

Brownsche Distanz Korrelation



Bachelorarbeit
am Fachbereich 12 Mathematik und Informatik
der Philipps Universität Marburg

vorgelegt von Viktor Bengs

26. September 2012
Betreuer: Prof. Dr. H. Holzmann

Inhaltsverzeichnis

Einleitung	2
Kapitel 1 Pearsonsche Korrelation & Charakteristische Funktion	4
1.1 Pearsonsche Korrelation	4
1.2 Charakteristische Funktionen	10
Kapitel 2 Distanz Kovarianz & Distanz Korrelation	24
2.1 Definiton und grundlegende Eigenschaften	24
2.2 Empirische Distanz Kovarianz und Korrelation	41
2.3 Unabhängigkeitstest	56
Kapitel 3 Brownsche Distanz Korrelation	64
3.1 Bedingter Erwartungswert und Brownsche Bewegung	64
3.2 Brownsche Distanz Korrelation	69
Kapitel 4 Simulation	75
Literaturverzeichnis	80

Einleitung

Diese Arbeit beschäftigt sich mit der Möglichkeit Abhängigkeiten zwischen Zufallsvariablen beliebiger Dimensionen zu analysieren. Dabei werden drei Möglichkeiten genau diskutiert: die klassische Korrelation nach Pearson und die in dieser Arbeit vorgestellte Distanz Korrelation und Brownsche Distanz Korrelation.

Die Abhängigkeit oder allgemein der Zusammenhang zwischen Zufallsgrößen sind Aufgaben, welche in vielen Bereichen, wie zum Beispiel der Zeitreihenanalyse oder der Finanzmathematik, große Anwendung finden, aber auch in anderen Bereichen, wie in klinischen Studien, der Metaanalyse oder der Mustererkennung eine wichtige Rolle spielen.

Der häufig genutzte Zugang über die Pearsonsche Korrelation ist eine Möglichkeit lineare Zusammenhänge zwischen Zufallsgrößen zu ermitteln. Da die Bestimmung der Pearsonschen Korrelation relativ einfach ist, wird diese Maßzahl häufig verwendet, obwohl die Korrelation einige Nachteile mit sich bringt. So zum Beispiel werden nichtlineare Zusammenhänge zwischen den Zufallsgrößen von der Korrelation nicht miterfasst, deswegen ist die Unabhängigkeit durch die Korrelation nicht eindeutig charakterisierbar. Die Charakterisierung der Unabhängigkeit ist aber eine wünschenswerte Anforderung an ein Maß zur Messung des Zusammenhangs zwischen Zufallsgrößen. Die in vorliegender Arbeit vorgestellten Messgrößen, die Brownsche Distanz Korrelation und die Distanz Korrelation, erfüllen genau diese Anforderung. Beide Größen sind, wie die Pearsonsche Korrelation, Maße zur Messung der Abhängigkeit zwischen Zufallsvariablen und haben außerdem noch einige interessante Eigenschaften, die den Eigenschaften der Pearsonschen Korrelation ähneln. Zudem kann die Brownsche Distanz Korrelation als eine Erweiterung beziehungsweise eine Verallgemeinerung der Pearsonschen Korrelation interpretiert werden.

Das erste Kapitel beschäftigt sich nochmal mit dem Zugang über den Pearsonschen Korrelationskoeffizienten um einerseits nochmal die Nachteile bzw. Grenzen dieser Maßzahl zu verdeutlichen und um andererseits parallele Eigenschaften, welche in Kapitel

2 erst deutlich werden, mit der Brownschen Distanz Korrelation aufzuzeigen. Außerdem werden die charakteristischen Funktionen eingeführt, da diese Funktionen grundlegend für die Distanz Korrelation sind und daher durchgängig in den weiteren Kapiteln Verwendung finden werden.

Im zweiten Kapitel definieren wir die Distanz Korrelation und untersuchen einige ihrer Eigenschaften, wobei die bereits erwähnten Gemeinsamkeiten beziehungsweise Unterschiede zur Pearsonschen Korrelation genauer betrachtet werden. Danach werden Schätzer für die Distanz Korrelation und Distanz Kovarianz eingeführt und deren Eigenschaften genauer analysiert. Anschließend wird ein Test auf Unabhängigkeit vorgestellt, welcher einen engen Zusammenhang zu dem zuvor erwähnten Schätzer der Distanz Kovarianz besitzt. Dieses Kapitel basiert hauptsächlich auf den Arbeiten von Székely, Rizzo und Bakirov aus dem Artikel „Measuring and Testing dependence by correlation of distances“, 2007.

Das dritte Kapitel befasst sich mit der Brownschen Distanz Korrelation. Die Hauptaussage dieses Kapitels ist, dass Brownsche Distanz Korrelation und Distanz Korrelation gleich sind und dass die Brownsche Distanz Korrelation eine Erweiterung der Pearsonschen Korrelation ist. Dieses Kapitel befasst sich überwiegend mit den Ergebnissen von Székely, Rizzo und Bakirov aus dem Artikel „Brownian Distance Covariance“, 2009.

Im vierten Kapitel sind einige Beispielanwendungen für die vorherigen Ergebnisse aus Kapitel 2 und Kapitel 3 gegeben.

Zur Berechnung von Daten und der Darstellung von Grafiken wurde die statistische Programmiersprache R verwendet.

Kapitel 1

Pearsonsche Korrelation & Charakteristische Funktion

1.1 Pearsonsche Korrelation

Für diesen Unterabschnitt definieren wir Voraussetzungen, die fortan für jeden Satz gelten außer wenn explizit andere Voraussetzungen angegeben sind: $(\Omega, \mathbf{A}, \mathbf{P})$ ist ein Maßraum und (Ω', \mathbf{A}') ist ein Messraum. $X, Y: \Omega \rightarrow \Omega'$ sind zwei Zufallsvariablen auf $(\Omega, \mathbf{A}, \mathbf{P})$ mit $\mathbb{E}(X^2) + \mathbb{E}(Y^2) < \infty$

Definition 1.1 (Kovarianz). Seien X und Y Zufallsvariablen, so heißt

$$\begin{aligned} Cov &:= Cov(X, Y) = \mathbb{E}((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))) \\ &= \int_{\Omega} \int_{\Omega} (x - \mathbb{E}(X))(y - \mathbb{E}(Y)) dP_X(x) dP_Y(y) \end{aligned}$$

die *Kovarianz* von X und Y .

Die Kovarianz ist eine Messgröße um den linearen Zusammenhang zwischen zwei Zufallsvariablen zu ermitteln. Wenn die Kovarianz negativ ist, dann gehen große Werte der einen Zufallsvariable größtenteils mit kleinen Werten der anderen Zufallsvariablen einher. Falls die Kovarianz positiv ist, gehen große Werte der einen Zufallsvariable größtenteils mit großen Werten der anderen Zufallsvariablen einher.

Beispiel 1.2. Seien $X \sim \text{Bin}(n, p_X)$ und $Y \sim \text{Bin}(n, p_Y)$ und $\widehat{X} = (X, Y) \sim \text{Mult}(n, (p_x, p_y))$.

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))) = \mathbb{E}(XY - \mathbb{E}(X)Y - \mathbb{E}(Y)X + \mathbb{E}(X)\mathbb{E}(Y)) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(\widehat{X}) - \mathbb{E}(X)\mathbb{E}(Y) \\ &= n(n-1)p_X p_Y - n^2 p_X p_Y = -n p_X p_Y \end{aligned}$$

Die Kovarianz besitzt vorteilhafte Eigenschaften, die im folgenden Satz zusammengefasst sind:

Satz 1.3. Seien X, X_1, X_2, \dots, X_n und Y Zufallsvariablen, dann gelten folgende Eigenschaften:

- (i) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
- (ii) $\text{Cov}(aX+b, Y) = a\text{Cov}(X, Y)$;
- (iii) $\text{Cov}(X+Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$;
- (iv) $\text{Cov}(X, X) = \text{Var}(X)$;
- (v) $|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}$;
- (vi) $\text{Var}(\sum_{i=1}^n X_i) = \sum_{i,j=1}^n \text{Cov}(X_i, X_j) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{Cov}(X_i, X_j)$.

Beweis. (i)

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))) = \mathbb{E}((Y - \mathbb{E}(Y)) \cdot (X - \mathbb{E}(X))) = \text{Cov}(Y, X).$$

(ii)

$$\text{Cov}(X, Y) = \mathbb{E}(a(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = a\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = a\text{Cov}(X, Y).$$

(iii)

$$\begin{aligned} \text{Cov}(X + Y, Z) &= \mathbb{E}((X + Y - \mathbb{E}(X + Y))(Z - \mathbb{E}(Z))) \\ &= \mathbb{E}(((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))(Z - \mathbb{E}(Z))) \\ &= \mathbb{E}((X - \mathbb{E}(X))(Z - \mathbb{E}(Z)) + (Y - \mathbb{E}(Y))(Z - \mathbb{E}(Z))) \\ &= \mathbb{E}((X - \mathbb{E}(X))(Z - \mathbb{E}(Z))) + \mathbb{E}((Y - \mathbb{E}(Y))(Z - \mathbb{E}(Z))) \\ &= \text{Cov}(X, Z) + \text{Cov}(Y, Z). \end{aligned}$$

(iv)

$$\text{Cov}(X, X) = \mathbb{E}((X - \mathbb{E}(X)) \cdot (X - \mathbb{E}(X))) = \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2)$$

$$\begin{aligned} &= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(X)^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \text{Var}(X). \end{aligned}$$

(v) Es reicht zu zeigen, dass $\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$ gilt.

Wir definieren $f(a) := \mathbb{E}((Y - aX)^2)$, dann ist $f'(a) = -2\mathbb{E}(XY) + 2a\mathbb{E}(X^2)$ und somit $a_{min} = \frac{\mathbb{E}(XY)}{\mathbb{E}(X^2)}$. Setzen wir nun a_{min} in f ein, erhalten wir:

$$0 \leq f(a_{min}) = \mathbb{E}(Y^2) - 2\frac{\mathbb{E}(XY)^2}{\mathbb{E}(X^2)} + \frac{\mathbb{E}(XY)^2}{\mathbb{E}(X^2)} \Leftrightarrow \mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2).$$

Außerdem gilt: $\mathbb{E}(|XY|) < \infty$ wegen $|XY| \leq (\max\{X, Y\})^2 \leq X^2 + Y^2$.

(vi) folgt aus (iii) und (iv). □

Bemerkung: Aus den Eigenschaften (i)-(iii) folgt, dass die Kovarianz eine symmetrische Bilinearform auf dem Raum der quadratisch integrierbaren Zufallsvariablen ist. Die Eigenschaft (v) bezeichnet man als die Cauchy-Schwarzsche Ungleichung.

Anmerkung: Oftmals hat man bei realen Aufgabenstellungen das Problem dass man die genauen Parameter einer Zufallsvariable nicht genau kennt, aber dafür eine bestimmte Anzahl an Beobachtungen der Ausprägungen der gesuchten Zufallsvariable gegeben ist. Aus diesen Beobachtungen ist es nun möglich erwartungstreue Schätzer für die gewünschten Parameter zu berechnen. Dabei gilt: Desto größer die Anzahl der Beobachtungen, desto besser ist die Approximation.

Definition 1.4 (Empirische Kovarianz). Seien x_1, x_2, \dots, x_n und y_1, y_2, \dots, y_n beobachtete Werte einer Zufallsvariable X bzw. Y . Die *empirische Kovarianz* ist gegeben durch:

$$s_{X,Y}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n),$$

wobei

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \quad \bar{y}_n = \frac{1}{n} \sum_{i=1}^n y_i$$

die empirischen Mittelwerte von x_1, \dots, x_n bzw. y_1, \dots, y_n sind.

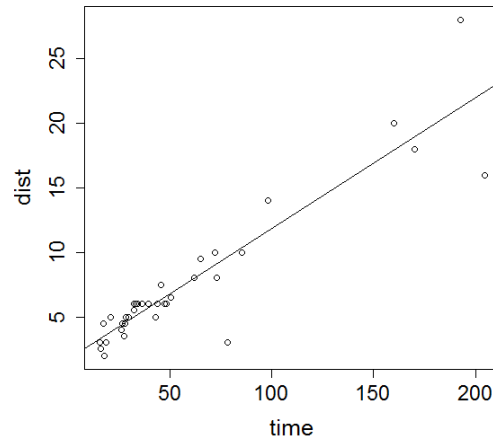


Abbildung 1.1: Regressionsgerade für Datensatz hills

Beispiel 1.5. Betrachten wir den Datensatz *hills* aus dem R-Paket **MASS** und untersuchen den Zusammenhang zwischen den Stichproben des Attributs *time* und des Attributs *dist*, so sehen wir graphisch in Abb.1.1, dass ein linearer Zusammenhang besteht.

Berechnen wir die empirische Kovarianz für die Werte erhalten wir einen Wert von $\text{cov}(\text{dist}, \text{time}) = 254.1944$ und schließen daraus, dass ein positiver linearer Zusammenhang zwischen den Stichproben besteht. Nun bleibt die Frage nach der Stärke des Zusammenhangs, die als nächstes geklärt wird.

Bei unseren ersten Vorüberlegungen und dem Beispiel oben haben wir festgestellt, dass das Vorzeichen der Kovarianz lediglich die Tendenz des linearen Zusammenhangs bestimmt, aber nicht den Grad des Zusammenhangs. Um zusätzlich den Grad zu bestimmen, ist eine Normierung der Kovarianz notwendig.

Definition 1.6 (Pearsonsche Korrelation). Seien X und Y zwei Zufallsvariablen mit endlichen zweiten Momenten, dann ist die *Korrelation* (nach Pearson) gegeben durch:

$$\rho := \rho_{X,Y} = \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}}.$$

Diese Maßzahl dient ähnlich wie die Kovarianz zur Bestimmung des linearen Zusammenhangs zweier Zufallsvariablen. Der Unterschied zur Kovarianz ist, dass die Korrelation eine standardisierte Größe ist, das heißt der Wertebereich ist auf das Intervall $[-1,1]$ eingeschränkt. Liegt der Wert bei 1, so haben die Zufallsvariablen einen perfekten

positiven Zusammenhang. Wenn der Wert bei -1 liegt, dementsprechend einen perfekten negativen linearen Zusammenhang. Ist der Wert 0, so liegt kein linearer Zusammenhang vor, dennoch ist es möglich das die Zufallsvariablen nichtlinear zusammenhängen. Diese Situation bleibt aber in der Regel von dieser Maßzahl unentdeckt. Dies gilt auch für die Kovarianz.

Bemerkung Ist $X = (X_1, \dots, X_d)$ dann wird die Kovarianz bzw. die Korrelation mit der sogenannten *Kovarianzmatrix* $Cov(X)$ bzw. *Korrelationsmatrix* $\rho(X)$ ausgedrückt. Es gilt

$$Cov(X) = (Cov(X_i, X_j))_{i,j=1}^d \text{ und } \rho(X) = (\rho(X_i, X_j))_{i,j=1}^d.$$

Beispiel 1.7. Seien $X \sim Bin(n, p_X)$ und $Y \sim Bin(n, p_Y)$ und $\hat{X} = (X, Y) \sim Mult(n, (p_x, p_y))$. In Beispiel 1.5 haben wir schon gesehen, dass $Cov(X, Y) = -np_X p_Y$. Daraus errechnen wir:

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{-np_X p_Y}{\sqrt{np_X(1-p_X)}\sqrt{np_Y(1-p_Y)}} = -\sqrt{\frac{p_X p_Y}{(1-p_X)(1-p_Y)}}.$$

Satz 1.8. Seien X und Y Zufallsvariablen

- (i) $\rho_{X,Y} = \rho_{Y,X}$;
- (ii) $\rho_{X,X} = 1$;
- (iii) $\rho_{X,Y} \in [-1, 1]$;
- (iv) $\rho_{bX+a, cY+d} = \frac{b}{|b|} \frac{c}{|c|} \rho_{X,Y}$ für $a, b, c, d \in \mathbb{R}$.

Beweis. (i) und (ii) folgen direkt aus der Definition der Korrelation.

(iii)

$$\rho_{X,Y}^2 = \frac{Cov(X, Y)^2}{Var(X)Var(Y)} \stackrel{\text{Satz 1.3(v)}}{\leq} \frac{Var(X)Var(Y)}{Var(X)Var(Y)} = 1.$$

(iv)

$$\rho_{bX+a, cY+d} = \frac{Cov(bX + a, cY + d)}{\sqrt{Var(bX + a)Var(cY + d)}} = \frac{bc \cdot Cov(X, Y)}{\sqrt{b^2 Var(X)c^2 Var(Y)}} = \frac{b}{|b|} \frac{c}{|c|} \rho_{X,Y}.$$

□

Bemerkung Eigenschaft (iv) besagt, dass ρ invariant unter positiven linearen Transformationen ist.

Definition 1.9 (Unkorreliert). Zwei Zufallsvariablen X und Y heißen *unkorreliert*, falls gilt $\rho_{X,Y} = 0$.

Unmittelbar aus der Definition können wir folgern, dass Zufallsvariablen unkorreliert sind, wenn sie unabhängig sind.

Satz 1.10. X, Y unabhängig $\implies X, Y$ unkorreliert.

Beweis.

Es reicht zu zeigen, dass die Kovarianz von X und Y gleich 0 ist:

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))) = \mathbb{E}(XY - \mathbb{E}(X)Y - \mathbb{E}(Y)X + \mathbb{E}(X)\mathbb{E}(Y)) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) \\ &= 2\mathbb{E}(X)\mathbb{E}(Y) - 2\mathbb{E}(X)\mathbb{E}(Y) = 0. \end{aligned}$$

□

Bemerkung: Aus diesem Satz folgern wir nun, dass wenn X und Y unabhängig sind, gilt:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + \underbrace{2\text{Cov}(X, Y)}_{=0} = \text{Var}(X) + \text{Var}(Y). \quad (1.1)$$

Wünschenswert wäre die Äquivalenz in Satz 1.10, da die Berechnung der Korrelation leicht durchführbar ist und man somit ein effektives Werkzeug hätte, um die Unabhängigkeit von Zufallsvariablen nachzuweisen. Jedoch impliziert die Unkorreliertheit nicht die Unabhängigkeit wie das folgende Beispiel zeigt.

Beispiel 1.11. Sei (Ω, P) ein Laplaceraum über $\Omega = \{0, 1, -1\}$ und $X(\omega) := |\omega|$, $Y(\omega) := \omega$, dann sind X und Y unkorreliert:

$$\mathbb{E}(XY) = \mathbb{E}(Y) = \frac{-1 + 0 + 1}{3} = 0 = \mathbb{E}(X) \cdot \mathbb{E}(Y) \implies \text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0,$$

aber X und Y sind nicht unabhängig:

$$P(X = Y = 0) = P(Y = 0) = \frac{1}{3} \neq \frac{1}{9} = \frac{1}{3} \cdot \frac{1}{3} = P(X = 0) \cdot P(Y = 0).$$

Wie bei der Kovarianz kann man auch bei der Korrelation einen Erwartungstreuen Schätzer für diese Maßzahl definieren. Dabei ist der Zusammenhang zwischen der empirischen Kovarianz und der empirischen Korrelation analog wie bei der Kovarianz und der Korrelation.

Definition 1.12 (Empirische Korrelation). Seien x_1, x_2, \dots, x_n und y_1, y_2, \dots, y_n beobachtete Werte einer Zufallsvariable X bzw. Y . Die empirische Korrelation ist gegeben durch:

$$\tilde{\rho}_{X,Y} = \frac{s_{X,Y}^2}{\sqrt{s_{X,X}^2 s_{Y,Y}^2}} = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_n)^2 \sum_{i=1}^n (y_i - \bar{y}_n)^2}}.$$

Beispiel 1.13. Betrachten wir noch einmal den Datensatz *hills* aus dem Paket **MASS** wie in Beispiel 1.5 und berechnen die empirische Korrelation:

$$\tilde{\rho}_{dist,time} = \text{cor}(dist,time) = 0.9195892.$$

Damit sehen wir nun auch rechnerisch das tatsächlich ein starker positiver linearer Zusammenhang zwischen *dist* und *time* besteht.

1.2 Charakteristische Funktionen

Die Charakteristische Funktion ist eine Möglichkeit die Verteilung einer Zufallsvariablen zu charakterisieren. Insbesondere die Unabhängigkeit, die einen engen Zusammenhang zu charakteristischen Funktionen besitzt, wird hier genau untersucht. Es werden einige Eigenschaften der Charakteristischen Funktion vorgestellt, die durchgängig in dieser Arbeit von Nutzen sein werden. Das Hauptziel dieses Unterabschnittes ist eine Äquivalenzbedingung zwischen Unabhängigkeit und Charakteristischen Funktionen und einen Schätzer für die charakteristische Funktion zu erhalten.

Definition 1.14 (Charakteristische Funktion). Sei X eine Zufallsvariable mit $\dim(X)=n$, dann heißt $\varphi_X : \mathbb{R}^n \rightarrow \mathbb{C}$ mit

$$\varphi_X(t) := \int_{\Omega} e^{it^T X} dP = \mathbb{E}(e^{it^T X}) \quad \forall t \in \mathbb{R}^n$$

die *charakteristische Funktion von X*.

Bemerkung

- (i) Wegen $|e^{it^T X}| = 1$ und der majorisierten Konvergenz ist $\varphi_X(t)$ für alle $t \in \mathbb{R}^n$ definiert;
- (ii) $\varphi_X(t)$ ist nur von der Verteilung P_X abhängig.

Lemma 1.15. Es gilt: $|e^{i\alpha} - 1| \leq |\alpha| \quad \forall \alpha \in \mathbb{R}$.

Beweis.

$$\begin{aligned} |e^{i\alpha} - 1| &= 2 \left| \frac{e^{-i\frac{\alpha}{2}}}{2} \right| \cdot |e^{i\alpha} - 1| = 2 \left| \frac{e^{-i\frac{\alpha}{2} + i\alpha} - e^{-i\frac{\alpha}{2}}}{2} \right| = 2 \left| \frac{e^{i\frac{\alpha}{2}} - e^{-i\frac{\alpha}{2}}}{2} \right| = 2 \left| \sin \frac{\alpha}{2} \right| \leq 2 \left| \frac{\alpha}{2} \right| \\ &= |\alpha|. \end{aligned}$$

□

Wir zeigen nun eine Reihe von besonderen Eigenschaften der charakteristischen Funktion.

Satz 1.16. Sei X eine Zufallsvariable mit $\dim(X)=n$ und bezeichne φ_X die Charakteristische Funktion von X . Es gilt:

- (i) $\varphi_X(0) = 1$;
- (ii) $|\varphi_X(t)| \leq 1 \quad \forall t \in \mathbb{R}^n$;
- (iii) $\overline{\varphi_X(t)} = \varphi_X(-t) = \varphi_{-X}(t)$;
- (iv) $X=0$ P-fast sicher $\Rightarrow \varphi_X \equiv 1$;
- (v) Für $A \in \mathbb{R}^{m \times n}$, $\beta \in \mathbb{R}^m$ und $Y := AX + \beta$ gilt: $\varphi_Y(t) = e^{it^T \beta} \varphi_X(A^T t)$;
- (vi) $\varphi_X(t)$ ist gleichmäßig stetig.

Beweis. (i)

$$\varphi_X(0) = \mathbb{E}(e^{i0^T X}) = \mathbb{E}(e^0) = \mathbb{E}(1) = 1.$$

(ii)

$$|\varphi_X(t)| = |\mathbb{E}(e^{it^T X})| \leq \mathbb{E}(|e^{it^T X}|) = \mathbb{E}(1) = 1.$$

(iii)

$$\overline{\varphi_X(t)} = \overline{\mathbb{E}(e^{it^T X})} = \mathbb{E}(\overline{e^{it^T X}}) = \mathbb{E}(e^{-it^T X}) = \mathbb{E}(e^{i(-t)^T X}) = \varphi_X(-t)$$

und

$$\varphi_X(-t) = \mathbb{E}(e^{-it^T X}) = \mathbb{E}(e^{it^T (-X)}) = \varphi_{-X}(t).$$

(iv)

$$\varphi_X(t) = \int_{\Omega} e^{it^T X} dP = \int_{\Omega} e^{it^T 0} dP = \int_{\Omega} 1 dP = 1 \quad \forall t \in \mathbb{R}^n.$$

(v)

$$\begin{aligned}\varphi_Y(t) &= \mathbb{E}(e^{it^T Y}) = \mathbb{E}(e^{it^T (AX+\beta)}) = \mathbb{E}(e^{i(A^T t)^T X} \cdot e^{it^T \beta}) \\ &= e^{it^T \beta} \cdot \mathbb{E}(e^{i(A^T t)^T X}) = e^{it^T \beta} \varphi_X(A^T t).\end{aligned}$$

(vi) Sei $\epsilon > 0$, wir wählen $C > 0$ so groß, dass $P(|X| > C) < \frac{\epsilon}{3}$ und $|(t_1 - t_2)^T| \leq \frac{\epsilon}{3C}$. Wir erhalten mit Lemma 1.15

$$\begin{aligned}|\varphi_X(t_1) - \varphi_X(t_2)| &= \left| \int_{\Omega} e^{it_1^T X} - e^{it_2^T X} dP \right| \leq \int_{\Omega} |e^{it_1^T X} - e^{it_2^T X}| dP \\ &= \int_{|X| \leq C} |e^{it_1^T X} - e^{it_2^T X}| dP + \int_{|X| > C} \underbrace{|e^{it_1^T X} - e^{it_2^T X}|}_{\leq 2} dP \\ &\leq \int_{|X| \leq C} |e^{it_1^T X} - e^{it_2^T X}| dP + \int_{|X| > C} 2 dP \\ &= \int_{|X| \leq C} \underbrace{|e^{it_2^T X}|}_{\leq 1} \cdot |e^{i(t_1-t_2)^T X} - 1| dP + 2 \cdot \underbrace{P(|X| > C)}_{< \frac{\epsilon}{3}} \\ &\leq \int_{|X| \leq C} |e^{i(t_1-t_2)^T X} - 1| dP + \frac{2}{3}\epsilon \\ &\leq \int_{|X| \leq C} |(t_1 - t_2)^T X| dP + \frac{2}{3}\epsilon \\ &\leq \underbrace{|(t_1 - t_2)^T|}_{\leq \frac{\epsilon}{3C}} C + \frac{2}{3}\epsilon \leq \epsilon.\end{aligned}$$

Damit folgt die Behauptung. □

Satz 1.17. Seien X und Y unabhängige Zufallsvariablen mit $\dim(X)=n$, $\dim(Y)=m$, dann gilt:

$$\varphi_{X,Y}(t, s) = \varphi_X(t) \cdot \varphi_Y(s) \quad \forall t \in \mathbb{R}^n, \forall s \in \mathbb{R}^m.$$

Beweis.

$$\varphi_{X,Y}(t, s) = \mathbb{E}(e^{i(t^T X + s^T Y)}) = \mathbb{E}(e^{it^T X} e^{is^T Y}) \stackrel{X,Y \text{ unabh.}}{=} \mathbb{E}(e^{it^T X}) \cdot \mathbb{E}(e^{is^T Y}) = \varphi_X(t) \cdot \varphi_Y(s).$$

□

Anmerkung Die in Satz 1.17 verwendete Charakteristische Funktion $\varphi_{X,Y}(t, s)$ ist die Charakteristische Funktion der gemeinsamen Verteilung von X und Y . Wenn die Umkehrung in Satz 1.17 gilt, erhalten wir eine Äquivalenzbedingung für die Unabhängigkeit. Die Umkehrung wird später gezeigt.

