
$$



$$r_{xy\bullet z}=\frac{r_{xy}-r_{xz}r_{yz}}{\sqrt{1-r_{xz}^2}\sqrt{1-r_{yz}^2}}=\frac{0.6-0.8\times0.7}{\sqrt{1-0.8^2}\sqrt{1-0.7^2}}=0.089$$

即，控制年龄之后，阅读能力和身高相关系数为0.089。

$z=\sqrt{100}\times0.089=0.89,\ p=0.37,$不显著。