

Chapter 1

Zusammenhangsstrukturen

G. Tutz sehr vorläufig
June 27, 2013

1.1	Korrelationskoeffizienten: marginale, bedingte und partielle Korrelation . . .	3
1.2	Korrelationskoeffizienten bei vektoriellen Größen	8
1.3	Die Konzentrationsmatrix	9
1.4	Assoziationsstrukturen und Graphen	10
1.5	Multiple Korrelation	11
1.6	Kanonische Korrelation	11

Untersucht man den Zusammenhang zweier Größen wie Mietpreis und Wohnungsgröße, stellt sich als erstes die Fragen, ob überhaupt ein Zusammenhang besteht und im weiteren – so ein Zusammenhang vorliegt – von welcher Art dieser ist. Interessanter und komplexer wird das Problem, wenn der Zusammenhang von mehr als zwei Variablen zu untersuchen ist. Dann stellt sich die Frage, wie zwei Variablen zusammenhängen unter Berücksichtigung der anderen Variablen. Der Zusammenhang ist ein durchaus anderer, wenn das Vorhandensein der weiteren Variablen in die Analyse einbezogen wird. Bekannt ist das Lehrbuchbeispiel von der positiven Korrelation zwischen Geburtenraten und Störchezahl, erhoben in den Nachkriegsjahren. Die positive Korrelation verschwindet erwartungsgemäß, wenn man den Zeitfaktor berücksichtigt. Über die Jahre des Erhebungszeitraums hatte sich sowohl die Geburtenrate als auch die Störcherate gleichsinnig verändert. Berücksichtigt man die Variable Zeit, ist die Korrelation Null, vernachlässigt man die Variable Zeit, erhält man eine positive Korrelation.

Im weiteren werden verschiedene Ansätze zur Berücksichtigung weiterer Variablen bei der Analyse von Zusammenhangsstrukturen dargestellt. Nach der Darstellung grundlegender Konzepte wird als erstes das Normalverteilungsmodell dargestellt.

1.1 Korrelationskoeffizienten: marginale, bedingte und partielle Korrelation

Im folgenden werden Varianten des klassischen Korrelationskoeffizienten dargestellt. Dieser ist geeignet für normalverteilte Größen und mißt den linearen Zusammenhang zwischen zwei Größen. Man geht daher von zumindest approximativ normalverteilten Größen aus, für die der lineare Zusammenhang ein adäquates Assoziationsmaß darstellt.

Der Einfachheit halber werden zuerst drei Variablen, X , Y und Z , betrachtet. Die entsprechenden Erwartungswerte werden mit $\mu_X = E(X)$, $\mu_Y = E(Y)$, $\mu_Z = E(Z)$, die Kovarianzen mit $\sigma_{XY} = \text{cov}(X, Y)$, $\sigma_{XZ} = \text{cov}(X, Z)$ und $\sigma_{YZ} = \text{cov}(Y, Z)$ bezeichnet. Man erhält damit für den Vektor (Y, X, Z)

$$\boldsymbol{\mu} = E\left(\begin{pmatrix} Y \\ X \\ Z \end{pmatrix}\right) = \begin{pmatrix} \mu_Y \\ \mu_X \\ \mu_Z \end{pmatrix}, \quad \boldsymbol{\Sigma} = \text{cov}\left(\begin{pmatrix} Y \\ X \\ Z \end{pmatrix}\right) = \begin{pmatrix} \sigma_Y^2 & \sigma_{YX} & \sigma_{YZ} \\ \sigma_{XY} & \sigma_X^2 & \sigma_{XZ} \\ \sigma_{ZY} & \sigma_{ZX} & \sigma_Z^2 \end{pmatrix}.$$

Marginale Korrelation

Das übliche (lineare) Zusammenhangsmaß ist die Korrelation, genauer die *marginale Korrelation*

$$\varrho_{YX} = \varrho(Y, X) = \frac{\text{cov}(Y, X)}{\sigma_Y \sigma_X},$$

die den linearen Zusammenhang der Variablen Y und X mißt ohne weitere Variablen einzubeziehen. Der Begriff marginale Korrelation bezieht sich darauf, daß dazu nur die gemeinsame Verteilung von Y und X benutzt wird. Völlig analog sind $\varrho(Y, Z)$ und $\varrho(X, Z)$ definiert.

Bedingte Korrelation

Unter der *bedingten Korrelation* von Y und X , gegeben $Z = z$, abgekürzt $\varrho(Y, X|Z = z)$ versteht man die Korrelation, die aus den bedingten Verteilungen von $(Y, X)|Z = z$ und $Y|Z = z$, $X|Z = z$ berechnet wird,

$$\varrho_{YX|Z=z} = \varrho(Y, X|Z = z) = \frac{\text{cov}(Y, X|Z = z)}{\sqrt{\text{var}(Y|Z = z)}\sqrt{\text{var}(X|Z = z)}}. \quad (1.1)$$

Man mißt also, wie stark der lineare Zusammenhang der bedingten Variablen Y und X für festgelegten Wert $Z = z$ ist. Die bedingten Verteilungen, die für die Bestimmung relevant sind, besitzen die Dichte

$$\begin{aligned} f_{XY|Z=z}(x, y) &= \frac{f_{XYZ}(x, y, z)}{f_Z(z)}, \\ f_{Y|Z=z}(y) &= \frac{f_{YZ}(y, z)}{f_Z(z)}, \\ f_{X|Z=z}(x) &= \frac{f_{XZ}(x, z)}{f_Z(z)}, \end{aligned}$$

wobei die Marginaldichten durch Indices gekennzeichnet sind (beispielsweise bezeichnet $f_{XZ}(x, z)$ die Dichte von (X, Z)) und von der Existenz der bedingten Dichte ausgegangen wird.