1.2. Charakteristische Funktionen

Außerdem folgt aus dem Satz folgendes: Sind X und Y unabhängig mit $\dim(X) = n = \dim(Y) \Rightarrow \varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t)$. Wählen wir $Z := X+Y = A \cdot (X, Y)^T$ mit $A := (I_n \ I_n)$, wobei I_n die $n \times n$ Einheitsmatrix ist, dann ist

$$\varphi_Z(t) \stackrel{\text{Satz 1.16(v)}}{=} \underbrace{e^{it^T \cdot 0}}_{=1} \varphi_{X,Y}(A^T t) = \varphi_{X,Y}(t, t) \stackrel{\text{Satz 1.17}}{=} \varphi_X(t) \cdot \varphi_Y(t).$$

Beispiel 1.18 (Normalverteilung). Da wir in einem späteren Zusammenhang die charakteristische Funktion der Normalverteilung benutzen werden, ermitteln wir diese hier zunächst. Sei $X \sim N(0, 1)$ mit $f_X(y) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{y^2}{2}}$, dann ermitteln wir die charakteristische Funktion:

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{ity} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{e^{ity - \frac{y^2}{2}}}_{:=A(t)} dy,$$

dann ist wegen $|\frac{A(t+h)-A(t)}{h}| \leq |\frac{e^{ihy}-1}{h}| \cdot |A(t)| \leq |A(t)| \frac{|y||h|}{|h|} = |A(t)||y|$ und dem Satz von der majorisierten Konvergenz die Funktion integrierbar, also führen wir die Berechnung der CF auf eine Differentialgleichung zurück:

$$\begin{aligned} \varphi'_X(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} iye^{ity - \frac{y^2}{2}} dy = \frac{i}{\sqrt{2\pi}} \left[\frac{-1}{y} e^{-\frac{y^2}{2}} ye^{ity} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} ite^{-\frac{y^2}{2}} \cdot e^{ity} dy \\ &= -\frac{t}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} \cdot e^{ity} dy = -t \int_{-\infty}^{\infty} e^{ity} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{y^2}{2}} dy = -t \cdot \varphi_X(t). \end{aligned}$$

Die Differentialgleichung besitzt die Lösung $\varphi_X(t) = e^{-\frac{t^2}{2}} \cdot const$.

Für $t=0$ erhalten wir $1 = \varphi_X(0) = e^{-\frac{0^2}{2}} \cdot const = 1 \cdot const \Rightarrow const = 1$ und somit $\Rightarrow \varphi_X(t) = e^{-\frac{t^2}{2}}$.

Ist $Y \sim N(\mu, \sigma^2)$ dann ist $Y = \sigma X + \mu$ also mit Satz 1.16(v)

$$\varphi_Y(t) = e^{it\mu} \cdot \varphi_X(\sigma^T t) = e^{it\mu - \frac{(\sigma t)^2}{2}}.$$

Für den multivariaten Fall: Sei $X \sim N(\mu, \Sigma)$ mit Kovarianzmatrix $\Sigma \in \mathbb{R}^{n \times n}$ dann ist:

$$\varphi_X(t) = e^{it^T \mu - \frac{t^T \Sigma t}{2}}.$$

Hier ein Überblick über einige charakteristische Funktionen:

Verteilung	charakteristische Funktion	Verteilung	charakteristische Funktion
Ber(p)	$pe^{it} + 1 - p$	Unif(a,b)	$\frac{e^{ibt} - e^{iat}}{i(b-a)t}$
Bin(n,p)	$(pe^{it} + 1 - p)^n$	$N(\mu, \sigma^2)$	$e^{\mu it} e^{-\frac{\sigma^2 t^2}{2}}$
NBin(r,p)	$(\frac{1-pe^{it}}{1-p})^{-r}$	χ_n^2	$(1 - 2it)^{-\frac{n}{2}}$
Poi(λ)	$e^{\lambda(e^{it}-1)}$	exp(λ)	$\frac{\lambda}{\lambda - it}$
Geo(p)	$\frac{p}{1-(1-p)e^{it}}$	Cauchy(a, b)	$e^{iat-b t }$

1.2. Charakteristische Funktionen

Wir wollen eine Darstellung der charakteristischen Funktion haben, die der Taylorentwicklung ähnelt, aber anstelle der Ableitungen der Funktion die Momente der Zufallsvariable benutzt. Dazu beweisen wir einen Satz über die Ableitungen der charakteristischen Funktionen und anschließend eine Aussage über das Restglied der Exponentialreihe.

Satz 1.19. Sei $X = (X_1, \dots, X_n)$ eine Zufallsvariable zu der alle gemischten Momente existieren, d.h. $\mathbb{E}(\prod_{j=1}^n X_j^{r_j}) < \infty$ mit $0 \leq r_j \leq n_j$, $j = 1, \dots, n$ und $r_j, n_j \in \mathbb{N}$, dann gilt:

(i) $D(t) := \frac{\partial^{r_1+\dots+r_n}}{\partial t_1^{r_1} \dots \partial t_n^{r_n}} \varphi_X(t)$ existiert $\forall t \in \mathbb{R}^n$ und ist gleichmäßig stetig in t ;

(ii) $D(0) = i^{\sum_{j=1}^n r_j} \cdot \mathbb{E}(\prod_{j=1}^n X_j^{r_j})$.

Beweis. Abkürzend bezeichnen wir $r := (r_1, \dots, r_n)$. Wegen $\int_{\Omega} |\prod_{j=1}^n X_j^{r_j}| dP < \infty \forall r_j \leq n_j$ $j = 1, \dots, n$ existiert

$$\psi_r(t) := i^{\sum_{j=1}^n r_j} \int_{\Omega} (\prod_{j=1}^n X_j^{r_j}) e^{it^T X} dP \quad \forall t \in \mathbb{R}^n.$$

Nun setzen wir $Y := \prod_{j=1}^n X_j^{r_j}$ und definieren die Dichten

$$dP^+ := \begin{cases} 0 & \text{falls } \int Y^+ dP = 0, \\ \frac{Y^+}{\int Y^+ dP} & \text{sonst} \end{cases}, \quad \text{und } dP^- := \begin{cases} 0 & \text{falls } \int Y^- dP = 0, \\ \frac{Y^-}{\int Y^- dP} & \text{sonst.} \end{cases}$$

Dann ist

$$\psi_r(t) = i^{\sum_{j=1}^n r_j} \left(\int_{\Omega} e^{it^T X} dP^+ + \int_{\Omega} e^{it^T X} dP^- \right)$$

als charakteristische Funktion aufgefasst gleichmäßig stetig.

Für $h \neq 0$ gilt

$$\frac{\psi_r(t_1 + h, \dots, t_n) - \psi_r(t_1, \dots, t_n)}{h} = i^{\sum_{j=1}^n r_j} \underbrace{\int_{\Omega} (\prod_{j=1}^n X_j^{r_j}) e^{it^T X} \left(\frac{e^{ihX_1} - 1}{h} \right) dP}_{:=g(h)}$$

Es gilt $|g(h)| = |(\prod_{j=1}^n X_j^{r_j}) e^{it^T X} \left(\frac{e^{ihX_1} - 1}{h} \right)| \leq |(\prod_{j=1}^n X_j^{r_j})| \cdot \left| \frac{e^{ihX_1} - 1}{h} \right| = |\prod_{j=1}^n X_j^{r_j}| \cdot |X_1|$

und wegen $\lim_{h \rightarrow 0} \frac{e^{ihX_1} - 1}{h} \stackrel{L'Hospital}{=} \lim_{h \rightarrow 0} \frac{iX_1 e^{ihX_1}}{1} = iX_1$ folgt

$$g(h) \rightarrow \left(\prod_{j=1}^n X_j^{r_j} \right) e^{it^T X} \cdot iX_1 \quad h \rightarrow 0.$$

Mit dem Satz von der majorisierten Konvergenz folgt, dass $\frac{\partial}{\partial t_1} \psi_r(t)$ existiert mit $\frac{\partial}{\partial t_1} \psi_r(t) = \psi_{(r_1+1, r_2, \dots, r_n)}(t)$. Wegen $\psi_0(t) = \varphi_X(t)$ ergibt sich durch iteratives Anwenden von dem obigen die Existenz von

$$\frac{\partial^{r_1 + \dots + r_n}}{\partial t_1^{r_1} \dots \partial t_n^{r_n}} \varphi_X(t) = D(t).$$

Weiter gilt $D(t)$ ist gleichmäßig stetig, da $\varphi_r(t)$ gleichmäßig stetig \Rightarrow (i). Setzen wir für $t=0$ in $D(t)$ ein erhalten wir:

$$D(0) = \psi_r(0) = i^{\sum_{j=1}^n r_j} \cdot \mathbb{E} \left(\prod_{j=1}^n X_j^{r_j} \right)$$

und somit folgt (ii). □

Lemma 1.20. Es gilt $e^{ix} - \sum_{k=0}^m \frac{(ix)^k}{k!} = \frac{x^{m+1}}{(m+1)!} \cdot \vartheta(x)$ mit $|\vartheta(x)| \leq 1 \quad \forall x \in \mathbb{R}$.

Mit den vorangegangenen Hilfsmitteln ist es nun möglich eine Art Taylorentwicklung für die charakteristische Funktion anzugeben.

Satz 1.21. Sei X eine Zufallsvariable mit $\mathbb{E}(|X|^{k+1}) < \infty$ für ein $k \in \mathbb{N}$, wobei $|\vartheta(x)| \leq 1$. Dann gilt:

$$\varphi_X(t) = \sum_{j=0}^k \frac{(it)^j}{j!} \mathbb{E}(|X|^j) + \vartheta(t) \frac{t^{k+1}}{(k+1)!} \mathbb{E}(|X|^{k+1}) \quad \forall t \in \mathbb{R}.$$

Beweis.

$$\begin{aligned} \varphi_X(t) &= \mathbb{E}(e^{itX}) \stackrel{\text{Lemma 1.20}}{=} \mathbb{E} \left(\sum_{j=0}^k \frac{(it)^j}{j!} (|X|^j) + \vartheta(t) \frac{t^{k+1}}{(k+1)!} (|X|^{k+1}) \right) \\ &= \sum_{j=0}^k \frac{(it)^j}{j!} \mathbb{E}(|X|^j) + \vartheta(t) \frac{t^{k+1}}{(k+1)!} \mathbb{E}(|X|^{k+1}). \end{aligned}$$

□

Bemerkung: Wir haben nun lediglich eine Taylorapproximation mittels Momenten für den univariaten Fall. Für unsere Zwecke ist der multivariate Fall auch von Interesse zum Beispiel bei der gemeinsamen charakteristischen Funktion $\varphi_{X,Y}(t, s)$. Wir folgern die Taylorapproximation von (X, Y) mittels Momenten aus der tatsächlichen Taylorentwicklung und Satz 1.19. Dazu benutzen wir eine mehrdimensionale Taylorentwicklung der charakteristischen Funktion um den Nullpunkt:

$$\begin{aligned}\varphi_{X,Y}(t, s) &= \sum_{j,k=0}^{\infty} \frac{\partial^{j+k}}{\partial t^j \partial s^k} \varphi_{X,Y}(0, 0) \frac{t^j s^k}{j!k!} \stackrel{\text{Satz 1.19}}{=} \sum_{j,k=0}^{\infty} i^{j+k} \mathbb{E}(X^j Y^k) \frac{t^j s^k}{j!k!} \\ &= \sum_{j,k=0}^{\infty} \frac{(it)^j}{j!} \frac{(is)^k}{k!} \mathbb{E}(X^j Y^k).\end{aligned}$$

Diese Taylorentwicklung ist nur möglich, falls alle gemischten Momente $\mathbb{E}(X^k Y^l)$ für $k, l \in \mathbb{N}$ existieren. Falls $n_X, n_Y \in \mathbb{N}$ die größten natürlichen Zahlen sind für die das gemischte Moment $\mathbb{E}(X^{n_X} Y^{n_Y})$ existiert, so gilt

$$\begin{aligned}\varphi_{X,Y}(t, s) &= \sum_{\substack{j \leq n_X \& k \leq n_Y \\ j+k \neq n_X \& j+k \neq n_Y}} \frac{(it)^j}{j!} \frac{(is)^k}{k!} \mathbb{E}(X^j Y^k) + \sum_{\substack{j+k=n_X \\ j \leq n_X \& k \leq n_Y}} \frac{(it)^j}{j!} \frac{(is)^k}{k!} \mathbb{E}(X^j Y^k) \vartheta_{j,k}(t, s) \\ &+ \sum_{\substack{j+k=n_Y \\ j \leq n_X \& k \leq n_Y}} \frac{(it)^j}{j!} \frac{(is)^k}{k!} \mathbb{E}(X^j Y^k) \vartheta_{j,k}(t, s)\end{aligned}$$

mit $|\vartheta_{j,k}(t, s)| \leq 1 \forall t, s \in \mathbb{R}, j \in \{1, \dots, n_X\}, k \in \{1, \dots, n_Y\}$. Diese Taylorentwicklung mittels Momenten können wir nun nutzen, um leichter charakteristische Funktionen zu berechnen.

Beispiel 1.22. Sei $X \sim N(0, 1)$, dann kann man mit der sogenannten Momentenmethode die charakteristische Funktion wie folgt berechnen:

Es gilt $\mathbb{E}(X^{2l+1}) = 0 \forall l \in \mathbb{N}_0$ und $\mathbb{E}(X^{2l}) = \frac{(2l)!}{2^l l!} \forall l \in \mathbb{N}_0$, denn

$$\begin{aligned}\mathbb{E}(X^{2j+1}) &= \int_{-\infty}^{\infty} t^{2j+1} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \left(\underbrace{\left[-\frac{1}{t} e^{-\frac{t^2}{2}} t^{2j+1} \right]_{-\infty}^{\infty}}_{=0} + \int_{-\infty}^{\infty} (2j) t^{2j-1} e^{-\frac{t^2}{2}} dt \right) \\ &= (2j) \int_{-\infty}^{\infty} t^{2j-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 2j \mathbb{E}(X^{2j-1}) = \dots = 2j \cdot \dots \cdot \underbrace{2 \cdot \mathbb{E}(X)}_{=0} = 0.\end{aligned}$$

Und

$$\begin{aligned}\mathbb{E}(X^{2(j+1)}) &= \int_{-\infty}^{\infty} t^{2j} t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \left(\underbrace{\left[-\frac{1}{t} e^{-\frac{t^2}{2}} t^{2(j+1)} \right]_{-\infty}^{\infty}}_{=0} + \int_{-\infty}^{\infty} (j+1) t^{2j} e^{-\frac{t^2}{2}} dt \right) \\ &= (j+1) \int_{-\infty}^{\infty} t^{2j} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = (j+1) \mathbb{E}(X^{2j}) = \frac{2(j+1)j}{2j} \mathbb{E}(X^{2j}) \\ &= \dots = \frac{(2(j+1))!}{2^{j+1}(j+1)!}.\end{aligned}$$

Und schließlich mit Satz 1.21 folgt:

$$\begin{aligned}\varphi_X(t) &= \sum_{j=0}^{\infty} \frac{(it)^j}{j!} \mathbb{E}(X^j) = \sum_{j=0}^{\infty} \frac{(it)^{2j}}{(2j)!} \mathbb{E}(X^{2j}) = \sum_{j=0}^{\infty} \frac{(it)^{2j} (2j)!}{(2j)! 2^j j!} = \sum_{j=0}^{\infty} \frac{\left(\frac{i^2 t^2}{2}\right)^j}{j!} \\ &= \sum_{j=0}^{\infty} \frac{\left(-\frac{t^2}{2}\right)^j}{j!} = e^{-\frac{t^2}{2}}.\end{aligned}$$

Um ein Äquivalenzkriterium zwischen charakteristischen Funktionen und der Unabhängigkeit zu erhalten ist es notwendig den Begriff der *Fourier-Transformierten* einer Funktion einzuführen sowie die sogenannte *Umkehrformel*.

Definition 1.23. Sei $h: \mathbb{R}^n \rightarrow \mathbb{C}$ lebesgue-integrierbar, dann heißt $\gamma: \mathbb{R}^n \rightarrow \mathbb{C}$ mit

$$\gamma(y) := \int_{\mathbb{R}^n} e^{it^T y} h(t) dt \quad \forall y \in \mathbb{R}^n$$

die *Fourier-Transformation* von h .

Anmerkung:

- (i) Wegen $|e^{it^T y}| \leq 1 \quad \forall y \in \mathbb{R}^n$ und weil h lebesgue-integrierbar ist, d.h. $\int_{\mathbb{R}^n} |h(t)| dt < \infty$ existiert $\gamma(y) \quad \forall y \in \mathbb{R}^n$;
- (ii) $\gamma(y)$ ist beschränkt, denn $|\gamma(y)| \leq \int_{\mathbb{R}^n} |e^{it^T y}| |h(t)| dt = \int_{\mathbb{R}^n} |h(t)| dt$;
- (iii) Wenn man h entsprechend $h = (Re h)^+ + i(Im h)^+ - (Re h)^- - i(Im h)^-$ aufteilt und die einzelnen Glieder als Dichten auffasst, so erhält man eine Summe aus charakteristischen Funktionen. Da die gleichmäßige Stetigkeit bei Summation erhalten bleibt, folgt das auch die Fourier-Transformation von h gleichmäßig stetig ist;

1.2. Charakteristische Funktionen

(iv) $dt := d\lambda^n = dt_1 \dots dt_n$, wobei λ^n das Lebesguemaß auf dem \mathbb{R}^n bezeichnet.

Satz 1.24. Sei $h(t)$ lebesgue-integrierbar mit der Fourier-Transformation $\gamma(y)$ und X eine Zufallsvariable, dann gilt:

$$\mathbb{E}(\gamma(X)) = \int_{\mathbb{R}^n} h(t) \cdot \varphi_X(t) dt.$$

Beweis. 1. Aus der Beschränktheit von $\gamma(y)$ folgt, dass $\mathbb{E}(\gamma(X))$ existiert.

2. Da $e^{it^T x}$ stetig ist, ist $e^{it^T x}$ auch $\mathfrak{B}(\mathbb{R}^n) \times \mathfrak{A}$ -messbar, zudem ist $h(t)$ $\mathfrak{B}(\mathbb{R}^n)$ -messbar und aufgefasst als Abbildung $h : \mathbb{R}^n \times \Omega \rightarrow \mathbb{C}$ auch $\mathfrak{B}(\mathbb{R}^n) \times \mathfrak{A}$ -messbar.

Insgesamt ist $h(t)e^{it^T y}$ $\mathfrak{B}(\mathbb{R}^n) \times \mathfrak{A}$ -messbar und insbesondere $\lambda^n \otimes P$ -integrierbar.

3. Mit dem Satz von Fubini folgt

$$\begin{aligned} \int_{\mathbb{R}^n} h(t) \cdot \varphi_X(t) dt &= \int_{\mathbb{R}^n} h(t) \cdot \int_{\Omega} e^{it^T X} dP dt \stackrel{\text{Fubini}}{=} \int_{\mathbb{R}^n \times \Omega} h(t) e^{it^T X} d\lambda^n \otimes dP \\ &= \int_{\Omega} \int_{\mathbb{R}^n} h(t) e^{it^T X} d\lambda^n dP = \mathbb{E}(\gamma(X)). \end{aligned}$$

□

Die Verteilung einer Zufallsvariable P_X könnten wir nun aus Satz 1.24 mittels der charakteristischen Funktionen gewinnen, falls zu jedem mehrdimensionalen Intervall $I = I_a^b = (a_1, b_1] \times \dots \times (a_n, b_n]$ eine lebesgue-integrierbare Funktion h existiert, sodass $\gamma(y) = \mathbb{1}_{I_a^b}(y)$ gilt. Das Problem dabei ist, dass γ stetig ist, aber $\mathbb{1}_I$ nicht (außer wenn $I = \emptyset$ oder $I = \mathbb{R}^n$). Daher muss man $\mathbb{1}_I$ geeignet durch γ_λ approximieren, dazu müssen wir h_λ finden, derart dass $\gamma_\lambda \rightarrow \mathbb{1}_I$ für $\lambda \rightarrow 0$.

Lemma 1.25. Sei $h_\lambda(t; y', y'') := \frac{e^{-ity'} - e^{-ity''}}{it} \cdot e^{-\lambda|t|}$ mit $t, \lambda, y', y'' \in \mathbb{R}$ und $y' \leq y''$, $\lambda > 0$, dann ist

$$\gamma_\lambda(y; y', y'') = 2 \arctan\left(\frac{y-y'}{\lambda}\right) - 2 \arctan\left(\frac{y-y''}{\lambda}\right).$$

Beweis. 1. Wegen

$$\begin{aligned} |h_\lambda(t; y', y'')| &= \left| \frac{e^{-ity'} - e^{-ity''}}{it} \cdot e^{-\lambda|t|} \right| = \frac{e^{-\lambda|t|}}{|t|} |e^{-it(y'-y'')} - 1| |e^{-ity''}| \\ &\leq \frac{|t||y' - y''|}{|t|} e^{-\lambda|t|} \end{aligned}$$

und dem Satz von der Majorisierten Konvergenz ist h_λ lebesgue-integrierbar in t .

2. Es gilt

$$\begin{aligned}
 \frac{|h_\lambda(t; y' + w, y'') - h_\lambda(t; y', y'')|}{|w|} &= \frac{|e^{-it(y'+w)} - e^{-ity''} - e^{-ity'} + e^{-ity''}| |e^{-\lambda|t|}}{|it||w|} \\
 &= \frac{e^{-\lambda|t|}}{|t||w|} \cdot | -e^{-ity'} e^{-itw} - e^{-ity'} | \\
 &= \frac{e^{-\lambda|t|}}{|t||w|} \cdot |e^{-itw} + 1| \underbrace{|e^{-ity'}|}_{=1} \\
 &\leq \frac{e^{-\lambda|t|}}{|t||w|} \cdot |t||w| = e^{-\lambda|t|},
 \end{aligned}$$

somit ist h_λ mit dem Satz von der Majorisierten Konvergenz differenzierbar in y' .

3.

$$\begin{aligned}
 \gamma'_\lambda &:= \frac{\partial}{\partial y'} \gamma_\lambda(y; y', y'') = \frac{\partial}{\partial y'} \int_{-\infty}^{\infty} h_\lambda(t; y', y'') \cdot e^{ity} dt \stackrel{\text{majo.Kvgz.}}{=} \int_{-\infty}^{\infty} -\frac{e^{-ity'}(it)}{it} e^{-\lambda|t|} e^{ity} dt \\
 &= \int_{-\infty}^{\infty} -e^{it(y-y')} e^{-\lambda|t|} dt = -2 \int_0^{\infty} \cos(t(y-y')) e^{-\lambda t} dt \\
 &= -\frac{2}{\lambda} \left(\underbrace{[-e^{-\lambda t} \cos(t(y-y'))]_0^{\infty}}_{=1} - \int_0^{\infty} e^{-\lambda t} \sin(t(y-y'))(y-y') dt \right) \\
 &= -\frac{2}{\lambda} \left(1 - \left(\underbrace{[-\frac{e^{-\lambda t}}{\lambda} \sin(t(y-y'))(y-y')]_0^{\infty}}_{=0} + \underbrace{\int_0^{\infty} \frac{e^{-\lambda t}}{\lambda} \cos(t(y-y'))(y-y')^2 dt}_{=-\frac{\gamma'_\lambda}{2\lambda}(y-y')^2} \right) \right) \\
 &= -\frac{2}{\lambda} (1 + \frac{\gamma'_\lambda}{2\lambda}(y-y')^2).
 \end{aligned}$$

Also insgesamt:

$$\gamma'_\lambda = -\frac{2}{\lambda} (1 + \frac{\gamma'_\lambda}{2\lambda}(y-y')^2) \Leftrightarrow \gamma'_\lambda (1 + (\frac{y-y'}{\lambda})^2) = -\frac{2}{\lambda} \Leftrightarrow \gamma'_\lambda = -\frac{2}{\lambda} (1 + (\frac{y-y'}{\lambda})^2)^{-1} =: (*).$$

Integrieren wir nun (*) nach y' erhalten wir:

$$\gamma_\lambda(y; y', y'') = 2 \arctan \left(\frac{y-y'}{\lambda} \right) + \text{Const}_\lambda(y', y''),$$

1.2. Charakteristische Funktionen

wobei $Const_\lambda(y'; y'')$ eine Konstante ist, die nur von y' , y'' und λ abhängt und wegen

$$\gamma_\lambda(y; y'', y') = \int_{-\infty}^{\infty} e^{ity} \underbrace{\frac{e^{-ity''} - e^{-ity'}}{it}}_{=0} e^{-\lambda|t|} dt = 0,$$

folgt $Const_\lambda(y', y'') = -2 \arctan\left(\frac{y-y''}{\lambda}\right)$ und insgesamt

$$\gamma_\lambda(y; y', y'') = 2 \arctan\left(\frac{y-y'}{\lambda}\right) - 2 \arctan\left(\frac{y-y''}{\lambda}\right).$$

□

Bemerkung: Mit dem Satz von Fubini können wir dieses Ergebnis nun auf höhere Dimensionen erweitern.

