

Kanonische Korrelationsanalyse

Anwendungsszenario und Grundzüge

Ausgangspunkt einer kanonischen Korrelationsanalyse ist die exploratorische Frage nach der Stärke des Zusammenhangs einer multivariaten unabhängigen Variablen x und einer multivariaten abhängigen Variable y . Im Kontext der kanonischen Korrelationsanalyse werden diese Variablen auch oft als *Prädiktoren* und *Kriterien*, respektive, bezeichnet. Dabei werden vorliegende Daten von sowohl der Prädiktoren als auch der Kriterien als Realisierungen von unabhängig und identisch verteilten m_x - bzw. m_y -dimensionalen Zufallsvektoren betrachtet. Um die Frage nach der Stärke des Zusammenhangs von Prädiktoren und Kriterien zu beantworten, betrachtet die kanonische Korrelationsanalyse Linearkombinationen der Komponenten der Prädiktoren und Kriterien. Wir bezeichnen die Linearkombinationen von x und y mit Parametervektoren $a \in \mathbb{R}^{m_x}$ und $b \in \mathbb{R}^{m_y}$ im Folgenden mit

$$\xi := a^T x \text{ und } v := b^T y \quad (1)$$

ξ und v sind dann als Linearkombinationen von Zufallsvariablen selbst Zufallsvariablen, deren Korrelation $\rho(\xi, v)$ berechnet werden kann. Die kanonische Korrelationsanalyse bestimmt dann die Parametervektoren a und b so, dass die Korrelation von ξ und v so groß wie möglich wird. Ist eine solche Parametervektorkombination und ihre entsprechende Korrelation, die dann als kanonische Korrelation bezeichnet gefunden, so kann ξ als bester Prädiktor und v als “am besten prädizierbares Kriterium” interpretiert werden.

Für Skalare α und β sind die Korrelationen $\rho(\xi, v)$ und $\rho(\alpha\xi, \beta v)$ allerdings, wie im Theorem zu Kovarianz und Korrelation bei linear-affinen Transformationen gezeigt, identisch. Die kanonische Korrelationsanalyse sucht deshalb speziell Parametervektoren a und b , für die $\rho(\xi, v)$ einerseits maximal ist und für die ξ und v gleichzeitig jeweils eine standardisierte Varianz von 1 haben. Da aufgrund des Theorem zu Kovarianz und Korrelation bei linear-affinen Transformationen die Varianzen zu verschiedenen skalaren Vielfachen von ξ und v verschieden sind, legt diese die Parametervektorwerte a und b , für die $\rho(\xi, v)$ maximal ist, eindeutig fest. Zur Bestimmung einer kanonischen Korrelation und der Parametervektoren von a und b ist man also auf ein Optimierungsproblem mit Nebenbedingungen geführt.

In unserer Darstellung kanonischen Korrelationsanalyse folgen wir @mardia1979. Dabei werden die Prädiktor- und Kriterienzufallsvektoren x und y in einem Zufallsvektor

$$z := \begin{pmatrix} x \\ y \end{pmatrix} \quad (2)$$

zusammengefasst, für den wir durchgängig annehmen wollen, dass $\mathbb{E}(z) = 0_m$ mit $m = m_x + m_y$ gilt. Im Anwendungskontext entspricht dies der Annahme, dass vor Durchführung der kanonischen Korrelationsanalyse die Stichprobenmittel des Stichprobenmittels von den beobachteten Daten vor Durchführung der kanonischen Korrelationsanalyse subtrahiert werden. Da, wie wir

sehen werden, die Schätzung der kanonischen Korrelation allerdings lediglich auf den Stichprobenkovarianzmatrizen beruht ist dieser Schritt verzichtbar.

Der mathematische Fokus der Entwicklung der kanonischen Korrelationsanalyse nach @mardia1979 liegt damit auf der Kovarianzmatrix $\mathbb{C}(z)$. Speziell ergeben sich die Kovarianzen von Linearkombinationen von x und y aus Matrixprodukten von $\mathbb{C}(z)$ und es können einige Matrixtheoreme, die in ?@sec-zufallsvektoren und ?@sec-eigenanalyse dokumentiert sind angewendet werden. Generell wird in der Entwicklung nach @mardia1979 ein Optimierungsansatz mithilfe von Lagrangefunktionen, wie in den Originalarbeiten von @hotelling1935 und @hotelling1936 gewählt, zugunsten der Eigenanalyse von Matrixprodukten supprimiert. Für die Entwicklung mit einem Lagrangeansatz verweisen wir auf zum Beispiel @anderson2003.