Ein interessanter und wichtiger Spezialfall tritt auf, wenn alle Variablen gemeinsam normalverteilt sind, d.h. $(Y, X, Z)^T \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Für diesen Fall hängt die bedingte Korrelation nicht von dem Wert z ab, so daß man die einfachere Bezeichnung $\varrho(Y, X|Z)$ wählt. Die bedingte Korrelation ergibt sich durch

$$\varrho(Y, X|Z) = \varrho(Y, X|Z = z) = \frac{\varrho_{YX} - \varrho_{YZ}\varrho_{XZ}}{\sqrt{1 - \varrho_{YZ}^2}\sqrt{1 - \varrho_{XZ}^2}} \quad (1.2)$$

Bemerkenswert ist, daß sich die marginale Korrelation im allgemeinen unterscheidet von der bedingten Korrelation, d.h.

$$\varrho(Y, X|Z) \neq \varrho(Y, X).$$

Beim Übergang von $\varrho(Y, X)$ zur bedingten Korrelation kann sich das Vorzeichen ändern, marginale Unkorreliertheit kann zu deutlich bedingter Korreliertheit führen und marginale Korreliertheit kann mit bedingter Korreliertheit einhergehen. Die fälschliche Unterstellung einer Gleichheit von marginaler und bedingter Korreliertheit ist verantwortlich für einen großen Teil statistisch "begründbarer" Fehlschlüsse und Fehlinterpretationen. Prinzipiell gilt, daß bei der Analyse der Zusammenhangsstruktur alle potentiell

wirksamen Merkmale zu berücksichtigen sind. Die marginale Korrelation ignoriert allerdings die Wirkung aller anderen Kovariablen. Daher die Tendenz zu Fehlinterpretationen. Im folgendem sind einige Fälle skizziert, in denen sich die beiden Korrelationen unterscheiden.

(1) *Bedingte Unkorreliertheit* ("Scheinkorrelation")

Da für normalverteilte Größen bedingte Unkorreliertheit zusammenfällt mit bedingter Unabhängigkeit, läßt sich hier untersuchen, ob Variablen auch einen bedingten Zusammenhang aufweisen. Es sind einfache Konstellationen konstruierbar, so daß $\varrho_{YX} \neq 0$, aber $\varrho_{YX|Z} = 0$. Man erhält also einen deutlichen Zusammenhang zwischen den Variablen Y und X , solange man nur diese beiden berücksichtigt, der Zusammenhang verschwindet jedoch, wenn das Vorhandensein der Variable Z in Form der bedingten Korrelation einbezogen wird. Dieser Typ liegt dem Störchebeispiel zugrunde, mit verschwindender Korrelation zwischen Störchezahl und Geburtenzahl, wenn die Zeit einbezogen wird.

(2) *Verdeckte Korrelation*

Im Gegensatz zur Scheinkorrelation kann der Fall auftreten, daß zwei Merkmale Y und X keine Korrelation aufweisen, obwohl die Variablen bedingt durchaus korreliert sind. Es ist dann zweifelhaft, die Merkmale als nicht zusammenhängend zu charakterisieren, da sie ja tatsächlich in Subpopulationen zusammenhängen. Formal ergibt sich dieses Phänomen aus (1.2) wenn $\varrho_{YX} = 0$ und $\varrho_{YZ}, \varrho_{XZ} \neq 0$ (z.B. $\varrho_{YX} = 0, \varrho_{YZ} = 0.5, \varrho_{XZ} = -0.8, \varrho(Y, X|Z) = 0.84$)

(3) *Gekippte Effekte*

Zwei Merkmale können positiv korreliert erscheinen, wenn man die marginale Korrelation zugrundelegt, obwohl in Subpopulationen der Zusammenhang eine negative Korrelation aufweist. Legt man Normalverteilung zugrunde, ergibt sich aus (1.2) beispielsweise für positive Korrelation $\varrho_{YX} = 0.4$ und $\varrho_{YZ} = \varrho_{XZ} = 0.8$ die negative bedingte Korrelation -0.66 . Der Übergang von der bedingten Korrelation zur marginalen Korrelation kann also eine Vorzeichenänderung zur Folge haben. Das hat für die Interpretation erhebliche Konsequenzen und manch Fehlinterpretation zur Folge (Rehabilitation). In der Analyse von Kontingenztafeln, ist dieses Phänomen als Simpsons Paradoxon bekannt (Simpson, 1951). Ein schönes Beispiel gibt Whittaker (1990), S. 46.

(4) *Suppressor-Variablen*

Betrachtet sei wieder der Zusammenhang zwischen Y und X . Die Variable Z wird Suppressor-Variable genannt, wenn der marginale Zusammenhang zwischen Y und X schwächer ist als der bedingte Zusammenhang, gegeben Z , d.h. (für positive Korrelation), wenn

$$\varrho(Y, X) < \varrho(X, Y|Z)$$

gilt. Der Effekt tritt beispielsweise auf, wenn $\varrho_{YZ} = 0$ bzw. $\varrho_{XZ} = 0$, da dann die bedingte Korrelation bestimmt ist durch

$$\varrho(X, Y|Z) = \frac{\varrho_{YX}}{\sqrt{1 - \varrho_{XZ}^2}} \text{ bzw. } \varrho(X, Y|Z) = \frac{\varrho_{YX}}{\sqrt{1 - \varrho_{YZ}^2}}.$$

Wird die Suppressor-Variable vernachlässigt, indem man die marginale Korrelation $\varrho(Y, X)$ betrachtet, ergibt die Analyse des Zusammenhangs einen schwächeren Zusammenhang als der unter Berücksichtigung der Wirkung von Z vorliegende.

Die Darstellung in Form eines bedingten Unabhängigkeitsgraphen findet sich in Abbildung ?? . Gegeben sind die drei Variablen, Y und X unabhängig, gegeben Z ($X \perp Y|Z$), Y und Z unabhängig, gegeben X ($Y \perp Z|X$) und X und Z unabhängig, gegeben Y ($X \perp Z|Y$). Verwendet wird dabei die übliche Abkürzung mit $\perp$ für die Unabhängigkeit und $|$ für die Bedingung.

Partielle Korrelation

Partielle Korrelation zielt darauf ab, zu bestimmen inwieweit die Korrelation zwischen zwei Variablen, Y und X , auf den linearen Einfluß einer dritten Variable Z zurückzuführen ist. Dabei wird nicht wie bei der bedingten Korrelation auf $Z = z$ bedingt, sondern versucht den linearen Anteil von Z herauszurechnen.