Wir erhalten

$$h_\lambda(t; y', y'') := e^{-\lambda \sum_{i=1}^n |t_i|} \prod_{i=1}^n \frac{e^{-it_i y'_i} - e^{-it_i y''_i}}{it_i}, \quad (1.2)$$

mit $t, y', y'' \in \mathbb{R}^n$ und für $\lambda > 0$

$$\gamma_\lambda(y; y', y'') := 2^n \prod_{i=1}^n \left(\arctan\left(\frac{y_i - y'_i}{\lambda}\right) - \arctan\left(\frac{y_i - y''_i}{\lambda}\right) \right). \quad (1.3)$$

Satz 1.26 (Umkehrformel). Sei $X = (X_1, \dots, X_n)$ eine Zufallsvariable und seien $y', y'' \in \mathbb{R}^n$ derart, dass die Komponenten y'_i, y''_i Stetigkeitsstellen von F_{X_i} sind $i = (1, \dots, n)$. Mit h_λ und γ_λ gemäß (1.2) bzw. (1.3) gilt

$$P(\{X \in I_{y'}^{y''}\}) = P_X(I_{y'}^{y''}) = \lim_{\lambda \searrow 0} \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \varphi_X(t) h_\lambda(t; y', y'') dt.$$

Beweis.

$$\begin{aligned} & \frac{1}{(2\pi)^n} \int_{\Omega} 2^n \prod_{i=1}^n \left(\arctan\left(\frac{X_i - y'_i}{\lambda}\right) - \arctan\left(\frac{X_i - y''_i}{\lambda}\right) \right) dP = \frac{1}{(2\pi)^n} \mathbb{E}(\gamma_\lambda(X; y', y'')) \\ & \stackrel{\text{Satz 1.24}}{=} \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \varphi_X(t) h_\lambda(t; y', y'') dt \end{aligned}$$

Wegen $|\frac{1}{(2\pi)^n} 2^n \prod_{i=1}^n (\arctan(\frac{x_i - y'_i}{\lambda}) - \arctan(\frac{x_i - y''_i}{\lambda}))| \leq 1 \quad \forall \lambda > 0$ und zusammen mit $P(X_i = y'_i) = 0 = P(X_i = y''_i)$ und dem Satz von der Majorisierten Konvergenz folgt

$$\begin{aligned}
 & \lim_{\lambda \searrow 0} \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \varphi_X(t) h_\lambda(t; y', y'') dt \\
 &= \lim_{\lambda \searrow 0} \frac{1}{(2\pi)^n} \int_{\Omega} 2^n \prod_{i=1}^n \left(\arctan \left(\frac{X_i - y'_i}{\lambda} \right) - \arctan \left(\frac{X_i - y''_i}{\lambda} \right) \right) dP \\
 &= \left(\frac{1}{2\pi} \right)^n \int_{\Omega} \prod_{i=1}^n 2\pi \mathbb{1}_{I_{y'_i}^{y''_i}}(X) dP = P_X(I_{y'}^{y''}).
 \end{aligned}$$

□

Satz 1.27. Die Verteilung von X ist eindeutig durch ihre charakteristische Funktion bestimmt.

Beweis. Die Menge $A := \{x \in \mathbb{R}^n \mid x_i \text{ ist Unstetigkeitsstelle von } F_{X_i}, i = 1, \dots, n\}$ ist abzählbar, da jedes F_{X_i} nur abzählbar viele Unstetigkeitsstellen besitzt. Definieren wir die Menge $B := \{x \in \mathbb{R}^n \mid x_i \text{ ist Stetigkeitsstelle von } F_{X_i}, i = 1, \dots, n\}$ dann gilt $\forall x \in \mathbb{R}^n$:

$$F_X(x) = \lim_{\substack{x_n \searrow x \\ x_n \in B}} F_X(x_n) = \lim_{\substack{x_n \searrow x \\ x_n \in B}} P(-\infty \leq X_i \leq x_{n_i}) = \lim_{\substack{x_n \searrow x \\ x_n \in B}} \lim_{\substack{y_j \rightarrow -\infty \\ y_j \leq x_n \\ y_j \in B}} P(y_{j_i} \leq X_i \leq x_{n_i}).$$

Mit Satz 1.26 folgt nun die Behauptung.

□

Satz 1.28. X und Y sind genau dann unabhängig, wenn gilt: $\varphi_{X,Y} = \varphi_X \cdot \varphi_Y$.

Beweis. „ \Rightarrow “ folgt aus Satz 1.17.

„ \Leftarrow “ Nach Satz 1.17 gilt für die Produktverteilung $P := P_X \otimes P_Y$ und

$$\varphi_P(t, s) = \varphi_{P_X}(t) \cdot \varphi_{P_Y}(s) = \varphi_{X,Y}(t, s).$$

$\stackrel{\text{Satz 1.27}}{\Rightarrow} P_{X,Y} = P = P_X \otimes P_Y \stackrel{\text{Def. Unabh.}}{\Rightarrow} X$ und Y unabhängig.

□

Für die charakteristische Funktion kann auch ein Erwartungstreuer Schätzer definiert werden:

Definition 1.29 (Empirische Charakteristische Funktion). Sei X eine reelle Zufallsvariable und X_1, \dots, X_d Stichproben der Zufallsvariable und bezeichne

$$F_X^d(t) := \frac{1}{d} \sum_{k=1}^d \mathbb{1}_{(-\infty, t]}(X_k)$$

die empirische Verteilungsfunktion, so heißt

$$\varphi_X^d(t) := \int_{\mathbb{R}^n} e^{it^T x} dF_X^d = \frac{1}{d} \sum_{k=1}^d e^{it^T X_k}$$

die *empirische charakteristische Funktion* von X .

Nach dem starken Gesetz der großen Zahlen gilt $F_X^d \rightarrow F_X$ P-f.s. Man kann sogar zeigen, dass F_X^d gleichmäßig gegen F_X konvergiert. Daraus folgt

$$\lim_{d \rightarrow \infty} \varphi_X^d(t) = \lim_{d \rightarrow \infty} \int_{\mathbb{R}^n} e^{it^T x} dF_X^d = \int_{\mathbb{R}^n} e^{it^T x} dF_X = \varphi_X.$$

Bemerkung Gegeben seien die Stichproben $(X^d, Y^d) = \{(X_k, Y_k) | k = 1, \dots, d\}$ eines Zufallsvektors (X, Y) mit $\dim(X) = n$ und $\dim(Y) = m$, dann ist die empirische Charakteristische Funktion von (X, Y) durch

$$\varphi_{X,Y}^d(t, s) := \int_{\mathbb{R}^{n+m}} e^{it^T x + is^T y} dF_{X,Y}^d = \frac{1}{d} \sum_{k=1}^d e^{it^T X_k + is^T Y_k}$$

gegeben.

Satz 1.30. Es gilt

- (i) $\mathbb{E}(\varphi_X^d(t)) = \varphi_X(t)$;
- (ii) $\mathbb{E}(|\varphi_X^d(t)|^2) = \frac{d-1}{d} |\varphi_X(t)|^2 + \frac{1}{d}$;
- (iii) $\mathbb{E}(|\varphi_X^d(t) - \varphi_X(t)|^2) = \frac{1}{d}(1 - |\varphi_X(t)|^2)^2$.

Beweis. (i)

$$\mathbb{E}(\varphi_X^d(t)) = \frac{1}{d} \sum_{k=1}^d \mathbb{E}(e^{it^T X_k}) = \frac{1}{d} \sum_{k=1}^d \varphi_X(t) = \varphi_X(t).$$

(ii) Es ist

$$\begin{aligned} \mathbb{E}(|\varphi_X^d(t)|^2) &= \mathbb{E}(\varphi_X^d(t) \varphi_X^d(-t)) = \mathbb{E} \left(\frac{1}{d} \sum_{k=1}^d e^{it^T X_k} \frac{1}{d} \sum_{l=1}^d e^{-it^T X_l} \right) \\ &= \mathbb{E} \left(\frac{1}{d^2} \sum_{k,l=1}^d e^{it^T (X_k - X_l)} \right) = \mathbb{E} \left(\frac{1}{d^2} \sum_{k,l=1; k \neq l}^d e^{it^T (X_k - X_l)} \right) + \frac{1}{d} \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{d^2} \sum_{k,l=1;k \neq j}^d \mathbb{E} \left(e^{it^T X_k} \right) \mathbb{E} \left(e^{-it^T X_l} \right) + \frac{1}{d} \\
 &\stackrel{(i)}{=} \frac{d^2 - d}{d^2} \varphi_X(t) \varphi_X(-t) + \frac{1}{d} \\
 &= \frac{d-1}{d} |\varphi_X(t)|^2 + \frac{1}{d}.
 \end{aligned}$$

(iii) Mit (i) und (ii) erhalten wir

$$\begin{aligned}
 \mathbb{E}(|\varphi_X^d(t) - \varphi_X(t)|^2) &= \mathbb{E}(|\varphi_X^d(t)|^2) - \mathbb{E}(\varphi_X^d(t))\varphi_X(-t) - \mathbb{E}(\varphi_X^d(-t))\varphi_X(t) + \mathbb{E}(|\varphi_X(t)|^2) \\
 &= \frac{d-1}{d} |\varphi_X(t)|^2 + \frac{1}{d} - |\varphi_X(t)|^2 - |\varphi_X(t)|^2 + |\varphi_X(t)|^2 \\
 &= \frac{d-1}{d} |\varphi_X(t)|^2 + \frac{1}{d} - |\varphi_X(t)|^2 \\
 &= \frac{1}{d} (1 - |\varphi_X(t)|^2).
 \end{aligned}$$

□

Kapitel 2

Distanz Kovarianz & Distanz Korrelation

In diesem Kapitel definieren wir die Distanz Kovarianz und die Distanz Korrelation. Diese Funktionen sind wie die Kovarianz und die Korrelation Maßzahlen zur Messung des Zusammenhangs zwischen zwei Zufallsvariablen. Die Vorteile sind, dass bei der Distanz Kovarianz\Korrelation Zufallsvariablen beliebiger Dimension zugelassen sind und nichtlineare Zusammenhänge auch miterfasst werden. Eine weitere wichtige Eigenschaft ist, dass durch diese Maße die Unabhängigkeit charakterisiert ist, denn die Distanz Kovarianz und die Distanz Korrelation nehmen den Wert 0 genau dann an, wenn die Zufallsvariablen unabhängig sind.

Zudem wird noch der enge Zusammenhang zwischen Pearsonscher Korrelation und Distanz Korrelation im normalverteilten bivariaten Fall genauer untersucht.

2.1 Definiton und grundlegende Eigenschaften

Wir entwickeln zuerst bestimmte Voraussetzungen an ein allgemeines Maß, welches den Zusammenhang zwischen Zufallsvariablen messen soll. Und definieren nachdem wir mögliche Probleme beseitigt haben die Distanz Kovarianz. Zudem zeigen wir einige Eigenschaften der Distanz Kovarianz bzw. der Distanz Korrelation, die ähnlich zu den Eigenschaften der Kovarianz bzw. der Korrelation im Kapitel 1 sind.

Eine allgemeine Kennzahl, die den Zusammenhang zwischen Zufallsvariablen misst, sollte folgende Eigenschaften besitzen:

- (i) Sie soll 0 sein wenn die Zufallsvariablen unabhängig sind;

- (ii) Zufallsvariablen verschiedener und beliebiger Dimensionen sollen vergleichbar sein;
- (iii) Sie soll skaleninvariant sein.

Wenn man diese Kennzahl nun als Abstand zwischen dem Produkt der charakteristischen Funktionen der Zufallsvariablen und der gemeinsamen charakteristischen Funktion der Zufallsvariablen definiert, wird die Maßzahl 0 wenn die Zufallsvariablen unabhängig sind. Das ist eine Folgerung aus Satz 1.28.

Um den Abstand zwischen den charakteristischen Funktionen zu messen, benötigen wir eine geeignete Norm, die in folgender Definition bestimmt wird:

Definition 2.1 (Gewichteter L^2 -Raum). Ein ω -gewichteter L^2 -Raum ist ein L^2 -Raum $L^2(\Omega, A, P)$ mit der Norm

$$\|f(t, s)\|_{\omega}^2 := \int_{\mathbb{R}^{n+m}} |f(t, s)|^2 \omega(t, s) dt ds$$

mit $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{C}$ und ω eine beliebige positive Funktion für die das Integral existiert. Die Funktion ω heißt *Gewichtungsfunktion*.

Durch

$$\begin{aligned} A^2(X, Y; \omega) &:= \|\varphi_{X,Y}(t, s) - \varphi_X(t)\varphi_Y(s)\|_{\omega}^2 \\ &= \int_{\mathbb{R}^{n+m}} |\varphi_{X,Y}(t, s) - \varphi_X(t)\varphi_Y(s)|^2 \omega(t, s) dt ds \end{aligned} \quad (2.1)$$

erhalten wir ein Maß, welches den Zusammenhang zwischen Zufallsvariablen misst und durch

$$B^2(X, Y; \omega) := \frac{A^2(X, Y; \omega)}{\sqrt{A^2(X, X; \omega)A^2(Y, Y; \omega)}}$$

erhält man eine standardisierte Version des Maßes. Das Problem ist nun ω treffend zu wählen, so dass die Voraussetzungen, die wir an diese Kennzahl stellen, erfüllt bleiben. Wählen wir für ω eine integrierbare Funktion, dann erhalten wir nach [13]

$$\lim_{\varepsilon \rightarrow 0} \frac{A^2(\varepsilon X, \varepsilon Y; \omega)}{\sqrt{A^2(\varepsilon X, \varepsilon X; \omega)A^2(\varepsilon Y, \varepsilon Y; \omega)}} = \rho^2(X, Y).$$

Also erhält man wieder die Pearsonsche Korrelation und die Forderung (i) kann nicht erfüllt werden. Die Wahl der Gewichtungsfunktion ω wird durch folgendes Lemma motiviert:

Lemma 2.2. Sei $0 < \alpha < 2$, dann ist $\forall x \in \mathbb{R}^n$

$$\int_{\mathbb{R}^n} \frac{1 - \cos(t^T x)}{\|t\|^{n+\alpha}} dt = C(n, \alpha) \|x\|_2^\alpha, \quad (2.2)$$

wobei

$$C(n, \alpha) = \frac{2\pi^{\frac{n}{2}} \Gamma(1 - \frac{\alpha}{2})}{\alpha 2^\alpha \Gamma(\frac{n+\alpha}{2})} > 0. \quad (2.3)$$

Beweis. Für $x=0$ ist die Aussage klar, denn $1 - \cos(t^T 0) = 0$ und $\|0\|_2 = 0$, sodass auf beiden Seiten 0 steht.

Sei also $x \neq 0$ und ohne Einschränkung soll gelten, dass $x = (a, 0, \dots, 0)^T$ mit $a > 0$, da jeder Vektor durch orthogonale Abbildung in diese Form gebracht werden kann und die Norm unter orthogonalen Abbildungen erhalten bleibt. Wir definieren

$$A := \int_{\mathbb{R}^{n-1}} \frac{dz_2 \cdots dz_n}{(1 + z_2^2 + \cdots + z_n^2)^{\frac{n+\alpha}{2}}}.$$

Nach [10] Formel 3.3.2.1, Seite 585, erhalten wir

$$A = \int_{\mathbb{R}^{n-1}} \frac{dz_2 \cdots dz_n}{(1 + z_2^2 + \cdots + z_n^2)^{\frac{n+\alpha}{2}}} = \frac{2\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} \int_0^\infty \frac{t^{n-2}}{(1+t^2)^{\frac{n+\alpha}{2}}} dt$$

und mit [10] Formel 2.2.4.24, Seite 298, erhalten wir

$$\begin{aligned} A &= \frac{2\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} \int_0^\infty \frac{t^{n-2}}{(1+t^2)^{\frac{n+\alpha}{2}}} dt \\ &= \frac{2\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} 1^{n-2-\frac{n+\alpha}{2}} B\left(n-1, \frac{n+\alpha}{2} - (n-1)\right) \\ &= \frac{\pi^{\frac{n-1}{2}} \Gamma(\frac{1+\alpha}{2})}{\Gamma(\frac{n+\alpha}{2})}, \end{aligned}$$

wobei B die eulersche Betafunktion bezeichnet mit $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ für $x, y \in \mathbb{C}$ und $\operatorname{Re}(x), \operatorname{Re}(y) > 0$.

Es ist

$$\frac{d}{da} \frac{1 - \cos(ax)}{x^{1+\alpha}} = x \frac{\sin(ax)}{x^{1+\alpha}}$$

und somit

$$\frac{d}{da} \left(\int_0^\infty \frac{1 - \cos(ax)}{x^{1+\alpha}} dx \right) = \int_0^\infty \frac{\sin(ax)}{x^\alpha} dx.$$

Mittels der Substitution $t=ax$ erhalten wir

$$\int_0^\infty \frac{\sin(ax)}{x^\alpha} dx = \int_0^\infty \frac{\sin(t)}{a(\frac{t}{a})^\alpha} dt = a^{\alpha-1} \int_0^\infty \frac{\sin(t)}{t} dt.$$

Und mit [10] Formel 2.5.3.13, Seite 387, erhalten wir

$$a^{\alpha-1} \int_0^\infty \frac{\sin(t)}{t} dt = a^{\alpha-1} \frac{\sqrt{\pi} \Gamma(1 - \frac{\alpha}{2})}{2^\alpha \Gamma(\frac{1+\alpha}{2})}.$$

Im Umkehrschluss ist

$$\int_0^\infty \frac{1 - \cos(ax)}{x^{1+\alpha}} dx = \frac{a^\alpha \sqrt{\pi} \Gamma(1 - \frac{\alpha}{2})}{\alpha 2^\alpha \Gamma(\frac{1+\alpha}{2})}.$$

Mit $x = (a, 0, \dots, 0)^T$, $t = (t_1, \dots, t_n)^T$ und der Substitution $g : (t_1, \dots, t_n) \mapsto (t_1, t_2/t_1, \dots, t_n/t_1)$ gilt

$$\begin{aligned} & \int_{\mathbb{R}^n} \frac{1 - \cos(t^T x)}{\|t\|_2^{n+\alpha} \|x\|_2^\alpha} dt \\ &= \int_{\mathbb{R}^n} \frac{1 - \cos(t_1 a)}{(t_1^2 + \dots + t_n^2)^{\frac{n+\alpha}{2}} |a|^\alpha} dt \\ &= \int_{\mathbb{R}^n} \frac{1 - \cos(t_1 a)}{(1 + (t_2/t_1)^2 + \dots + (t_n/t_1)^2)^{\frac{n+\alpha}{2}} |t|^{n+\alpha} |a|^\alpha} dt \\ &= \int_{\mathbb{R}^n} \frac{1 - \cos(t_1 a) |t_1|^{n-1}}{(1 + t_2^2 + \dots + t_n^2)^{\frac{n+\alpha}{2}} |t_1|^{n+\alpha} |a|^\alpha} dt \\ &= \underbrace{\int_{\mathbb{R}^{n-1}} \frac{dt_2 \cdots dt_n}{(1 + t_2^2 + \dots + t_n^2)^{\frac{n+\alpha}{2}}}}_{=A} \int_{\mathbb{R}} \frac{1 - \cos(t_1 a)}{|t_1|^{1+\alpha} |a|^\alpha} dt. \end{aligned}$$

Wegen $a > 0$ folgt

$$C(n, \alpha) \|x\|_2^\alpha = \int_{\mathbb{R}^n} \frac{1 - \cos(t^T x)}{\|t\|^{n+\alpha}} dt$$

$$\begin{aligned}
 \Leftrightarrow C(n, \alpha) &= A \cdot \int_{-\infty}^{\infty} \frac{1 - \cos(z_1 a)}{|z_1|^{1+\alpha} |a|^\alpha} dz_1 = A \cdot \frac{2}{a^\alpha} \int_0^{\infty} \frac{1 - \cos(z_1 a)}{z_1^{1+\alpha}} dz_1 \\
 &= A \cdot \frac{2a^\alpha \sqrt{\pi} \Gamma(1 - \frac{\alpha}{2})}{a^\alpha \alpha 2^\alpha \Gamma(\frac{1+\alpha}{2})} = \frac{\pi^{\frac{n-1}{2}} \Gamma(\frac{1+\alpha}{2})}{\Gamma(\frac{n+\alpha}{2})} \cdot \frac{2 \sqrt{\pi} \Gamma(1 - \frac{\alpha}{2})}{\alpha 2^\alpha \Gamma(\frac{1+\alpha}{2})} \\
 &= \frac{2\pi^{\frac{n}{2}} \Gamma(1 - \frac{\alpha}{2})}{\alpha 2^\alpha \Gamma(\frac{\alpha+n}{2})} > 0.
 \end{aligned}$$

□

Mit Lemma 2.2 und der Tatsache, dass bei der Integration von charakteristischen Funktionen die Sinus-Komponenten wegfallen und dass die Funktion nicht integrierbar sein darf, erhalten wir die Gewichtungsfunktion

$$\omega(t, s; \alpha) = (C(n, \alpha) C(m, \alpha) \|t\|_2^{n+\alpha} \|s\|_2^{m+\alpha})^{-1} \quad 0 < \alpha < 2.$$

Für die weitere Arbeit wird $\alpha = 1$ gesetzt und für ω gilt fortan:

$$\omega(t, s) := \omega(t, s; 1) = (c_n c_m \|t\|_2^{1+n} \|s\|_2^{1+m})^{-1}, \quad (2.4)$$

wobei

$$c_n := C(n, 1) = \frac{\pi^{\frac{1+n}{2}}}{\Gamma(\frac{1+n}{2})}.$$

Anmerkung Dass diese Gewichtungsfunktion nicht integrierbar ist, wird schnell deutlich: Wir nehmen an, die Gewichtungsfunktion sei integrierbar (wir betrachten den Fall $n=m=1$), dann ist

$$\begin{aligned}
 \int_{\mathbb{R}^2} \frac{1}{c_1^2 |t|^2 |s|^2} dt ds &= \frac{1}{c_1^2} \int_{\mathbb{R}} \frac{1}{t^2} dt \int_{\mathbb{R}} \frac{1}{s^2} ds \\
 &= \frac{1}{c_1^2} 2 \int_{\mathbb{R}_+} \frac{1}{t^2} dt 2 \int_{\mathbb{R}_+} \frac{1}{s^2} ds \\
 &= \frac{1}{\pi^2} 2 \left[-\frac{1}{t} \right]_0^\infty 2 \left[-\frac{1}{s} \right]_0^\infty \not\downarrow.
 \end{aligned}$$

Und für die Gewichtungsfunktion höherer Dimension folgt die Nicht-Integrierbarkeit analog.

Vereinfachend schreiben wir

$$d\omega := \frac{dt ds}{c_n c_m \|t\|_2^{1+n} \|s\|_2^{1+m}}.$$

Bevor wir nun die Distanz Kovarianz endgültig definieren, müssen wir noch zeigen, dass der Integrand in 2.1 tatsächlich existiert. Dazu benötigen wir einige Hilfsaussagen, die später auch noch Verwendung finden werden.

Satz 2.3 (Additionstheorem und Produkttheorem). Es gilt $\forall x, y \in \mathbb{R}$

- (i) $\sin(x) \sin(y) = \frac{1}{2}(\cos(x - y) - \cos(x + y));$
- (ii) $\cos(x) \cos(y) = \frac{1}{2}(\cos(x - y) + \cos(x + y));$
- (iii) $\sin(x) \cos(y) = \frac{1}{2}(\sin(x - y) + \sin(x + y));$
- (iv) $\cos(x \pm y) = \cos(x) \cos(y) \mp \sin(x) \sin(y);$
- (v) $\sin(x \pm y) = \sin(x) \cos(y) \pm \sin(y) \cos(x);$
- (vi) $\cos(x) \cos(y) = 1 - (1 - \cos(x)) - (1 - \cos(y)) + (1 - \cos(x))(1 - \cos(y)).$

Für die Existenz von dem Integrand in 2.1 ist die Voraussetzung $\mathbb{E}(\|X\|_2) + \mathbb{E}(\|Y\|_2) < \infty$ hinreichend, denn

$$\begin{aligned}
 & |\varphi_{X,Y}(t, s) - \varphi_X(t)\varphi_Y(s)|^2 = |\varphi_{X,Y}(t, s) - 2\varphi_X(t)\varphi_Y(s) + \varphi_X(t)\varphi_Y(s)|^2 \\
 & = \left| \mathbb{E}(e^{it^T X + is^T Y}) - \mathbb{E}(e^{it^T X})\varphi_Y(s) - \mathbb{E}(e^{is^T Y})\varphi_X(t) + \varphi_X(t)\varphi_Y(s) \right|^2 \\
 & = \left| \mathbb{E}(e^{it^T X + is^T Y} - e^{it^T X}\varphi_Y(s) - e^{is^T Y}\varphi_X(t) + \varphi_X(t)\varphi_Y(s)) \right|^2 \\
 & = \left| \mathbb{E}((e^{it^T X} - \varphi_X(t))(e^{is^T Y} - \varphi_Y(s))) \right|^2 \\
 & \leq \mathbb{E} \left(\left| e^{it^T X} - \varphi_X(t) \right|^2 \right) \mathbb{E} \left(\left| e^{is^T Y} - \varphi_Y(s) \right|^2 \right) \tag{*} \\
 & = \mathbb{E} \left(\underbrace{\left| e^{it^T X} \right|^2}_{=1} - e^{it^T X} \overline{\varphi_X(t)} - \overline{e^{it^T X}} \varphi_X(t) + |\varphi_X(t)|^2 \right) \\
 & \cdot \mathbb{E} \left(\underbrace{\left| e^{is^T Y} \right|^2}_{=1} - e^{is^T Y} \overline{\varphi_Y(s)} - \overline{e^{is^T Y}} \varphi_Y(s) + |\varphi_Y(s)|^2 \right) \\
 & = (1 - |\varphi_X(t)|^2)(1 - |\varphi_Y(s)|^2).
 \end{aligned}$$

Außerdem gilt mit X' , welches eine unabhängige identisch verteilte Kopie von X ist, dass

$$\begin{aligned}
 |\varphi_X(t)|^2 &= |\mathbb{E}(\cos(tX)) + i\mathbb{E}(\sin(tX))|^2 = \mathbb{E}(\cos(tX))^2 + \mathbb{E}(\sin(tX))^2 \\
 &= \mathbb{E}(\cos(tX))\mathbb{E}(\cos(tX')) + \mathbb{E}(\sin(tX))\mathbb{E}(\sin(tX')) \\
 &= \mathbb{E}(\cos(tX)\cos(tX')) + \mathbb{E}(\sin(tX)\sin(tX')) \\
 &= \mathbb{E}(\cos(tX - tX')).
 \end{aligned} \tag{2.5}$$

Für Y und Y' , welches eine unabhängige identische verteilte Kopie ist, gilt analog

$$|\varphi_Y(s)|^2 = \mathbb{E}(\cos(sY - sY')).$$

Und damit erhalten wir mit dem Satz von Fubini, Lemma 2.2 und den obigen Überlegungen

$$\begin{aligned}
 &\int_{\mathbb{R}^{n+m}} |\varphi_{X,Y}(t,s) - \varphi_X(t)\varphi_Y(s)|^2 d\omega \stackrel{(*)}{\leq} \int_{\mathbb{R}^{n+m}} \frac{(1 - |\varphi_X(t)|^2)(1 - |\varphi_Y(s)|^2)}{c_n \|t\|_2^{1+n} c_m \|s\|_2^{1+m}} dt ds \\
 &= \int_{\mathbb{R}^n} \frac{1 - |\varphi_X(t)|^2}{c_n \|t\|_2^{1+n}} dt \int_{\mathbb{R}^m} \frac{1 - |\varphi_Y(s)|^2}{c_m \|s\|_2^{1+m}} ds \\
 &= \int_{\mathbb{R}^n} \frac{1 - \mathbb{E}(\cos(t^T(X - X')))}{c_n \|t\|_2^{1+n}} dt \int_{\mathbb{R}^m} \frac{1 - \mathbb{E}(\cos(s^T(Y - Y')))}{c_m \|s\|_2^{1+m}} ds \\
 &= \mathbb{E} \left(\int_{\mathbb{R}^n} \frac{1 - \cos(t^T(X - X'))}{c_n \|t\|_2^{1+n}} dt \right) \mathbb{E} \left(\int_{\mathbb{R}^m} \frac{1 - \cos(s^T(Y - Y'))}{c_m \|s\|_2^{1+m}} ds \right) \\
 &= \mathbb{E}(\|X - X'\|_2) \mathbb{E}(\|Y - Y'\|_2) < \infty.
 \end{aligned} \tag{2.6}$$

Damit definieren wir zuerst die Distanz Varianz:

Definition 2.4 (Distanz Varianz). Sei X eine Zufallsvariable mit $\mathbb{E}(X) < \infty$, dann heißt die Quadratwurzel aus

$$DVar(X)^2 := \int_{\mathbb{R}^{n+n}} \frac{|\varphi_{X,X}(t,s) - \varphi_X(t)\varphi_X(s)|^2}{c_n^2 \|t\|_2^{1+n} \|s\|_2^{1+n}} dt ds$$

die *Distanz Varianz* von X .