Im Folgenden diskutieren wir in Section zunächst zwei Theoreme, die direkt durch die kanonische Korrelationsanalyse motiviert sind und die, zusammen mit den oben erwähnten Theoremen in ?@sec-zufallsvektoren und ?@sec-eigenanalyse das analytische und probabilistische Fundament der kanonischen Korrelationsanalyse bilden. In Section führen wir mit den *kanonischen Korrelationen*, den *kanonischen Variaten* und den *kanonischen Koeffizientenvektoren* dann die zentralen Begriffe der kanonischen Korrelationsanalyse ein. In Section diskutieren wir dann, wie diese Größen auf Basis der Stichprobenkovarianzmatrix eines Datensatzes von Prädiktoren- und Kriterienrealisierungen geschätzt werden können. In Section schließlich wenden wir die kanonische Korrelationsanalyse auf das unten skizzierte Anwendungsbeispiel an.

Anwendungsbeispiel

Als konkretes Anwendungsbeispiel für eine kanonische Korrelationsanalyse betrachten wir einen simulierten Datensatz zur Effektivität einer Psychotherapie bei Depressionen. Dabei seien pro Patient:in vier Variablen verfügbar: Zum einen als Maße für die Therapiequalität die Dauer der Psychotherapie (DUR) und die klinische Erfahrung der behandelnden Psychotherapeut:in (EXP); zum anderen als Maße für die Reduktion der Depressionssymptomatik sowohl BDI Score und Glukokortikoidplasmalevel Differenzwerte zwischen Beginn und Ende der Therapie (dBDI und dGLU), wobei positive Werte eine Verringerung der Depressionssymptomatik anzeigen sollen. Wir stellen uns vor, dass man im Sinne einer exploratorischen Analyse daran interessiert ist, inwieweit die Variablen DUR (x_1) und EXP (x_2) als Prädiktoren (unabhängige Variablen) mit den Variablen dBDI (y_1) und dGLU (y_2) als Kriterien (abhängige Variablen) zusammenhängen. Im Sinne obiger Bezeichner gilt hier also $m_x = m_y = 2$. Table 1 zeigt dazu einen Beispieldatensatz für $n = 20$ Patient:innen.

Table 1: Beispielhafte Prädiktoren- und Kriterienrealisierungen im Kontext der Psychotherapie bei Depressionen

DUR	EXP	dBDI	dGLU
27.9	7.8	35.5	6.1
15.3	9.3	25.0	4.0
17.4	2.1	19.7	1.7
21.5	6.5	28.8	2.6
28.2	1.3	29.4	1.9
14.0	2.7	17.2	0.9
28.0	3.9	32.9	2.0
28.9	0.1	28.3	4.1
23.2	3.8	25.8	3.9
22.6	8.7	31.3	3.8
11.2	3.4	14.4	2.1
14.1	4.8	18.4	2.0
13.5	6.0	19.1	5.0
23.7	4.9	28.0	2.6
17.7	1.9	20.3	2.1
25.4	8.3	34.8	4.4
20.0	6.7	27.6	4.0
24.4	7.9	31.9	3.9
29.8	1.1	32.2	1.0
17.6	7.2	24.6	1.9

Mathematischer Hintergrund

Folgendes Theorem bildet das analytische Fundament der kanonischen Korrelationsanalyse.

Theorem 0.1 (Maximierung quadratischer Formen mit Nebenbedingungen). $A \in \mathbb{R}^{m \times m}$, $B \in \mathbb{R}^{m \times m}$ *pd* 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

$$\max_x x^T A x \text{ unter der Nebenbedingung } x^T B x = 1. \quad (3)$$

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

$$y := B^{1/2}x \Leftrightarrow x = B^{-1/2}y \quad (4)$$

Dann kann mit der symmetrischen Matrix

$$K := B^{-1/2}AB^{-1/2} \in \mathbb{R}^{m \times m} \quad (5)$$

das Optimierungsproblem (3) geschrieben werden als

$$\max_y y^T K y \text{ unter der Nebenbedingung } y^T y = 1. \quad (6)$$

Dies gilt, weil

$$\max_x x^T A x \Leftrightarrow \max_y (B^{-1/2} y)^T A (B^{-1/2} y) \Leftrightarrow \max_y y^T B^{-1/2} A B^{-1/2} y \Leftrightarrow \max_y y^T K y \quad (7)$$

und

$$x^T B x = 1 \Leftrightarrow y^T B^{-1/2} B B^{-1/2} y = 1 \Leftrightarrow y^T y = 1. \quad (8)$$