Dazu bestimmt man die linearen Regressionsmodelle von Y auf Z und X auf Z

$$\begin{aligned} Y &= \beta_0 + \beta Z + \varepsilon_{Y \cdot Z}, \\ X &= \gamma_0 + \gamma Z + \varepsilon_{X \cdot Z}. \end{aligned}$$

Etwas genauer bestimmt man für das lineare Regressionsmodell $Y = \beta_0 + \beta Z + \varepsilon_{Y \cdot Z}$ die Parameter β_0, β so, daß $\hat{Y} = \beta_0 + \beta Z$ die *beste lineare Prognose* darstellt, d.h. β_0 und β werden so gewählt, daß die erwartete quadratische Abweichung zwischen Y und $\hat{Y} = \beta_0 + \beta Z$ minimal wird,

$$E\{(Y - \beta_0 - \beta Z)^2\} \xrightarrow{\beta_0, \beta} \min.$$

Entsprechend minimiert man für die lineare Regression $X = \gamma_0 + \gamma Z + \varepsilon_{X \cdot Z}$

$$E\{(X - \gamma_0 - \gamma Z)^2\} \xrightarrow{\gamma_0, \gamma} \min.$$

Die Existenz der Momente vorausgesetzt, läßt sich zeigen, daß das erste Minimierungsproblem zu

$$\beta = \frac{\sigma_{YZ}}{\sigma_Z^2}, \quad \beta_0 = \mu_Y - \beta \mu_Z, \quad (1.3)$$

das zweite zu

$$\gamma = \frac{\sigma_{XZ}}{\sigma_X^2}, \quad \gamma_0 = \mu_X - \gamma \mu_Z, \quad (1.4)$$

führt.

Die *partielle Korrelation* zwischen Y und X nach Elimination des linearen Anteils von Z ist dann bestimmt durch die Korrelation der Residuen $\varepsilon_{Y \cdot Z}, \varepsilon_{X \cdot Z}$

$$\rho_{YX \cdot Z} = \frac{\text{cov}(\varepsilon_{Y \cdot Z}, \varepsilon_{X \cdot Z})}{\sqrt{\text{var}(\varepsilon_{Y \cdot Z})} \sqrt{\text{var}(\varepsilon_{X \cdot Z})}.$$

Sie stellt die Korrelation zwischen Y und X dar, die verbleibt, wenn der lineare Einfluß von Z auf Y und von Z auf X berücksichtigt ist. Man spricht davon, daß der lineare Einfluß von Z *herauspartialisiert* wurde.

Einfache Ableitung zeigt, daß unter Verwendung der Werte $\beta_0, \beta, \gamma_0, \gamma$ aus (1.3) und (1.4) für die Residuen gilt

$$\begin{aligned} E(\varepsilon_{Y \cdot Z}) &= 0, & \text{var}(\varepsilon_{Y \cdot Z}) &= \sigma_Y^2(1 - \rho_{YZ}^2) \\ E(\varepsilon_{X \cdot Z}) &= 0, & \text{var}(\varepsilon_{X \cdot Z}) &= \sigma_X^2(1 - \rho_{XZ}^2) \\ \text{cov}(\varepsilon_{Y \cdot Z}, \varepsilon_{X \cdot Z}) &= \sigma_Y \sigma_X (\rho_{YX} - \rho_{XZ} \rho_{YZ}) \end{aligned}$$

so daß sich durch Einsetzen

$$\rho_{XY \cdot Z} = \frac{\rho_{YX} - \rho_{XZ} \rho_{YZ}}{\sqrt{1 - \rho_{YZ}^2} \sqrt{1 - \rho_{XZ}^2}}$$

ergibt. Wie der Vergleich mit (1.2) zeigt, ist für normalverteilte Merkmale die bedingte Korrelation identisch mit der partiellen Korrelation. Normalverteilte Merkmale sind in dieser Beziehung ein Spezialfall, da für sie die beste lineare Prognose auch das beste Prognosemodell darstellt. Die *beste Prognose* $\hat{Y}$ durch Z ist generell bestimmt durch die Eigenschaft

$$E\{(Y - \hat{Y})^2\} \longrightarrow \min,$$

wobei über alle von Z abhängigen Zufallsvariablen $\hat{Y} = \hat{Y}(Z)$ mit existierenden Varianzen minimiert wird. Das Resultat ist die bedingte Erwartung $E(Y|Z)$, die für die Normalverteilung linear ist, d.h. man erhält

$$\hat{Y}(Z) = E(Y|Z) = \beta_0 + \beta Z$$

mit $\beta_0 = \mu_Y - \beta \mu_Z, \beta = \sigma_{YZ}/\sigma_Z^2$. Völlig analog erhält man $E(X|Z) = \gamma_0 + \gamma Z$ mit $\gamma_0 = \mu_X - \gamma \mu_Z, \gamma = \sigma_{XZ}/\sigma_X^2$. Die Regression besitzt daher die Form

$$Y = E(Y|Z) + \varepsilon_{YZ} = \beta_0 + \beta Z + \varepsilon_{Y \cdot Z}$$

und darüberhinaus sind Y und ε_{YZ} unabhängig.

Marginale Korrelation von Y und X

$$\rho_{YX} = \frac{\text{cov}(Y, X)}{\sigma_Y \sigma_X}$$

Bedingte Korrelation von Y und X , gegeben $Z = z$

$$\rho_{YX|Z=z} = \frac{\text{cov}(Y, X|Z=z)}{\sqrt{\text{var}(Y|Z=z)} \sqrt{\text{var}(X|Z=z)}}$$

Partielle Korrelation von Y und X nach Elimination des linearen Anteils von Z

$$\rho_{XY \cdot Z} = \frac{\text{cov}(\varepsilon_{Y \cdot Z}, \varepsilon_{X \cdot Z})}{\sqrt{\text{var}(\varepsilon_{Y \cdot Z})} \sqrt{\text{var}(\varepsilon_{X \cdot Z})}}$$

Für (Y, X, Z) gemeinsam normalverteilt, gilt

$$\rho_{YX|Z=z} = \rho_{XY \cdot Z}$$

Partielle Korrelation und Regression

Die Koeffizienten eines linearen Regressionsmodells stellen einen Zusammenhang zwischen abhängiger Größe und Prädiktor dar, der im wesentlichen die bedingten Korrelationen wiedergibt. Für das Regressionsmodell

$$Y = \beta_0 + \beta_X X + \beta_Z Z + \varepsilon_{Y \cdot XZ}$$

ergeben sich die Regressionskoeffizienten

$$\beta_X = \rho_{YX \cdot Z} \frac{\sqrt{\text{var}(\varepsilon_{Y \cdot Z})}}{\text{var}(\varepsilon_{X \cdot Z})}, \quad \beta_Z = \rho_{YZ \cdot X} \frac{\sqrt{\text{var}(\varepsilon_{Y \cdot X})}}{\text{var}(\varepsilon_{Z \cdot X})},$$

wobei $\varepsilon_{Y \cdot Z}, \varepsilon_{X \cdot Z}, \varepsilon_{Y \cdot X}, \varepsilon_{Z \cdot X}$ die Störterme der entsprechenden univariaten Regressionsmodelle sind. Die Koeffizienten entsprechen damit bis auf Skalierungseffekte den partiellen Korrelationen mit entsprechenden Konsequenzen für die Interpretation.