Wir weisen nun einige Eigenschaften der Distanz Varianz nach.

Satz 2.5 (Eigenschaften der Distanz Varianz). Seien X und Y Zufallsvektoren mit $\dim(X) = \dim(Y) = n$ und $\mathbb{E}(X) + \mathbb{E}(Y) < \infty$, dann gilt:

- (i) $DVar(X) = 0 \Rightarrow X \stackrel{f.s.}{=} \mathbb{E}(X)$;
- (ii) $DVar(a + bCX) = |b|DVar(X) \quad \forall a \in \mathbb{R}^n, b \in \mathbb{R}$ und eine orthogonale Matrix $C \in \mathbb{R}^{n \times n}$;
- (iii) Sind X und Y unabhängig, dann gilt $DVar(X + Y) \leq DVar(X) + DVar(Y)$;
- (iv) Gilt $DVar(X + Y) = DVar(X) + DVar(Y)$, so ist X oder Y eine Konstante.

Beweis. (i) Ist $DVar(X)=0$, dann ist auch

$$DVar(X)^2 = \int_{\mathbb{R}^{2n}} \frac{|\varphi_{X,X}(t, s) - \varphi_X(t)\varphi_X(s)|^2}{c_n^2 \|t\|_2^{1+n} \|s\|_2^{1+n}} dt ds = 0.$$

Es gilt also $\varphi_{X,X}(t, s) = \varphi_X(t+s) = \varphi_X(t)\varphi_X(s) \Leftrightarrow \mathbb{E}(e^{i(t^T+s^T)X}) = \mathbb{E}(e^{it^T X})\mathbb{E}(e^{is^T X})$
 $\forall t, s \in \mathbb{R}^n$.

Aus dieser Gleichung folgt $\varphi_X(t) = e^{it^T v}$ für einen konstanten Vektor $v \in \mathbb{R}^n$, also gilt fast sicher $X = v$ und somit $\mathbb{E}(X) = X$ fast sicher.

- (ii) Mit Satz 1.16 (v) und $Y := bCX + a$ gilt $e^{i(t+s)^T a} \varphi_{X,X}(bC^T t, bC^T s) = \varphi_{Y,Y}(t, s)$ und $e^{it^T a} \varphi_X(bC^T t) = \varphi_Y(t)$, hiermit gilt

$$\begin{aligned} & |\varphi_{Y,Y}(t, s) - \varphi_Y(t)\varphi_Y(s)|^2 \\ &= \left| e^{i(t+s)^T a} \int_{\mathbb{R}^n} e^{i(t^T bCx + s^T bCx)} dF_X(x) - e^{it^T a} \int_{\mathbb{R}^n} e^{i(t^T bCx)} dF_X(x) e^{is^T a} \int_{\mathbb{R}^n} e^{i(s^T bCx)} dF_X(x) \right|^2 \\ &= \underbrace{\left| e^{i(t+s)^T a} \right|^2}_{=1} \left| \int_{\mathbb{R}^n} e^{i(t^T bCx + s^T bCx)} dF_X(x) - \int_{\mathbb{R}^n} e^{i(t^T bCx)} dF_X(x) \int_{\mathbb{R}^n} e^{i(s^T bCx)} dF_X(x) \right|^2 \\ &= |\varphi_{X,X}(bC^T t, bC^T s) - \varphi_X(bC^T t)\varphi_X(bC^T s)|^2. \end{aligned}$$

Also hat der Vektor a keine Auswirkung auf die $DVar$. Weil C eine orthogonale Matrix ist, gilt $|\det(C)| = 1$ und C^{-1} ist auch orthogonal mit $\|C^{-1}x\|_2 = \|Cx\|_2 = \|x\|_2$. Mit der Transformationsformel erhalten wir

$$DVar(Y)^2 = \int_{\mathbb{R}^{2n}} \frac{|\varphi_{Y,Y}(t, s) - \varphi_Y(t)\varphi_Y(s)|^2}{c_n^2 \|t\|_2^{1+n} \|s\|_2^{1+n}} dt ds$$

2.1. Definiton und grundlegende Eigenschaften

$$\begin{aligned}
&= \int_{\mathbb{R}^{2n}} \frac{|\varphi_{X,X}(bC^T t, bC^T s) - \varphi_X(bC^T t)\varphi_X(bC^T s)|^2}{c_n^2 \|t\|_2^{1+n} \|s\|_2^{1+n}} dt ds \\
&= \int_{\mathbb{R}^{2n}} \left| \frac{1}{\det(bC)} \right|^2 \frac{|\varphi_{X,X}(t, s) - \varphi_X(t)\varphi_X(s)|^2}{c_n^2 \|C^{-1}t/b\|_2^{1+n} \|C^{-1}s/b\|_2^{1+n}} dt ds \\
&= \int_{\mathbb{R}^{2n}} \left| \frac{b^{n+1}}{b^n} \right|^2 \underbrace{\left| \frac{1}{\det(C)} \right|^2}_{=1} \frac{|\varphi_{X,X}(t, s) - \varphi_X(t)\varphi_X(s)|^2}{c_n^2 \|t\|_2^{1+n} \|s\|_2^{1+n}} dt ds \\
&= |b|^2 \int_{\mathbb{R}^{2n}} \frac{|\varphi_{X,X}(t, s) - \varphi_X(t)\varphi_X(s)|^2}{c_n^2 \|t\|_2^{1+n} \|s\|_2^{1+n}} dt ds = |b|^2 \|\varphi_{X,X}(t, s) - \varphi_X(t)\varphi_X(s)\|_{\omega}^2.
\end{aligned}$$

Also

$$DVar(Y)^2 = DVar(bCX + a)^2 = |b|^2 DVar(X)^2 \Leftrightarrow DVar(bCX + a) = |b| DVar(X).$$

(iii) Es gilt

$$\begin{aligned}
\|\varphi_{X,X}(t, s) (\varphi_{Y,Y}(t, s) - \varphi_Y(t)\varphi_Y(s))\|_{\omega}^2 &= \int_{\mathbb{R}^{2n}} |\varphi_{X,X}(t, s) (\varphi_{Y,Y}(t, s) - \varphi_Y(t)\varphi_Y(s))|^2 d\omega \\
&\leq \int_{\mathbb{R}^{2n}} |(\varphi_{Y,Y}(t, s) - \varphi_Y(t)\varphi_Y(s))|^2 d\omega \\
&= \|\varphi_{Y,Y}(t, s) - \varphi_Y(t)\varphi_Y(s)\|_{\omega}^2
\end{aligned}$$

und analog

$$\|\varphi_Y(t)\varphi_Y(s) (\varphi_{X,X}(t, s) - \varphi_X(t)\varphi_X(s))\|_{\omega} \leq \|(\varphi_{X,X}(t, s) - \varphi_X(t)\varphi_X(s))\|_{\omega}.$$

Damit erhalten wir, da X und Y unabhängig sind

$$\begin{aligned}
DVar(X + Y) &= \|\varphi_{X+Y, X+Y}(t, s) - \varphi_{X+Y}(t)\varphi_{X+Y}(s)\|_{\omega} \\
&= \|\varphi_{X,X}(t, s)\varphi_{Y,Y}(t, s) - \varphi_X(t)\varphi_Y(t)\varphi_X(s)\varphi_Y(s)\|_{\omega} \\
&= \|\varphi_{X,X}(t, s) (\varphi_{Y,Y}(t, s) - \varphi_Y(t)\varphi_Y(s)) \\
&\quad + \varphi_Y(t)\varphi_Y(s) (\varphi_{X,X}(t, s) - \varphi_X(t)\varphi_X(s))\|_{\omega} \\
&\leq \|\varphi_{X,X}(t, s) (\varphi_{Y,Y}(t, s) - \varphi_Y(t)\varphi_Y(s))\|_{\omega} \quad (*) \\
&\quad + \|\varphi_Y(t)\varphi_Y(s) (\varphi_{X,X}(t, s) - \varphi_X(t)\varphi_X(s))\|_{\omega} \\
&\leq \|(\varphi_{Y,Y}(t, s) - \varphi_Y(t)\varphi_Y(s))\|_{\omega} \\
&\quad + \|(\varphi_{X,X}(t, s) - \varphi_X(t)\varphi_X(s))\|_{\omega} \\
&= DVar(X) + DVar(Y).
\end{aligned}$$

(iv) In der obigen Ungleichung (*) gilt Gleichheit nur wenn X oder Y zu sich selbst unabhängig wären, dies ist aber der Fall wenn X beziehungsweise Y eine Konstante ist.

□

Definition 2.6 (Distanz Kovarianz). Seien X und Y Zufallsvariablen mit $\dim(X)=n$ und $\dim(Y)=m$ und $\mathbb{E}(X) + \mathbb{E}(Y) < \infty$, dann heißt die Quadratwurzel aus

$$DCov(X, Y)^2 := \int_{\mathbb{R}^{n+m}} \frac{|\varphi_{X,Y}(t, s) - \varphi_X(t)\varphi_Y(s)|^2}{c_n c_m \|t\|_2^{1+n} \|s\|_2^{1+m}} dt ds$$

die *Distanz Kovarianz* von X und Y.

Durch die Defintion der Distanz Kovarianz und Distanz Varianz folgt unmittelbar die Positivität der Maßzahlen, denn es ist

$$DCov(X, Y) = \|\varphi_{X,Y} - \varphi_X \varphi_Y\|_{\omega} \geq 0$$

wegen den Eigenschaften einer Norm.

Nun möchten wir einige Eigenschaften der Distanz Kovarianz näher untersuchen und Gemeinsamkeiten und Unterschiede zur Kovarianz feststellen.

Satz 2.7 (Eigenschaften der Distanz Kovarianz). Seien X und Y Zufallsvektoren mit $\dim(X)=n$ und $\dim(Y)=m$ und $\mathbb{E}(X) + \mathbb{E}(Y) < \infty$, dann gilt:

- (i) $DCov(X, X) = DVar(X)$;
- (ii) $DCov(a_1 + b_1 C_1 X, a_2 + b_2 C_2 Y) = \sqrt{|b_1 b_2|} DCov(X, Y)$ für beliebige $b_1, b_2 \in \mathbb{R}$, $\forall a_1 \in \mathbb{R}^n, \forall a_2 \in \mathbb{R}^m$ und für beliebige orthogonale Matrizen $C_1 \in \mathbb{R}^{n \times n}, C_2 \in \mathbb{R}^{m \times m}$;
- (iii) Sind $(X_1, Y_1), (X_2, Y_2)$ unabhängige Zufallsvektoren, dann gilt $DCov(X_1 + X_2, Y_1 + Y_2) \leq DCov(X_1, Y_1) + DCov(X_2, Y_2)$;
- (iv) $DCov(X_1 + X_2, Y_1 + Y_2) = DCov(X_1, Y_1) + DCov(X_2, Y_2) \Leftrightarrow X_1$ und Y_1 sind Konstanten oder X_2 und Y_2 sind Konstanten oder X_1, X_2, Y_1, Y_2 sind unabhängig;
- (v) $DCov(X, Y)^2 \leq 2\sqrt{\mathbb{E}(\|X - \mathbb{E}(X)\|_2^2)\mathbb{E}(\|Y - \mathbb{E}(Y)\|_2^2)}$.

Beweis. (i) folgt direkt aus den Defintionen.

(ii) Mit Satz 1.16 (v) und $Y_1 := b_1 C_1 X + a_1, Y_2 := b_2 C_2 Y + a_2$ gilt $e^{it^T a_1 + is^T a_2} \varphi_{X,Y}(b_1 C_1^T t, b_2 C_2^T s) = \varphi_{Y_1, Y_2}(t, s)$ und $e^{it^T a_1} \varphi_X(b_1 C_1^T t) = \varphi_{Y_1}(t)$ und für Y und Y_2 analog. Hiermit gilt

2.1. Definiton und grundlegende Eigenschaften

$$\begin{aligned}
& |\varphi_{Y_1, Y_2}(t, s) - \varphi_{Y_1}(t)\varphi_{Y_2}(s)|^2 \\
&= \left| e^{it^T a_1 + is^T a_2} \int_{\mathbb{R}^{n+m}} e^{it^T b_1 C_1 x + is^T b_2 C_2 y} dF_{X, Y}(x, y) - e^{it^T a_1} \int_{\mathbb{R}^n} e^{it^T b_1 C_1 x} dF_X(x) e^{is^T a_2} \int_{\mathbb{R}^m} e^{is^T b_2 C_2 y} dF_Y(y) \right|^2 \\
&= \underbrace{\left| e^{it^T a_1 + is^T a_2} \right|^2}_{=1} \left| \int_{\mathbb{R}^{n+m}} e^{it^T b_1 C_1 x + is^T b_2 C_2 y} dF_{X, Y}(x, y) - \int_{\mathbb{R}^n} e^{it^T b_1 C_1 x} dF_X(x) \int_{\mathbb{R}^m} e^{is^T b_2 C_2 y} dF_Y(y) \right|^2 \\
&= \left| \varphi_{X, Y}(b_1 C_1^T t, b_2 C_2^T s) - \varphi_X(b_1 C_1^T t) \varphi_Y(b_2 C_2^T s) \right|^2,
\end{aligned}$$

also haben die Vektoren a_1 und a_2 keine Auswirkungen auf die Distanz Kovarianz. Mit einem ähnlichen Vorgehen wie in Satz 2.5 (ii) erhalten wir

$$\begin{aligned}
DCov(X, Y)^2 &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \frac{|\varphi_{Y_1, Y_2}(t, s) - \varphi_{Y_1}(t)\varphi_{Y_2}(s)|^2}{c_n \|t\|_2^{1+n} c_m \|s\|_2^{1+m}} dt ds \\
&= \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \frac{|\varphi_{X, Y}(b_1 C_1^T t, b_2 C_2^T s) - \varphi_X(b_1 C_1^T t) \varphi_Y(b_2 C_2^T s)|^2}{c_n \|t\|_2^{1+n} c_m \|s\|_2^{1+m}} dt ds \\
&= \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \left| \frac{1}{\det(b_1 C_1)} \right| \left| \frac{1}{\det(b_2 C_2)} \right| \frac{|\varphi_{X, Y}(t, s) - \varphi_X(t)\varphi_Y(s)|^2}{c_n \|C_1^{-1}t/b_1\|_2^{1+n} c_m \|C_2^{-1}s/b_2\|_2^{1+m}} dt ds \\
&= \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \left| \frac{b_1^{n+1} b_2^{m+1}}{b_1^n b_2^m} \right| \underbrace{\left| \frac{1}{\det(C_1)} \right| \left| \frac{1}{\det(C_2)} \right|}_{=1} \frac{|\varphi_{X, Y}(t, s) - \varphi_X(t)\varphi_Y(s)|^2}{c_n \|t\|_2^{1+n} c_m \|s\|_2^{1+m}} dt ds \\
&= |b_1 b_2| \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \frac{|\varphi_{X, Y}(t, s) - \varphi_X(t)\varphi_Y(s)|^2}{c_n \|t\|_2^{1+n} c_m \|s\|_2^{1+m}} dt ds \\
&= |b_1 b_2| \|\varphi_{X, Y}(t, s) - \varphi_X(t)\varphi_Y(s)\|_{\omega}^2.
\end{aligned}$$

(iii) Mit einem ähnlichen Vorgehen wie im Beweis von Satz 2.5 (iii) folgt, da (X_1, Y_1) und (X_2, Y_2) unabhängig sind

$$\begin{aligned}
DCov(X_1 + X_2, Y_1 + Y_2) &= \|\varphi_{X_1 + X_2, Y_1 + Y_2}(t, s) - \varphi_{X_1 + X_2}(t)\varphi_{Y_1 + Y_2}(s)\|_{\omega} \\
&= \|\varphi_{X_1, Y_1}(t, s)\varphi_{X_2, Y_2}(t, s) - \varphi_{X_1}(t)\varphi_{X_2}(t)\varphi_{Y_1}(s)\varphi_{Y_2}(s)\|_{\omega} \\
&= \|\varphi_{X_1, Y_1}(t, s)(\varphi_{X_2, Y_2}(t, s) - \varphi_{X_2}(t)\varphi_{Y_2}(s)) \\
&\quad + \varphi_{X_2}(t)\varphi_{Y_2}(s)(\varphi_{X_1, Y_1}(t, s) - \varphi_{X_1}(t)\varphi_{Y_1}(s))\|_{\omega} \\
&\leq \|(\varphi_{X_2, Y_2}(t, s) - \varphi_{X_2}(t)\varphi_{Y_2}(s))\|_{\omega} \\
&\quad + \|(\varphi_{X_1, Y_1}(t, s) - \varphi_{X_1}(t)\varphi_{Y_1}(s))\|_{\omega} \\
&= DCov(X_1, Y_1) + DCov(X_2, Y_2).
\end{aligned}$$

(iv)

„ \Leftarrow “ Aus (ii) folgt die Behauptung, wenn X_1 und Y_1 oder X_2 und Y_2 Konstanten sind. Wenn X_1, X_2, Y_1, Y_2 unabhängig sind, folgt die Behauptung unmittelbar, denn es gilt $DCov(X_1, Y_1) = DCov(X_2, Y_2) = DCov(X_1 + X_2, Y_1 + Y_2) = 0$.

„ \Rightarrow “ 1. Wenn X_1 und Y_1 oder X_2 und Y_2 keine Konstanten sind, dann gilt die Gleichung

$$\begin{aligned} & \| \varphi_{X_1, Y_1}(t) (\varphi_{X_2, Y_2}(t, s) - \varphi_{X_2}(t) \varphi_{Y_2}(s)) + \varphi_{X_2}(t) \varphi_{Y_2}(s) (\varphi_{X_1, Y_1}(t, s) - \varphi_{X_1}(t) \varphi_{Y_1}(t)) \|_{\omega} \\ &= \| (\varphi_{X_2, Y_2}(t, s) - \varphi_{X_2}(t) \varphi_{Y_2}(s)) \|_{\omega} \\ &+ \| (\varphi_{X_1, Y_1}(t, s) - \varphi_{X_1}(t) \varphi_{Y_1}(t)) \|_{\omega} := (*) \end{aligned}$$

nur wenn X_1, X_2, Y_1, Y_2 unabhängig sind.

2. Wenn X_1, X_2, Y_1, Y_2 nicht unabhängig sind und X_2 und Y_2 keine Konstanten sind, dann gilt die Gleichung (*) nur wenn X_1 und Y_1 Konstanten sind.

3. Wenn X_1, X_2, Y_1, Y_2 nicht unabhängig sind und X_1 und Y_1 keine Konstanten sind, dann gilt die Gleichheit nur wenn X_2 und Y_2 Konstanten sind.

(v) Nach (2.6) ist $DCov(X, Y)^2 \leq \mathbb{E}(\|X - X'\|_2) \mathbb{E}(\|Y - Y'\|_2)$ und somit

$$\begin{aligned} \mathbb{E}(\|X - X'\|_2) &\leq \sqrt{\mathbb{E}(\|X - X'\|_2^2)} = \sqrt{\mathbb{E}(\|(X - \mathbb{E}(X)) - (X' - \mathbb{E}(X'))\|_2^2)} \\ &\leq \sqrt{\mathbb{E}(\|X - \mathbb{E}(X)\|_2^2) + \mathbb{E}(\|X' - \mathbb{E}(X')\|_2^2)} = \sqrt{2\mathbb{E}(\|X - \mathbb{E}(X)\|_2^2)}. \end{aligned}$$

Analog folgt $\mathbb{E}(\|Y - Y'\|_2) \leq \sqrt{2\mathbb{E}(\|Y - \mathbb{E}(Y)\|_2^2)}$. Somit ist

$$DCov(X, Y)^2 \leq \mathbb{E}(\|X - X'\|_2) \mathbb{E}(\|Y - Y'\|_2) \leq 2\sqrt{\mathbb{E}(\|X - \mathbb{E}(X)\|_2^2) \mathbb{E}(\|Y - \mathbb{E}(Y)\|_2^2)}.$$

□

Bemerkung Man sieht anhand der Eigenschaften der Distanz Kovarianz, dass diese Unterschiede zur Kovarianz aufweist. Zum Beispiel ist die Funktion nicht bilinear, dennoch kann die Wurzel vom Betrag eines Skalars herausgezogen werden und Konstanten, welche an die Zufallsvariable addiert werden, haben keine Auswirkungen auf die Maßzahl. Dies zählt also zu den Gemeinsamkeiten der Kovarianz und der Distanz Kovarianz, ebenso wie die Tatsache, dass beide Funktionen zur Messung des Zusammenhangs zwischen Zufallsvariablen genutzt werden.

Wie bei der Kovarianz können wir auch für die Distanz Kovarianz eine standardisierte Größe angeben:

Definition 2.8. Seien X und Y Zufallsvektoren mit $\mathbb{E}(X) + \mathbb{E}(Y) < \infty$, dann heißt

$$DCor(X, Y) := \begin{cases} \frac{DCov(X, Y)}{\sqrt{DVar(X)DVar(Y)}} & \text{falls } DVar(X)DVar(Y) > 0, \\ 0 & \text{falls } DVar(X)DVar(Y) = 0, \end{cases}$$

die *Distanz Korrelation*.

Da diese Maßzahl analog zur Korrelation definiert ist, können wir auch sehr ähnliche Eigenschaften für diese feststellen wie in Satz 1.8.

Satz 2.9 (Eigenschaften der Distanz Korrelation). Seien X und Y Zufallsvektoren mit $\dim(X)=n$, $\dim(Y)=m$ und $\mathbb{E}(X) + \mathbb{E}(Y) < \infty$, dann gelten folgende Eigenschaften:

- (i) $DCor(X, Y) = DCor(Y, X)$;
- (ii) $DCor(X, X) = 1$;
- (iii) $DCor(X, Y) \in [0, 1]$;
- (iv) $DCor(X, Y) = 0 \Leftrightarrow X$ und Y sind unabhängig;
- (v) $DCor(a_1 + b_1C_1X, a_2 + b_2C_2Y) = DCor(X, Y)$ für beliebige $b_1, b_2 \in \mathbb{R}$,
 $\forall a_1 \in \mathbb{R}^n, \forall a_2 \in \mathbb{R}^m$ und für beliebige orthogonale Matrizen $C_1 \in \mathbb{R}^{n \times n}, C_2 \in \mathbb{R}^{m \times m}$.

Beweis. (i) und (ii) folgen unmittelbar aus der Definition.

(iii) Da $DCov \geq 0$ und $DVar \geq 0$ folgt $DCor = \frac{DCov(X, Y)}{\sqrt{DVar(X)DVar(Y)}} \geq 0$.

$DCor \leq 1$ wird später mittels der Brownschen Distanz Kovarianz gezeigt.

(iv) „ \Leftarrow “ folgt unmittelbar, da X und Y unabhängig und somit $DCov(X, Y)=0$.

“ \Rightarrow “ Die einzige Möglichkeit, dass $DCor(X, Y)$ Null ist, ist wenn der Zähler

$$DCov(X, Y)^2 = \|\varphi_{X, Y}(t, s) - \varphi_X(t)\varphi_Y(s)\|_{\omega}^2$$

von $DCor^2$ Null ist. Dies ist aber nur der Fall wenn X und Y unabhängig sind.

(v) Mit Satz 2.5 (ii) und Satz 2.7 (ii) ist

$$\begin{aligned} DCor(a_1 + b_1C_1X, a_2 + b_2C_2Y) &= \frac{DCov(a_1 + b_1C_1X, a_2 + b_2C_2Y)}{\sqrt{DVar(a_1 + b_1C_1X)DVar(a_2 + b_2C_2Y)}} \\ &= \frac{\sqrt{|b_1b_2|}DCor(X, Y)}{\sqrt{|b_1|DVar(X)|b_2|DVar(Y)}} = DCor(X, Y). \end{aligned}$$

□

Bemerkung Eigenschaft (v) besagt, dass die Distanz Korrelation skaleninvariant ist. Die Skaleninvarianz war eine Anforderung an eine Kennzahl, welche zur Messung des Zusammenhangs zwischen Zufallsvariablen dient, die wir am Anfang des Kapitels gefordert hatten.

