Multivariate Datenanalyse

MSc Psychologie WiSe 2021/22

Prof. Dr. Dirk Ostwald

(12) Kanonische Korrelationsanalyse

Vorbemerkungen

Modellformulierung

Modellschätzung

Modellevaluation

Vorbemerkungen

Datenanalyseszenarien

UV	AV	Datenanalysemethoden
Univariat	Univariat	Korrelation, Einfache Regression, T-Tests
Univariat	Multivariat	T^2-Tests, MANOVA
Multivariat	Univariat	Multiple Korrelation, Multiple Regression, ALM, SVMs, NNs
Multivariat	Multivariat	Kanonische Korrelation, MALM, NNs

Vorbemerkungen

Datenanalyseszenarien

UV	AV
x_1	y_1
x_{11}	y_{11}
x_{12}	y_{12}
x_{13}	y_{13}
$\vdots$	$\vdots$
$\vdots$	$\vdots$
$\vdots$	$\vdots$
x_{1n}	y_{1n}

Korrelation
Einfache Regression
T-Tests

UV			AV
x_1	$\cdots$	x_m	y_1
x_{11}	$\cdots$	x_{m1}	y_{11}
x_{12}	$\cdots$	x_{m2}	y_{12}
x_{13}	$\cdots$	x_{m3}	y_{13}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{1n}	$\cdots$	x_{mn}	y_{1n}

Multiple Korrelation
Multiple Regression
ALM, SVMs, NNs

Datenanalyseszenarien

UV	AV		
x_1	y_1	$\cdots$	y_m
x_{11}	y_{12}	$\cdots$	y_{m1}
x_{12}	y_{13}	$\cdots$	y_{m2}
x_{13}	y_{14}	$\cdots$	y_{m3}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{1n}	y_{1n}	$\cdots$	y_{mn}

T^2-Tests
MANOVA

UV			AV		
x_1	$\cdots$	x_{m_x}	y_1	$\cdots$	y_{m_y}
x_{11}	$\cdots$	x_{m_x1}	y_{11}	$\cdots$	y_{m_y1}
x_{12}	$\cdots$	x_{m_x2}	y_{12}	$\cdots$	y_{m_y2}
x_{13}	$\cdots$	x_{m_x3}	y_{13}	$\cdots$	y_{m_y3}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{1n}	$\cdots$	x_{m_xn}	y_{1n}	$\cdots$	y_{m_yn}

Kanonische Korrelation
MALM, NNs

Überblick und Notation

Wir folgen in der Darstellung Mardia, Kent, and Bibby (1979), Chapter 10.

Die Datenvektoren $x_{1i}, ..., x_{m_x i}, i = 1, ..., n$ werden als u.i.v. Realisierungen eines Zufallsvektors X interpretiert. Die Datenvektoren $y_{1i}, ..., y_{m_y i}, i = 1, ..., n$ werden als u.i.v. Realisierungen eines Zufallsvektors Y interpretiert.

Die "erste kanonische Korrelation" ist die maximale Korrelation von Linearkombinationen von X und Y; wir bezeichnen die Linearkombinationen von X und Y mit Vektoren $a \in \mathbb{R}^{m_x}$ und $b \in \mathbb{R}^{m_y}$ mit

$$\xi = a^T X = a_1 X_1 + \cdots + a_{m_x} X_{m_x} \text{ und } \upsilon = b^T Y = b_1 Y_1 + \cdots + b_{m_y} Y_{m_y} \tag{1}$$

ξ und υ sind dann als Linearkombinationen von Zufallsvariablen selbst Zufallsvariablen; Die Korrelation von ξ und υ bezeichnen wir mit $\rho(\xi, \upsilon)$

Wenn die Zufallsvektoren X als unabhängige Variable und Y als abhängige Variable interpretiert werden, dann kann $\xi = a^T X$ als "bester Prädiktor" und $\upsilon = b^T Y$ als "am besten prädizierbares Kriterium" interpretiert werden. Kanonische Korrelationsanalyse fragt damit nach Parametern $a \in \mathbb{R}^{m_x}$ und $b \in \mathbb{R}^{m_y}$ für die $\rho(\xi, \upsilon)$ maximal ist.

Für Skalare $\alpha, \beta \in \mathbb{R}$ sind die Korrelationen $\rho\left(a^T X, b^T Y\right)$ und $\rho\left(\left(\alpha a^T\right) X, \left(\beta b^T\right) Y\right)$ allerdings identisch (siehe unten). Man sucht deshalb Parameter $a \in \mathbb{R}^{m_x}$ und $b \in \mathbb{R}^{m_y}$ für die $\rho(\xi, \upsilon)$ maximal ist und für die $a^T X$ und $b^T Y$ jeweils eine Varianz von 1 haben, also $\mathbb{V}(\xi) = \mathbb{V}(\upsilon) = 1$ gilt.

Wahrscheinlichkeitstheorie

Definition (Korrelation)

Die Korrelation zweier Zufallsvariablen ξ und υ ist definiert als

$$\rho(\xi, \upsilon) := \frac{\mathbb{C}(\xi, \upsilon)}{\mathbb{S}(\xi)\mathbb{S}(\upsilon)}, \tag{2}$$

wobei

$$\mathbb{C}(\xi, \upsilon) := \mathbb{E}((\xi - \mathbb{E}(\xi))(\upsilon - \mathbb{E}(\upsilon))) \tag{3}$$

die Kovarianz von ξ und υ und

$$\mathbb{S}(\xi) := \sqrt{\mathbb{V}(\xi)} = \sqrt{\mathbb{C}(\xi, \xi)} \text{ und } \mathbb{S}(\upsilon) := \sqrt{\mathbb{V}(\upsilon)} = \sqrt{\mathbb{C}(\upsilon, \upsilon)} \tag{4}$$

die Standardabweichungen von ξ und υ, respektive, bezeichnen.

Bemerkungen

- $\rho \in [-1, 1]$.
- Wenn ξ und υ unabhängig sind, dann gilt $\rho(\xi, \upsilon) = 0$.
- $\rho(\xi, \upsilon)$ misst dem Grad des linear-affinen Zusammenhangs $\upsilon = a\xi + b$.

Wahrscheinlichkeitstheorie

Theorem (Kovarianz und Korrelation bei linear-affinen Transformationen)

$\rho(\xi, \upsilon)$ sei die Korrelation von ξ und υ und es seien $\alpha, \beta, \gamma, \delta \in \mathbb{R}$. Dann gelten

$$\mathbb{C}(\alpha\xi + \beta, \gamma\upsilon + \delta) = \alpha\beta\mathbb{C}(\xi, \upsilon) \tag{5}$$

und

$$\rho(\alpha\xi + \beta, \gamma\upsilon + \delta) = \rho(\xi, \upsilon). \tag{6}$$

Bemerkung

- Die Varianzen von $a^T X$ und $b^T y$ und die Varianzen von Linearkombinationen von X und Y mit beliebigen skalaren Vielfachen von a und b sind im Sinne der ersten Aussage zur Kovarianz verschieden.
- Insbesondere gilt einen m_x-dimensionalen Zufallsvektoren X und einen m_y-dimensionalen Zufallsvektor Y, $a \in \mathbb{R}^{m_x}, b \in \mathbb{R}^{m_y}$, $\alpha, \beta \in \mathbb{R}$ und die Linearkombinationen $\xi := a^T X$ und $\upsilon := b^T Y$ also auch

$$\begin{aligned} & \rho(\xi, \upsilon) = \rho(\alpha\xi, \beta\upsilon) \\ \Leftrightarrow\ & \rho(a^T X, b^T Y) = \rho(\alpha(a^T X), \beta(b^T Y)) \\ \Leftrightarrow\ & \rho(a^T X, b^T Y) = \rho((\alpha a^T)X), (\beta b^T)Y). \end{aligned} \tag{7}$$

Die Korrelation von $a^T X$ und $b^T y$ und die Korrelationen von Linearkombinationen von X und Y mit beliebigen skalaren Vielfachen von a und b sind also gleich.