Weil K eine symmetrische Matrix ist, existiert die Orthonormalzerlegung (vgl. **sec-eigenanalyse**)

$$K = Q \Lambda Q^T, \quad (9)$$

wobei die Spalten der orthogonalen Matrix Q die Eigenvektoren von K und die Diagonalelemente von Λ die zugehörigen Eigenwerte von K sind. Mit der orthogonalen Matrix Q aus obiger Orthonormalzerlegung sei nun

$$z := Q^T y \Leftrightarrow y := Q z. \quad (10)$$

Dann kann das Optimierungsproblem (6) geschrieben werden als

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \text{ unter der Nebenbedingung } z^T z = 1, \quad (11)$$

weil

$$\max_y y^T K y \Leftrightarrow \max_z (Q z)^T K (Q z) \Leftrightarrow \max_z z^T Q^T Q \Lambda Q^T Q z \Leftrightarrow \max_z z^T \Lambda z \Leftrightarrow \max_z \sum_{i=1}^m \lambda_i z_i^2 \quad (12)$$

und

$$y^T y = 1 \Leftrightarrow (Q z)^T Q z = 1 \Leftrightarrow z^T Q^T Q z = 1 \Leftrightarrow z^T z = 1. \quad (13)$$

Die Eigenwerte von K seien nun absteigend sortiert, also $\lambda_1 \geq \dots \geq \lambda_m$. Dann gilt für das Optimierungsproblem (11), dass

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \leq \lambda_1, \quad (14)$$

weil

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \leq \max_z \sum_{i=1}^m \lambda_1 z_i^2 = \lambda_1 \max_z \sum_{i=1}^m z_i^2 = \lambda_1 \quad (15)$$

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

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 = \lambda_1, \quad (16)$$

für $z := e_1 = (1, 0, \dots, 0)^T$. Zusammenfassend heißt das, dass $z = e_1$ eine Lösung des Optimierungsproblem (11) ist und das λ_1 das entsprechende Maximum ist. Damit ergibt sich aber sofort, dass dann

$$y = Q z = Q e_1 = q_1 \text{ und } x = B^{-1/2} q_1 \quad (17)$$

Lösungen der äquivalenten Optimierungsprobleme (6) und (3), 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

$$B^{-1/2}B^{-1/2}A = B^{-1}A \quad (18)$$

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

$$\max_x x^T A x \text{ unter der Nebenbedingung } x^T B x = 1. \quad (19)$$

ist. □

Nach Wortlaut des Theorems gilt also für die Funktion

$$f : \mathbb{R}^m \rightarrow \mathbb{R}, x \mapsto f(x) := x^T A x, \quad (20)$$

dass

$$v_1 = \operatorname{argmax}_x x^T A x \text{ unter der Nebenbedingung } x^T B x = 1 \quad (21)$$

und dass

$$\lambda_1 = \max_x x^T A x \text{ unter der Nebenbedingung } x^T B x = 1. \quad (22)$$

Das folgende Theorem setzt die für die kanonischen Korrelationsanalyse zentralen Größen der Varianzen von Linearkombinationen von Zufallsvektoren und der Korrelation von Linearkombinationen von Zufallsvektoren in Bezug zur Kovarianzmatrix der gemeinsamen Verteilung der Zufallsvektoren und bildet das probabilistische Fundament der kanonischen Korrelationsanalyse.

Theorem 0.2 (Linearkombinationen von Zufallsvektorpartitionen).

Es sei

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \text{ mit } \mathbb{E}(z) = 0_m \text{ und } \mathbb{C}(z) = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \quad (23)$$

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

$$\xi := a^T x \text{ und } v := b^T y \quad (24)$$

als Linearkombinationen der Komponenten von x und y definiert. Dann gelten

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

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

$$\begin{aligned}
\mathbb{V}(\xi) &= \mathbb{E}(\xi\xi) - \mathbb{E}(\xi)\mathbb{E}(\xi) \\
&= \mathbb{E}((a^T x)(a^T x)) - \mathbb{E}(a^T x)\mathbb{E}(a^T x) \\
&= \mathbb{E}((a^T x)(a^T x)^T) - \mathbb{E}(a^T x)\mathbb{E}(a^T x) \\
&= \mathbb{E}(a^T x x^T a) - \mathbb{E}(a^T x)\mathbb{E}(a^T x) \\
&= a^T \mathbb{E}(x x^T) a - a^T \mathbb{E}(x) a^T \mathbb{E}(x) \\
&= a^T \mathbb{E}(x x^T) a - a^T 0_{m_x} a^T 0_{m_x} \\
&= a^T \Sigma_{xx} a.
\end{aligned} \tag{25}$$