Die Distanz Korrelation besitzt einen engen Zusammenhang zur Korrelation im Fall von bivariaten Normalverteilungen, den wir nun genauer untersuchen. Dazu nehmen wir an, dass $X, Y \sim N(0, 1)$ mit $Cov(X, Y) = \rho(X, Y) = \rho$.

Satz 2.10. Es gilt

(i) $DCor(X, Y) \leq |\rho|$;

(ii) $DCor^2(X, Y) = \frac{\rho \arcsin(\rho) + \sqrt{1-\rho^2} - \rho \arcsin(\rho/2) - \sqrt{4-\rho^2+1}}{1+\pi/3-\sqrt{3}}$;

(iii) $\inf_{\rho \neq 0} \frac{DCor(X, Y)}{|\rho|} = \lim_{\rho \rightarrow 0} \frac{DCor(X, Y)}{|\rho|} = \frac{1}{2(1+\pi/3-\sqrt{3})^{1/2}} \approx 0.89066$.

Beweis. Nach Beispiel 1.18 folgt, dass $\varphi_X(t) = e^{-t^2/2}$ und $\varphi_{X,Y}(t, s) = e^{-(t^2+s^2)/2-\rho ts}$. Wir definieren die Funktion

$$\begin{aligned} F(\rho) &:= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\varphi_{X,Y}(t, s) - \varphi_X(t)\varphi_Y(s)|^2 \frac{dt}{t^2} \frac{ds}{s^2} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| e^{-(t^2+s^2)/2-\rho ts} - e^{-t^2/2} e^{-s^2/2} \right|^2 \frac{dt}{t^2} \frac{ds}{s^2}. \end{aligned}$$

Damit folgt wegen $c_1 = \pi$, dass $DCov(X, Y)^2 = \frac{F(\rho)}{c_1^2} = \frac{F(\rho)}{\pi^2}$ und dass

$$\begin{aligned} DVar(X)^2 &= 1/\pi^2 \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\varphi_{X,X}(t, s) - \varphi_X(t)\varphi_X(s)|^2 \frac{dt}{t^2} \frac{ds}{s^2} \\ &= 1/\pi^2 \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| e^{-(t+s)^2/2} - e^{-t^2/2} e^{-s^2/2} \right|^2 \frac{dt}{t^2} \frac{ds}{s^2} \\ &= 1/\pi^2 \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| e^{-(t^2+s^2)/2-ts} - e^{-t^2/2} e^{-s^2/2} \right|^2 \frac{dt}{t^2} \frac{ds}{s^2} = 1/\pi^2 \cdot F(1). \end{aligned}$$

2.1. Definiton und grundlegende Eigenschaften

Also erhalten wir die Darstellung $DCor^2(X, Y) = \frac{F(\rho)}{F(1)}$.

(i)

$$\begin{aligned}
 F(\rho) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| e^{-(t^2+s^2)/2-\rho ts} - e^{-t^2/2} e^{-s^2/2} \right|^2 \frac{dt}{t^2} \frac{ds}{s^2} \\
 &= \int_{\mathbb{R}^2} e^{-t^2-s^2-2\rho ts} - 2e^{-(t^2+s^2)-\rho ts} + e^{-(t^2+s^2)} \frac{dt}{t^2} \frac{ds}{s^2} \\
 &= \int_{\mathbb{R}^2} e^{-t^2-s^2} (1 - 2e^{-\rho ts} + e^{-2\rho ts}) \frac{dt}{t^2} \frac{ds}{s^2} \\
 &= \int_{\mathbb{R}^2} e^{-t^2-s^2} \sum_{l=2}^{\infty} \frac{2^l - 2}{l!} (-\rho ts)^l \frac{dt}{t^2} \frac{ds}{s^2} \\
 &= \int_{\mathbb{R}^2} e^{-t^2-s^2} \sum_{k=1}^{\infty} \frac{2^{2k} - 2}{(2k)!} (-\rho ts)^{2k} \frac{dt}{t^2} \frac{ds}{s^2} \\
 &= \rho^2 \underbrace{\sum_{k=1}^{\infty} \frac{2^{2k} - 2}{(2k)!} \rho^{2(k-1)} \int_{\mathbb{R}^2} e^{-t^2-s^2} (ts)^{2(k-1)} dt ds}_{:=G(\rho)}
 \end{aligned}$$

Somit ist $F(\rho) = \rho^2 G(\rho)$ und damit $G(1) = F(1)$. Für $G(\rho)$ gilt $G(\rho) \leq G(1) \forall \rho \in [0, 1]$, woraus nun folgt, dass

$$DCor^2(X, Y) = \frac{F(\rho)}{F(1)} = \rho^2 \frac{G(\rho)}{G(1)} \leq \rho^2 \frac{G(1)}{G(1)} = \rho^2$$

gilt und deswegen auch $DCor(X, Y) \leq |\rho|$.

(ii) Es ist $F(0) = F'(0) = 0$ und somit $F(\rho) = \int_0^\rho \int_0^x F''(z) dz dx$, wobei

$$F''(z) = \frac{d^2}{dz^2} \int_{\mathbb{R}^2} e^{-t^2-s^2} (1 - 2e^{-zts} + e^{-2zts}) \frac{dt}{t^2} \frac{ds}{s^2} = 4V(z) - 2V\left(\frac{z}{2}\right),$$

mit

$$V(z) = \int_{\mathbb{R}^2} e^{-t^2-s^2-2zts} dt ds.$$

Denn es ist

2.1. Definiton und grundlegende Eigenschaften

$$\begin{aligned}
 F''(z) &= \frac{d^2}{dz^2} \int_{\mathbb{R}^2} e^{-t^2-s^2} (1 - 2e^{-zts} + e^{-2zts}) \frac{dt}{t^2} \frac{ds}{s^2} \\
 &= \frac{d}{dz} \underbrace{\int_{\mathbb{R}^2} e^{-t^2-s^2} (2tse^{-zts} - 2tse^{-2zts}) \frac{dt}{t^2} \frac{ds}{s^2}}_{=F'(z)} \\
 &= \int_{\mathbb{R}^2} e^{-t^2-s^2} (-2t^2s^2e^{-zts} + 4t^2s^2e^{-2zts}) \frac{dt}{t^2} \frac{ds}{s^2} \\
 &= -2 \int_{\mathbb{R}^2} e^{-t^2-s^2-zts} dt ds + 4 \int_{\mathbb{R}^2} e^{-t^2-s^2-2zts} dt ds = -2V\left(\frac{z}{2}\right) + 4V(z)
 \end{aligned}$$

und

$$\int_0^\rho \int_0^x F''(z) dz dx = \int_0^\rho F'(x) - \underbrace{F'(0)}_{=0} dx = F(\rho) - \underbrace{F(0)}_{=0} = F(\rho).$$

Die Eigenwerte von $A := \begin{pmatrix} 1 & z \\ z & 1 \end{pmatrix}$ sind $1 \pm z$ und außerdem gilt

$$-(t, s)A(t, s)^T = -(t, s) \begin{pmatrix} 1 & z \\ z & 1 \end{pmatrix} \begin{pmatrix} t \\ s \end{pmatrix} = -t^2 - s^2 - 2zts.$$

Für A gibt es eine Orthogonale Matrix C, sodass $A = C^T \Lambda C$, wobei $\Lambda = \begin{pmatrix} 1-z & 0 \\ 0 & 1+z \end{pmatrix}$.

Weiter gilt $\int_{\mathbb{R}} e^{-t^2\lambda} dt = \left(\frac{\pi}{\lambda}\right)^{1/2}$ und somit erhalten wir mittels Transformationsregel

$$\begin{aligned}
 V(z) &= \int_{\mathbb{R}^2} e^{-t^2-s^2-2zts} dt ds = \int_{\mathbb{R}^2} e^{-(t,s)A(t,s)^T} dt ds = \int_{\mathbb{R}^2} e^{-(t,s)C^T \Lambda C(t,s)^T} dt ds \\
 &= \int_{\mathbb{R}^2} e^{-(t,s)\Lambda(t,s)^T} dt ds = \int_{\mathbb{R}^2} e^{-t^2(1-z)-s^2(1+z)} dt ds \\
 &= \int_{\mathbb{R}} e^{-t^2(1-z)} dt \int_{\mathbb{R}} e^{-s^2(1+z)} ds = \left(\frac{\pi}{1-z}\right)^{1/2} \left(\frac{\pi}{1+z}\right)^{1/2} \\
 &= \frac{\pi}{\sqrt{1-z^2}}.
 \end{aligned}$$

Nun folgt

$$F(\rho) = \int_0^\rho \int_0^x \left(\frac{4\pi}{\sqrt{1-z^2}} - \frac{2\pi}{\sqrt{1-z^2/4}} \right) dz dx$$

$$\begin{aligned}
 &= \int_0^\rho |4\pi \arcsin(z) - 4\pi \arcsin(z/2)|_0^x dx \\
 &= 4\pi \int_0^\rho \arcsin(x) - \arcsin(x/2) dx \\
 &= 4\pi \left[\sqrt{1-x^2} + x \arcsin(x) - \sqrt{4-x^2} - x \arcsin(x/2) \right]_0^\rho \\
 &= 4\pi(\rho \arcsin(\rho) + \sqrt{1-\rho^2} - \rho \arcsin(\rho/2) - \sqrt{4-\rho^2} + 1).
 \end{aligned}$$

Also ist $F(1) = 4\pi(\arcsin(1) - \arcsin(1/2) - \sqrt{3} + 1) = 4\pi(1 + \pi/3 - \sqrt{3})$ und es ist

$$DCor(X, Y)^2 = \frac{F(\rho)}{F(1)} = \frac{\rho \arcsin(\rho) + \sqrt{1-\rho^2} - \rho \arcsin(\rho/2) - \sqrt{4-\rho^2} + 1}{1 + \pi/3 - \sqrt{3}}.$$

(iii) Wir benutzen die Regel von L'Hospital und verifizieren das äquivalente Ergebnis

$$\lim_{\rho \rightarrow 0} \frac{DCor(X, Y)^2}{\rho^2} = \frac{1}{4(1 + \pi/3 - \sqrt{3})}.$$

Es ist

$$\begin{aligned}
 &\frac{d}{d\rho} \rho \arcsin(\rho) + \sqrt{1-\rho^2} - \rho \arcsin(\rho/2) - \sqrt{4-\rho^2} + 1 \\
 &= \arcsin(\rho) + \frac{\rho}{\sqrt{1-\rho^2}} - \frac{\rho}{\sqrt{1-\rho^2}} - \arcsin(\rho/2) - \frac{\rho}{2\sqrt{1-\rho^2/4}} + \frac{\rho}{2\sqrt{1-\rho^2/4}} \quad (1) \\
 &= \arcsin(\rho) - \arcsin(\rho/2)
 \end{aligned}$$

und somit folgt mittels zweimaligen Anwendens der Regel von L'Hospital:

$$\begin{aligned}
 \lim_{\rho \rightarrow 0} \frac{DCor(X, Y)^2}{\rho^2} &\stackrel{(1)}{=} \lim_{\rho \rightarrow 0} \frac{\arcsin(\rho) - \arcsin(\rho/2)}{2\rho(1 + \pi/3 - \sqrt{3})} = \lim_{\rho \rightarrow 0} \frac{\frac{1}{\sqrt{1-\rho^2}} - \frac{1}{2\sqrt{1-\rho^2/4}}}{2(1 + \pi/3 - \sqrt{3})} \\
 &= \frac{1}{4(1 + \pi/3 - \sqrt{3})}.
 \end{aligned}$$

□

Berechnet man mittels Satz 2.10 (ii) die Werte von $DCor^2$ und ρ in jeweiliger Abhängigkeit, so erhält man Abbildung 2.1. Die unterbrochene Linie stellt den Wert von ρ in Abhängigkeit zu $DCor^2$ dar und die durchgezogene Linie repräsentiert den Wert von $DCor^2$ in Abhängigkeit zu ρ . Man sieht hier gut die Eigenschaft (i) von Satz 2.10, da die unterbrochene Linie stets über der durchgezogenen liegt. Das beide Werte im Nullpunkt übereinstimmen, folgt insbesondere aus der Tatsache, dass wenn X und Y normalverteilt sind, dass $\rho(X, Y) = 0 \Leftrightarrow X$ und Y sind unabhängig. Also repräsentieren Korrelation und Distanz Korrelation im bivariaten normalverteilten Fall Unabhängigkeit.

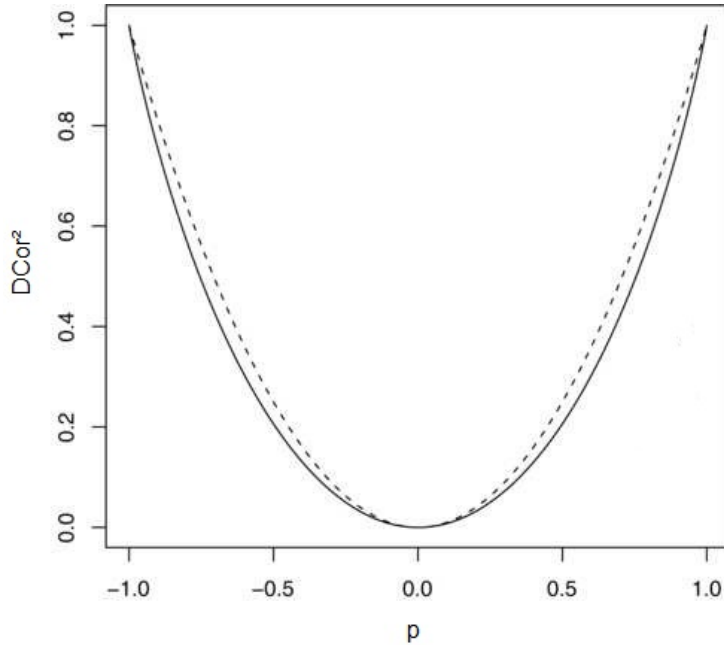


Abbildung 2.1: Korrelation und Distanz Korrelation im normalverteilten bivariaten Fall

2.2 Empirische Distanz Kovarianz und Korrelation

Wie bei der Kovarianz und Korrelation kann auch zu der Distanz Kovarianz\Korrelation ein Schätzer definiert werden. Wir werden neben der Definition noch einige Eigenschaften dieser Schätzer vorstellen.

Definition 2.11 (Empirische Distanz Kovarianz). Seien $(X^d, Y^d) = \{(X_i, Y_i) | i = 1, \dots, d\}$ Stichproben des Zufallsvektors (X, Y) mit $\dim(X)=n$ und $\dim(Y)=m$.

Sei $a_{kl} := \|X_k - X_l\|_2$, $b_{kl} := \|Y_k - Y_l\|_2$ und

$$A_{kl} := a_{kl} - \bar{a}_k - \bar{a}_l + \bar{a}_{..}, \quad k, l = 1, \dots, d,$$

wobei

$$\bar{a}_k := \frac{1}{d} \sum_{l=1}^d a_{kl}, \quad \bar{a}_l := \frac{1}{d} \sum_{k=1}^d a_{kl}, \quad \bar{a}_{..} := \frac{1}{d^2} \sum_{l,k=1}^d a_{kl}.$$

Analog wird $B_{kl} := b_{kl} - \bar{b}_k - \bar{b}_l + \bar{b}_{..}$, $k, l = 1, \dots, d$ definiert. Die Quadratwurzel aus

$$DCov_d^2(X^d, Y^d) := \frac{1}{d^2} \sum_{k,l=1}^d A_{kl} B_{kl}$$

heißt *empirische Distanz Kovarianz*. Analog erhält man einen Schätzer für die Distanz Varianz durch

$$DVar_d^2(X^d) = DCov_d^2(X^d, X^d) = \frac{1}{d^2} \sum_{k,l=1}^d A_{kl}^2,$$

folglich heißt die Quadratwurzel aus dieser Zahl die *empirische Distanz Varianz*.

Beispiel 2.12. Betrachten wir den Datensatz *hills* aus dem Paket **MASS** und untersuchen mittels der Distanz Kovarianz den Zusammenhang zwischen den Stichproben bezüglich der Attribute *time* und *dist* erhalten wir den Wert $dcov(dist, time) = 8.930865$. Die empirische Kovarianz, welche wir in einem früherem Beispiel berechnet haben, ist $Cov(dist, time) = 254.1944$.

Somit kann man also auch mit der Distanz Kovarianz einen Zusammenhang zwischen den Stichproben vermuten.

Berechnet man nun die empirische Distanz Varianz und die empirische Standardabweichung, so erhält man für *dist* und *time*:

	dist	time
emp. Distanz Varianz	3.03603	30.032
emp. Standardabw.	5.523936	50.04072

Da die Distanz Kovarianz über die Norm von charakteristischen Funktionen definiert ist, kommt die Frage auf, ob die empirische Distanz Kovarianz nicht auch einen Bezug zu den charakteristischen Funktionen besitzt. Tatsächlich gibt es einen anderen Zugang zu der empirischen Distanz Kovarianz, welcher die empirischen charakteristischen Funktionen berücksichtigt und sogar mit dem Zugang von obiger Definition übereinstimmt. Dazu benötigen wir vorher einige Hilfsaussagen der folgenden Lemmata.

Lemma 2.13. Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ eine stetige ungerade Funktion, dann gilt

$$\int_{-z}^z f(x) dx = 0 \quad \forall z \in \mathbb{R}_+^n \setminus \{0\}.$$

Beweis. Die Stammfunktion F einer stetigen ungeraden Funktion f ist gerade, somit gilt

$$\int_{-z}^z f(x) dx = |F(x)|_{-z}^z = F(z) - F(-z) = F(z) - F(z) = 0.$$

□

Lemma 2.14. Seien X und Y Zufallsvektoren mit $\dim(X)=n$ und $\dim(Y)=m$, dann gilt

$$\int_{\mathbb{R}^n} \frac{1 - e^{it^T X}}{\|t\|_2^{1+n}} dt = c_n \|X\|_2.$$

Beweis. Sinus ist eine ungerade Funktion und $\frac{1}{\|t\|_2^{1+n}}$ ist eine gerade Funktion. Das Produkt einer ungeraden und einer geraden Funktion ist ungerade. Mit Lemma 2.13 erhalten wir

$$\int_{\mathbb{R}^n} \frac{\sin it^T X}{\|t\|_2^{1+n}} dt = 0.$$

Und somit

$$\int_{\mathbb{R}^n} \frac{1 - e^{it^T X}}{\|t\|_2^{1+n}} dt = \int_{\mathbb{R}^n} \frac{1 - \cos t^T X}{\|t\|_2^{1+n}} dt - i \underbrace{\int_{\mathbb{R}^n} \frac{\sin t^T X}{\|t\|_2^{1+n}} dt}_{=0} = c_n \|X\|_2.$$

□

Satz 2.15. Seien (X^d, Y^d) Stichproben des Zufallsvektors (X, Y) mit $\dim(X)=n$, $\dim(Y)=m$ und $\mathbb{E}(\|X\|_2) + \mathbb{E}(\|Y\|_2) < \infty$ und φ_X^d, φ_Y^d und $\varphi_{X,Y}^d$ die zugehörigen empirischen charakteristischen Funktionen wie in Definition 1.29. Dann gilt

$$DCov_d^2(X^d, Y^d) = \|\varphi_{X,Y}^d(t, s) - \varphi_X^d(t)\varphi_Y^d(s)\|_\omega^2.$$

Beweis. Einfachheitshalber betrachten wir den Fall $\dim(X) = \dim(Y) = 1$. Es ist

$$\begin{aligned} |\varphi_{X,Y}^d(t, s) - \varphi_X^d(t)\varphi_Y^d(s)|^2 &= \varphi_{X,Y}^d(t, s)\overline{\varphi_{X,Y}^d(t, s)} - \varphi_X^d(t, s)\overline{\varphi_X^d(t)\varphi_Y^d(s)} \\ &\quad - \overline{\varphi_{X,Y}^d(t, s)\varphi_X^d(t)\varphi_Y^d(s)} + \varphi_X^d(t)\varphi_Y^d(s)\overline{\varphi_X^d(t)\varphi_Y^d(s)}. \end{aligned}$$

Sei $c_{X,Y}^k := tX_k + sY_k$ und $c_{X,Y}^l := tX_l + sY_l$, dann ist mit Satz 2.3

$$\begin{aligned} \varphi_{X,Y}^d(t, s)\overline{\varphi_{X,Y}^d(t, s)} &= \frac{1}{d} \sum_{k=1}^d e^{ic_{X,Y}^k} \frac{1}{d} \sum_{k=1}^d e^{-ic_{X,Y}^k} \\ &= \frac{1}{d} \sum_{k=1}^d \cos(c_{X,Y}^k) + i \cdot \sin(c_{X,Y}^k) \cdot \frac{1}{d} \sum_{k=1}^d \cos(c_{X,Y}^k) - i \cdot \sin(c_{X,Y}^k) \\ &= \frac{1}{d^2} \sum_{k,l=1}^d \cos(c_{X,Y}^k) \cos(c_{X,Y}^l) + \sin(c_{X,Y}^k) \sin(c_{X,Y}^l) + R \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{d^2} \sum_{k,l=1}^d \cos(t(X_k - X_l) + s(Y_k - Y_l)) + R \\
 &= \frac{1}{d^2} \sum_{k,l=1}^d \cos(t(X_k - X_l)) \cos(s(Y_k - Y_l)) - \underbrace{\sin(t(X_k - X_l)) \sin(s(Y_k - Y_l))}_{:=R_1} + R \\
 &= \frac{1}{d^2} \sum_{k,l=1}^d \cos(t(X_k - X_l)) \cos(s(Y_k - Y_l)) - R_1, \\
 &\quad \underbrace{\hspace{10em}}_{:=S_1}
 \end{aligned}$$

wobei

$$\begin{aligned}
 R &= \sum_{k,l=1}^d i \cdot \sin(c_{X,Y}^k) \cos(c_{X,Y}^l) - i \cdot \sin(c_{X,Y}^k) \cos(c_{X,Y}^l) \\
 &= i \sum_{k,l=1}^d \sin(c_{X,Y}^k - c_{X,Y}^l) + \sin(c_{X,Y}^k + c_{X,Y}^l).
 \end{aligned}$$

Nach Lemma 2.13 ist $\int R_1 d\omega = 0$. Daher ist R_1 der Teil der bei der Auswertung des Integrals redundant ist.

Analog erhalten wir

$$\begin{aligned}
 \varphi_{X,Y}^d(t, s) \overline{\varphi_X^d(t) \varphi_Y^d(s)} &= \frac{1}{d^3} \sum_{k,l,m=1}^d \cos((X_k - X_l)t) \cos((Y_k - Y_m)s) - R_2, \\
 &\quad \underbrace{\hspace{10em}}_{:=S_2} \\
 \varphi_X^d(t) \varphi_Y^d(s) \overline{\varphi_X^d(t) \varphi_Y^d(s)} &= \frac{1}{d^2} \sum_{k,l=1}^d \cos((X_k - X_l)t) \frac{1}{d^2} \sum_{k,l=1}^d \cos((Y_k - Y_l)s) - R_3 \\
 &\quad \underbrace{\hspace{10em}}_{:=S_3}
 \end{aligned}$$

mit $\int R_i d\omega = 0$ für $i = 2, 3$ (da diese ähnlich wie R_1 aus Summen von Sinus-Funktionen bestehen).