Wahrscheinlichkeitstheorie

Beweis

Es gilt zunächst

$$
\begin{aligned}
\mathbb{C}(\alpha\xi + \beta, c\upsilon + d) &= \mathbb{E}((\alpha\xi + \beta - \mathbb{E}(\alpha\xi + \beta))(\gamma\upsilon + \delta - \mathbb{E}(\gamma\upsilon + \delta))) \\
&= \mathbb{E}((\alpha\xi + \beta - \alpha\mathbb{E}(\xi) - \beta)(\gamma\upsilon + \delta - \gamma\mathbb{E}(\gamma\upsilon) - \delta)) \\
&= \mathbb{E}(\alpha(\xi - \mathbb{E}(\xi))(\gamma(\upsilon - \gamma\mathbb{E}(\upsilon))) \\
&= \mathbb{E}(\alpha\gamma((\xi - \mathbb{E}(\xi))(\upsilon - \gamma\mathbb{E}(\upsilon)))) \\
&= \alpha\gamma\mathbb{C}(\xi, \upsilon)
\end{aligned} \tag{8}
$$

Also folgt

$$
\begin{aligned}
\rho(\alpha\xi + \beta, \gamma\upsilon + \delta) &= \frac{\mathbb{C}(\alpha\xi + \beta, \gamma\upsilon + \delta)}{\sqrt{\mathbb{V}(\alpha\xi + \beta)}\sqrt{\mathbb{V}(\gamma\upsilon + \delta)}} \\
&= \frac{\alpha\gamma\mathbb{C}(\xi, \upsilon)}{\sqrt{\alpha^2\mathbb{V}(\xi)}\sqrt{\gamma^2\mathbb{V}(\upsilon)}} \\
&= \frac{\alpha\gamma\mathbb{C}(\xi, \upsilon)}{\alpha\mathbb{S}(\xi)c\mathbb{S}(\upsilon)} \\
&= \frac{\mathbb{C}(\xi, \upsilon)}{\mathbb{S}(\xi)\mathbb{S}(\upsilon)} \\
&= \rho(\xi, \upsilon).
\end{aligned} \tag{9}
$$

Lineare Algebra

Definition (Symmetrische Quadratwurzel einer Matrix)

$A \in \mathbb{R}^{m \times m}$ sei eine invertierbare symmetrische Matrix mit positiven Eigenwerten. Dann sind für $r \in \mathbb{N}^0$ und $s \in \mathbb{N}$ die rationalen Potenzen von A einer orthonormalen Matrix $Q \in \mathbb{R}^{m \times m}$ der Eigenvektoren von A und einer Diagonalmatrix $\Lambda = \text{diag}(\lambda_i) \in \mathbb{R}^{m \times m}$ der zugehörigen Eigenwerte $\lambda_1, ..., \lambda_m$ von A definiert als

$$A^{r/s} = Q\Lambda^{r/s}Q^T \text{ mit } \Lambda^{r/s} = \text{diag}\left(\lambda_i^{r/s}\right). \tag{10}$$

Der Spezialfall $r := 1, s := 2$ wird als symmetrische Quadratwurzel von A bezeichnet und hat die Form

$$A^{1/2} = Q\Lambda^{1/2}Q^T \text{ mit } \Lambda^{1/2} = \text{diag}\left(\lambda_i^{1/2}\right). \tag{11}$$

Bemerkungen

- Offenbar gilt
$$\left(A^{1/2}\right)^2 = Q\Lambda^{1/2}Q^TQ\Lambda^{1/2}Q^T = Q\Lambda^{1/2}\Lambda^{1/2}Q^T = Q\Lambda Q^T = A. \tag{12}$$
- Weiterhin gilt
$$\left(A^{-1/2}\right)^2 = Q\Lambda^{-1/2}Q^TQ\Lambda^{-1/2}Q^T = Q\Lambda^{-1}Q^T = A^{-1}. \tag{13}$$
  Die vorletzte Gleichung mag überraschen, aber es gilt ja zum Beispiel
$$4^{-1/2} \cdot 4^{-1/2} = \frac{1}{\sqrt{4}} \cdot \frac{1}{\sqrt{4}} = \frac{1}{4} = 4^{-1}. \tag{14}$$

Lineare Algebra

Theorem (Eigenwerte und Eigenvektoren von Matrixprodukten)

Für $A \in \mathbb{R}^{n \times m}$ und $B \in \mathbb{R}^{m \times n}$ sind die Eigenwerte von $AB \in \mathbb{R}^{n \times n}$ und $BA \in \mathbb{R}^{m \times m}$ gleich. Weiterhin gilt, dass für einen Eigenvektor v zu einem von Null verschiedenen Eigenwert λ von AB $w := Bv$ ein Eigenvektor von BA ist.

Bemerkungen

- Für einen Beweis verweisen wir auf Mardia, Kent, and Bibby (1979), S. 468.

```
A   = matrix(1:6, nrow = 2,  byrow = T)          # Matrix A \in \mathbb{R}^{2 x 3}
B   = matrix(1:6, ncol = 2,  byrow = T)          # Matrix B \in \mathbb{R}^{3 x 2}
EAB = eigen(A %*% B)                             # Eigenanalyse von AB \in \mathbb{R}^{2 \times 2}
EBA = eigen(B %*% A)                             # Eigenanalyse von BA \in \mathbb{R}^{3 \times 3}
w   = B %*% EAB$vectors[,1]                      # Eigenvektor von BA
cat("Eigenwerte von AB :"  , EAB$values[1:2],
    "\nEigenwerte von BA :", EBA$values[1:2],
    "\nBAw mit w = Bv    :", B %*% A %*% w,
    "\nlw mit w = Bv     :", EBA$values[1] * w)
```

```
> Eigenwerte von AB : 85.6 0.421
> Eigenwerte von BA : 85.6 0.421
> BAw mit w = Bv    : -191 -417 -642
> lw mit w = Bv     : -191 -417 -642
```

Lineare Algebra

Theorem (Eigenwert und Eigenvektor eines Matrixvektorprodukts)

Für $A \in \mathbb{R}^{n \times m}$, $B \in \mathbb{R}^{p \times n}$, $a \in \mathbb{R}^m$ und $b \in \mathbb{R}^p$ gilt, dass der einzige von Null verschiedene Eigenwert von $Aab^TB \in \mathbb{R}^{n \times n}$ gleich b^TBAa mit zugehörigem Eigenvektor Aa ist.

Bemerkungen

- Für einen Beweis verweisen wir auf Mardia, Kent, and Bibby (1979), S. 468.

```
A      = matrix(1:6, nrow = 2,  byrow = T)     # Matrix A \in \mathbb{R}^{2 x 3}
B      = matrix(1:8, ncol = 2,  byrow = T)     # Matrix B \in \mathbb{R}^{4 x 2}
a      = matrix(1:3, nrow = 3,  byrow = T)     # Vektor a \in \mathbb{R}^{3 x 1}
b      = matrix(1:4, nrow = 4,  byrow = T)     # Vektor b \in \mathbb{R}^{4 x 1}
EAabTB = eigen(A %*% a %*% t(b) %*% B)         # Eigenanalyse von Aab^TB \in \mathbb{R}^{4 x 4}
cat("Eigenwerte von AabTB :", EAabTB$values,
    "\nbTBAa              :", t(b) %*% B %*% A %*% a,
    "\nAa                 :", A %*% a,
    "\n(AabTB)Aa          :",(A %*% a %*% t(b) %*% B) %*% A %*% a,            # Mv
    "\n(bTBAa)Aa          :",as.vector((t(b) %*% B %*% A %*% a)) * (A %*% a)) # = \lambda v
```

```
> Eigenwerte von AabTB : 2620 0
> bTBAa                : 2620
> Aa                   : 14 32
> (AabTB)Aa            : 36680 83840
> (bTBAa)Aa            : 36680 83840
```

Lineare Algebra

Theorem (Maximierung quadratischer Formen mit Nebenbedingungen)