Der Beweis zur Varianz von v folgt dann analog. Mit der Definition der Korrelation von Zufallsvariablen und mit $\mathbb{V}(\xi) = \mathbb{V}(v) = 1$ und dem Kovarianzverschiebungssatz gilt

$$\begin{aligned}
\rho(\xi, v) &= \frac{\mathbb{C}(\xi, v)}{\sqrt{\mathbb{V}(\xi)}\sqrt{\mathbb{V}(v)}} \\
&= \frac{\mathbb{C}(\xi, v)}{\sqrt{1}\sqrt{1}} \\
&= \mathbb{C}(\xi, v) \\
&= \mathbb{E}(\xi v) - \mathbb{E}(\xi)\mathbb{E}(v) \\
&= \mathbb{E}((a^T x)(b^T y)) - \mathbb{E}(a^T x)\mathbb{E}(b^T y) \\
&= \mathbb{E}((a^T x)(b^T y)^T) - \mathbb{E}(a^T x)\mathbb{E}(b^T y) \\
&= \mathbb{E}(a^T x y^T b) - \mathbb{E}(a^T x)\mathbb{E}(b^T y) \\
&= a^T \mathbb{E}(x y^T) b - a^T \mathbb{E}(x) b^T \mathbb{E}(y) \\
&= a^T \mathbb{E}(x y^T) b - a^T 0_{m_x} b^T 0_{m_y} \\
&= a^T \Sigma_{xy} b.
\end{aligned} \tag{26}$$

□

Die Varianz der Zufallsvariable $a^T x$ ergibt sich also als die “quadrierte” Linearkombination von Σ_{xx} und die Varianz der Zufallsvariable $b^T y$ ergibt sich als die “quadrierte” Linearkombination von Σ_{yy} . Die Korrelation der Zufallsvariablen $a^T x$ und $b^T y$ schließlich ergibt sich “quadrierte” Linearkombination von Σ_{xy} .

Modellformulierung

Mit den beiden in Section diskutierte Theoremen ist es nun möglich, die Begriffe der *kanonischen Koeffizientenvektoren*, der *kanonischen Variate*, und schließlich der *kanonischen Korrelationen* zu definieren.

Definition 0.1 (Kanonische Koeffizientenvektoren, Variate, Korrelationen).

Es seien

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \text{ mit } \mathbb{E}(z) := 0_m \text{ und } \mathbb{C}(z) := \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \quad (27)$$

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

$$K := \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \in \mathbb{R}^{m_x \times m_y} \quad (28)$$

mit der Singulärwertzerlegung

$$K = A \Lambda B^T, \quad (29)$$

wobei

$$A := (\alpha_1 \ \dots \ \alpha_k) \in \mathbb{R}^{m_x \times m_y} \text{ und } B := (\beta_1 \ \dots \ \beta_k) \in \mathbb{R}^{m_y \times m_y} \quad (30)$$

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

$$\Lambda := \text{diag}(\lambda_1^{1/2}, \dots, \lambda_k^{1/2}) \in \mathbb{R}^{m_y \times m_y}, \quad (31)$$

die Diagonalmatrix der Quadratwurzeln der zugehörigen absteigend geordneten Eigenwerte bezeichnet. Schließlich seien für $i = 1, \dots, k$

$$a_i := \Sigma_{xx}^{-1/2} \alpha_i \in \mathbb{R}^{m_x} \text{ und } b_i := \Sigma_{yy}^{-1/2} \beta_i \in \mathbb{R}^{m_y}. \quad (32)$$

Dann heißen für $i = 1, \dots, k$

- (1) $a_i \in \mathbb{R}^{m_x}$ und $b_i \in \mathbb{R}^{m_y}$ die *iten kanonischen Koeffizientenvektoren*,
- (2) die Zufallsvektoren $\xi_i := a_i^T x$ und $v_i := b_i^T y$ die *iten kanonischen Variaten* und
- (3) $\rho_i := \lambda_i^{1/2}$ die *ite kanonische Korrelation*.

Ihre Bedeutsamkeit erlangen diese Begriffe durch ihre Eigenschaften, die wir im folgenden Theorem zusammenfassen.

Theorem 0.3 (Eigenschaften kanonischer Korrelationen und Variaten). *Es seien*

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \text{ mit } \mathbb{E}(z) := 0_m \text{ und } \mathbb{C}(z) := \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \quad (33)$$

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

$$\phi_r = \max_{a,b} a^T \Sigma_{xy} b \quad (34)$$

unter den Nebenbedingungen

$$a^T \Sigma_{xx} a = 1, \quad b^T \Sigma_{yy} b = 1, \quad a_i^T \Sigma_{xx} a = 0 \text{ für } i = 1, \dots, r-1 \quad (35)$$

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

Proof. Wir betrachten das restringierte Optimierungsproblem

$$\phi_r^2 = \max_{a,b} (a^T \Sigma_{xy} b)^2 \text{ u.d.N. } a^T \Sigma_{xx} a = 1, b^T \Sigma_{yy} b = 1, a_i^T \Sigma_{xx} a = 0, i = 1, \dots, r-1 \quad (36)$$

Wir folgen @mardia1979, S. 284 und gehen schrittweise vor, d.h. wir lösen das restringierte Optimierungsproblem

$$\phi_r^2 = \max_a \left(\max_b (a^T \Sigma_{xy} b)^2 \text{ u.d.N. } b^T \Sigma_{yy} b = 1 \right) \text{ u.d.N. } a^T \Sigma_{xx} a = 1, a_i^T \Sigma_{xx} a = 0, i = 1, \dots, r-1 \quad (37)$$

von innen nach außen.