Üblicherweise werden β_X und β_Z interpretiert als die Veränderung in Y , wenn X bzw. Z um eine Einheit zunehmen. Eine Quelle der Fehlinterpretation ist hier, daß häufig die Berücksichtigung der jeweils anderen Variablen vergessen wird. Dies sei an einem kleinen Beispiel veranschaulicht:

Example 1.1: Mietdaten

Das Regressionsmodell

$$\text{Nettomiete} = \beta_0 + \text{Anzahl Räume} \beta_A + \varepsilon_{NA}$$

liefert für einen Datensatz $\hat{\beta}_A = 88.58$. Gültigkeit des Modells vorausgesetzt, würde der Mietpreis im Mittel um 88.58 Euro steigen, wenn die Wohnung einen weiteren Raum hat. Nimmt man die Wohnfläche in das Regressionsmodell auf und betrachtet

$$\text{Nettomiete} = \beta_0 + \text{Anzahl Räume} \beta_A + \text{Wohnfläche} \beta_W + \varepsilon_{N \cdot AW}$$

erhält man $\hat{\beta}_A = -41.68$. Wohnungen werden demnach im Durchschnitt um 41.68 Euro billiger, wenn die Raumzahl steigt. Diese Interpretation ist unzulässig verkürzt. Die Nettomiete nimmt mit wachsender Raumzahl nur ab bei festgehaltener Wohnfläche, genauer, wenn der lineare Effekt der Wohnfläche berücksichtigt ist. $\square$

1.2 Korrelationskoeffizienten bei vektoriellen Größen

Bezeichne im folgenden $\mathbf{y}^T = (y_1, \dots, y_q)$, $\mathbf{x}^T = (x_1, \dots, x_p)$, $\mathbf{z}^T = (z_1, \dots, z_s)$ Vektoren von Merkmalen. Der gesamte Zufallsvektor besitzt Erwartungswert und Varianz

$$\boldsymbol{\mu} = E\left\{\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{z} \end{pmatrix}\right\} = \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_z \end{pmatrix}, \quad \boldsymbol{\Sigma} = \text{cov}\left\{\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{z} \end{pmatrix}\right\} = \begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yz} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xz} \\ \boldsymbol{\Sigma}_{zy} & \boldsymbol{\Sigma}_{zx} & \boldsymbol{\Sigma}_{zz} \end{pmatrix},$$

wobei die Vektoren und Teilmatrizen entsprechend dimensioniert sind.

Marginale und bedingte Kovarianzen

Die *marginale Kovarianz* ist dann bereits durch die Teilmatrix gegeben, z.B. für $\mathbf{y}$ und $\mathbf{x}$ gilt

$$\text{cov}(\mathbf{y}, \mathbf{x}) = \boldsymbol{\Sigma}_{yx}.$$

Daraus erhält man die *marginale Korrelationsmatrix*

$$\mathbf{K} = (k_{ij}) \text{ mit } k_{ij} = \sigma_{y_i x_j} / (\sigma_{y_i} \sigma_{x_j}).$$

Die *Matrix der bedingten Kovarianzen* beispielsweise für $\mathbf{y}$ und $\mathbf{x}$ gegeben $\mathbf{z} = \tilde{\mathbf{z}}$ ergibt sich durch

$$\boldsymbol{\Sigma}_{\mathbf{yx}|\mathbf{z}=\tilde{\mathbf{z}}} = (\text{cov}(y_i, x_j | \mathbf{z} = \tilde{\mathbf{z}}))_{ij}.$$

Partielle Korrelation

Bei der Bestimmung der partiellen Korrelation betrachtet man den Fall skalarer Merkmale y und x . Um den linearen Anteil von $\mathbf{z}$ aus der Korrelation zwischen y und x herauszurechnen, betrachtet man wieder die linearen Regressionsmodelle von y auf $\mathbf{z}$ und x auf $\mathbf{z}$. Der Einfachheit halber wird für die abhängigen Variablen $E(y) = E(x) = 0$ vorausgesetzt. Die Regression besitzt dann die Form

$$y = \mathbf{z}^T \boldsymbol{\beta} + \varepsilon_{y \cdot \mathbf{z}},$$

$$x = \mathbf{z}^T \boldsymbol{\gamma} + \varepsilon_{x \cdot \mathbf{z}},$$

wobei entsprechend der optimalen linearen Prognose $\boldsymbol{\beta}$ (bzw. $\boldsymbol{\gamma}$) so gewählt wird, daß $E(y - \mathbf{z}^T \boldsymbol{\beta})^2$ (bzw. $E(x - \mathbf{z}^T \boldsymbol{\gamma})^2$) minimiert wird. Existenz der Momente vorausgesetzt erhält man

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_{\mathbf{z}}^{-1} \boldsymbol{\Sigma}_{\mathbf{zy}}, \quad \boldsymbol{\gamma} = \boldsymbol{\Sigma}_{\mathbf{z}}^{-1} \boldsymbol{\Sigma}_{\mathbf{zx}}.$$

Die Assoziation der Residuen $\varepsilon_{y \cdot \mathbf{z}} = y - \mathbf{z}^T \boldsymbol{\beta}$, $\varepsilon_{x \cdot \mathbf{z}} = x - \mathbf{z}^T \boldsymbol{\gamma}$ ergibt die partielle Kovarianz

$$\begin{aligned} \text{cov}(y, x | \mathbf{z}) &= \text{cov}(\varepsilon_{y \cdot \mathbf{z}}, \varepsilon_{x \cdot \mathbf{z}}) \\ &= \sigma_{yx} - \boldsymbol{\Sigma}_{y\mathbf{z}} \boldsymbol{\Sigma}_{\mathbf{z}}^{-1} \boldsymbol{\Sigma}_{\mathbf{z}y}, \end{aligned}$$

wobei die letzte Form sich durch einfaches Berechnen ergibt.