$A \in \mathbb{R}^{m \times m}$, $B \in \mathbb{R}^{m \times m}$ p.d. seien symmetrische Matrizen und λ_1 sei der größte Eigenwert von $B^{-1}A$ mit assoziiertem Eigenvektor $v_1 \in \mathbb{R}^m$. Dann ist λ_1 eine Lösung des Optimierungsproblems

$$\max_x x^T Ax \text{ unter der Nebenbedingung } x^T Bx = 1. \tag{15}$$

Bemerkungen

- Das Theorem ist direkt durch die kanonische Korrelationsanalyse motiviert.
- Nach Wortlaut des Theorems gilt also

$$v_1 = \arg\max_x x^T Ax \text{ unter der Nebenbedingung } x^T Bx = 1 \tag{16}$$

- Nach Wortlaut des Theorems gilt weiterhin

$$\lambda_1 = \max_x x^T Ax \text{ unter der Nebenbedingung } x^T Bx = 1. \tag{17}$$

Vorbemerkungen

Lineare Algebra

Beweis

$B^{1/2}$ sei die symmetrische Quadratwurzel von B und es sei

$$y := B^{1/2}x \Leftrightarrow x = B^{-1/2}y \tag{18}$$

Dann kann mit der symmetrischen Matrix

$$K := B^{-1/2}AB^{-1/2} \in \mathbb{R}^{m \times m} \tag{19}$$

das Optimierungsproblem (15) geschrieben werden als

$$\max_y y^T Ky \text{ unter der Nebenbedingung } y^T y = 1. \tag{20}$$

Dies gilt, weil

$$\max_x x^T Ax \Leftrightarrow \max_y \left(B^{-1/2}y\right)^T A \left(B^{-1/2}y\right) \Leftrightarrow \max_y y^T B^{-1/2}AB^{-1/2}y \Leftrightarrow \max_y y^T Ky \tag{21}$$

und

$$x^T Bx = 1 \Leftrightarrow y^T B^{-1/2}BB^{-1/2}y = 1 \Leftrightarrow y^T y = 1. \tag{22}$$

Weil K eine symmetrische Matrix ist, existiert die Orthonormalzerlegung (vgl. (2) Matrizen)

$$K = Q\Lambda Q^T, \tag{23}$$

wobei die Spalten der orthogonalen Matrix Q die Eigenvektoren von K und die Diagonalemente von Λ die zugehörigen Eigenwerte von K sind.

Lineare Algebra

Beweis (fortgeführt)

Mit der orthogonalen Matrix Q aus obiger Orthornomalzerlegung sei nun

$$z := Q^T y \Leftrightarrow y := Qz. \tag{24}$$

Dann kann das Optimierungsproblem (20) geschrieben werden als

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \text{ unter der Nebenbedingung } z^T z = 1, \tag{25}$$

weil

$$\max_y y^T K y \Leftrightarrow \max_z (Qz)^T K(Qz) \Leftrightarrow \max_z z^T Q^T Q\Lambda Q^T Qz \Leftrightarrow \max_z z^T \Lambda z \Leftrightarrow \max_z \sum_{i=1}^m \lambda_i z_i^2 \tag{26}$$

und

$$y^T y = 1 \Leftrightarrow (Qz)^T Qz = 1 \Leftrightarrow z^T Q^T Qz = 1 \Leftrightarrow z^T z = 1. \tag{27}$$

Lineare Algebra

Beweis (fortgeführt)

Die Eigenwerte von K seien nun absteigend sortiert, also $\lambda_1 \geq \cdots \geq \lambda_m$. Dann gilt für das Optimierungsproblem (25), dass

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \leq \lambda_1, \tag{28}$$

weil

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \leq \max_z \sum_{i=1}^m \lambda_1 z_i^2 = \lambda_1 \max_z \sum_{i=1}^m z_i^2 = \lambda_1 \tag{29}$$

wobei sich die letzte Gleichung aus der Nebenbedingung $z^T z = 1$ ergibt. Schließlich gilt

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 = \lambda_1, \tag{30}$$

für $z := e_1 = (1, 0, ..., 0)^T$. Zusammenfassend heißt das, dass $z = e_1$ eine Lösung des Optimierungsproblem (25) ist und das λ_1 das entsprechende Maximum ist.

Lineare Algebra

Beweis (fortgeführt)

Damit ergibt sich aber sofort, dass dann

$$y = Qz = Qe_1 = q_1 \text{ und } x = B^{-1/2} q_1 \tag{31}$$

Lösungen der äquivalenten Optimierungsprobleme (20) und (15), respektive, sind. Nach Konstruktion ist q_1 ein Eigenvektor von $B^{-1/2}AB^{-1/2}$ und nach obigem Theorem zu Eigenwerten und Eigenvektoren von Matrixprodukten damit auch ein Eigenvektor von

$$B^{-1/2}B^{-1/2}A = B^{-1}A \tag{32}$$

und die zugehörigen Eigenwerte sind gleich. Damit aber folgt, dass der größte Eigenwert von $B^{-1}A$ und sein assoziierter Eigenvektor eine Lösung von

$$\max_x x^T Ax \text{ unter der Nebenbedingung } x^T Bx = 1. \tag{33}$$

ist. $\square$

Vorbemerkungen

Modellformulierung

Modellschätzung

Modellevaluation

Überblick

Zur Entwicklung der kanonischen Korrelationsanalyse werden X und Y als

$$Z := \begin{pmatrix} X \\ Y \end{pmatrix} \tag{34}$$

zusammengefasst.

Wir nehmen durchgängig an, dass $\mathbb{E}(Z) = 0_m$ mit $m = m_x + m_y$.

Der mathematische Fokus ist auf der Kovarianzmatrix $\mathbb{C}(Z)$.

- Kovarianzen von Linearkombinationen von X und Y ergeben sich aus Matrixprodukten von $\mathbb{C}(Z)$.
- Die Matrixtheoreme aus den Vorbemerkungen können auf diese Matrixprodukte angewendet werden.

Generell wird im folgenden ein restringierter Optimierungsansatz mithilfe der Lagrangefunktion zugunsten der Eigenanalyse von Matrixprodukten supprimiert. Für den Lagrangeansatz, siehe zum Beispiel Anderson (2003), Kapitel 12.

Theorem (Kovarianzmatrizen von Zufallsvektoren)

Z sei ein m-dimensionaler Zufallsvektor mit Erwartungswert $\mathbb{E}(Z) = 0_m$ und es sei

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} \text{ mit } \mathbb{E}(Z) := 0_m \tag{35}$$

ein $m_x + m_y$-dimensionaler Zufallsvektor und sein Erwartungswertvektor, respektive. Dann kann die $m \times m$ Kovarianzmatrix Z geschrieben werden als

$$\mathbb{C}(Z) = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{36}$$

wobei

$$\begin{aligned}
\Sigma_{xx} &:= \mathbb{E}\left(XX^T\right) \in \mathbb{R}^{m_x \times m_x} \\
\Sigma_{xy} &:= \mathbb{E}\left(XY^T\right) \in \mathbb{R}^{m_x \times m_y} \\
\Sigma_{yx} &:= \mathbb{E}\left(YX^T\right) \in \mathbb{R}^{m_y \times m_x} \\
\Sigma_{yy} &:= \mathbb{E}\left(YY^T\right) \in \mathbb{R}^{m_x \times m_y}
\end{aligned} \tag{37}$$

Modellformulierung

Beweis

Nach Definition der Kovarianzmatrix eines Zufallsvektors (vgl. (3) Wahrscheinlichkeitshtheorie) gilt

$$
\begin{aligned}
\mathbb{C}(Z) &= \mathbb{E}\left((Z - \mathbb{E}(Z))(Z - \mathbb{E}(Z))^T\right) \\
&= \mathbb{E}\left((Z - 0_m)(Z - 0_m)^T\right) \\
&= \mathbb{E}\left(ZZ^T\right) \\
&= \mathbb{E}\left(\begin{pmatrix} X \\ Y \end{pmatrix}\begin{pmatrix} X^T & Y^T \end{pmatrix}\right) \\
&= \mathbb{E}\left(\begin{pmatrix} XX^T & XY^T \\ YX^T & YY^T \end{pmatrix}\right) \\
&= \begin{pmatrix} \mathbb{E}\left(XX^T\right) & \mathbb{E}\left(XY^T\right) \\ \mathbb{E}\left(YX^T\right) & \mathbb{E}\left(YY^T\right) \end{pmatrix} \\
&= \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}
\end{aligned}
\tag{38}
$$

□

Modellformulierung

Theorem (Linearkombinationen von Zufallsvektorpartitionen)

Es sei

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} \text{ mit } \mathbb{E}(X) = 0_m \text{ und } \mathbb{C}(Z) = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \tag{39}$$

ein m-dimensionaler partitionierter Zufallsvektor sowie sein Erwartungswertvektor und seine Kovarianzmatrix, respektive. Weiterhin seien für $a \in \mathbb{R}^{m_x}$ und $b \in \mathbb{R}^{m_y}$ die Zufallsvariablen

$$\xi := a^T X \text{ und } \upsilon := b^T Y \tag{40}$$

als Linearkombinationen der Komponenten von X und Y definiert. Dann gelten

(1) $\mathbb{V}(\xi) = a^T \Sigma_{xx} a$

(2) $\mathbb{V}(\upsilon) = b^T \Sigma_{yy} b$

(2) $\rho(\xi, \upsilon) = a^T \Sigma_{xy} b$, wenn $\mathbb{V}(\xi) = 1$ und $\mathbb{V}(\upsilon) = 1$.