Schritt (1)

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

$$\max_b (a^T \Sigma_{xy} b)^2 \text{ u.d.N. } b^T \Sigma_{yy} b = 1 \quad (38)$$

Dieses Optimierungsproblem kann geschrieben werden als

$$\max_b b^T \Sigma_{yx} a a^T \Sigma_{xy} b \text{ u.d.N. } b^T \Sigma_{yy} b = 1, \quad (39)$$

weil gilt, dass

$$(a^T \Sigma_{xy} b)^2 = (a^T \Sigma_{xy} b) (a^T \Sigma_{xy} b) = (a^T \Sigma_{xy} b)^T a^T \Sigma_{xy} b = b^T \Sigma_{yx} a a^T \Sigma_{xy} b. \quad (40)$$

Das Optimierungsproblem (39) kann nun mithilfe des Theorems zur Maximierung quadratischer Formen mit Nebenbedingungen gelöst werden. Im Sinne dieses Theorems setzen wir dazu

$$A := \Sigma_{yx} a a^T \Sigma_{xy} \text{ und } B := \Sigma_{yy}. \quad (41)$$

Dann hat (39) die Form

$$\max_b b^T A b \text{ unter der Nebenbedingung } b^T B b = 1, \quad (42)$$

Das Maximum von (42) entspricht nach dem Theorem zur Maximierung quadratischer Formen mit Nebenbedingungen dem größten Eigenwert von

$$B^{-1} A = \Sigma_{yy}^{-1} \Sigma_{yx} a a^T \Sigma_{xy} \quad (43)$$

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

$$A := \Sigma_{yy}^{-1}\Sigma_{yx}, \quad b := a, \quad B := \Sigma_{xy} \quad (44)$$

und erhalten für den betreffenden Eigenwert

$$\lambda_a = b^T B A a = a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a. \quad (45)$$

als Lösung (Maximum) des restringierten Optimierungsproblems

$$\max_b (a^T \Sigma_{xy} b)^2 \text{ u.d.N. } b^T \Sigma_{yy} b = 1 \quad (46)$$

Schritt (2)

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

$$\phi_r^2 = \max_a a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \text{ u.d.N. } a^T \Sigma_{xx} a = 1, a_i^T \Sigma_{xx} a = 0, i = 1, \dots, r-1 \quad (47)$$

Dazu halten wir zunächst fest, dass (47) mit den Definitionen von α_i und K in der Definition der Kanonischen Koeffizientenvektoren, Variaten, und Korrelationen geschrieben werden kann als

$$\phi_r^2 = \max_{\alpha} \alpha^T K K^T \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \alpha_i^T \alpha = 0, i = 1, \dots, r-1, \quad (48)$$

denn

$$\begin{aligned} \phi_r^2 &= \max_a a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \text{ u.d.N. } a^T \Sigma_{xx} a = 1, a_i^T \Sigma_{xx} a = 0 \Leftrightarrow \\ \phi_r^2 &= \max_{\alpha} a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \text{ u.d.N. } \alpha^T \Sigma_{xx}^{-1/2} \Sigma_{xx} \Sigma_{xx}^{-1/2} \alpha = 1, \alpha_i^T \Sigma_{xx}^{-1/2} \Sigma_{xx} \Sigma_{xx}^{-1/2} \alpha = 0 \\ \phi_r^2 &= \max_{\alpha} \alpha^T \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1/2} \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \alpha_i^T \alpha = 0 \\ \phi_r^2 &= \max_{\alpha} \alpha^T \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yy}^{-1/2} \Sigma_{yy}^{1/2} \Sigma_{yx} \Sigma_{xx}^{-1/2} \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \alpha_i^T \alpha = 0 \\ \phi_r^2 &= \max_{\alpha} \alpha^T K K^T \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \alpha_i^T \alpha = 0 \end{aligned} \quad (49)$$

Dabei sind nach der betreffenden Definition die α_i die Eigenvektoren von $K K^T$ mit den $i = 1, \dots, r-1$ größten Eigenwerten. Nach dem Theorem zur Maximierung quadratischer Formen mit Nebenbedingungen ist die Lösung von (48) der größte Eigenwert von $K K^T$ mit seinem assoziierten Eigenvektor. Die Nebenbedingung $\alpha_i^T \alpha = 0$ schränkt diese Wahl auf den r -ten größten Eigenwert und seinen assoziierten Eigenvektor α_r ein. Mit der Definition von Eigenwerten und Eigenvektoren gilt also

$$\phi_r^2 = \alpha_r^T K K^T \alpha_r = \alpha_r^T \lambda_r \alpha_r = \lambda_r \alpha_r^T \alpha_r = \lambda_r. \quad (50)$$