Die Varianz von $\varepsilon_{y \cdot \mathbf{z}}$ (die auch der partiellen Varianz $\text{var}(y | \mathbf{z}) = \text{cov}(y, y | \mathbf{z})$ entspricht) ergibt sich durch einfache Umformungen zu

$$\text{var}(\varepsilon_{y \cdot \mathbf{z}}) = \sigma_y^2 - \boldsymbol{\Sigma}_{y\mathbf{z}} \boldsymbol{\Sigma}_{\mathbf{z}}^{-1} \boldsymbol{\Sigma}_{\mathbf{z}y},$$

und entsprechend

$$\text{var}(\varepsilon_{x \cdot \mathbf{z}}) = \sigma_x^2 - \boldsymbol{\Sigma}_{x\mathbf{z}} \boldsymbol{\Sigma}_{\mathbf{z}}^{-1} \boldsymbol{\Sigma}_{\mathbf{z}x}.$$

Daraus erhält man die partielle Korrelation

$$\rho_{xy \cdot \mathbf{z}} = \frac{\text{cov}(\varepsilon_{y \cdot \mathbf{z}}, \varepsilon_{x \cdot \mathbf{z}})}{\sqrt{\text{var}(\varepsilon_{y \cdot \mathbf{z}})} \sqrt{\text{var}(\varepsilon_{x \cdot \mathbf{z}})}} = \frac{\sigma_{yx} - \boldsymbol{\Sigma}_{y\mathbf{z}} \boldsymbol{\Sigma}_{\mathbf{z}}^{-1} \boldsymbol{\Sigma}_{\mathbf{z}y}}{\sqrt{\sigma_y^2 - \boldsymbol{\Sigma}_{y\mathbf{z}} \boldsymbol{\Sigma}_{\mathbf{z}}^{-1} \boldsymbol{\Sigma}_{\mathbf{z}y}} \sqrt{\sigma_x^2 - \boldsymbol{\Sigma}_{x\mathbf{z}} \boldsymbol{\Sigma}_{\mathbf{z}}^{-1} \boldsymbol{\Sigma}_{\mathbf{z}x}}}$$

Für vektorielle Größen $\mathbf{y}$ und $\mathbf{x}$ läßt sich bei Zugrundelegung multivariater Regressionsmodelle völlig analog die partielle Kovarianzmatrix

$$\text{cov}(\mathbf{y}, \mathbf{x} | \mathbf{z}) = \boldsymbol{\Sigma}_{\mathbf{y}, \mathbf{x}} - \boldsymbol{\Sigma}_{\mathbf{y} \mathbf{z}} \boldsymbol{\Sigma}_{\mathbf{z}}^{-1} \boldsymbol{\Sigma}_{\mathbf{z} \mathbf{y}}$$

ableiten.

Partielle Korrelation und Regression

Betrachtet man das Regressionsmodell

$$Y = \mathbf{x}^T \boldsymbol{\beta} + \varepsilon_{Y \cdot \mathbf{x}}$$

ergibt sich wiederum ein enger Zusammenhang zwischen Koeffizientenvektor $\boldsymbol{\beta}^T = (\beta_1, \dots, \beta_p)$ und den partiellen Korrelationen zwischen y und einzelnen Prädiktoren. Man erhält

$$\beta_i = \varrho(y, x_i | \mathbf{x}_{\setminus i}) \frac{\sqrt{\varepsilon_{y \cdot \mathbf{x}}}}{\sqrt{\varepsilon_{x_i \cdot \mathbf{x}_{\setminus i}}}}$$

Insbesondere ist zu bemerken, daß Prädiktoren vernachlässigbar sind ($\beta_i = 0$), wenn die partielle Korrelation zwischen der abhängigen Variablen und dem entsprechenden Prädiktor unter Berücksichtigung des linearen Einflusses der restlichen Prädiktoren verschwindet ($\varrho(y, x_i | \mathbf{x}_{\setminus i}) = 0$).

1.3 Die Konzentrationsmatrix

Ausgangspunkt ist ein Merkmalsvektor $\mathbf{x}^T = (x_1, \dots, x_p)$ mit Erwartungswert $\boldsymbol{\mu}$ und Kovarianzmatrix $\text{cov}(\mathbf{x}) = \boldsymbol{\Sigma}$. Die Inverse der Kovarianz tritt in der Normalverteilungsdichte explizit auf. Die Elemente dieser Matrix besitzen eine einfache Interpretation, die die Grundlage für bedingte Unabhängigkeitsgraphen darstellt.

Die Matrix $\text{cov}(\mathbf{x})^{-1} = \boldsymbol{\Sigma}^{-1}$ heißt *Konzentrations-* oder *Präzisionsmatrix*. Bezeichne c_{ij} die zugehörigen Elemente, $\boldsymbol{\Sigma}^{-1} = (c_{ij})$. Für diese Elemente lässt sich zeigen

- Die Diagonalelemente von $\boldsymbol{\Sigma}^{-1}$ entsprechen den invertierten partiellen Varianzen,

$$c_{ii} = \frac{1}{\text{var}(x_i | \mathbf{x}_{\setminus i})},$$

wobei $\mathbf{x}_{\setminus i}$ den Vektor $\mathbf{x}$ ohne die i te Komponente bezeichnet, $\mathbf{x}_{\setminus i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p)$. Das spezifische Diagonalelement c_{ii} ist damit die partielle Varianz von x_i unter Berücksichtigung des linearen Effekts der anderen Variablen.

- Die Elemente neben der Diagonalen entsprechen der mit den zugehörigen partiellen Varianzen skalierten (negativen) partiellen Korrelation, für $i \neq j$ gilt

$$c_{ij} = -\sqrt{c_{ii}c_{jj}}\varrho(x_i, x_j | \mathbf{x}_{\setminus ij}),$$

bzw.

$$\varrho(x_i, x_j | \mathbf{x}_{\setminus ij}) = -\frac{\sqrt{c_{ij}}}{\sqrt{c_{ii}c_{jj}}},$$

wobei $\mathbf{x}_{\setminus ij}$ den Vektor $\mathbf{x}$ ohne die Komponenten x_i, x_j bezeichnet. Bis auf Skalierung gibt also das Element c_{ij} die partielle Korrelation von x_i und x_j nach Elimination des linearen Einflusses aller übrigen Variablen wieder.

(Vgl. Whittaker, 1990, p. 142 ff.).

Die partiellen Korrelationen ergeben sich somit in analoger Art und Weise wie die marginalen Korrelationen durch Normieren der Elemente einer Matrix durch ihre Diagonalelemente. Während man die marginale Korrelation unmittelbar aus der Kovarianzmatrix erhält, d.h.

$$\boldsymbol{\Sigma} = (\sigma_{ij}) \quad \text{liefert} \quad \varrho(x_i, x_j) = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}},$$

gilt für die partielle Korrelation

$$\boldsymbol{\Sigma}^{-1} = (c_{ij}) \quad \text{liefert} \quad \varrho(x_i, x_j | \mathbf{x}_{\setminus ij}) = -\frac{c_{ij}}{\sqrt{c_{ii}c_{jj}}}.$$

Insbesondere läßt sich aus der Konzentrationsmatrix ablesen, ob zwei Merkmale partiell unkorreliert sind. Dies ist der Fall, wenn an der entsprechenden Stelle neben der Diagonalen 0 auftritt.