Bemerkungen

- Die Varianz der Zufallsvariable $a^T X$ ergibt sich als "doppelte Linearkombination" von Σ_{xx}.
- Die Varianz der Zufallsvariable $b^T Y$ ergibt sich als "doppelte Linearkombination" von Σ_{yy}.
- Die Korrelation der Zufallsvariablen $a^T X$ und $b^T Y$ ergibt sich "doppellte Linearkombination" von Σ_{xy}.

Beweis von (1) und (2)

Wir betrachten zunächst die Varianz von ξ. Mit dem Varianzverschiebungssatz gilt

$$
\begin{aligned}
\mathbb{V}(\xi) &= \mathbb{E}\left(\xi\xi\right) - \mathbb{E}(\xi)\mathbb{E}(\xi) \\
&= \mathbb{E}\left((a^TX)(a^TX)\right) - \mathbb{E}\left(a^TX\right)\mathbb{E}\left(a^TX\right) \\
&= \mathbb{E}\left((a^TX)(a^TX)^T\right) - \mathbb{E}\left(a^TX\right)\mathbb{E}\left(a^TX\right) \\
&= \mathbb{E}\left(a^TXX^Ta\right) - \mathbb{E}\left(a^TX\right)\mathbb{E}\left(a^TX\right) \\
&= a^T\mathbb{E}\left(XX^T\right)a - a^T\mathbb{E}(X)a^T\mathbb{E}(X) \\
&= a^T\mathbb{E}\left(XX^T\right)a - a^T0_{m_x}a^T0_{m_x} \\
&= a^T\Sigma_{xx}a.
\end{aligned}
\tag{41}
$$

Der Beweis zur Varianz von υ folgt dann analog.

Beweis von (3)

Mit der Definition der Korrelation von Zufallsvariablen und mit $\mathbb{V}(\xi) = \mathbb{V}(\upsilon) = 1$ und dem Kovarianzverschiebungssatz gilt

$$
\begin{aligned}
\rho(\xi, \upsilon) &= \frac{\mathbb{C}(\xi, \upsilon)}{\sqrt{\mathbb{V}(\xi)}\sqrt{\mathbb{V}(\upsilon)}} \\
&= \frac{\mathbb{C}(\xi, \upsilon)}{\sqrt{1}\sqrt{1}} \\
&= \mathbb{C}(\xi, \upsilon) \\
&= \mathbb{E}(\xi\upsilon) - \mathbb{E}(\xi)\mathbb{E}(\upsilon) \\
&= \mathbb{E}\left((a^TX)(b^TY)\right) - \mathbb{E}(a^TX)\mathbb{E}(b^TY) \\
&= \mathbb{E}\left((a^TX)(b^TY)^T\right) - \mathbb{E}(a^TX)\mathbb{E}(b^TY) \\
&= \mathbb{E}\left(a^TXY^Tb\right) - \mathbb{E}(a^TX)\mathbb{E}(b^TY) \\
&= a^T\mathbb{E}\left(XY^T\right)b - a^T\mathbb{E}(X)b^T\mathbb{E}(Y) \\
&= a^T\mathbb{E}\left(XY^T\right)b - a^T0_{m_x}b^T0_{m_y} \\
&= a^T\Sigma_{xy}b.
\end{aligned}
\tag{42}
$$

□

Definition (Kanonische Koeffizientenvektoren, Variate, Korrelationen)

Es seien

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} \text{ mit } \mathbb{E}(Z) := 0_m \text{ und } \mathbb{C}(Z) := \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{43}$$

ein m-dimensionaler partitionierter Zufallsvektor sowie sein Erwartungswert und seine Kovarianzmatrix, respektive. Weiterhin sei

$$K := \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \in \mathbb{R}^{m_x \times m_y} \tag{44}$$

mit der Singulärwertzerlegung

$$K = A \Lambda B^T, \tag{45}$$

wobei

$$A := \begin{pmatrix} \alpha_1 & \cdots & \alpha_k \end{pmatrix} \in \mathbb{R}^{m_x \times m_y} \text{ und } B := \begin{pmatrix} \beta_1 & \cdots & \beta_k \end{pmatrix} \in \mathbb{R}^{m_y \times m_y} \tag{46}$$

die orthogonalen Matrix der Eigenvektoren von KK^T und die orthogonale Matrix der Eigenvektoren von $K^T K$, respektive, bezeichnen und

$$\Lambda := \operatorname{diag}\left(\lambda_1^{1/2}, \ldots, \lambda_k^{1/2}\right) \in \mathbb{R}^{m_y \times m_y}, \tag{47}$$

die Diagonalmatrix der Quadratwurzeln der zugehörigen Eigenvektoren bezeichnet. Schließlich seien für $i = 1, \ldots, k$

$$a_i := \Sigma_{xx}^{-1/2} \alpha_i \in \mathbb{R}^{m_x} \text{ und } b_i := \Sigma_{yy}^{-1/2} \beta_i \in \mathbb{R}^{m_y}. \tag{48}$$

Dann heißen für $i = 1, \ldots, k$

(1) $a_i \in \mathbb{R}^{m_x}$ und $b_i \in \mathbb{R}^{m_y}$ die *$i$ten kanonischen Koeffizientenvektoren*,

(2) die Zufallsvektoren $\xi_i := a_i^T X$ und $\upsilon_i := b_i^T Y$ die $i$ten *$i$ten kanonischen Variaten* und

(3) $\rho_i := \lambda_i^{1/2}$ die *$i$te kanonische Korrelation*.

Theorem (Eigenschaften kanonischer Korrelationen und Variaten)

Es seien

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} \text{ mit } \mathbb{E}(Z) := 0_m \text{ und } \mathbb{C}(Z) := \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{49}$$

ein m-dimensionaler partitionierter Zufallsvektor sowie sein Erwartungswert und seine Kovarianzmatrix, respektive. Weiterhin seien für $i = 1, ..., k$ die kanonischen Koeffizientenvektoren a_i, b_i, die kanonischen Variaten ξ, υ_i und die kanonischen Korrelationen ρ_i definiert wie oben. Dann gilt, dass für $1 \le r \le k$ das Maximum des $r$ten restringierten Optimierungsproblems

$$\phi_r = \max_{a,b} a^T \Sigma_{xy} b \tag{50}$$

unter den Nebenbedingungen

$$a^T \Sigma_{xx} a = 1, \quad b^T \Sigma_{yy} b = 1, \quad a_i^T \Sigma_{xx} a = 0 \text{ für } i = 1, ..., r-1 \tag{51}$$

(1) den Wert $\phi_r = \rho_r$ hat und (2) bei $a = a_r$ und $b = b_r$ angenommen wird.