Wir berechnen nun $|\varphi_{X,Y}^d(t, s) - \varphi_X^d(t) \varphi_Y^d(s)|^2$ und verwenden dafür Satz 2.3(vi). Für den besseren Lesefluss definieren wir $u := (X_k - X_l)t$, $v := (Y_k - Y_l)s$, $w := (Y_k - Y_m)s$ und $z := (Y_m - Y_j)s$ und erhalten

$$\begin{aligned}
 S_1 &= \frac{1}{d^2} \sum_{k,l=1}^d 1 - (1 - \cos(u)) - (1 - \cos(v)) + (1 - \cos(u))(1 - \cos(v)) \\
 &= 1 - \frac{1}{d^2} \sum_{k,l=1}^d (1 - \cos(u)) - \frac{1}{d^2} \sum_{k,l=1}^d (1 - \cos(v)) + \frac{1}{d^2} \sum_{k,l=1}^d (1 - \cos(u))(1 - \cos(v)),
 \end{aligned}$$

$$\begin{aligned}
 S_2 &= \frac{1}{d^3} \sum_{k,l,m=1}^d 1 - (1 - \cos(u)) - (1 - \cos(w)) + (1 - \cos(u))(1 - \cos(w)) \\
 &= 1 - \frac{1}{d^2} \sum_{k,l=1}^d (1 - \cos(u)) - \frac{1}{d^2} \sum_{k,m=1}^d (1 - \cos(w)) + \frac{1}{d^3} \sum_{k,l,m=1}^d (1 - \cos(u))(1 - \cos(w)),
 \end{aligned}$$

$$\begin{aligned}
 S_3 &= \frac{1}{d^4} \sum_{k,l,m,j=1}^d 1 - (1 - \cos(u)) - (1 - \cos(z)) + (1 - \cos(u))(1 - \cos(z)) \\
 &= 1 - \frac{1}{d^2} \sum_{k,l=1}^d (1 - \cos(u)) - \frac{1}{d^2} \sum_{m,j=1}^d (1 - \cos(z)) + \frac{1}{d^4} \sum_{k,l,m,j=1}^d (1 - \cos(u))(1 - \cos(z)).
 \end{aligned}$$

Insgesamt

$$\begin{aligned}
 S_1 - 2S_2 + S_3 &= \\
 &= \frac{1}{d^2} \sum_{k,l=1}^d (1 - \cos(u))(1 - \cos(v)) - 2 \frac{1}{d^3} \sum_{k,l,m=1}^d (1 - \cos(u))(1 - \cos(w)) \\
 &\quad + \frac{1}{d^4} \sum_{k,l,m,j=1}^d (1 - \cos(u))(1 - \cos(z)).
 \end{aligned}$$

Wenn $\|\varphi_{X,Y}^d(t,s) - \varphi_X^d(t)\varphi_Y^d(s)\|_\omega^2$ ausgerechnet wird, müssen Integrale der Form $\int (1 - \cos(u))(1 - \cos(w))d\omega$ ausgewertet werden. Nun erhalten wir mit dem Satz von Fubini und Lemma 2.2

$$\begin{aligned}
 &\int_{\mathbb{R}^2} \frac{(1 - \cos((X_k - X_l)t))(1 - \cos((Y_k - Y_l)s))}{c_1^2 t^2 s^2} dt ds \\
 &= \int_{\mathbb{R}} \frac{(1 - \cos((X_k - X_l)t))}{c_1 t^2} dt \cdot \int_{\mathbb{R}} \frac{(1 - \cos((Y_k - Y_l)s))}{c_1 s^2} ds \\
 &= \|X_k - X_l\|_2 \|Y_k - Y_l\|_2.
 \end{aligned}$$

Insgesamt gilt

$$\|\varphi_{X,Y}^d(t,s) - \varphi_X^d(t)\varphi_Y^d(s)\|_\omega^2 = T_1 - 2T_2 + T_3 \tag{2.7}$$

mit

$$T_1 = \frac{1}{d^2} \sum_{k,l=1}^d \|X_k - X_l\|_2 \|Y_k - Y_l\|_2,$$

2.2. Empirische Distanz Kovarianz und Korrelation

$$T_2 = \frac{1}{d^3} \sum_{k,l,m=1}^d \|X_k - X_l\|_2 \|Y_k - Y_m\|_2,$$

$$T_3 = \frac{1}{d^2} \sum_{k,l=1}^d \|X_k - X_l\|_2 \frac{1}{d^2} \sum_{k,l=1}^d \|Y_k - Y_l\|_2.$$

Jetzt muss noch $DCov_d^2(X^d, Y^d) = T_1 - 2T_2 + T_3$ verifiziert werden. Mit den Bezeichnungen aus Definition 2.11 erhalten wir

$$\begin{aligned} d^2 DCov_d^2(X^d, Y^d) &= \sum_{k,l=1}^d \left\{ \begin{array}{cccc} a_{kl}b_{kl} & -a_{kl}\bar{b}_k & -a_{kl}\bar{b}_l & a_{kl}\bar{b}_{..} \\ -\bar{a}_k b_{kl} & \bar{a}_k \bar{b}_k & \bar{a}_k \bar{b}_l & -\bar{a}_k \bar{b}_{..} \\ -\bar{a}_l b_{kl} & \bar{a}_l \bar{b}_k & \bar{a}_l \bar{b}_l & -\bar{a}_l \bar{b}_{..} \\ \bar{a}_{..} b_{kl} & -\bar{a}_{..} \bar{b}_k & -\bar{a}_{..} \bar{b}_l & \bar{a}_{..} \bar{b}_{..} \end{array} \right\} \\ &= \sum_{k,l=1}^d a_{kl}b_{kl} - \sum_{k=1}^d a_k \bar{b}_k - \sum_{l=1}^d a_l \bar{b}_l + a_{..} \bar{b}_{..} \\ &\quad - \sum_{k=1}^d \bar{a}_k b_k + d \sum_{k=1}^d \bar{a}_k \bar{b}_k + \sum_{k,l=1}^d \bar{a}_k \bar{b}_l - d \sum_{k=1}^d \bar{a}_k \bar{b}_{..} \\ &\quad - \sum_{l=1}^d \bar{a}_l b_l + \sum_{k,l=1}^d \bar{a}_l \bar{b}_k + d \sum_{l=1}^d \bar{a}_l \bar{b}_l - d \sum_{l=1}^d \bar{a}_l \bar{b}_{..} \\ &\quad + \bar{a}_{..} b_{..} - d \sum_{k=1}^d \bar{a}_{..} \bar{b}_k - d \sum_{l=1}^d \bar{a}_{..} \bar{b}_l + d^2 \bar{a}_{..} \bar{b}_{..}, \end{aligned}$$

mit $a_k = d\bar{a}_k$, $a_l = d\bar{a}_l$, $b_k = d\bar{b}_k$ und $b_l = d\bar{b}_l$. Es ist

$$T_1 = \frac{1}{d^2} \sum_{k,l=1}^d a_{kl}b_{kl},$$

$$T_2 = \frac{1}{d^3} \sum_{k,l,m=1}^d a_{kl}b_{km} = \frac{1}{d^3} \sum_{k=1}^d a_k b_k = \frac{1}{d} \sum_{k=1}^d \bar{a}_k \bar{b}_k,$$

$$T_3 = \frac{1}{d^2} \sum_{k,l=1}^d a_{kl} \frac{1}{d^2} \sum_{k,l=1}^d b_{kl} = \bar{a}_{..} \bar{b}_{..},$$

$$d^2 T_3 = d^2 \bar{a}_{..} \bar{b}_{..} = \frac{a_{..}}{d} \sum_{l=1}^d \frac{b_l}{d} = \sum_{k,l=1}^d \bar{a}_k \bar{b}_l,$$

und somit erhalten wir

$$\begin{aligned}
 d^2 DCov_d^2(X^d, Y^d) &= d^2 T_1 - d^2 T_2 - d^2 T_2 + d^2 T_3 \\
 &\quad - d^2 T_2 + d^2 T_2 + d^2 T_3 - d^2 T_3 \\
 &\quad - d^2 T_2 + d^2 T_3 + d^2 T_2 - d^2 T_3 \\
 &\quad + d^2 T_3 - d^2 T_3 - d^2 T_3 + d^2 T_3 = d^2 (T_1 - 2T_2 + T_3).
 \end{aligned}$$

□

Bemerkung In Formel (2.7) ist eine weitere Möglichkeit angegeben die empirische Distanz Kovarianz auszurechnen. Tatsächlich ist dieser Rechenweg aufwändiger als die Berechnung aus Definition 2.11. So braucht die Berechnung zum Beispiel beim Datensatz *hills* über Formel (2.7) ca.1.1 Sekunden im Durchschnitt und die Berechnung über die Definition ca.0.05 Sekunden durchschnittlich und das bei einem Stichprobenumfang von nur 35.

Satz 2.16. Es gilt

- (i) $DCov_d^2(X^d, Y^d) \geq 0$ und $DVar_d^2(X^d) \geq 0$;
- (ii) $DVar_d(X^d) = 0 \Leftrightarrow X_1 = \dots = X_d$.

Beweis. (i) folgt direkt aus Satz 2.15, da $\|\cdot\|_\omega \geq 0$.

(ii) „ \Leftarrow “ folgt unmittelbar aus $X_1 = \dots = X_d$, da $A_{kl} = 0$ für alle $k, l = 1, \dots, d$. „ \Rightarrow “ Sei $DVar_d(X^d) = 0$, dann ist $A_{kl} = 0$ für alle $k, l = 1, \dots, d$. Hiermit gilt

$$\begin{aligned}
 0 = A_{kk} &= \underbrace{a_{kk}}_{=0} - \bar{a}_{k\cdot} - \bar{a}_{\cdot k} + \bar{a}_{\cdot\cdot} = -\bar{a}_{k\cdot} - \bar{a}_{\cdot k} + \bar{a}_{\cdot\cdot} \\
 &= -\frac{1}{d} \sum_{l=1}^d a_{kl} - \frac{1}{d} \sum_{l=1}^d a_{lk} + \frac{1}{d^2} \sum_{k,l=1}^d a_{kl}.
 \end{aligned}$$

Wegen der Positivität von a_{kl} bzw. a_{lk} folgt

$$\bar{a}_{k\cdot} = \bar{a}_{\cdot k} = \bar{a}_{\cdot\cdot}/2.$$

Und damit folgt nun

$$0 = A_{kl} = a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{\cdot\cdot} = a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{k\cdot} + \bar{a}_{\cdot l} = a_{kl} = \|X_k - X_l\|_2$$

und daraus folgt $X_1 = \dots = X_d$. □

Wir möchten nun die Konvergenz der empirischen Distanz Kovarianz gegen die Distanz Kovarianz nachweisen. Dazu benötigen wir einige Hilfsmittel aus der Statistik.

Definition 2.17 (U-parameter). $\theta = \theta(F)$ heißt *U-Parameter* der Verteilung F , falls eine messbare Funktion $h : \mathbb{R}^n \rightarrow \mathbb{R}$ existiert mit

$$\theta(F) = \theta_h(F) = \int_{\mathbb{R}^n} h dF^n.$$

Die Funktion h heißt *Kern*. F^n bezeichnet das Produktmaß $F \times \dots \times F$ auf \mathbb{R}^n und n bezeichnet die Ordnung des Kerns.

Beispiel 2.18. Der Erwartungswert $\mathbb{E}(X)$ einer Zufallsvariable X mit Verteilung F ist ein U-Parameter mit Kern $h(x) := x$, denn

$$\mathbb{E}(X) = \int x dF = \int h dF.$$

Analog ist jedes k -te Moment $\mathbb{E}(X^k)$ einer Zufallsvariable X mit Verteilung F ein U-Parameter mit Kern $h(x) := x^k$, $k \in \mathbb{N}$.

Definition 2.19 (U-Statistik und V-Statistik). Sei $\theta_h(F)$ ein U-Parameter und X_1, \dots, X_d eine Stichprobe, dann heißt

$$U_h(X_1, \dots, X_d) := \frac{(d-n)!}{d!} \sum \{h(X_{i_1}, \dots, X_{i_d}) \mid \{i_j\} \text{ verschieden, } 1 \leq i_j \leq d\}$$

die *U-Statistik* für $\theta_h(F)$. Und

$$V_h(X_1, \dots, X_d) := \frac{1}{d^n} \sum \{h(X_{i_1}, \dots, X_{i_d}) \mid 1 \leq i_j \leq d \text{ für alle } j\}$$

heißt die *V-Statistik* für $\theta_h(F)$.

Definition 2.20. Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ mit $f = (f_1, \dots, f_n)$ und $f_i : \mathbb{R} \rightarrow \mathbb{R}$. Dann heißt f *Produktfunktion* (oder kurz Produkt), falls

$$f_1 \otimes \dots \otimes f_n := f(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n)$$

gilt. Ist weiter F eine Verteilung auf \mathbb{R} , dann heißt die Produktfunktion $f_1 \otimes \dots \otimes f_n$ *F-integrierbar*, falls jedes f_i messbar ist und $\int |f_i| dF < \infty$.

Eine messbare Funktion $h : \mathbb{R}^n \rightarrow \mathbb{R}$ nennt man *beschränkt durch ein F-integrierbares Produkt*, wenn $|h| \leq f_1 \otimes \dots \otimes f_n$ für ein F-integrierbares Produkt $f_1 \otimes \dots \otimes f_n$.

Definition 2.21. Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum (Ω', \mathfrak{A}') ein Messraum und T eine Indexmenge. Man nennt X einen *stochastischen Prozess* (oder kurz Prozess), wenn es eine Familie von Zufallsvariablen $X_t : \Omega \rightarrow \Omega'$ $t \in T$ gibt, sodass $X : \Omega \times T \rightarrow \Omega'$, $(\omega, t) \mapsto X_t(\omega)$ und X_t $\mathfrak{A} - \mathfrak{A}'$ -messbar ist für alle $t \in T$.

Für jedes feste $\omega \in \Omega$ heißt die Funktion $X(\omega) : T \rightarrow \Omega'$, $t \mapsto X_t(\omega)$, der *Pfad* des stochastischen Prozesses X .

Die Funktion

$$\mu_X(t) := \mathbb{E}(X(t))$$

heißt *Erwartungswertfunktion* des stochastischen Prozesses,

$$C_X(t, s) := \text{Cov}(X_t, X_s)$$

heißt (*Auto-*)*Kovarianzfunktion* des stochastischen Prozesses und

$$R_X(t, s) := \mathbb{E}(X(t)\overline{X(s)})$$

heißt (*Auto-*)*Korrelationsfunktion* des stochastischen Prozesses.

Bemerkung Ein Stochastischer Prozess $X : \Omega \times T \rightarrow \Omega'$, $(\omega, t) \mapsto X_t(\omega)$ ist als Produktabbildung aufgefasst im Allgemeinen nicht messbar. Die Eigenschaft, dass X_t $\mathfrak{A} - \mathfrak{A}'$ -messbar ist für alle $t \in T$ besagt, dass für jedes feste $t \in T$ X_t eine Zufallsvariable ist. T wird oftmals als Zeiteinheit aufgefasst, also $T = \mathbb{N}$ oder $T = \mathbb{R}$.

Für die Korrelations- und Kovarianzfunktion gilt:

$$\begin{aligned} C_X(t, s) &= \text{Cov}(X_t, X_s) = \mathbb{E}((X(t) - \mathbb{E}(X(t)))(\overline{X(s) - \mathbb{E}(X(s))})) \\ &= \mathbb{E}(X(s)\overline{X(t)}) - \mathbb{E}(X(t))\overline{\mathbb{E}(X(s))} = R_X(t, s) - \mu_X(t)\overline{\mu_X(s)}. \end{aligned}$$

Definition 2.22. Ein stochastischer Prozess $(X_k)_{k \in \mathbb{N}}$ heißt *ergodisch*, falls

$$\mathbb{E}(X_{t_i}) = \lim_{d \rightarrow \infty} \frac{1}{d} \sum_{j=1}^d X_{t_j} \quad \forall i \in \mathbb{N}.$$

$(X_k)_{k \in \mathbb{N}}$ heißt *stationär*, falls

$$F_{X_{t_1}, \dots, X_{t_d}}(z_1, \dots, z_n) = F_{X_{t_1+t}, \dots, X_{t_d+t}}(z_1, \dots, z_d) \quad \forall t_1, \dots, t_d, t, d \in \mathbb{N}.$$

Satz 2.23. Sei $(X_k)_{k \in \mathbb{N}}$ ein ergodischer stationärer Prozess und sei $h : \mathbb{R}^n \rightarrow \mathbb{R}$ beschränkt durch ein F-integrierbares Produkt. Dann gilt

$$\lim_{d \rightarrow \infty} |U_h(X_1, \dots, X_d) - V_h(X_1, \dots, X_d)| = 0 \quad P - \text{fast sicher.}$$

Beweis. Siehe [1]. □

Satz 2.24 (Starkes Gesetz der großen Zahlen für U-Statistiken). Sei $(X_k)_{k \in \mathbb{N}}$ ein ergodischer stationärer Prozess und sei $h : \mathbb{R}^n \rightarrow \mathbb{R}$ beschränkt durch ein F-integrierbares Produkt. Wenn eine der beiden folgenden Eigenschaften gilt:

- (i) F ist diskret;
- (ii) h ist F^n fast überall stetig;

dann ist

$$\lim_{d \rightarrow \infty} U_h(X_1, \dots, X_d) = \theta_h(F) \quad P - \text{fast sicher.}$$

Beweis. Siehe [4]. □

Bemerkung Wegen Satz 2.23 gilt das Starke Gesetz der großen Zahlen für U-Statistiken auch analog für V-Statistiken und wird auch als *Starkes Gesetz der großen Zahlen für V-Statistiken* bezeichnet.

Satz 2.25. Seien (X^d, Y^d) Stichproben des Zufallsvektors (X, Y) mit $\dim(X)=n$, $\dim(Y)=m$ und $\mathbb{E}(X) + \mathbb{E}(Y) < \infty$, dann gilt

$$\lim_{d \rightarrow \infty} DCov_d(X^d, Y^d) = DCov(X, Y) \quad \text{fast sicher.}$$

Beweis. Verkürzend bezeichnen wir

$$\xi_d(t, s) := \frac{1}{d} \sum_{k=1}^d e^{it^T X_k + is^T Y_k} - \frac{1}{d} \sum_{k=1}^d e^{it^T X_k} \frac{1}{d} \sum_{k=1}^d e^{is^T Y_k}$$

und erhalten somit, dass $DCov_d^2(X^d, Y^d) = \|\xi_d(t, s)\|_\omega^2$ und per Definition ist $\xi_d(t, s)$ eine V-Statistik. Außerdem definieren wir

$$u_k := e^{it^T X_k} - \varphi_X(t) \quad \text{und} \quad v_k := e^{is^T Y_k} - \varphi_Y(s).$$

Wir erhalten

$$\begin{aligned} & \frac{1}{d} \sum_{k=1}^d u_k v_k - \frac{1}{d} \sum_{k=1}^d u_k \frac{1}{d} \sum_{k=1}^d v_k \\ &= \frac{1}{d} \sum_{k=1}^d e^{it^T X_k + is^T Y_k} - \varphi_X(t) e^{is^T Y_k} - \varphi_Y(s) e^{it^T X_k} + \varphi_X(t) \varphi_Y(s) \\ & \quad - \left(\frac{1}{d^2} \sum_{k, l=1}^d e^{it^T X_k + is^T Y_l} - \varphi_X(t) e^{is^T Y_k} - \varphi_Y(s) e^{it^T X_k} + \varphi_X(t) \varphi_Y(s) \right) \end{aligned}$$

$$= \frac{1}{d} \sum_{k=1}^d e^{it^T X_k + is^T Y_k} - \frac{1}{d} \sum_{k=1}^d e^{it^T X_k} \frac{1}{d} \sum_{k=1}^d e^{is^T Y_k} = \xi_d(t, s).$$

Für jedes $\delta > 0$ definieren wir die Menge $D(\delta) := \{(t, s) | \delta \leq \|t\|_2 \leq 1/\delta, \delta \leq \|s\|_2 \leq 1/\delta\}$ und

$$DCov_{d,\delta}^2 := \int_{D(\delta)} |\xi_d(t, s)|^2 d\omega.$$

Für jedes $\delta > 0$ ist die Gewichtungsfunktion $\omega(t, s)$ auf $D(\delta)$ beschränkt und somit ist $DCov_{d,\delta}$ beschränkt. Für jedes $\delta > 0$ folgt nun mit dem Starken Gesetz der Großen Zahlen für V-Statistiken und dem Satz von der majorisierten Konvergenz

$$\begin{aligned} \lim_{d \rightarrow \infty} DCov_{d,\delta}^2 &= \lim_{d \rightarrow \infty} \int_{D(\delta)} |\xi_d(t, s)|^2 d\omega \int_{D(\delta)} \lim_{d \rightarrow \infty} |\xi_d(t, s)|^2 d\omega \\ &= \underbrace{\int_{D(\delta)} |\varphi_{X,Y}(t, s) - \varphi_X(t)\varphi_Y(s)|^2 d\omega}_{=: DCov_{,\delta}^2} \text{ fast sicher.} \end{aligned}$$

Außerdem gilt

$$\lim_{\delta \rightarrow 0} DCov_{,\delta}^2 = DCov^2.$$

Zu zeigen ist nun

$$\limsup_{\delta \rightarrow 0} \limsup_{d \rightarrow \infty} |DCov_{d,\delta}^2 - DCov_d^2| = 0.$$

Für jedes $\delta > 0$ gilt

$$\begin{aligned} |DCov_{d,\delta}^2 - DCov_d^2| &\leq \int_{\|t\|_2 < \delta} |\xi_d(t, s)|^2 d\omega + \int_{\|t\|_2 > 1/\delta} |\xi_d(t, s)|^2 d\omega \\ &\quad + \int_{\|s\|_2 < \delta} |\xi_d(t, s)|^2 d\omega + \int_{\|s\|_2 > 1/\delta} |\xi_d(t, s)|^2 d\omega. \end{aligned} \tag{1}$$

Wegen $|x + y|^2 \leq 2|x|^2 + 2|y|^2$ und der Cauchy-Schwarzschen Ungleichung erhalten wir

$$\begin{aligned} |\xi_d(t, s)|^2 &= \left| \frac{1}{d} \sum_{k=1}^d u_k v_k - \frac{1}{d} \sum_{k=1}^d u_k \frac{1}{d} \sum_{k=1}^d v_k \right|^2 \leq 2 \left| \frac{1}{d} \sum_{k=1}^d u_k v_k \right|^2 + 2 \left| \frac{1}{d} \sum_{k=1}^d u_k \frac{1}{d} \sum_{k=1}^d v_k \right|^2 \\ &\leq \frac{2}{d} \sum_{k=1}^d |u_k|^2 \frac{1}{d} \sum_{k=1}^d |v_k|^2 + \frac{2}{d} \sum_{k=1}^d |u_k|^2 \frac{1}{d} \sum_{k=1}^d |v_k|^2 = \frac{4}{d} \sum_{k=1}^d |u_k|^2 \frac{1}{d} \sum_{k=1}^d |v_k|^2 \end{aligned}$$

Betrachten wir nun den ersten Summand in (1), dann ist

$$\int_{\|t\|_2 < \delta} |\xi_d(t, s)|^2 d\omega \leq \frac{4}{d} \sum_{k=1}^d \int_{\|t\|_2 < \delta} \frac{|u_k|^2 dt}{c_n \|t\|_2^{1+n}} \frac{1}{d} \sum_{k=1}^d \int_{\mathbb{R}^m} \frac{|v_k|^2 ds}{c_m \|s\|_2^{1+m}}.$$

Nun ist

$$\begin{aligned} |v_k|^2 &= \left| e^{is^T Y_k} - \varphi_Y(s) \right|^2 = \left| e^{is^T Y_k} \right|^2 + |\varphi_Y(s)|^2 - e^{is^T Y_k} \overline{\varphi_Y(s)} - e^{-is^T Y_k} \varphi_Y(s) \\ &= 1 + |\varphi_Y(s)|^2 - e^{is^T Y_k} \overline{\varphi_Y(s)} - e^{-is^T Y_k} \varphi_Y(s) \end{aligned}$$

und

$$\begin{aligned} |u_k|^2 &= \left| e^{it^T X_k} - \varphi_X(t) \right|^2 = \left| e^{it^T X_k} \right|^2 + |\varphi_X(t)|^2 - e^{it^T X_k} \overline{\varphi_X(t)} - e^{-it^T X_k} \varphi_X(t) \\ &= 1 + |\varphi_X(t)|^2 - e^{it^T X_k} \overline{\varphi_X(t)} - e^{-it^T X_k} \varphi_X(t). \end{aligned}$$

Daraus folgt mit analogen Schritten wie in der Gleichung (2.6) und Lemma 2.14

$$\begin{aligned} \int_{\mathbb{R}^m} \frac{|v_k|^2 ds}{c_m \|s\|_2^{1+m}} &= \int_{\mathbb{R}^m} \frac{1 + |\varphi_Y(s)|^2 - e^{is^T Y_k} \overline{\varphi_Y(s)} - e^{-is^T Y_k} \varphi_Y(s)}{c_m \|s\|_2^{1+m}} ds \\ &= \int_{\mathbb{R}^m} \frac{1 + |\varphi_Y(s)|^2 - e^{is^T Y_k} \overline{\varphi_Y(s)} - e^{-is^T Y_k} \varphi_Y(s) + 1 - 1}{c_m \|s\|_2^{1+m}} ds \\ &= - \int_{\mathbb{R}^m} \frac{1 - |\varphi_Y(s)|^2}{c_m \|s\|_2^{1+m}} ds + \mathbb{E}_Y \left(\int_{\mathbb{R}^m} \frac{1 - e^{is(Y_k - Y)}}{c_m \|s\|_2^{1+m}} ds \right) + \mathbb{E}_Y \left(\int_{\mathbb{R}^m} \frac{1 - e^{is(Y - Y_k)}}{c_m \|s\|_2^{1+m}} ds \right) \\ &= -\mathbb{E}(\|Y - Y'\|_2) + \mathbb{E}_Y(\|Y_k - Y\|_2) + \mathbb{E}_Y(\|Y - Y_k\|_2) \\ &= 2\mathbb{E}_Y(\|Y_k - Y\|_2) - \underbrace{\mathbb{E}(\|Y - Y'\|_2)}_{>0} \\ &\leq 2\mathbb{E}(\|Y_k\|_2) + \mathbb{E}(\|Y\|_2) = 2(\|Y_k\|_2 + \mathbb{E}(\|Y\|_2)). \end{aligned}$$

\mathbb{E}_Y bezeichnet den Erwartungswert bezüglich Y und Y' ist eine unabhängige identisch verteilte Kopie von Y . Wir definieren nun für $z = (z_1, \dots, z_n) \in \mathbb{R}^n$ die Funktion

$$G(y) := \int_{\|z\|_2 < y} \frac{1 - \cos(z_1)}{\|z\|_2^{1+n}} dz,$$

welche durch c_n beschränkt ist, denn mit $e_1 = (1, 0, \dots, 0)^T$ folgt

$$G(y) = \int_{\|z\|_2 < y} \frac{1 - \cos(z_1)}{\|z\|_2^{1+n}} dz \leq \int_{\mathbb{R}^n} \frac{1 - \cos(z^T e_1)}{\|z\|_2^{1+n}} dz = c_n \underbrace{\|e_1\|_2}_{=1} = c_n.$$

Außerdem gilt $\lim_{y \rightarrow 0} G(y) = 0$. Wir können Lemma 2.2 auf folgende Weise modifizieren:

$$\int_{\|t\|_2 < \delta} \frac{1 - \cos(t^T x) dt}{\|t\|_2^{1+n}} = \|x\|_2 G(\|x\|_2 \delta).$$

Mit der Substitution $z = t\|x\|_2$ und wenn x mittels orthogonalen Projektionen in eine geeignete Form gebracht wird, erhalten wir

$$\int_{\|t\|_2 < \delta} \frac{1 - \cos(t^T x) dt}{\|t\|_2^{1+n}} = \|x\|_2 \cdot \int_{\|z\|_2 < \delta\|x\|_2} \frac{1 - \cos(z_1) dz}{\|z\|_2^{1+n}} = \|x\|_2 G(\|x\|_2 \delta). \quad (*)$$

Mit (*) erhalten wir nun

$$\begin{aligned} & \int_{\|t\|_2 < \delta} \frac{|u_k|^2 dt}{c_n \|t\|_2^{1+n}} \\ &= - \int_{\|t\|_2 < \delta} \frac{1 - |\varphi_X(t)|^2 dt}{c_n \|t\|_2^{1+n}} + \mathbb{E}_X \left(\int_{\|t\|_2 < \delta} \frac{1 - e^{it(X_k - X)} dt}{c_n \|t\|_2^{1+n}} \right) + \mathbb{E}_X \left(\int_{\|t\|_2 < \delta} \frac{1 - e^{it(X - X_k)} dt}{c_n \|t\|_2^{1+n}} \right) \\ &= 2\mathbb{E}_X(\|X_k - X\|_2) G(\|X_k - X\|_2 \delta) - \mathbb{E}(\|X - X'\|_2) G(\|X - X'\|_2 \delta) \\ &\leq 2\mathbb{E}_X(\|X_k - X\|_2) G(\|X_k - X\|_2 \delta), \end{aligned}$$

wobei \mathbb{E}_X den Erwartungswert bezüglich X bezeichnet und X' eine unabhängige identisch verteilte Kopie von X ist. Insgesamt gilt also

$$\int_{\|t\|_2 < \delta} |\xi_d(t, s)|^2 d\omega \leq 4 \frac{1}{d} \sum_{k=1}^d 2(\|Y_k\|_2 + \mathbb{E}(\|Y\|_2)) \frac{1}{d} \sum_{k=1}^d 2\mathbb{E}_X(\|X_k - X\|_2) G(\|X_k - X\|_2 \delta).$$