Die Aussage läßt sich auch genereller für Gruppen von Variablen formulieren. Dazu betrachtet man den partitionierten Vektor $(\mathbf{y}^T, \mathbf{x}^T, \mathbf{z}^T)$. Die Inverse in der zugehörigen Kovarianzmatrix sei partitioniert in die entsprechenden Anteile durch

$$\{\text{cov}\left(\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{z} \end{pmatrix}\right)\}^{-1} = \begin{pmatrix} C^{yy} & C^{yx} & C^{yz} \\ C^{xy} & C^{xx} & C^{xz} \\ C^{zy} & C^{zx} & C^{zz} \end{pmatrix}$$

Dann ergibt sich, daß

$$C^{yx} = \iff \text{cov}(\mathbf{y}, \mathbf{x} | \mathbf{z}) =$$

d.h. partielle Kovarianz von $\mathbf{y}$ und $\mathbf{x}$ verschwindet, wenn der entsprechenden Block der inversen Kovarianz verschwindet.

1.4 Assoziationsstrukturen und Graphen

Elemente der Graphentheorie

Ein Graph besteht aus zwei Mengen, der Menge der *Knoten* K (vertices) und der Menge der *Kanten* E (edges). Die Kanten repräsentieren Verbindungen zwischen Knoten, sind also immer als Paare von Knoten darstellbar. Bezeichnet man der Einfachheit halber die Knoten durch natürliche Zahlen, $K = \{1, \dots, p\}$, dann ist eine Kante ein Tupel (i, j) , $i, j \in K$. Der *Graph* ist bestimmt durch

$$G = (K, E),$$

wobei formal $E \subset K \times K$. In einer Kante (i, j) wird i als Elternteil, j als Kind bezeichnet. Wenn E sowohl (i, j) als (j, i) enthält ist die Kante *ungerichtet* und wird als einfache Verbindung zwischen i und j dargestellt, andernfalls als gerichtete Verbindung von i nach j . Der Graph heißt ungerichtet, wenn alle Verbindungen ungerichtet sind. Das zu einem Graphen gehörende Diagramm stellt die Knoten als Kreise und die Kanten als (gerichtet oder ungerichtete) Verbindungen zwischen Knoten dar. Knoten heißen verbunden (benachbart), wenn die ungerichtete Kante in E enthalten ist.

Ein *Pfad* ist eine Sequenz von Knoten $i_1, i_2, \dots, i_m$, so daß alle aufeinanderfolgenden Knoten (i_j, i_{j+1}) im Graphen enthalten sind. Der Graph wird als *gerichtet*, von i_1 nach i_m , betrachtet. Der Pfad wird als *Zyklus* bezeichnet, wenn Anfangs- und Endknoten identisch sind, d.h. $i_1 = i_m$.

Bedingte Unabhängigkeitsgraphen

Sei nun $\mathbf{x}^T = (x_1, \dots, x_p)$ ein Vektor von Zufallsvariablen und bezeichne $K = \{1, \dots, k\}$ die Kanten der Komponenten. Ein (*bedingter*) *Unabhängigkeitsgraph* ist ein ungerichteter Graph $G = (K, E)$ mit der Eigenschaft, daß die Kanten (i, j) nicht in E enthalten ist, genau dann wenn

x_i und x_j bedingt unabhängig sind, gegeben der Rest der Variablen,

kurz $x_i \perp x_j | x_{K \setminus \{i, j\}}$.

Tabelle 1.1 zeigt einige Graphen für drei Variablen und die zugehörigen bedingten Unabhängigkeiten.

Das vollständige Fehlen von Kanten zwischen Gruppen von Variablen läßt sich als Unabhängigkeit interpretieren. Partitioniert man die Kanten K in die zwei Untermengen A und B , so daß kein Pfad zwischen irgendwelchen Variablen aus A und B im Graphen enthalten ist, dann sind alle Variablen aus A unabhängig von Variablen aus B , d.h. $x_i \perp x_j$ für alle $i \in A, j \in B$. Darüberhinaus gilt: $\{x_i, i \in A\}$ ist unabhängig von $\{x_i, i \in B\}$. (Block independence lemma)

Eine weitere zentrale Aussage für die Interpretation von Graphen ist das *Separationstheorem*. Ausgangspunkt dafür sind drei paarweise disjunkte Mengen A, B, C von Variablen.

Separationstheorem: Es gilt, daß alle in A enthaltenen Variablen bedingt unabhängig sind von den in B enthaltenen Variablen, gegeben die Variablen in C , genau dann, wenn jeder im Graphen enthaltene Pfad zwischen einer Variablen aus A und einer Variablen aus B mindestens eine Variable aus C enthält.

Bedingte Unabhängigkeit zwischen A und B gegeben C liegt also dann vor, wenn jeder Pfad zwischen A und B durch die Menge C führt. Man nennt zwei Knoten i, j auch separiert durch C wenn jeder Pfad, der i und j verbindet mindestens ein Element von C enthält.

1.5 Multiple Korrelation

Die multiple Korrelation erfaßt die Stärke des Zusammenhangs eines skalaren Merkmals y mit einer vektoriellen Größe $\mathbf{x}$. Betrachtet man wieder für zentrierte Variable y ($E(y) = 0$) das lineare Regressionsmodell

$$y = \mathbf{x}^T \boldsymbol{\beta} + \varepsilon_{y \cdot \mathbf{x}},$$

wobei $\boldsymbol{\beta}$ durch die beste lineare Prognose durch Minimierung von $E(y - \mathbf{x}^T \boldsymbol{\beta})^2$ bestimmt ist, erhält man

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_{\mathbf{x}^{-1} \boldsymbol{\Sigma}_{\mathbf{x}y}}.$$

Es ergibt sich

$$\begin{aligned} \text{var}(\mathbf{x}^T \boldsymbol{\beta}) &= \boldsymbol{\Sigma}_{y\mathbf{x}} \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} \boldsymbol{\Sigma}_{\mathbf{x}y} = \text{cov}(y, \mathbf{x}^T \boldsymbol{\beta}) \\ \text{var}(y|\mathbf{x}) &= \text{var}(\varepsilon_{y \cdot \mathbf{x}}) = \sigma_y^2 - \boldsymbol{\Sigma}_{y\mathbf{x}} \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} \boldsymbol{\Sigma}_{\mathbf{x}y} \end{aligned}$$

und man erhält die Zerlegung

$$\text{var}(y) = \text{var}(\mathbf{x}^T \boldsymbol{\beta}) + \text{var}(y|\mathbf{x}).$$