Bemerkungen

- ϕ_1 ist die größtmögliche Korrelation von $\xi = a^T X$ und $\upsilon = b^T Y$ unter den Nebenbedingungen
  - $\mathbb{V}(\xi) = a^T \Sigma_{xx} a = 1$ und $\mathbb{V}(\upsilon) = b^T \Sigma_{yy} b = 1$
- ϕ_r ist die größtmögliche Korrelation von $\xi = a^T X$ und $\upsilon = b^T Y$ unter den Nebenbedingungen
  - $\mathbb{V}(\xi) = a^T \Sigma_{xx} a = 1$ und $\mathbb{V}(\upsilon) = b^T \Sigma_{yy} b = 1$
  - $\mathbb{C}(\xi_i, \xi) = a_i^T \Sigma_{xx} a = 0$ für die ersten $i = 1, ..., r-1$ kanonischen Variaten ξ_i

Modellformulierung

Beweis

Wir betrachten das restringierte Optimierungsproblem

$$\phi_r^2 = \max_{a,b} \left(a^T \Sigma_{xy} b\right)^2 \text{ u.d.N. } a^T \Sigma_{xx} a = 1,\ b^T \Sigma_{yy} b = 1,\ a_i^T \Sigma_{xx} a = 0, i = 1, \ldots, r-1 \tag{52}$$

Wir folgen Mardia, Kent, and Bibby (1979), S. 284 und gehen schrittweise vor, d.h. wir lösen das restringierte Optimierungsproblem

$$\phi_r^2 = \max_{a} \left(\max_{b} \left(a^T \Sigma_{xy} b\right)^2 \text{ u.d.N.} b^T \Sigma_{yy} b = 1\right) \text{ u.d.N. } a^T \Sigma_{xx} a = 1,\ a_i^T \Sigma_{xy} a = 0, i = 1, \ldots, r-1 \tag{53}$$

von innen nach außen.

Schritt (1)

Wir wählen wir zunächst ein festes $a \in \mathbb{R}^m$ und betrachten das restringierte Optimierungsproblem

$$\max_{b} \left(a^T \Sigma_{xy} b\right)^2 \text{ u.d.N. } b^T \Sigma_{yy} b = 1 \tag{54}$$

Dieses Optimierungsproblem kann geschrieben werden als

$$\max_{b} b^T \Sigma_{yx} a a^T \Sigma_{xy} b \text{ u.d.N. } b^T \Sigma_{yy} b = 1, \tag{55}$$

weil gilt, dass

$$\left(a^T \Sigma_{xy} b\right)^2 = \left(a^T \Sigma_{xy} b\right) \left(a^T \Sigma_{xy} b\right) = \left(a^T \Sigma_{xy} b\right)^T a^T \Sigma_{xy} b = b^T \Sigma_{yx} a a^T \Sigma_{xy} b. \tag{56}$$

Beweis (fortgeführt)

Das Optimierungsproblem (55) kann nun mithilfe des Theorems zur Maximierung quadratischer Formen mit Nebenbedingungen gelöst werden. Im Sinne dieses Theorems setzen wir dazu

$$A := \Sigma_{yx} a a^T \Sigma_{xy} \text{ und } B := \Sigma_{yy}. \tag{57}$$

Dann hat (55) die Form

$$\max_b b^T A b \text{ unter der Nebenbedingung } b^T B b = 1, \tag{58}$$

Das Maximum von (58) entspricht nach dem Theorem zur Maximierung quadratischer Formen mit Nebenbedingungen dem größten Eigenwert von

$$B^{-1} A = \Sigma_{yy}^{-1} \Sigma_{yx} a a^T \Sigma_{xy} \tag{59}$$

Der größte Eigenwert von $\Sigma_{yy}^{-1} \Sigma_{yx} a a^T \Sigma_{xy}$ wiederum kann mithilfe des Theorems zum Eigenwert und Eigenvektor eines Matrixvektorprodukts bestimmt werden. Im Sinne dieses Theorems setzen wir dazu

$$A := \Sigma_{yy}^{-1} \Sigma_{yx}, \quad b := a, \quad B := \Sigma_{xy} \tag{60}$$

und erhalten für den betreffenden Eigenwert

$$\lambda_a = b^T B A a = a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a. \tag{61}$$

als Lösung (Maximum) des restringierten Optimierungsproblems

$$\max_b \left(a^T \Sigma_{xy} b\right)^2 \text{ u.d.N. } b^T \Sigma_{yy} b = 1 \tag{62}$$

Beweis (fortgeführt)

Schritt (2)

Basierend auf Schritt (1) verbleibt die Lösung des restringierten Optimierungsproblem

$$\phi_r^2 = \max_a a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \text{ u.d.N. } a^T \Sigma_{xx} a = 1,\ a_i^T \Sigma_{xx} a = 0, i = 1, ..., r-1 \tag{63}$$

Dazu halten wir zunächst fest, dass (63) mit den Definitionen von α_i und K in der Definition der Kanonischen Koeffizientenvektoren, Variaten, und Korrelationen geschrieben werden kann als

$$\phi_r^2 = \max_\alpha \alpha^T K K^T \alpha \text{ u.d.N. } \alpha^T \alpha = 1,\ \alpha_i^T \alpha = 0, i = 1, ..., r-1, \tag{64}$$

denn

$$\begin{aligned}
\phi_r^2 &= \max_a a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \text{ u.d.N. } a^T \Sigma_{xx} a = 1, a_i^T \Sigma_{xx} a = 0 \Leftrightarrow \\
\phi_r^2 &= \max_a a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \text{ u.d.N. } \alpha^T \Sigma_{xx}^{-1/2} \Sigma_{xx} \Sigma_{xx}^{-1/2} \alpha = 1, \alpha_i^T \Sigma_{xx}^{-1/2} \Sigma_{xx} \Sigma_{xx}^{-1/2} \alpha = 0 \\
\phi_r^2 &= \max_a \alpha^T \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1/2} \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \alpha_i^T \alpha = 0 \\
\phi_r^2 &= \max_a \alpha^T \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \Sigma_{yy}^{-1/2} \Sigma_{yx} \Sigma_{xx}^{-1/2} \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \alpha_i^T \alpha = 0 \\
\phi_r^2 &= \max_a \alpha^T K K^T \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \alpha_i^T \alpha = 0
\end{aligned} \tag{65}$$

Modellformulierung

Beweis (fortgeführt)

Dabei sind nach der betreffenden Definition die α_i die Eigenvektoren von KK^T mit den $i = 1, ..., r-1$ größten Eigenwerten. Nach dem Theorem zur Maximierung quadratischer Formen mit Nebenbedingungen ist die Lösung von (64) der größte Eigenwert von KK^T mit seinem assoziierten Eigenvektor. Die Nebenbedingung $\alpha_i^T \alpha = 0$ schränkt diese Wahl auf den rt-größten Eigenwert und seinen assoziierten Eigenvektor α_r ein. Mit der Definition von Eigenwerten und Eigenvektoren gilt also

$$\phi_r^2 = \alpha_r^T KK^T \alpha_r = \alpha_r^T \lambda_r \alpha_r = \lambda_r \alpha_r^T \alpha_r = \lambda_r. \tag{66}$$

Wir haben also gezeigt, dass das restringierte Optimierungsproblem des Theorems den Maximumwert $\phi_r = \lambda_r^{1/2}$ hat. Es bleibt zu zeigen, dass dieser Maximumwert für a_r und b_r angenommen wird.

Schritt (3)

Einsetzen von a_r und b_r in $a^T \Sigma_{xy} b$ ergibt mit

$$K = A\Lambda B^T \Leftrightarrow KB = A\Lambda B^T B \Leftrightarrow KB = A\Lambda \Leftrightarrow K\beta_r = \alpha_r \lambda_r^{1/2} \tag{67}$$

dass

$$a_r^T \Sigma_{xy} b_r = \alpha_r^T \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \beta_r = \alpha_r^T K \beta_r = \alpha_r^T \alpha_r \lambda_r^{1/2} = \rho_r \tag{68}$$

Also nimmt $a^T \Sigma_{xy} b$ bei a_r und b_r seinen restringierten Maximalwert λ_r an.