Mit dem Starken Gesetz der Großen Zahlen erhalten wir

$$\limsup_{d \rightarrow \infty} \int_{\|t\|_2 < \delta} |\xi_d(t, s)|^2 d\omega \leq 4 \cdot 2 \cdot 2\mathbb{E}(\|Y\|_2) \cdot 2\mathbb{E}(\|X_1 - X_2\|_2) G(\|X_1 - X_2\|_2 \delta).$$

Mit dem Satz von der majorisierten Konvergenz folgt

$$\limsup_{\delta \rightarrow 0} \limsup_{d \rightarrow \infty} \int_{\|t\|_2 < \delta} |\xi_d(t, s)|^2 = 0 \text{ fast sicher.}$$

Nun schätzen wir den zweiten Summanden in (1) ab: Es gilt $|u_k|^2 \leq 4$ und somit

$\frac{1}{d} \sum_{k=1}^d |u_k|^2 \leq 4$. Noch eine Abschätzung, die wir benötigen ist

$$\int_{\|t\|_2 > 1/\delta} \frac{dt}{c_n \|t\|_2^{1+n}} \leq \frac{\delta^{1+n}}{c_n}.$$

Somit ist

$$\begin{aligned}
 \int_{\|t\|_2 > 1/\delta} |\xi_d(t, s)|^2 d\omega &\leq \frac{4}{d} \sum_{k=1}^d \int_{\|t\|_2 > 1/\delta} \frac{|u_k|^2 dt}{c_n \|t\|_2^{1+n}} \frac{1}{d} \sum_{k=1}^d \int_{\mathbb{R}^m} \frac{|v_k|^2 ds}{c_m \|s\|_2^{1+m}} \\
 &\leq 16 \int_{\|t\|_2 > 1/\delta} \frac{dt}{c_n \|t\|_2^{1+n}} \frac{1}{d} \sum_{k=1}^d \int_{\mathbb{R}^m} \frac{|v_k|^2 ds}{c_m \|s\|_2^{1+m}} \\
 &\leq 16\delta^{n+1} \frac{1}{c_n d} \sum_{k=1}^d \int_{\mathbb{R}^m} \frac{|v_k|^2 ds}{c_m \|s\|_2^{1+m}} \\
 &\leq 16\delta^{n+1} \frac{2}{c_n d} \sum_{k=1}^d (\|Y_k\|_2 + \mathbb{E}(\|Y\|_2)).
 \end{aligned}$$

Mit dem Starken Gesetz der Großen Zahlen gilt

$$\limsup_{d \rightarrow \infty} 16\delta^{n+1} \frac{2}{c_n d} \sum_{k=1}^d (\|Y_k\|_2 + \mathbb{E}(\|Y\|_2)) = \frac{32}{c_n} \cdot \delta^{n+1} \cdot 2\mathbb{E}(Y)$$

und somit

$$\limsup_{\delta \rightarrow 0} \limsup_{d \rightarrow \infty} \int_{\|t\|_2 > 1/\delta} |\xi_d(t, s)|^2 d\omega \leq \limsup_{\delta \rightarrow 0} \frac{32}{c_n} \cdot \delta^{n+1} \cdot 2\mathbb{E}(Y) = 0 \text{ fast sicher.}$$

Nun kann man für die restlichen Summanden in (1) analog argumentieren und somit folgt die Behauptung. \square

Völlig analog zu der Definition der empirischen Korrelation durch die empirische Kovarianz und empirische Varianz wie in Definition 1.12, kann man die empirische Distanz Korrelation durch die empirische Distanz Kovarianz und empirische Distanz Varianz definieren:

Definition 2.26 (Empirische Distanz Korrelation). Seien (X^d, Y^d) Stichproben der Zufallsvektoren X und Y mit $\dim(X)=n$, $\dim(Y)=m$ und $\mathbb{E}(X) + \mathbb{E}(Y) < \infty$, dann heißt

$$DCor_d(X^d, Y^d) := \begin{cases} \frac{DCov_d(X^d, Y^d)}{\sqrt{DVar_d(X^d)DVar_d(Y^d)}} & DVar_d(X^d)DVar_d(Y^d) > 0, \\ 0 & DVar_d(X^d)DVar_d(Y^d) = 0, \end{cases}$$

die *empirische Distanz Korrelation*.

Satz 2.27. Es gilt

- (i) $0 \leq DCor_d(X^d, Y^d) \leq 1$;
- (ii) Ist $DCor_d(X^d, Y^d) = 1$, dann existiert ein Vektor a , ein Skalar b und eine orthogonale Matrix C , sodass $Y^d = bCX^d + a$ gilt.

Beweis. (i) folgt analog zu Satz 2.9 (iii).

(ii) Einfachheitshalber gehen wir davon aus, dass $\dim(X)=\dim(Y)=n$ ist. Nach der Cauchy-Schwarzschen Ungleichung gilt für

$$\frac{1}{d^4} \left(\sum_{k,l=1}^d A_{k,l} B_{k,l} \right)^2 \leq \frac{1}{d^2} \left(\sum_{k,l=1}^d A_{k,l}^2 \right) \frac{1}{d^2} \left(\sum_{k,l=1}^d B_{k,l}^2 \right)$$

die Gleichheit nur wenn $A_{k,l} = \varepsilon B_{k,l}$ für $\varepsilon \in \mathbb{R}, \forall k, l$. Also gilt $DCor_d(X^d, Y^d) = 1$ nur wenn $A_{k,l} = \varepsilon B_{k,l} \forall k, l$.

Sei $|\varepsilon| = 1$, dann ist $\|X_k - X_l\|_2 = \|Y_k - Y_l\|_2 + d_k + d_l \forall k, l$ für gewisse d_k, d_l . Für $k = l$ erhält man $d_k = d_l = 0 \forall k$, da $\|X_k - X_l\|_2 = \|Y_k - Y_l\|_2 = 0$.

X^d und Y^d sind isometrisch, also kann man X^d aus Y^d durch Shift-,Rotations- und Reflektionsoperatoren erhalten und damit ist $Y^d = bCX^d + a$ für einen Vektor a , ein Skalar $b = \varepsilon$ und eine Orthogonalmatrix C .

Sei nun $|\varepsilon| \neq 1$ und $\varepsilon \neq 0$, dann kann man εX^d aus Y^d durch Shift-,Rotations- und Reflektionsoperatoren erhalten und damit ist $Y^d = bCX^d + a$ für einen Vektor a , ein Skalar $b = \varepsilon$ und eine Orthogonalmatrix C .

□

Beispiel 2.28. Für den Datensatz *hills* aus dem Paket **MASS** ist

$DCor_d(dist, time) = 0.9352944$. Zur Erinnerung: Die empirische Korrelation ist $Cor(dist, time) = 0.9195892$. Also ermittelt die Distanz Korrelation auch einen engen linearen Zusammenhang zwischen den Attributen.

Korollar 2.29. Seien (X^d, Y^d) Stichproben des Zufallsvektors (X, Y) mit $\dim(X)=n, \dim(Y)=m$ und $\mathbb{E}(X) + \mathbb{E}(Y) < \infty$, dann gilt

$$\lim_{d \rightarrow \infty} DCor_d(X^d, Y^d) = DCor(X, Y) \text{ fast sicher.}$$

Beweis. Folgt aus Satz 2.25 und der Definition der empirischen Distanz Korrelation. □

2.3 Unabhängigkeitstest

In diesem Abschnitt beschäftigen wir uns damit wie ein multivariater Test auf Unabhängigkeit aufbauend auf der empirischen Distanz Kovarianz abgeleitet werden kann. Bevor wir den Unabhängigkeitstest vorstellen, benötigen wir noch einige Hilfsmittel.

Definition 2.30 (Verteilungskonvergenz). Eine Folge von Verteilungsfunktionen $(F_k)_{k \in \mathbb{N}}$ heißt *verteilungskonvergent* gegen eine Funktion $F : \mathbb{R}^n \rightarrow [0, 1]$, wenn gilt

- (i) F ist rechtsstetig;
- (ii) alle Unstetigkeitsstellen von F liegen auf abzählbar vielen Hyperebenen $\{y \in \mathbb{R}^n | y_\nu = y_{\nu, \rho}\}$ mit den Unstetigkeitskoordinaten $y_{\nu, \rho} \in \mathbb{R}$;
- (iii) für $y \in \mathbb{R}^n$ für die F stetig in y ist, gilt $\lim_{k \rightarrow \infty} F_k(y) = F(y)$.

Man schreibt dafür abkürzend $F_k \xrightarrow{v} F$. Eine Folge von Zufallsvariablen $(X_k)_{k \in \mathbb{N}}$ heißt *verteilungskonvergent* gegen eine Zufallsvariable X , falls $F_{X_k} \xrightarrow{v} F_X$. Man schreibt dafür auch abkürzend $X_k \xrightarrow{v} X$.

Anmerkung Die Starke Konvergenz sowie die Konvergenz in Wahrscheinlichkeit implizieren die Verteilungskonvergenz, aber nicht umgekehrt! In [2] Satz 5.1. ist ein Beweis dazu gegeben.

Beispiel 2.31. Sei $X_k = (-1)^k X$ mit $X \sim N(0, 1)$, wegen der Symmetrie der Normalverteilung gilt $-X \sim N(0, 1)$. Damit gilt $X_k \xrightarrow{v} X$.

Satz 2.32. Seien $X, X_1, X_2 \dots$ Zufallsvariablen mit Verteilungsfunktionen $F_X, (F_{X_k})_{k \in \mathbb{N}}$ und charakteristischen Funktionen $\varphi_X, (\varphi_{X_k})_{k \in \mathbb{N}}$ und gelte $\lim_{k \rightarrow \infty} \varphi_{X_k}(t) = \varphi_X(t) \forall t \in \mathbb{R}^n$.

Dann gilt $F_{X_k} \xrightarrow{v} F_X$.

Gilt andererseits $F_{X_k} \xrightarrow{v} F_X$, so folgt $\lim_{k \rightarrow \infty} \varphi_{X_k}(t) = \varphi_X(t) \forall t \in \mathbb{R}^n$.

Beweis. Siehe [2] Satz 23.8. □

Satz 2.33. Seien $X, X_1, X_2 \dots$ Zufallsvariablen, dann sind äquivalent:

- (i) $X_k \xrightarrow{v} X$;
- (ii) $\lim_{k \rightarrow \infty} \mathbb{E}(f(X_k)) = E(f(X))$ für $f : \mathbb{R}^n \rightarrow \mathbb{R}$ stetig und beschränkt.

Beweis. Siehe [2] Satz 5.1. □

Definition 2.34. Für reelle Verteilungsfunktionen F und G heißt

$$d(F, G)_{levy} := \inf\{\varepsilon > 0 \mid G(x - \varepsilon) - \varepsilon \leq F(x) \leq G(x + \varepsilon) + \varepsilon, \forall x \in \mathbb{R}\}$$

die *Lévy-Metrik*. Für zwei Zufallsvariablen X und Y ist die Lévy-Metrik definiert durch

$$d(X, Y)_{levy} := d(F_X, F_Y)_{levy}.$$

Satz 2.35. Für eine Folge von reellen Verteilungsfunktionen $(F_d)_{d \in \mathbb{N}}$ und eine reelle Verteilungsfunktion F gilt:

$$F_d \xrightarrow{v} F \Leftrightarrow d(F_d, F)_{levy} \rightarrow 0.$$

Definition 2.36. Sei $X = (X_t)_{t \in T}$ ein stochastischer Prozess und T eine beliebige Indexmenge. X ist ein *Gaußscher Prozess*, falls $(X_{t_1}, \dots, X_{t_n})$ normalverteilt ist für alle $t_i \in T$, $t_1 < \dots < t_n$ und $n \in \mathbb{N}$.

Bemerkung Mit Satz 1.27 gilt, dass die charakteristische Funktion der endlichen Verteilung $(X_{t_1}, \dots, X_{t_n})$ den Gaußschen Prozess eindeutig bestimmt. Für jede endliche Indexmenge $t_1, \dots, t_n, t_i \in T$ ist die charakteristische Funktion gegeben durch

$$\varphi_{X_{t_1} \dots X_{t_n}}(\omega_1, \dots, \omega_n) = \exp\left(i \sum_{k=1}^n \mu_X(t_k) \omega_k - \frac{1}{2} \sum_{l=1}^n \sum_{k=1}^n Cov_X(t_l, t_k) \omega_l \omega_k\right),$$

wobei $\mu_X(t) = \mathbb{E}(X(t))$ und $Cov_X(t, s) = \mathbb{E}(X_t \overline{X_s})$. Ein Gaußscher Prozess ist also durch seine Erwartungswert- und Kovarianzfunktion eindeutig bestimmt.

Satz 2.37 (Multivariater Zentraler Grenzwertsatz). Seien $(X_t)_{t \in \mathbb{N}}$ unabhängige identisch verteilte Zufallsvektoren mit $\mathbb{E}(|X_1|^2) < \infty$ und $\mathbb{E}(X_1) = \mu$ und $Cov(X_1) = \Sigma$. Dann gilt

$$\frac{1}{\sqrt{d}} \sum_{t=1}^d (X_t - \mu) \xrightarrow{v} N(0, \Sigma), \quad d \rightarrow \infty.$$

Beweis. Siehe [3] Theorem 29.5. □

Die für den Unabhängigkeitstest genutzte Teststatistik lautet $dDCov_d^2/T_2$, wobei $T_2 = \bar{a} \cdot \bar{b}$ wie in Definition 2.11 definiert ist.

Satz 2.38. Seien X und Y unabhängige Zufallsvektoren mit $\dim(X)=n$, $\dim(Y)=m$ und $\mathbb{E}(\|X\|_2) + \mathbb{E}(\|Y\|_2) < \infty$ und $\zeta(\cdot)$ ein komplexer Gaußscher Prozess mit $\mathbb{E}(\zeta) = 0$ und Kovarianzfunktion

$$C_\zeta(u, u_0) = (\varphi_X(t - t_0) - \varphi_X(t) \overline{\varphi_X(t_0)}) (\varphi_Y(s - s_0) - \varphi_Y(s) \overline{\varphi_Y(s_0)}).$$

mit $u = (t, s), u_0 = (t_0, s_0) \in \mathbb{R}^n \times \mathbb{R}^m$. Dann gilt

$$\lim_{d \rightarrow \infty} dDCov_d^2 \stackrel{v}{=} \|\zeta(t, s)\|_\omega^2.$$

2.3. Unabhängigkeitstest

Beweis. Wir definieren den Prozess

$$\zeta_d(u) = \zeta_d(t, s) = \sqrt{d}\xi_d(t, s) = \sqrt{d}(\varphi_{X,Y}^d(t, s) - \varphi_X^d(t)\varphi_Y^d(s)).$$

Dann gilt, da X und Y unabhängig sind $\mathbb{E}(\zeta_d(u)) = 0$, $\mathbb{E}(\zeta_d(u)\overline{\zeta_d(u_0)}) = \frac{d-1}{d}C_\zeta(u, u_0)$, denn mit Satz 1.30

$$\begin{aligned}\mathbb{E}(\zeta_d(u)) &= \sqrt{d}(\mathbb{E}(\varphi_{X,Y}^d(t, s)) - \mathbb{E}(\varphi_X^d(t)\varphi_Y^d(s))) = \sqrt{d}(\varphi_{X,Y}(t, s) - \varphi_X(t)\varphi_Y(s)) \\ &= \sqrt{d}(\varphi_X(t)\varphi_Y(s) - \varphi_X(t)\varphi_Y(s)) = 0\end{aligned}$$

und

$$\begin{aligned}\mathbb{E}(\zeta_d(u)\overline{\zeta_d(u_0)}) &= d(\mathbb{E}(\varphi_{X,Y}^d(t, s)\varphi_{X,Y}^d(-t_0, -s_0)) - \mathbb{E}(\varphi_{X,Y}^d(t, s)\varphi_X^d(-t_0)\varphi_Y^d(-s_0)) \\ &\quad - \mathbb{E}(\varphi_X^d(t)\varphi_Y^d(s)\varphi_{X,Y}^d(-t_0, -s_0)) + \mathbb{E}(\varphi_X^d(t)\varphi_Y^d(s)\varphi_X^d(-t_0)\varphi_Y^d(-s_0))) \\ &= \frac{d-1}{d}(\varphi_X(t-t_0)\varphi_Y(s-s_0) - \varphi_X(t-t_0)\varphi_Y(s)\overline{\varphi_Y(s_0)}) \\ &\quad - \varphi_Y(s-s_0)\varphi_X(t)\overline{\varphi_X(t_0)} + \varphi_X(t)\overline{\varphi_X(t_0)}\varphi_Y(s)\overline{\varphi_Y(s_0)}) \\ &= \frac{d-1}{d}(\varphi_X(t-t_0) - \varphi_X(t)\overline{\varphi_X(t_0)})(\varphi_Y(s-s_0) - \varphi_Y(s)\overline{\varphi_Y(s_0)}) \\ &= \frac{d-1}{d}C_\zeta(u, u_0).\end{aligned}$$

Somit

$$\mathbb{E}(|\zeta_d(u)|^2) = \mathbb{E}(\zeta_d(u)\overline{\zeta_d(u)}) = \frac{d-1}{d}C_\zeta(u, u) = \frac{d-1}{d}(1 - |\varphi_X(t)|^2)(1 - |\varphi_Y(s)|^2). \quad (2.8)$$

Sei $\varepsilon > 0$, $D(\delta) := \{(t, s) | \delta \leq \|t\|_2 \leq 1/\delta, \delta \leq \|s\|_2 \leq 1/\delta\}$ und $(D_k)_{k=1}^N$ eine Partition von $D(\delta)$ mit $N = N(\varepsilon)$ messbaren Mengen mit einem Durchmesser von höchstens ε . Für $\delta > 0$ sei

$$Q_d(\delta) := \sum_{k=1}^N \int_{D_k} |\zeta_d(u)|^2 d\omega.$$

Und sei Q eine Zufallsvariable mit $Q(\delta) := \sum_{k=1}^N \int_{D_k} |\zeta(u)|^2 d\omega$. Wir zeigen nun, dass für $(Q_d)_{d \in \mathbb{N}}$ folgendes gilt:

- (i) $Q_d(\delta) \xrightarrow{v} Q$;
- (ii) $\mathbb{E}(|Q_d(\delta) - \int |\zeta_d|^2|) \leq \delta$;
- (iii) $\mathbb{E}(|Q(\delta) - \int |\zeta|^2|) \leq \delta$.

2.3. Unabhängigkeitstest

Mittels (i)-(iii), Satz 2.35 und

$$d(\|\zeta\|_\omega^2, \|\zeta_d\|_\omega^2)_{levy} \leq d(\|\zeta\|_\omega^2, Q(\delta))_{levy} + d(Q(\delta), Q_d(\delta))_{levy} + d(Q_d(\delta), \|\zeta_d\|_\omega^2)_{levy}$$

folgt nun $dDCov_d^2 = \|\zeta_d\|_\omega \xrightarrow{v} \|\zeta\|_\omega$.

Also müssen wir (i)-(iii) verifizieren.

(i) folgt aus dem Multivariaten Zentralen Grenzwertsatz und dem obigen.

(ii) Sei $M > 0$ fest und

$$\beta(\varepsilon) = \sup_{u, u_0} \mathbb{E}(|\zeta_d(u)|^2 - |\zeta_d(u_0)|^2|),$$

wobei das Supremum über alle $u = (t, s)$ und $u_0 = (t_0, s_0)$ genommen wird, sodass $\max\{|t|, |t_0|, |s|, |s_0|\} < M$ und $|t - t_0| + |s - s_0| < \varepsilon^2$. Somit ist $\lim_{\varepsilon \rightarrow 0} \beta(\varepsilon) = 0$ für jedes feste $M > 0$. Und für jedes feste $\delta > 0$ gilt

$$\mathbb{E} \left| \int_{D(\delta)} |\zeta_d(u)|^2 d\omega - Q_d(\delta) \right| \leq \beta(\varepsilon) \int_{D(\delta)} |\zeta_d(u)|^2 d\omega \xrightarrow{\varepsilon \rightarrow 0} 0. \quad (1)$$

Außerdem gilt

$$\begin{aligned} & \left| \int_{D(\delta)} |\zeta_d(u)|^2 d\omega - \int_{\mathbb{R}^{n+m}} |\zeta_d(u)|^2 d\omega \right| \\ & \leq \int_{\|t\|_2 < \delta} |\zeta_d(u)|^2 d\omega + \int_{\|t\|_2 > 1/\delta} |\zeta_d(u)|^2 d\omega \\ & \quad + \int_{\|s\|_2 < \delta} |\zeta_d(u)|^2 d\omega + \int_{\|s\|_2 > 1/\delta} |\zeta_d(u)|^2 d\omega. \end{aligned} \quad (2)$$

Aus dem Beweis von Satz 2.25 folgt

$$|\zeta_d(u)|^2 \leq \frac{4}{d} \sum_{k=1}^d |u_k|^2 \sum_{k=1}^d |v_k|^2,$$

$$\int_{\mathbb{R}^m} \frac{|v_k|^2 ds}{c_m \|s\|_2^{1+m}} \leq 2\mathbb{E}_Y(\|Y_k - Y\|_2) \quad \text{und} \quad \int_{\|t\|_2 < \delta} \frac{|u_k|^2 dt}{c_n \|t\|_2^{1+n}} \leq 2\mathbb{E}_X(\|X_k - X\|_2) G(\|X_k - X\|_2 \delta)$$

mit $u_k := e^{it^T X_k} - \varphi_X(t)$, $v_k := e^{is^T Y_k} - \varphi_Y(s)$ und $G(y) := \int_{\|z\|_2 < y} \frac{1 - \cos(z_1)}{\|z\|_2^{1+n}} dz$.

Also insgesamt

$$\int_{\|t\|_2 < \delta} |\zeta_d(u)|^2 d\omega \leq \frac{4}{d} \sum_{k=1}^d 2\mathbb{E}_Y(\|Y_k - Y\|_2) 2 \sum_{k=1}^d \mathbb{E}_X(\|X_k - X\|_2) G(\|X_k - X\|_2 \delta).$$

2.3. Unabhängigkeitstest

Eine weitere Ungleichung aus dem Beweis von Satz 2.25

$$\int_{\|t\|_2 > 1/\delta} |\zeta_d(u)| d\omega \leq 32\delta^{n+1}/c_n \sum_{k=1}^d \mathbb{E}_Y(\|Y_k - Y\|_2).$$

Es folgt

$$\begin{aligned} & \mathbb{E} \left(\int_{\|t\|_2 < \delta} |\zeta_d(u)|^2 d\omega + \int_{\|t\|_2 > 1/\delta} |\zeta_d(u)|^2 d\omega \right) \\ & \leq \mathbb{E} \left(\frac{4}{d} \sum_{k=1}^d 2\mathbb{E}_Y(\|Y_k - Y\|_2) 2 \sum_{k=1}^d \mathbb{E}_X(\|X_k - X\|_2) G(\|X_k - X\|_2 \delta) \right) \\ & \quad + \mathbb{E} \left(32\delta^{n+1}/c_n \sum_{k=1}^d \mathbb{E}_Y(\|Y_k - Y\|_2) \right) \\ & = 16d\mathbb{E}(\|X_1 - X_2\|_2) G(\|X_1 - X_2\|_2 \delta) \mathbb{E}(\|Y_1 - Y_2\|_2) + 32d\delta^{n+1}/c_n \mathbb{E}(\|Y_1 - Y_2\|_2) \xrightarrow{\delta \rightarrow 0} 0 \end{aligned}$$

und analog

$$\mathbb{E} \left(\int_{\|s\|_2 < \delta} |\zeta_d(u)|^2 d\omega + \int_{\|s\|_2 > 1/\delta} |\zeta_d(u)|^2 d\omega \right) \xrightarrow{\delta \rightarrow 0} 0.$$

Mit (1) und (2) folgt (ii).

(iii) Es können analoge Ungleichungen für Q und ζ wie für Q_d und ζ_d in(ii) aufgestellt werden, sodass (iii) gilt. □

Bemerkung Aus dem obigen Satz folgt, dass $DCov_d$ unter Unabhängigkeit von X und Y nicht Erwartungstreu ist: Wir haben $DCov_d = \left\| 1/\sqrt{d}\zeta_d \right\|_\omega$ und wegen der Unabhängigkeit ist $DCov(X, Y) = 0$. Somit

$$\begin{aligned} \mathbb{E}(DCov_d^2) &= \mathbb{E} \left(\left\| 1/\sqrt{d}\zeta_d \right\|_\omega^2 \right) = \mathbb{E} \left(\frac{1}{d} \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \frac{|\zeta_d|^2 dt ds}{c_n c_m \|t\|_2^{1+n} \|s\|_2^{1+m}} \right) \\ &= \frac{1}{d} \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \frac{\mathbb{E}(|\zeta_d|^2) dt ds}{c_n c_m \|t\|_2^{1+n} \|s\|_2^{1+m}} \\ &= \frac{d-1}{d^2} \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \frac{(1 - |\varphi_X(t)|^2)(1 - |\varphi_Y(s)|^2) dt ds}{c_n c_m \|t\|_2^{1+n} \|s\|_2^{1+m}} \end{aligned}$$

$$\begin{aligned}
 &= \frac{d-1}{d^2} \mathbb{E}(\|X - X'\|_2) \mathbb{E}(\|Y - Y'\|_2) \\
 &= \frac{d-1}{d^2} (\mathbb{E}(\|X - X'\|_2) \mathbb{E}(\|Y - Y'\|_2) + DCov(X, Y)).
 \end{aligned}$$

Korollar 2.39. Seien X und Y Zufallsvektoren mit $\mathbb{E}(\|X\|_2) + \mathbb{E}(\|Y\|_2) < \infty$, dann gilt:

- (i) Sind X und Y unabhängig und $Q \stackrel{v}{=} \sum_{j=1}^{\infty} \lambda_j Z_j^2$, wobei Z_j unabhängige standardnormalverteilte Zufallsvariablen sind und λ_j positive Konstanten sind, die nur von der Verteilung (X, Y) abhängen. So ist

$$\lim_{d \rightarrow \infty} dDCov_d^2/T_2 \stackrel{v}{=} Q \text{ und } \mathbb{E}(Q) = 1.$$

- (ii) Sind X und Y abhängig, dann gilt $\lim_{d \rightarrow \infty} dDCov_d^2/T_2 \stackrel{P}{=} \infty$.