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

$$K = A\Lambda B^T \Leftrightarrow KB = A\Lambda B^T B \Leftrightarrow KB = A\Lambda \Leftrightarrow K\beta_r = \alpha_r \lambda_r^{1/2} \quad (51)$$

dass

$$a_r^T \Sigma_{xy} b_r = \alpha_r^T \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \beta_r = \alpha_r^T K \beta_r = \alpha_r^T \alpha_r \lambda_r^{1/2} = \rho_r \quad (52)$$

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

ϕ_1 ist also die größtmögliche Korrelation von

$$\xi = a^T x \text{ und } v = b^T y \quad (53)$$

unter den Nebenbedingungen

$$\mathbb{V}(\xi) = 1 \text{ und } \mathbb{V}(v) = 1 \quad (54)$$

und erfüllt damit die Forderungen an *die* kanonische Korrelatione. ϕ_r mit $r > 1$ ist die größtmögliche Korrelation von

$$\xi = a^T x \text{ und } v = b^T y \quad (55)$$

unter den Nebenbedingungen

$$\mathbb{V}(\xi) = 1, \mathbb{V}(v) = 1 \text{ und } \mathbb{C}(\xi_i, \xi) = 0 \text{ für die kanonischen Variate } \xi_i \text{ mit } i = 1, \dots, r-1. \quad (56)$$

Simulationsbeispiel

Wir betrachten das Beispiel (vgl. @uurtio2018)

$$p(x) = N(x; 0_4, I_4) \text{ und } p(y|x) = N(y; Lx, G) \quad (57)$$

mit

$$L := \begin{pmatrix} 0.0 & 0.0 & 1.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & -1.0 \end{pmatrix} \text{ und } G := \begin{pmatrix} 0.2 & 0.0 & 0.0 \\ 0.0 & 0.4 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix} \quad (58)$$

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

$$\begin{aligned} y_1 &= x_3 + \varepsilon_1 \\ y_2 &= x_1 + \varepsilon_2 \\ y_3 &= -x_4 + \varepsilon_3 \end{aligned} \quad (59)$$

mit

$$x_1 \sim N(0, 1), x_3 \sim N(0, 1), x_4 \sim N(0, 1) \quad (60)$$

und

$$\varepsilon_1 \sim N(0, 0.2), \varepsilon_2 \sim N(0, 0.4), \varepsilon_3 \sim N(0, 0.3) \quad (61)$$

Mit dem Theorem zu gemeinsamen Normalverteilungen (vgl. [?@sec-normalverteilungen](#)) ergibt sich, dass

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N(0_7, \Sigma) \quad (62)$$

mit

$$\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}, \quad (63)$$

wobei

$$\Sigma_{xx} = I_4, \quad \Sigma_{xy} = L^T, \quad \Sigma_{yx} = L \text{ und } \Sigma_{yy} = G + LL^T. \quad (64)$$

Explizit ergibt sich also

$$\Sigma = \begin{pmatrix} I_4 & L^T \\ L & G + LL^T \end{pmatrix} = \begin{pmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & -1.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 1.2 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.4 & 0.0 \\ 0.0 & 0.0 & 0.0 & -1.0 & 0.0 & 0.0 & 1.3 \end{pmatrix} \quad (65)$$

Folgender **R** Code definiert zunächst den Kovarianzmatrixparameter der gemeinsamen Verteilung von x und y .

```
# R Pakete für Matrizenrechnung
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
```

Anhand von Definition 0.1 bestimmt folgender **R** Code dann basierend auf obigem Kovarianzmatrixparameter die kanonischen Korrelationen und kanonischen Koeffizientenvektoren.