$\square$

Simulationsbeispiel

Wir betrachten das Beispiel (vgl. Uurtio et al. (2018))

$$p(X) = N(x; 0_4, I_4) \text{ und } p(Y|X) = N(y; LX, G) \tag{69}$$

mit

$$L := \begin{pmatrix} 0.0 & 0.0 & 1.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & -1.0 \end{pmatrix} \text{ und } G := \begin{pmatrix} 0.2 & 0.0 & 0.0 \\ 0.0 & 0.4 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix} \tag{70}$$

Hier gilt offenbar $m_x = 4, m_y = 3, m = 7$ und

$$\begin{aligned} Y_1 &= \quad X_3 + \varepsilon_1 \\ Y_2 &= \quad X_1 + \varepsilon_2 \\ Y_3 &= -X_4 + \varepsilon_3 \end{aligned} \tag{71}$$

mit

$$X_1 \sim N(0, 1), X_3 \sim N(0, 1), X_4 \sim N(0, 1) \tag{72}$$

und

$$\varepsilon_1 \sim N(0, 0.2), \varepsilon_2 \sim N(0, 0.4), \varepsilon_3 \sim N(0, 0.3) \tag{73}$$

Simulationsbeispiel

Mit dem Theorem zu gemeinsamen Normalverteilungen (vgl. Einheit (3) Matrizen) ergibt sich, dass

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N(0_7, \Sigma) \tag{74}$$

mit

$$\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}, \tag{75}$$

wobei

$$\Sigma_{xx} = I_4, \quad \Sigma_{xy} = L^T, \quad \Sigma_{yx} = L \text{ und } \Sigma_{yy} = G + LL^T. \tag{76}$$

Explizit ergibt sich also

$$\Sigma = \begin{pmatrix} I_4 & L^T \\ L & G + LL^T \end{pmatrix} = \begin{pmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & -1.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 1.2 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.4 & 0.0 \\ 0.0 & 0.0 & 0.0 & -1.0 & 0.0 & 0.0 & 1.3 \end{pmatrix} \tag{77}$$

Simulationsbeispiel

```r
# R Pakete für Matrizenrechnung
library(matlib)
library(expm)

# Modellparameter
L = matrix(c(0,0,1, 0,
             1,0,0, 0,
             0,0,0,-1),
           nrow  = 3,
           byrow = T)
G = diag(c(0.2,0.4,0.3))

# Kovarianzmatrixpartition
Sigma_xx = diag(4)
Sigma_xy = t(L)
Sigma_yx = L
Sigma_yy = G + L %*% t(L)
Sigma    = rbind(cbind(Sigma_xx, Sigma_xy), cbind(Sigma_yx, Sigma_yy))
print(Sigma)
```

```
>      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
> [1,]    1    0    0    0  0.0  1.0  0.0
> [2,]    0    1    0    0  0.0  0.0  0.0
> [3,]    0    0    1    0  1.0  0.0  0.0
> [4,]    0    0    0    1  0.0  0.0 -1.0
> [5,]    0    0    1    0  1.2  0.0  0.0
> [6,]    1    0    0    0  0.0  1.4  0.0
> [7,]    0    0    0   -1  0.0  0.0  1.3
```

Simulationsbeispiel

```
# Evaluation der iten kanonischen Koeffizientenvektoren und Korrelationen
K      = sqrtm(inv(Sigma_xx)) %*% Sigma_xy %*% sqrtm(inv(Sigma_yy)) # K
ALB    = svd(K)                                                     # K = A\LambdaV
A      = ALB$u                                                      # A
Lambda = ALB$d                                                      # Lambda
B      = ALB$v                                                      # B
rho    = Lambda                                                     # \rho_i = \lambda_i^{1/2}
a      = sqrtm(inv(Sigma_xx)) %*% A                                 # a_i = \Sigma_{xx}^{-1/2}\alpha_i
b      = sqrtm(inv(Sigma_yy)) %*% B                                 # b_i = \Sigma_{yy}^{-1/2}\beta_i
```

Die kanonische Korrelationen und kanonischen Koeffizientenvektoren ergeben sich zu

```
> rho_1 =  0.913 , a_1^T = ( 0 0 -1 0 ), b_1^T = ( -0.913 0 0 )
> rho_2 =  0.877 , a_2^T = ( 0 0 0 1 ) , b_2^T = ( 0 0 -0.877 )
> rho_3 =  0.845 , a_3^T = ( -1 0 0 0 ), b_3^T = ( 0 -0.845 0 )
```

Vorbemerkungen

Modellformulierung

Modellschätzung

Modellevaluation

Definition (Schätzer kanonischer Korrelationen und Koeffizientenvektoren)

Für $i = 1, ..., n$ seien

$$Z_i = \begin{pmatrix} X_i \\ Y_i \end{pmatrix} \text{ mit } \mathbb{E}(Z_i) := 0_m \text{ und } \mathbb{C}(Z_i) := \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{78}$$

unabhängig und identisch verteilte m-dimensionale partitionierte Zufallsvektoren sowie ihr Erwartungswert und ihre Kovarianzmatrix, respektive, und

$$C := \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{79}$$

sei ihre Stichprobenkovarianzmatrix. Dann sind für $i = 1, ..., k := \min\{m_x, m_y\}$

$$\hat{a}_i := C_{xx}^{-1/2} \hat{\alpha}_i \in \mathbb{R}^{m_x}, \quad \hat{b}_i := C_{yy}^{-1/2} \hat{\beta}_i \in \mathbb{R}^{m_y} \text{ und } \hat{\rho}_i := \hat{\lambda}_i^{1/2} \tag{80}$$

Schätzer der $i$ten kanonischen Koeffizientenvektoren und kanonischen Korrelationen, respektive. Dabei sind mit

$$\hat{K} := C_{xx}^{-1/2} C_{xy} C_{yy}^{-1/2} \in \mathbb{R}^{m_x \times m_y} \tag{81}$$

$\hat{\alpha}_i$ und $\hat{\lambda}_i$ der $i$te Eigenvektor und sein zugehöriger Eigenwert von $\hat{K}\hat{K}^T$ und $\hat{\beta}_i$ der entsprechende Eigenkvektor von $\hat{K}^T\hat{K}$.

Bemerkungen

- Zur Modellschätzung wird $\mathbb{C}(Z)$ also durch C ersetzt.

Simulationsbeispiel

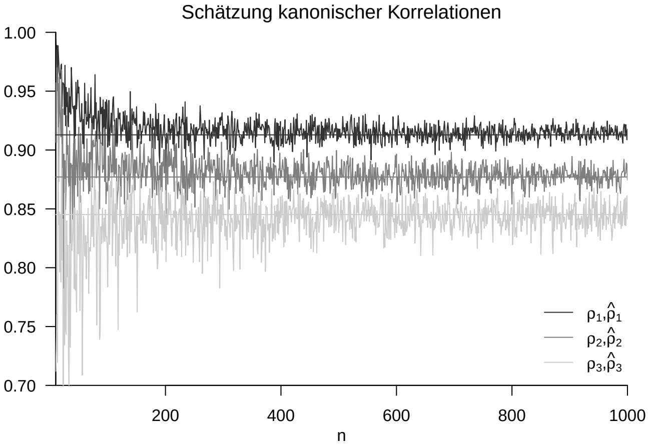
```r
# R Pakete
library(MASS)
library(matlib)
library(expm)

# Modellparameter
m_x      = 4
m_y      = 3
k        = min(m_x,m_y)
L        = matrix(c(0,0,1,0,1,0,0,0,0,0,0,-1), nrow = 3,byrow = 3)
G        = diag(c(0.2,0.4,0.3))
Sigma_xx = diag(4)
Sigma_xy = t(L)
Sigma_yx = L
Sigma_yy = G + L %*% t(L)
Sigma    = rbind(cbind(Sigma_xx, Sigma_xy), cbind(Sigma_yx, Sigma_yy))
K        = sqrtm(inv(Sigma_xx)) %*% Sigma_xy %*% sqrtm(inv(Sigma_yy))
ALB      = svd(K)
A        = ALB$u
Lambda   = ALB$d
B        = ALB$v
rho      = Lambda
a        = sqrtm(inv(Sigma_xx)) %*% A
b        = sqrtm(inv(Sigma_yy)) %*% B
```