Der quadrierte *multiple Korrelationskoeffizient* R^2 ist bestimmt als die Korrelation zwischen y und $\mathbf{x}^T \boldsymbol{\beta}$,

$$R^2 = \varrho(y, \mathbf{x}^T \boldsymbol{\beta})^2 = \frac{\text{cov}(y, \mathbf{x}^T \boldsymbol{\beta})}{\sqrt{\text{var}(y)} \sqrt{\text{var}(\mathbf{x}^T \boldsymbol{\beta})}}.$$

Einfache Umformung ergibt

$$R^2 = \frac{\text{var}(\mathbf{x}^T \boldsymbol{\beta})}{\text{var}(y)} = 1 - \frac{\text{var}(y|\mathbf{x})}{\text{var}(y)},$$

das auch als Anteil der durch die Regression erklärte Varianz interpretiert wird. R^2 wird auch als *Determinations- oder Bestimmtheitsmaß* bezeichnet.

Für eindimensionales x erhält man

$$\text{var}(y|x) = \sigma_y^2 - \sigma_{yx}^2 / \sigma_x^2 = \sigma_y^2 (1 - \varrho_{yx}^2)$$

und entsprechend ist mit

$$R^2 = \varrho_{yx}^2$$

der Determinationskoeffizient identisch dem quadrierten Korrelationskoeffizienten.

1.6 Kanonische Korrelation

Bei der kanonischen Korrelation werden zwei Gruppen von Merkmalen betrachtet, deren maximale Korrelation bestimmt wird. Die beiden Gruppen von Variablen werden durch die Vektoren $\mathbf{y}$ und $\mathbf{x}$ der Länge q bzw. p repräsentiert mit der üblichen Bezeichnungsweise $\boldsymbol{\mu}_{\mathbf{y}} = E(\mathbf{y})$, $\boldsymbol{\mu}_{\mathbf{x}} = E(\mathbf{x})$ für die Erwartungswerte und $\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}} = \text{cov}(\mathbf{x}, \mathbf{y})$.

Die kanonische Korrelation ist bestimmt durch die maximale Korrelation von Linearkombinationen von $\mathbf{x}$ und $\mathbf{y}$, d.h. man sucht Gewichtsvektoren $\mathbf{a}$, $\mathbf{b}$, so daß die Korrelation $\varrho(\mathbf{a}^T \mathbf{x}, \mathbf{b}^T \mathbf{y})$ maximal wird.

Dazu betrachtet man

$$\varrho(\mathbf{a}, \mathbf{b}) = \varrho(\mathbf{a}^T \mathbf{x}, \mathbf{b}^T \mathbf{y}) = \frac{\text{cov}(\mathbf{a}^T \mathbf{x}, \mathbf{b}^T \mathbf{y})}{\sqrt{\text{var}(\mathbf{a}^T \mathbf{x})} \sqrt{\text{var}(\mathbf{b}^T \mathbf{y})}} = \frac{\mathbf{a}^T \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}} \mathbf{b}}{\sqrt{\mathbf{a}^T \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}} \mathbf{a}} \sqrt{\mathbf{b}^T \boldsymbol{\Sigma}_{\mathbf{y}\mathbf{y}} \mathbf{b}}}.$$

Das Maximierungsproblem ist offensichtlich nicht eindeutig lösbar, da die Korrelation dieselbe bleibt, wenn man $\mathbf{a}$ (bzw. $\mathbf{b}$) mit einer Konstanten λ_a (bzw. λ_b) multipliziert. Es gilt $\varrho(\lambda_a \mathbf{a}^T \mathbf{x}, \lambda_b \mathbf{b}^T \mathbf{y}) = \varrho(\mathbf{a}^T \mathbf{x}, \mathbf{b}^T \mathbf{y})$, da, technisch ausgedrückt, die Korrelation ein skaleninvariantes Maß ist. Die Vektoren $\mathbf{a}$ und $\mathbf{b}$ sind damit nur bis auf eine Konstante eindeutig bestimmt und zusätzliche Normierungsbedingungen sind notwendig um Eindeutigkeit zu erreichen. Anstatt die naheliegende Bedingung $\|\mathbf{a}\| = \|\mathbf{b}\| = 1$ zu

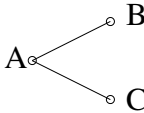
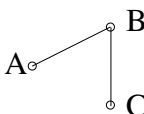
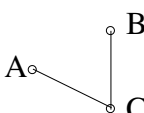
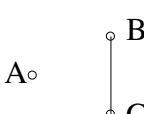
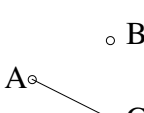
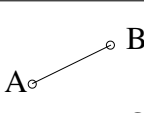
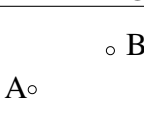
AB/AC	
X_B, X_C bedingt unabhängig, gegeben X_A	
AB/BC	
X_A, X_C bedingt unabhängig, gegeben X_B	
AC/BC	
X_A, X_B bedingt unabhängig, gegeben X_C	
A/BC	
X_A unabhängig von (X_B, X_C)	
AC/B	
(X_A, X_C) unabhängig von X_B	
AB/C	
(X_A, X_B) unabhängig von X_C	
$A/B/C$	
X_A, X_B, X_C sind unabhängig	

TABLE 1.1: Unabhängigkeitsgraphen

fordern, wählt man meist die Nebenbedingungen $\mathbf{a}^T \Sigma_{xx} \mathbf{a} = \text{var}(\mathbf{a}^T \mathbf{x}) = 1$, $\mathbf{b}^T \Sigma_{yy} \mathbf{b} = \text{var}(\mathbf{b}^T \mathbf{y}) = 1$. Das Maximierungsproblem stellt sich damit in der Form

$$\varrho(\mathbf{a}^T \mathbf{x}, \mathbf{b}^T \mathbf{y}) \xrightarrow{\mathbf{a}, \mathbf{b}} \max$$

mit der Nebenbedingung $\text{var}(\mathbf{a}^T \mathbf{x}) = \text{var}(\mathbf{b}^T \mathbf{y}) = 1$.