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

rho_1 = 0.9128709 , a_1^T = ( 0 0 -1 0 ) , b_1^T = ( -0.9128709 0 0 )
rho_2 = 0.877058 , a_2^T = ( 0 0 0 1 ) , b_2^T = ( 0 0 -0.877058 )
rho_3 = 0.8451543 , a_3^T = ( -1 0 0 0 ) , b_3^T = ( 0 0 -0.8451543 0 )
```

Modellschätzung

Zur Schätzung einer kanonischen Korrelationsanalyse wird die Kovarianzmatrix $\mathbb{C}(z)$ des gemeinsamen Zufallsvektors z von Prädiktoren und Kriterien durch ihr Stichprobenäquivalent C ersetzt. Dies ist die Aussage folgender Definition.

Definition 0.2 (Schätzer der kanonischen Korrelationsanalyse). Für $i = 1, \dots, n$ seien

$$z_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix} \text{ mit } \mathbb{E}(z_i) := 0_m \text{ und } \mathbb{C}(z_i) := \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \quad (66)$$

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

$$C := \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \quad (67)$$

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

$$\hat{a}_i := C_{xx}^{-1/2} \hat{\alpha}_i \in \mathbb{R}^{m_x}, \quad \hat{b}_i := C_{yy}^{-1/2} \hat{\beta}_i \in \mathbb{R}^{m_y} \text{ und } \hat{\rho}_i := \hat{\lambda}_i^{1/2} \quad (68)$$

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

$$\hat{K} := C_{xx}^{-1/2} C_{xy} C_{yy}^{-1/2} \in \mathbb{R}^{m_x \times m_y} \quad (69)$$

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

Wir verzichten auf eine Diskussion der Güte dieser Schätzung.

Simulationsbeispiel

Mithilfe folgenden **R** Codes verdeutlichen wir uns die Schätzung kanonischer Korrelationen und Koeffizientenvektoren in dem oben betrachteten Beispiel. Dazu generieren wir Realisierungen der Prädiktoren und Kriterien bei Stichprobenumfängen zwischen $n = 100$ und $n = 1000$. Figure 1 visualisiert die wahren, aber unbekannten, kanonischen Korrelationen ρ_1, ρ_2, ρ_3 und ihre für jede Simulation basierend auf der Stichprobenkovarianzmatrix des realisierten Datensatzes geschätzte Äquivalente $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3$. Die Variabilität der Schätzung nimmt mit zunehmenden Stichprobenumfang ab. Figure 2 visualisiert die Absolutwerte des wahren, aber unbekannten, ersten kanonischen Koeffizientenvektors a_1 sowie ihre entsprechenden Schätzungen als \hat{a}_1 . Auch hier nimmt die Variabilität der Schätzung mit zunehmenden Stichprobenumfang ab. Allerdings ist zu beachten, dass es sich hierbei um die Absolutwerte des kanonischen Koeffizientenvektors handelt und das Vorzeichen abhängig von der Schätzung auch im Widerspruch zum wahren, aber unbekannten, Wert des kanonischen Koeffizientenvektors stehen kann.

```
# R Pakete
library(MASS)
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,-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(solve(Sigma_xx)) %*% Sigma_xy %*% sqrtm(solve(Sigma_yy))
ALB = svd(K)
A = ALB$u
Lambda = ALB$d
B = ALB$v
rho = Lambda
a = sqrtm(solve(Sigma_xx)) %*% A
b = sqrtm(solve(Sigma_yy)) %*% B
```

```
# 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(solve(C_xx)) %*% C_xy %*% sqrtm(solve(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(solve(C_xx)) %*% A_hat
  b_hat = sqrtm(solve(C_yy)) %*% B_hat
  rho_hat[i,] = as.matrix(Lambda_hat)
  a_1_hat[i,] = a_hat[,1]
}
```

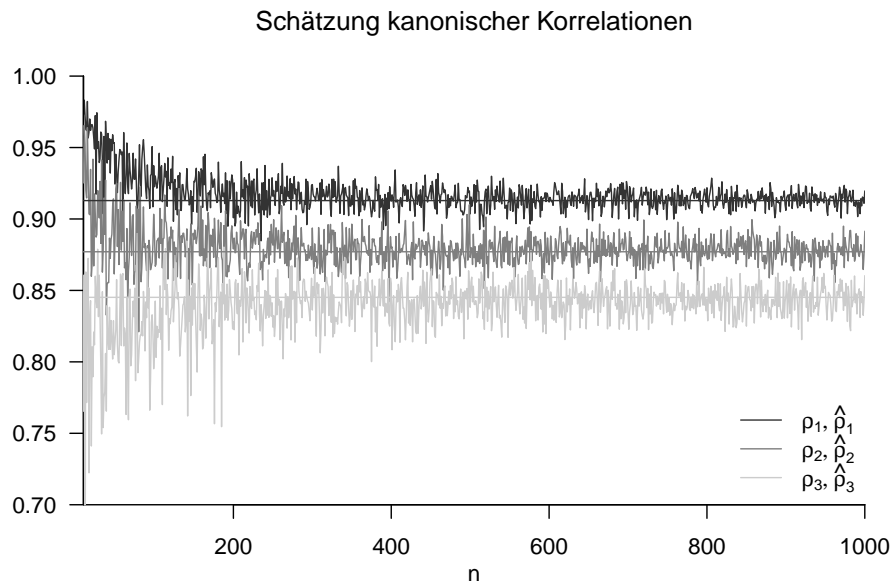


Figure 1: Schätzung kanonischer Korrelationen im Simulationsbeispiel

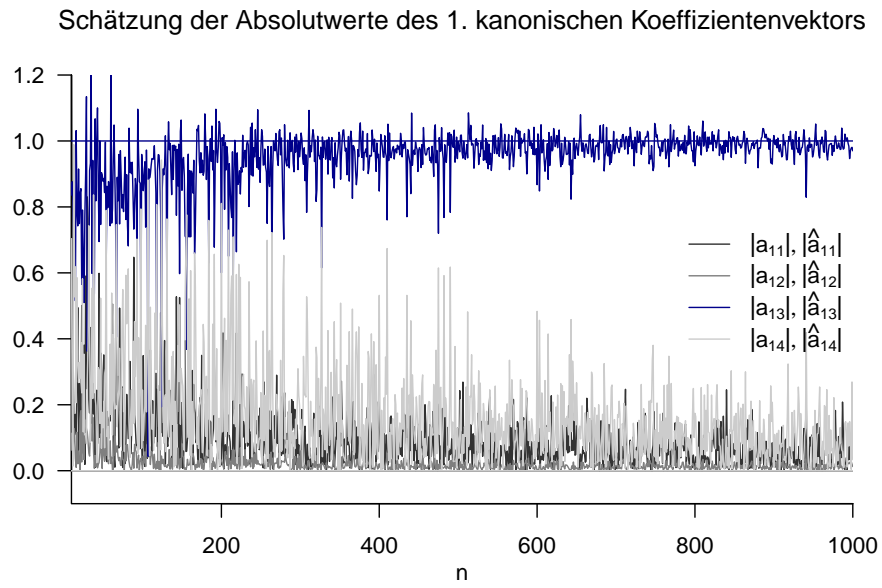


Figure 2: Schätzung der Absolutwerte des ersten kanonischen Koeffizientenvektors im Simulationsbeispiel

Anwendungsbeispiel

Zuletzt wollen wir die konkrete Berechnung einer kanonischen Korrelationsanalyse im Kontext des in Section diskutierten Anwendungsbeispiels demonstrieren. Folgender **R** Code implementiert die Berechnung der kanonischen Korrelationen und kanonischen Koeffizientenvektoren für diesen Datensatz.