Simulationsbeispiel

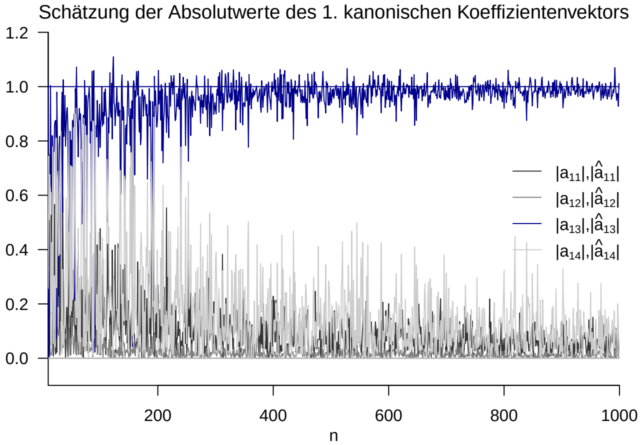
```
# Simulationen
n       = 1e1:1e3
rho_hat = matrix(rep(NaN, length(n)*k)  , nrow = k)
a_1_hat = matrix(rep(NaN, length(n)*m_x), nrow = m_x)
for(i in 1:length(n)){

    # Datengeneration
    Y           = t(mvrnorm(n[i],rep(0, m_x+m_y),Sigma))
    I_n         = diag(n[i])
    J_n         = matrix(rep(1,n[i]^2), nrow = n[i])

    # Stichprobenkovarianzmatrixpartition
    C           = (1/(n[i]-1))*(Y %*% (I_n-(1/n[i])*J_n) %*% t(Y))
    C_xx        = C[1:m_x,1:m_x]
    C_xy        = C[1:m_x,(m_x+1):(m_x+m_y)]
    C_yx        = C[(m_x+1):(m_x+m_y),1:m_x]
    C_yy        = C[(m_x+1):(m_x+m_y),(m_x+1):(m_x+m_y)]

    # Kanonische Korrelationsanalyse
    K_hat        = sqrtm(inv(C_xx)) %*% C_xy %*% sqrtm(inv(C_yy))
    ALB_hat      = svd(K_hat)
    A_hat        = ALB_hat$u
    Lambda_hat   = ALB_hat$d
    B_hat        = ALB_hat$v
    a_hat        = sqrtm(inv(C_xx)) %*% A_hat
    b_hat        = sqrtm(inv(C_yy)) %*% B_hat
    rho_hat[,i]  = as.matrix(Lambda_hat)
    a_1_hat[,i]  = a_hat[,1]
}
```

Modellschätzung

Simulationsbeispiel



Simulationsbeispiel



Modellschätzung

Anwendungsbeispiel

Wir betrachten erneut den Datensatz nach Rudolf and Buse (2020) Kapitel 4

Wir betrachten die psychodiagnostischen Daten der $n = 45$ Studierenden

- X_1 Intelligenztestscore
- X_2 Mathematiktestscore
- Y_1 Gewissenhaftigkeitscore
- Y_2 Verträglichkeitscore

als 45 unabhängige Realisierungen eines 4-dimensionaler Zufallsvektors Z. Wir sind also an den kanonischen Korrelationen der Kognitionstestwerte (Intelligenz, Mathematik) mit den Charaktertestwerten (Gewissenhaftigkeit, Veträglichkeit) interessiert.

Anwendungsbeispiel

Daten der ersten 12 Proband:innen

x1	x2	y1	y2
54	44	31	60
60	20	33	31
67	36	26	54
41	39	31	26
66	57	34	56
51	28	42	23
51	46	34	40
37	46	36	31
57	54	28	49
47	12	34	41
50	67	27	53
42	63	26	36

Anwendungsbeispiel

Kanonische Korrelationsanalyse

```r
# Datenpräprozessierung
D          = read.spss(file.path(getwd(), "12_Daten", "studienerfolg.sav"), to.data.frame = T)
x          = as.matrix(cbind(D$X1, D$X2))
y          = as.matrix(cbind(D$X3, D$X4))
n          = nrow(x)
m_x        = ncol(x)
m_y        = ncol(y)
Y          = rbind(t(x),t(y))

# Stichprobenkovarianzmatrixpartition
I_n        = diag(n)
J_n        = matrix(rep(1,n^2), nrow = n)
C          = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))
C_xx       = C[1:m_x,1:m_x]
C_xy       = C[1:m_x,(m_x+1):(m_x+m_y)]
C_yx       = C[(m_x+1):(m_x+m_y),1:m_x]
C_yy       = C[(m_x+1):(m_x+m_y),(m_x+1):(m_x+m_y)]

# Kanonische Korrelationsanalyse
K_hat      = sqrtm(inv(C_xx)) %*% C_xy %*% sqrtm(inv(C_yy))
ALB_hat    = svd(K_hat)
A_hat      = ALB_hat$u
Lambda_hat = ALB_hat$d
B_hat      = ALB_hat$v
a_hat      = sqrtm(inv(C_xx)) %*% A_hat
b_hat      = sqrtm(inv(C_yy)) %*% B_hat
rho_hat    = as.matrix(Lambda_hat)
```

```
> rho_hat_1 :  0.245
> a_hat_1   :  0.0152 0.0497
> b_hat_1   :  0.143 0.0412
> rho_hat_2 :  0.133
> a_hat_2   :  0.0905 -0.0132
> b_hat_2   :  -0.0484 0.0764
```

Modellschätzung

Anwendungsbeispiel

Kanonische Korrelationsanalyse mir R's `cancor()` Funktion

```
# Datenpräprozessierung
D           = read.spss(file.path(getwd(), "12_Daten", "studienerfolg.sav"), to.data.frame = T)
x           = as.matrix(cbind(D$X1, D$X2))
y           = as.matrix(cbind(D$X3, D$X4))
cca         = cancor(x,y)
```

```
> $cor
> [1] 0.245 0.133
>
> $xcoef
>         [,1]     [,2]
> [1,] 0.00229  0.01364
> [2,] 0.00749 -0.00199
>
> $ycoef
>         [,1]    [,2]
> [1,] 0.02161 -0.0073
> [2,] 0.00621  0.0115
>
> $xcenter
> [1] 59.0 57.2
>
> $ycenter
> [1] 36.4 39.1
```

Modellschätzung

Anwendungsbeispiel

Die geschätzte maximale Korrelation von Linearkombinationen von (x_1, x_2) und (y_1, y_2) ist 0.25.

- (x_1, x_2) und (y_1, y_2) sind multivariat also "gering bis mäßig" korreliert.

Basierend auf der simulationsvalidierten Schätzung ergibt sich

- $\xi = 0.02X_1 + 0.05X_2$ als "bester Prädiktor"
- $\upsilon = 0.09Y_1 - 0.01Y_2$ als "am besten prädizierbares Kriterium"

"Mathematikfähigkeiten" scheinen zur Prädiktion der betrachteten "Charaktereigenschaften" etwas wichtiger als "Intelligenz"; bei den betrachteten "Charaktereigenschaften" trägt "Gewissenhaftigkeit" mehr zum Kriterium bei als "Verträglichkeit".

Vorbemerkungen

Modellformulierung

Modellschätzung

Modellevaluation

Überblick

Ein Ziel der Modellevaluation bei der Kanonischen Korrelationsanalyse kann das Testen von

$$H_0 : \Sigma_{xy} = 0_{m_x m_y} \tag{82}$$

sein. Diese Nullhypothese besagt, dass zwischen keiner der Variablen $X_1, ..., X_{m_x}$ und $Y_1, ..., Y_{m_y}$ eine lineare Abhängigkeit besteht. Dies impliziert, dass alle kanonischen Korrelationen gleich 0 sind, denn es gilt

$$\Sigma_{xy} = 0_{m_x m_y} \Rightarrow K = 0_{m_x m_y} \Rightarrow \Lambda = 0_{m_y m_y}. \tag{83}$$

Die Alternativhypothese zu dieser Nullhypothese lautet

$$H_1 : \Sigma_{xy_{i,j}} \neq 0 \text{ für mindestens ein Paar } (i,j) \text{ mit } 1 \leq i \leq m_x \text{ und } 1 \leq j \leq m_y. \tag{84}$$

Die Alternativhypothese besagt also, dass mindestens eine der Variablen $X_1, ..., X_{m_x}$ und eine der Variablen $Y_1, ..., Y_{m_y}$ linear abhängig sind und damit nicht alle kanonischen Korrelationen gleich null sind.