Mit diesen Nebenbedingungen ist das Maximierungsproblem unmittelbar äquivalent zu

$$\text{cov}(\mathbf{a}^T \mathbf{x}, \mathbf{b}^T \mathbf{y}) = \mathbf{a}^T \Sigma_{xy} \mathbf{b} \xrightarrow{\mathbf{a}, \mathbf{b}} \max.$$

Durch einfache Umformungen läßt sich zeigen, daß man äquivalenterweise den bedingten quadrierten Abstand der Linearkombination $\mathbf{a}^T \mathbf{x}$, $\mathbf{b}^T \mathbf{y}$ minimieren kann, d.h. man kann genauso fordern

$$\text{E}(\mathbf{a}^T \mathbf{x} - \mathbf{b}^T \mathbf{y})^2 \xrightarrow{\mathbf{a}, \mathbf{b}} \min$$

mit der Nebenbedingung $\text{var}(\mathbf{a}^T \mathbf{x}) = \text{var}(\mathbf{b}^T \mathbf{y}) = 1$.

Um die Äquivalenz zu zeigen, muß man nur $\text{E}(\mathbf{a}^T \mathbf{x} - \mathbf{b}^T \mathbf{y})^2$ umformen.

Das Problem läßt sich mit der Methode der Lagrange Multiplikationen zur Maximierung unter Nebenbedingungen lösen. Man erhält eine Folge von Gewichtsvektoren und Korrelationen

$$(\mathbf{a}_i, \mathbf{b}_i, \varrho_i), \quad i = 1, \dots, s = \min\{p, q\}$$

wobei s das Minimum der Länge der Vektoren $\mathbf{y}$ und $\mathbf{x}$ bezeichnet, und $\varrho_i(\mathbf{a}_i^T \mathbf{x}, \mathbf{b}_i^T \mathbf{y})$ in absteigender Reihenfolge geordnet ist

$$\varrho_1 > \varrho_2 > \dots > \varrho_s.$$

Die Lösung des einfachen Maximierungsproblems (??) ist gegeben durch (a_1, b_1, ϱ_1) . Die weiteren Gewichtsvektoren und Korrelation sind so konzipiert, daß

$$\text{cov}(\mathbf{a}_i \mathbf{x}, \mathbf{a}_j \mathbf{x}) = \text{cov}(\mathbf{b}_i \mathbf{y}, \mathbf{b}_j \mathbf{y}) = 0 \quad \text{für } i \neq j.$$

Dazu betrachtet man eine etwas schärfere Formulierung des Maximierungsproblems, die im Detail später behandelt wird.

Man bezeichnet:

$\mathbf{a}_i, \mathbf{b}_i$ als die i ten *Korrelationsvektoren*,

$u_i = \mathbf{a}_i^T \mathbf{x}, v_i = \mathbf{b}_i^T \mathbf{y}$ als i te *Korrelationsvariablen* und

$\varrho_i = \varrho(u_i, v_i)$ als i te *kanonische Korrelationskoeffizienten*.

Wie in Abb. 1.1 dargestellt, ist der Effekt für die transformierte Variable

u_i, v_i sind maximal korreliert,

u_i, u_j sind unkorreliert,

v_i, v_j sind unkorreliert,

u_i, v_j sind unkorreliert.

Für die Korrelation von $\mathbf{u}^T = (u_1, \dots, u_s)$ und $\mathbf{v}^T = (v_1, \dots, v_s)$ ergibt sich

$$\text{cor} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 0 & \varrho_1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & \varrho_s \\ \varrho_1 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & \varrho_s & 0 & \cdots & 1 \end{pmatrix}.$$

Schärfere Formulierung und Lösungen

Eine schärfere Formulierung der kanonischen Korrelation ist das Maximierungsproblem: Suche Paare von Gewichtsvektoren $(\mathbf{a}_i, \mathbf{b}_i)$, $i = 1, \dots, s = \min\{p, q\}$ so daß

$$\varrho(\mathbf{a}_i^T \mathbf{x}, \mathbf{b}_i^T \mathbf{y}) \xrightarrow{\mathbf{a}_i, \mathbf{b}_i} \max$$

mit den Nebenbedingungen

$$\text{var}(\mathbf{a}_i^T \mathbf{x}) = \text{var}(\mathbf{b}_i^T \mathbf{y}) = 1, \quad i = 1, \dots, s$$

$$\text{cov}(\mathbf{a}_i^T \mathbf{x}, \mathbf{a}_j^T \mathbf{x}) = \text{cov}(\mathbf{b}_i^T \mathbf{y}, \mathbf{b}_j^T \mathbf{y}) = 0, \quad j = 1, \dots, i-1.$$

Die erste Nebenbedingung entspricht der Normierung aller Gewichtsvektoren um Eindeutigkeit zu erreichen. Die zweite Bedingung fordert, daß die Lösung $(\mathbf{a}_i, \mathbf{b}_i)$ dergestalt ist, daß $\mathbf{a}_i^T \mathbf{x}$ unkorreliert ist mit den vorhergehenden $\mathbf{a}_1^T \mathbf{x}, \dots, \mathbf{a}_{i-1}^T \mathbf{x}$ und $\mathbf{b}_i^T \mathbf{y}$ unkorreliert mit $\mathbf{b}_1^T \mathbf{y}, \dots, \mathbf{b}_{i-1}^T \mathbf{y}$.

Aus der Betrachtung des Maximierungsproblems mit Lagrange Multiplikatoren erhält man, daß $(\mathbf{a}_i, \mathbf{b}_i)$ Lösungen des kombinierten Eigenwertproblems

$$(\Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} - \lambda_i \Sigma_{xx}) \mathbf{a}_i = 0$$

$$(\Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} - \lambda_i \Sigma_{yy}) \mathbf{b}_i = 0$$

sind.

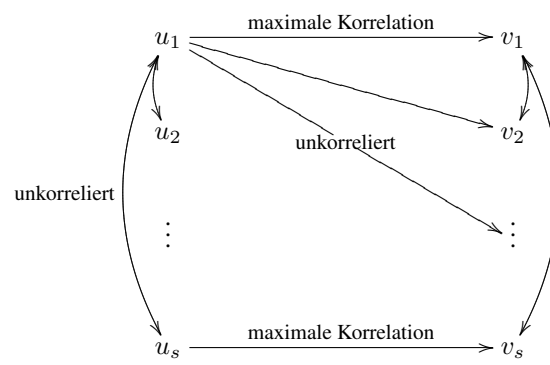


FIGURE 1.1: Kanonische Korrelation