```
# R Paket
library(expm)

# Datenpräprozessierung
fname = "../data/504-kanonische-korrelationsanalyse.csv"
D = read.table(fname, sep = ",", header = TRUE)
x = as.matrix(cbind(D$dUR, D$EXP))
y = as.matrix(cbind(D$dBDI, D$dGLU))
n = nrow(x)
m_x = ncol(x)
m_y = ncol(y)
Y = t(cbind(x,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(solve(C_xx)) %*% C_xy %*% sqrtm(solve(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(solve(C_xx)) %*% A_hat
b_hat = sqrtm(solve(C_yy)) %*% B_hat
rho_hat = as.matrix(Lambda_hat)
```

```
rho_hat_1 : 0.9950575
a_hat_1 : -0.1623409 -0.173979
b_hat_1 : -0.1554175 -0.05025419
rho_hat_2 : 0.5010358
a_hat_2 : -0.06026274 0.3118808
b_hat_2 : -0.08128072 0.7773036
```

Neben der Implementation mithilfe einer Singulärwertzerlegung bietet **R** auch eine direkte Bestimmung mithilfe der `cancor()` Funktion an. Folgender **R** Code demonstriert das entsprechende Vorgehen.

```
# Datenpräprozessierung
fname = "../data/504-kanonische-korrelationsanalyse.csv"
D = read.table(fname, sep = ",", header = TRUE)
x = as.matrix(cbind(D$dUR, D$EXP))
y = as.matrix(cbind(D$dBDI, D$dGLU))
cca = cancor(x,y)
```

```
rho_hat_1 : 0.9950575
rho_hat_2 : 0.5010358
```

Man findet also, dass die geschätzte maximale Korrelation einer Linearkombinationen der Prädiktorvariablen `dUR` und `EXP` mit einer Linearkombination der Kriterien `dBDI` und `dGLU` mit $\hat{\rho}_1 = 0.99$ sehr hoch ist. Man kann daraus schließen, dass in diesem Fall die Prädiktorvariablen

gemeinschaftlich hoch mit den Kriterien assoziiert sind. Es ergeben sich hier insbesondere die Linearkombinationen

$$\xi = 0.16 \text{ DUR} + 0.17 \text{ EXP} \text{ und } v = 0.15 \text{ dBDI} + 0.05 \text{ dGLU} \quad (70)$$

als bester Prädiktor und als am besten prädizierbares Kriterium, respektive. Die Dauer der Psychotherapie und die Erfahrung der behandelnden Psychotherapeut:in scheinen, bei der aktuellen Datenskalisierung zur bestmöglichen Prädiktion der Therapiegüte also in etwa gleichbedeutend, bei dem bestprädizierbarem Kriterium der Therapieeffizienz trägt bei der aktuellen Datenskalisierung die BDI Score Reduktion etwas mehr bei als die Glukokortikoidplasmalevel Reduktion bei.

Literaturhinweise

Die kanonische Korrelationsanalyse geht zurück auf @hotelling1935 und @hotelling1936.