Wie bei der einfaktoriellen Varianzanalyse (und generell im Kontext multivariater allgemeiner lineare Modelle) können kritische Wert-basierte Tests der obigen Nullhypothese mit verschiedenen Teststatistiken (*Wilks'* Λ, *Pillai Statisik*) konstruiert werden, der Verteilungen wiederum nur in Speziallfällen analytisch beschrieben sind und ansonsten mit f-Verteilungen approximiert werden.

Wir betrachten hier lediglich das Testen der obigen Nullhypothese mit der Wilk's Λ Statistik.

Theorem (Wilks' Λ für die kanonische Korrelationsanalyse)

Es seien das Modell der kanonischen Korrelationsanalyse, die Partition der Stichprobenkovarianzmatrix, und die Schätzer der kanonischen Korrelationen definiert wie oben. Dann hat die die Wilks' Λ Teststatistik die Form

$$\Lambda = \frac{|C|}{|C_{xx}||C_{yy}|} = \prod_{i=1}^{k} \left(1 - \hat{\rho}_i^2\right), \tag{85}$$

wobei $|\cdot|$ die Determinante bezeichnet. Weiterhin ist für

$$H_0 : \Sigma_{xy} = 0_{m_x m_y} \tag{86}$$

die Statistik

$$\tau := \frac{1 - \Lambda^{1/t}}{\Lambda^{1/t}} \frac{\nu_2}{\nu_1} \tag{87}$$

mit

$$\nu_1 := m_x m_y \text{ und } \nu_2 := wt - \frac{1}{2} m_x m_y + 1 \tag{88}$$

sowie

$$w := n - \frac{1}{2}(m_x + m_y + 3) \text{ und } t := \sqrt{\frac{m_x^2 m_y^2 - 4}{m_x^2 + m_y^2 - 5}} \tag{89}$$

approximativ f-verteilt mit den Freiheitsgradparametern ν_1 und ν_2.

Bemerkungen

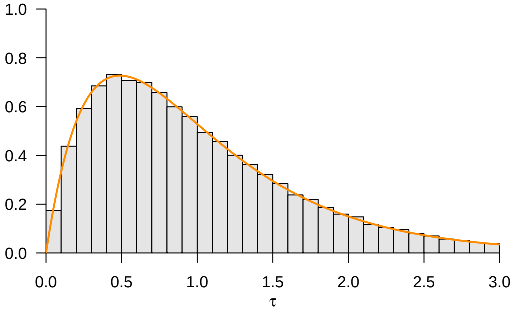
- Wir verzichten auf Beweise aller Aussagen dieses Theorems.
- Λ wird klein, wenn die absoluten Werte der Einträge in C_{xy} groß werden.
- Man denke in diesem Zusammenhang an das Berechnen der Determinante einer 2×2 Matrix.
- Λ wird klein, wenn zumindest ein $\hat{\rho}_i$ groß ist
- Für $\hat{\rho}_i = 0$ für alle $i = 1, ..., k$ gilt $\Lambda = 1$; für $\hat{\rho}_i = 1$ für alle $i = 1, ..., k$ gilt $\Lambda = 0$
- Kleine Werte von Λ und damit große Werte von τ sprechen also gegen H_0.

Simulationsbeispiel

```r
# Modellparameter
library(MASS)
m_x      = 2
m_y      = 2
Sigma_xx = diag(m_x)
Sigma_xy = matrix(rep(0,m_x*m_y), nrow = 2)            # H_0 : \Sigma_{xy} = 0_{m_xm_y}
Sigma_yx = Sigma_xy
Sigma_yy = diag(m_y)
Sigma    = rbind(cbind(Sigma_xx, Sigma_xy),
                 cbind(Sigma_yx, Sigma_yy))
# Testszenarioparameter
n        = 45                                          # Anzahl Datenpunkte
w        = n-((1/2)*(m_x+m_y+3))
t        = sqrt((m_x^2*m_y^2 - 4)/((m_x^2+m_y^2)-5))
nu_1     = m_x*m_y
nu_2     = w*t-((1/2)*m_x*m_y) + 1
alpha_0  = 0.05                                        # Signifikanzlevel
kW       = qf(1-alpha_0,nu_1,nu_2)                     # kritischer Wert

# Simulationen
sim      = 1e5                                         # Anzahl an Simulationen
tau      = rep(NaN, sim)                               # Teststatistikarray
phi      = rep(0,   sim)                               # Testarray
for(s in 1:sim){
    Y      = t(mvrnorm(n,rep(0, m_x+m_y),Sigma))       # Datengeneration
    I_n    = diag(n)
    J_n    = matrix(rep(1,n^2), nrow = n)
    C      = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y)) # Stichprobenkovarianzmatrix
    C_xx   = C[1:m_x,1:m_x]
    C_xy   = C[1:m_x,(m_x+1):(m_x+m_y)]
    C_yy   = C[(m_x+1):(m_x+m_y),(m_x+1):(m_x+m_y)]
    Lambda = det(C)/(det(C_xx)*det(C_yy))              # Wilk's Lambda
    tau[s] = ((1-Lambda^(1/t))/Lambda^(1/t))*(nu_2/nu_1) # \tau
    phi[s] = tau[s] > kW                               # Test
}
```

Simulationsbeispiel

$$n = 45, m_x = 2, m_y = 2, \nu_1 = 4, \nu_2 = 82, k = 2.48, \hat{\alpha} = 0.0498$$



Simulationsbeispiel

```
# Datenpräprozessierung
library(foreign)
D          = read.spss(file.path(getwd(), "12_Daten", "studienerfolg.sav"), to.data.frame = T)
x          = as.matrix(cbind(D$X1, D$X2))
y          = as.matrix(cbind(D$X3, D$X4))
n          = nrow(x)
m_x        = ncol(x)
m_y        = ncol(y)
Y          = rbind(t(x),t(y))

# Testszenarioparameter
w        = n-((1/2)*(m_x+m_y+3))
t        = sqrt((m_x^2*m_y^2 - 4)/((m_x^2+m_y^2)-5))
nu_1     = m_x*m_y
nu_2     = w*t-((1/2)*m_x*m_y) + 1
alpha_0  = 0.05
k        = qf(1-alpha_0,nu_1,nu_2)

# Stichprobenkovarianzmatrixpartition
I_n        = diag(n)
J_n        = matrix(rep(1,n^2), nrow = n)
C          = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))
C_xx       = C[1:m_x,1:m_x]
C_xy       = C[1:m_x,(m_x+1):(m_x+m_y)]
C_yx       = C[(m_x+1):(m_x+m_y),1:m_x]
C_yy       = C[(m_x+1):(m_x+m_y),(m_x+1):(m_x+m_y)]

# Nullhypothesentest
Lambda     = det(C)/(det(C_xx)*det(C_yy))
tau        = ((1-Lambda^(1/t))/Lambda^(1/t))*(nu_2/nu_1)
phi        = tau > k
pval       = 1 - pf(tau,nu_1,nu_2)
```

```
> Lambda : 0.924
> tau    : 0.831
> k      : 2.48
> phi    : 0
> p      : 0.509
```

$\Rightarrow H_0 : \Sigma_{xy} = 0_{m_x m_y}$ wird nicht verworfen.

References

Anderson, T. W. 2003. *An Introduction to Multivariate Statistical Analysis*. 3rd ed. Wiley Series in Probability and Statistics. Hoboken, N.J: Wiley-Interscience.

Mardia, K. V., J. T. Kent, and J. M. Bibby. 1979. *Multivariate Analysis*. Probability and Mathematical Statistics. London ; New York: Academic Press.

Rudolf, Matthias, and Johannes Buse. 2020. *Multivariate Verfahren*. Göttingen: Hogrefe.

Uurtio, Viivi, João M. Monteiro, Jaz Kandola, John Shawe-Taylor, Delmiro Fernandez-Reyes, and Juho Rousu. 2018. "A Tutorial on Canonical Correlation Methods." *ACM Computing Surveys* 50 (6): 1–33. https://doi.org/10.1145/3136624.