Beweis. (i) Da X und Y unabhängig sind, folgt $\mathbb{E}(\zeta_d) = \mathbb{E}(\zeta) = 0$, wobei ζ_d, ζ wie in Satz 2.38 definiert sind. Nach [8], Kapitel 1, Abschnitt 2, hat die quadratische Norm eines Gaußschen Prozesses $\|\zeta\|_{\omega}$ mit $\mathbb{E}(\|\zeta\|_{\omega}) = 0$ die Darstellung

$$\|\zeta\|_{\omega}^2 \stackrel{v}{=} \sum_{j=1}^{\infty} \lambda_j Z_j, \quad (1)$$

wobei Z_j unabhängige standardnormalverteilte Zufallsvariablen sind und λ_j positive Konstanten sind, die nur von der Verteilung von (X, Y) abhängen. Nach Satz 2.38 und der Unabhängigkeit von X und Y gilt, dass $dDCov_d^2 \stackrel{v}{=} \|\zeta\|_{\omega}^2$. Weiter gilt mit (2.6) und (2.8)

$$\begin{aligned}
 \mathbb{E}(\|\zeta\|_{\omega}^2) &= \int_{\mathbb{R}^{n+m}} R(u, u) d\omega = \int_{\mathbb{R}^{n+m}} (1 - |\varphi_X(t)|^2)(1 - |\varphi_Y(s)|^2) d\omega \\
 &= \mathbb{E}(\|X - X'\|_2 \|Y - Y'\|_2),
 \end{aligned} \quad (2)$$

wobei X' eine unabhängige identisch verteilte Kopie von X ist und Y' analog. Mit dem Starken Gesetz der großen Zahlen für V-Statistiken gilt

$$\lim_{d \rightarrow \infty} T_2 = \mathbb{E}(\|X - X'\|_2 \|Y - Y'\|_2) \text{ f.s.}$$

Mit Satz 2.38 und (1) gilt

$$\begin{aligned}
 dDCov_d^2/T_2 &\stackrel{v}{=} \|\zeta\|_{\omega}^2 / (\mathbb{E}(\|X - X'\|_2 \|Y - Y'\|_2)) \\
 &\stackrel{v}{=} \sum_{j=1}^{\infty} \underbrace{(\lambda_j / (\mathbb{E}(\|X - X'\|_2 \|Y - Y'\|_2)))}_{=: \lambda'_j} Z_j = \sum_{j=1}^{\infty} \lambda'_j Z_j =: Q.
 \end{aligned}$$

Damit hat Q die gewünschte Darstellung, da $\lambda'_j \geq 0 \forall j$ und nur von der Verteilung von (X, Y) abhängen und Z_j unabhängige standardnormal-verteilte Zufallsvariablen sind. Außerdem gilt mit Satz 2.33 und (2)

$$\begin{aligned} \mathbb{E}(Q) &= \mathbb{E} \left(\frac{\|\zeta\|_\omega}{\mathbb{E}(\|X - X'\|_2 \|Y - Y'\|_2)} \right) = \frac{1}{\mathbb{E}(\|X - X'\|_2 \|Y - Y'\|_2)} \mathbb{E}(\|\zeta\|_\omega) \\ &= \frac{1}{\mathbb{E}(\|X - X'\|_2 \|Y - Y'\|_2)} \mathbb{E}(\|X - X'\|_2 \|Y - Y'\|_2) = 1. \end{aligned}$$

(ii) Da X und Y abhängig sind, gilt $DCov(X, Y) > 0$ und nach Satz 2.25 ist $\lim_{d \rightarrow \infty} DCov_d \stackrel{f.s.}{=} DCov$ und somit gilt $dDCov_d^2 \xrightarrow{P} \infty$. Nach dem Starken Gesetz der großen Zahlen für V-Statistiken ist $\lim_{d \rightarrow \infty} T_2 \stackrel{f.s.}{=} \mathbb{E}(\|X - X'\|_2 \|Y - Y'\|_2) = const.$ Also insgesamt $dDCov_d^2/T_2 \xrightarrow{P} \infty$. □

Anmerkung Korollar 2.39 zeigt, dass die Teststatistik $dDCov_d^2/T_2$ konsistent ist.

Mit dem folgenden Satz können wir einen kritischen Wert für die Prüfgröße angeben. Außerdem zeigen wir, dass dieser kritische Wert eine scharfe Grenze ist.

Satz 2.40. Sei $T(X, Y, \alpha, d)$ der Test der die Unabhängigkeit von X und Y ablehnt, falls

$$\frac{dDCov_d^2(X^d, Y^d)}{T_2} > \chi_{1-\alpha}^2(1),$$

wobei $\chi_{1-\alpha}^2(1)$ das $1 - \alpha$ Quantil der Chi-Quadrat-Verteilung ist mit Freiheitsgrad 1 und $\alpha(X, Y, d)$ das Signifikanzlevel von $T(X, Y, \alpha, d)$ ist. Gilt $\mathbb{E}(\|X\|_2) + \mathbb{E}(\|Y\|_2) < \infty$, dann ist für alle $0 < \alpha \leq 0.215$

(i) $\lim_{d \rightarrow \infty} \alpha(X, Y, d) \leq \alpha;$

(ii) $\sup_{X, Y} \left\{ \lim_{d \rightarrow \infty} \alpha(X, Y, d) \mid DCov(X, Y) = 0 \right\} = \alpha.$

Beweis. (i) Dies ist ein Spezialfall von den Ergebnissen aus [12], Seite 186 ff. Es gilt für eine quadratische Form von standardnormal-verteilten Zufallsvariablen Q mit $\mathbb{E}(Q) = 1$, dass

$$\lim_{d \rightarrow \infty} P \left(\frac{dDCov_d^2(X^d, Y^d)}{T_2} \geq \chi_{1-\alpha}^2(1) \right) = P(Q \geq \chi_{1-\alpha}^2(1)) \leq \alpha$$

für $0 < \alpha \leq 0.215$.

(ii) Sind X und Y Bernoulli-verteilte Zufallsvariablen, dann gilt $DCor_d = |\tilde{\rho}|$.

2.3. Unabhängigkeitstest

Wegen der Unabhängigkeit von X und Y und mittels dem Zentralen Grenzwertsatz ist $\sqrt{d}|\tilde{\rho}| \xrightarrow{v} N(0, 1)$ also auch $d\tilde{\rho}^2 \xrightarrow{v} N(0, 1)^2$ und somit $dDCor_d^2 \xrightarrow{v} N(0, 1)^2$. Daraus folgt für die quadratische Form Q , dass $Q = Z_1^2$ mit $Z \sim N(0, 1)$. α wird erreicht, denn

$$P(Q \geq \chi_{1-\alpha}^2(1)) = P(Z_1^2 \geq \chi_{1-\alpha}^2(1)) = \alpha.$$

□

Kapitel 3

Brownsche Distanz Korrelation

In diesem Kapitel definieren wir die Brownsche Distanz Kovarianz und die Brownsche Distanz Korrelation. Diese Funktionen sind wie die Distanz Kovarianz\Korrelation Maßzahlen zur Messung des Zusammenhangs zwischen zwei Zufallsvariablen. Die Brownsche Distanz Kovarianz kann analog zur Distanz Kovarianz den Zusammenhang zwischen Zufallsvariablen beliebiger Dimensionen messen und die Unabhängigkeit von Zufallsvariablen eindeutig charakterisieren. Zum Abschluss des Kapitels zeigen wir, dass die Brownsche Distanz Kovarianz mit der Distanz Kovarianz übereinstimmt.

3.1 Bedingter Erwartungswert und Brownsche Bewegung

Zunächst beschäftigen wir uns mit der Brownschen Bewegung und den bedingten Erwartungswerten. Diese sind wesentliche Voraussetzungen für den Umgang mit der Brownschen Distanz Korrelation.

Definition 3.1 (Bedingter Erwartungswert). Sei X eine Zufallsvariable und $\mathfrak{B} \subset \mathfrak{A}$ eine σ -Algebra über Ω , dann heißt $Z := \mathbb{E}(X|\mathfrak{B})$ *bedingter Erwartungswert* von X unter \mathfrak{B} , wenn Z eine \mathfrak{B} -messbare Zufallsvariable ist mit $\int_B Z dP = \int_B X dP \quad \forall B \in \mathfrak{B}$.

Eine Zufallsvariable X mit $X \geq 0$ oder X eigentlich integrierbar heißt *zulässig*. Für ein zulässiges X existiert stets der bedingte Erwartungswert $\mathbb{E}(X|\mathfrak{B})$ für jede σ -Algebra $\mathfrak{B} \subset \mathfrak{A}$ über Ω .

Bemerkung Für $\mathfrak{B} = \{\emptyset, \Omega\}$ gilt $\mathbb{E}(X|\mathfrak{B}) = \mathbb{E}(X)$ und für $\mathfrak{B} = \mathfrak{A}$ gilt $\mathbb{E}(X|\mathfrak{B}) = X$.

Satz 3.2. Seien X und Y zulässig, dann gilt für jede σ -Algebra $\mathfrak{B} \subset \mathfrak{A}$ über Ω :

- (i) $\mathbb{E}(\mathbb{E}(X|\mathfrak{B})) = \mathbb{E}(X)$;
- (ii) $X \stackrel{P-f.s.}{=} Y \Rightarrow \mathbb{E}(X|\mathfrak{B}) \stackrel{P-f.s.}{=} \mathbb{E}(Y|\mathfrak{B})$;
- (iii) $\mathbb{E}(\alpha X|\mathfrak{B}) \stackrel{P-f.s.}{=} \alpha \mathbb{E}(X|\mathfrak{B})$;
- (iv) $\mathbb{E}(X + Y|\mathfrak{B}) \stackrel{P-f.s.}{=} \mathbb{E}(X|\mathfrak{B}) + \mathbb{E}(Y|\mathfrak{B})$;
- (v) $X \stackrel{P-f.s.}{\leq} Y \Rightarrow \mathbb{E}(X|\mathfrak{B}) \stackrel{P-f.s.}{\leq} \mathbb{E}(Y|\mathfrak{B})$;
- (vi) X \mathfrak{B} -messbar $\Rightarrow \mathbb{E}(XY|\mathfrak{B}) \stackrel{P-f.s.}{=} X \cdot \mathbb{E}(Y|\mathfrak{B})$;
- (vii) X \mathfrak{B} -messbar $\Rightarrow \mathbb{E}(X|\mathfrak{B}) \stackrel{P-f.s.}{=} X$.
- (viii) $\mathbb{E}(Y \cdot \mathbb{E}(X|\mathfrak{B})|\mathfrak{B}) \stackrel{P-f.s.}{=} \mathbb{E}(X|\mathfrak{B}) \cdot \mathbb{E}(Y|\mathfrak{B})$;

Beweis. (i)-(iv) folgen direkt aus der Definition.

(v) Sei ohne Einschränkung $X \leq Y$, dann existiert ein zulässiges Z mit $X + Z = Y$ und somit

$$\mathbb{E}(Y|\mathfrak{B}) = \mathbb{E}(X + Z|\mathfrak{B}) = \mathbb{E}(X|\mathfrak{B}) + \underbrace{\mathbb{E}(Z|\mathfrak{B})}_{\geq 0} \geq \mathbb{E}(X|\mathfrak{B}).$$

(vi) 1.Fall: $X \geq 0$ und $Y \geq 0$

Für $d \in \mathbb{N}$ definieren wir $X_d := \frac{1}{2^d} \lfloor 2^d X \rfloor$. Dann gilt $X_d \nearrow X$ und $X_d \mathbb{E}(Y|\mathfrak{B}) \nearrow X \mathbb{E}(Y|\mathfrak{B})$. Mit dem Satz von der monotonen Konvergenz ist

$$\lim_{d \rightarrow \infty} \mathbb{E}(\mathbb{1}_B X_d \mathbb{E}(Y|\mathfrak{B})) = \mathbb{E}(\lim_{d \rightarrow \infty} \mathbb{1}_B X_d \mathbb{E}(Y|\mathfrak{B})) = \mathbb{E}(\mathbb{1}_B X \mathbb{E}(Y|\mathfrak{B})) \quad \forall B \in \mathfrak{B}.$$

Weiter gilt

$$\begin{aligned} \mathbb{E}(\mathbb{1}_B X_d \mathbb{E}(Y|\mathfrak{B})) &= \sum_{k=1}^{\infty} \mathbb{E}(\mathbb{1}_B \mathbb{1}_{X_d=k2^{-d}} k 2^{-d} \mathbb{E}(Y|\mathfrak{B})) \\ &= \sum_{k=1}^{\infty} \mathbb{E}(\mathbb{1}_B \mathbb{1}_{X_d=k2^{-d}} k 2^{-d} Y) \\ &= \mathbb{E}(\mathbb{1}_B X_d Y) \xrightarrow{d \rightarrow \infty} \mathbb{E}(\mathbb{1}_B X Y). \end{aligned}$$

Insgesamt also $\mathbb{E}(\mathbb{1}_B X \mathbb{E}(Y|\mathfrak{B})) = \mathbb{E}(\mathbb{1}_B X Y) \Leftrightarrow \int_B X \mathbb{E}(Y|\mathfrak{B}) dP = \int_B X Y dP$.

2.Fall: X und Y sind beliebig aber Integrierbar.

Folgt analog zu Fall 1 durch algebraische Induktion.

(vii) folgt aus (vi) für $X \equiv 1$ und (viii) folgt ebenso aus (vi).

□

Definition 3.3. Seien $\mathfrak{B}_1, \mathfrak{B}_2 \subset \mathfrak{A}$ σ -Algebren, dann heißen \mathfrak{B}_1 und \mathfrak{B}_2 *unabhängig*, falls B_1 und B_2 unabhängig sind $\forall B_i \in \mathfrak{B}_i$ $i=1,2$. Sei X eine Zufallsvariable, dann heißen \mathfrak{B}_1 und X unabhängig, falls $\sigma(X) = X^{-1}(\mathfrak{B}(\mathbb{R}))$ und \mathfrak{B}_1 unabhängig sind.

Anmerkung Insbesondere sind zwei Zufallsvariablen X und Y genau dann unabhängig, wenn $\sigma(X)$ und $\sigma(Y)$ unabhängig sind. Für den bedingten Erwartungswert wird auch oftmals die Konvention $\mathbb{E}(X|Y) := \mathbb{E}(X|\sigma(Y))$ verwendet.

Satz 3.4. Sei X zulässig und $\mathfrak{B} \subset \mathfrak{A}$ eine σ -Algebra unabhängig von X , dann gilt $\mathbb{E}(X|\mathfrak{B}) = \mathbb{E}(X)$.

Beweis. $\mathbb{E}(X)$ ist messbar bezüglich \mathfrak{B} . Für $B \in \mathfrak{B}$ sind $\mathbb{1}_B$ und X unabhängig, daraus folgt wenn X integrierbar ist

$$\int_B X dP = \int_{\Omega} \mathbb{1}_B X dP = \int_{\Omega} \mathbb{1}_B dP \int_{\Omega} X dP = \int_{\Omega} \mathbb{1}_B \mathbb{E}(X) dP = \int_B \mathbb{E}(X) dP.$$

Für positives X folgt die Behauptung analog durch algebraische Induktion.

□

Definition 3.5 (Bedingte Verteilung). Sei X eine Zufallsvariable und $\mathfrak{B} \subset \mathfrak{A}$ eine σ -Algebra, dann heißt $P_X(\cdot|\mathfrak{B})(\cdot) : \Omega \times \mathfrak{B} \rightarrow \mathbb{R} \cup \{-\infty, \infty\}$ die *bedingte Verteilung* von X unter \mathfrak{B} , falls

- (i) $P_X(\cdot|\mathfrak{B})(\omega)$ ist ein Wahrscheinlichkeitsmaß auf $\mathfrak{B}' \forall \omega \in \Omega$;
- (ii) $P_X(B'|\mathfrak{B})(\cdot)$ ist \mathfrak{B} -messbar $\forall B' \in \mathfrak{B}'$;
- (iii) $P_X(B'|\mathfrak{B}) \stackrel{P-f.s.}{=} P(X^{-1}(B')|\mathfrak{B}) \quad \forall B' \in \mathfrak{B}'$.

Bemerkung Seien die Voraussetzungen wie in der Definition oben und zudem X eine reelle Zufallsvariable, dann existiert die bedingte Verteilung $P_X(B'|\mathfrak{B})$.

Satz 3.6. Sei X eine P -integrierbare Zufallsvariable zum Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ und $\mathfrak{B} \subset \mathfrak{A}$ eine σ -Algebra über Ω . Sei $f : \mathbb{R} \rightarrow \mathbb{R}$ messbar mit $\mathbb{E}(f(X)) \leq \infty$. Dann gilt

- (i) $\mathbb{E}(f(X)|\mathfrak{B}) \stackrel{P}{=} \int_{\mathbb{R}} f(x) P_X(dx|\mathfrak{B})$;

(ii) $\mathbb{E}(X|\mathfrak{B}) \stackrel{P}{=} \int_{\mathbb{R}} x P_X(dx|\mathfrak{B});$

(iii) Ist f konkav, so gilt $f(\mathbb{E}(X)) \leq \mathbb{E}(f(X))$ und insbesondere $f(\mathbb{E}(X|\mathfrak{B})) \stackrel{P}{\leq} \mathbb{E}(f(X)|\mathfrak{B})$.
(Jensensche Ungleichung)

Beweis. (i) Für $f = \mathbb{1}_B$ $B \in \mathfrak{B}(\mathbb{R})$ gilt die Aussage unmittelbar. Durch algebraische Induktion folgt die Aussage für f messbar.

(ii) folgt aus (i) durch $f = id_{\mathbb{R}}$.

(iii) siehe [7], Satz 8.19, Seite 172.

□

Definition 3.7 (Filtrierung & Adaption). Eine Familie von σ -Algebren $(\mathfrak{F}_t)_{t \in T}$ heißt *Filtrierung*, falls $\forall s, t \in T$ mit $s < t$ $\mathfrak{F}_s \subseteq \mathfrak{F}_t$ gilt. Ein stochastischer Prozess $(X_t)_{t \in T}$ heißt *adaptiert* bezüglich der Filtrierung $(\mathfrak{F}_t)_{t \in T}$, wenn X_t \mathfrak{F}_t -messbar ist $\forall t \in T$.

Beispiel 3.8. $(X_t)_{t \in T}$ sei ein stochastischer Prozess, dann ist $(\mathfrak{A}_t)_{t \in T}$ mit $\mathfrak{A}_t := \sigma(X_s | s \leq t)$ eine Filtration und $(X_t)_{t \in T}$ ist an $(\mathfrak{A}_t)_{t \in T}$ adaptiert. Diese Filtrierung nennt man auch die *kanonische Filtrierung*.

Definition 3.9 (Brownsche Bewegung \ Wiener-Prozess). Ein adaptierter Stochastischer Prozess $(W_t)_{t \in T}$ bezüglich der Filtrierung $(\mathfrak{F}_t)_{t \in T}$, heißt *Brownsche Bewegung* oder *Wiener-Prozess*, wenn gilt:

(i) $W_0 = 0;$

(ii) $(W_t)_{t \in T}$ hat unabhängige Zuwächse;

(iii) $W_t - W_s \sim N(0, t - s) \forall 0 \leq s < t;$

(iv) Die Pfade $t \rightarrow W_t(\omega)$, $t \in T$ sind stetig $\forall \omega$.

Bemerkung Durch die Unabhängigkeit und Normalverteilung der Zuwächse ist insbesondere jede Brownsche Bewegung $(W_t)_{t \in T}$ ein Gaußscher Prozess mit Erwartungswertfunktion $\mathbb{E}(W_t) = 0 \forall t \in T$ und Kovarianzfunktion $Cov(W_s, W_t) = \min(s, t) \forall s, t \in T$.

Definition 3.10. Eine Brownsche Bewegung $(W_t)_{t \in T}$ heißt *zweiseitige Brownsche Bewegung*, falls gilt

$$W_t := \begin{cases} W_t^+ & \text{falls } t \geq 0, \\ W_{-t}^- & \text{falls } t < 0. \end{cases}$$

Für diese Arbeit sei nun $(W_t)_{t \in T}$ eine zweiseitige Brownsche Bewegung mit

$$\mathbb{E}(W_t) = 0 \quad \forall t \in T \quad \text{und} \quad \text{Cov}(W_s, W_t) = |s| + |t| - |s - t| = 2\min(s, t) \quad \forall s, t \in T. \quad (3.1)$$

Und für die Indexmenge T wird aus Einfachheitsgründen die Menge der natürlichen Zahlen benutzt.

Satz 3.11. Für $(W_t)_{t \in T}$ gilt

- (i) $\mathbb{E}(W(t)^2) = 2|t|$;
- (ii) $\mathbb{E}(W(t)^4) = 3\mathbb{E}(W(t)^2)^2 = 12|t|^2$.

Beweis. (i) Da $\mathbb{E}(W_t) = 0$ ist $\text{Cov}(W_s, W_t) = \mathbb{E}(W_s W_t)$. Also gilt

$$\mathbb{E}(W_t^2) = \text{Cov}(W_t, W_t) = |t| + |t| - |t - t| = 2|t|.$$

(ii) Für $X \sim N(0, \sigma)$ gilt $\mathbb{E}(X^4) = 3\sigma^2$. Es folgt, da $W(t)$ normalverteilt ist mit Erwartungswert 0, dass

$$\mathbb{E}((W_t)^4) = 3\mathbb{E}(W(t)^2)^2 \stackrel{(i)}{=} 12|t|^2.$$

□

Definition 3.12 (Kovarianzquadrat). Seien X und Y Zufallsvektoren mit $\mathbb{E}(\|X\|_2^2) + \mathbb{E}(\|Y\|_2^2) < \infty$ und X', Y' jeweils unabhängig identisch-verteilte Kopien, dann ist das *Kovarianzquadrat* durch

$$\text{Cov}(X, Y)^2 = \mathbb{E}^2((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}((X - \mathbb{E}(X))(X' - \mathbb{E}(X'))(Y - \mathbb{E}(Y))(Y' - \mathbb{E}(Y')))$$

gegeben.

Definition 3.13 (Die U -zentrierte Version von X). Sei X eine Zufallsvariable und $U = (U_t)_{t \in \mathbb{N}}$ ein von X unabhängiger stochastischer Prozess, dann heißt

$$X_U := U(X) - \int_{-\infty}^{\infty} U(t) dF_X(t) = U(X) - \mathbb{E}(U(X)|U)$$

die *U -zentrierte Version von X* , soweit der bedingte Erwartungswert existiert.

Anmerkung Ist $(U_t)_{t \in \mathbb{N}}$ der Identitätsprozess, also $U_t(\omega) = id(\omega) \quad \forall t \in \mathbb{N} \quad \forall \omega \in \Omega$, dann erhält man $X_U = X_{id} = X - \mathbb{E}(X)$. Wir haben in Definition 3.13 genutzt, dass X und U unabhängig sind und mit Satz 3.6 die kompaktere Schreibweise gefolgt.

3.2 Brownsche Distanz Korrelation

Definition 3.14 (Brownsche Distanz Kovarianz). Seien X und Y Zufallsvektoren mit $\mathbb{E}(\|X\|_2^2) + \mathbb{E}(\|Y\|_2^2) < \infty$ und $(W_t)_{t \in \mathbb{N}}, (W'_t)_{t \in \mathbb{N}}$ zweiseitige Brownsche Bewegungen mit den Eigenschaften aus (3.1), dann heißt die Quadratwurzel aus

$$BCov^2(X, Y) := Cov_W^2(X, Y) := \mathbb{E}(X_W X'_W Y_{W'} Y'_{W'})$$

die *Brownsche Distanz Kovarianz*, wobei W und W' nicht von (X, X', Y, Y') abhängen.

Bemerkung Wenn man für $(W_t)_{t \in \mathbb{N}}$ und $(W'_t)_{t \in \mathbb{N}}$ den Identitätsprozess für $BCov$ einsetzt, erhält man $BCov = |Cov|$. Denn

$$\begin{aligned} BCov^2(X, Y) &= Cov_W^2(X, Y) = \mathbb{E}(X_{id} X'_{id} Y_{id} Y'_{id}) \\ &= \mathbb{E}((X - \mathbb{E}(X))(X' - \mathbb{E}(X'))(Y - \mathbb{E}(Y))(Y' - \mathbb{E}(Y'))) \\ &= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))^2 \\ &\Leftrightarrow BCov(X, Y) = |Cov(X, Y)|. \end{aligned}$$

Achtung: Der Identitätsprozess ist keine zweiseitige Brownsche Bewegung, da der Identitätsprozess kein Gaußscher Prozess ist.

Definition 3.15 (Brownsche Distanz Varianz). Sei X ein Zufallsvektor mit $\mathbb{E}(\|X\|_2^2) < \infty$ und $(W_t)_{t \in \mathbb{N}}$ eine zweiseitige Brownsche Bewegung mit den Eigenschaften aus (3.1), dann heißt

$$BVar(X) := Cov_W(X, X) = \mathbb{E}(X_W^2)$$

die *Brownsche Distanz Varianz*, wobei W nicht von X abhängt.

Bemerkung Wenn man für $(W_t)_{t \in \mathbb{N}}$ den Identitätsprozess bei $BVar$ einsetzt, erhält man analog wie oben $BVar(X) = Var(X)$.

Definition 3.16 (Brownsche Distanz Korrelation). Seien X und Y Zufallsvektoren mit $\mathbb{E}(\|X\|_2^2) + \mathbb{E}(\|Y\|_2^2) < \infty$ und $(W_t)_{t \in \mathbb{N}}, (W'_t)_{t \in \mathbb{N}}$ zweiseitige Brownsche Bewegungen mit den Eigenschaften aus (3.1), dann heißt

$$BCor(X, Y) := \begin{cases} \frac{BCov(X, Y)}{\sqrt{BVar(X)BVar(Y)}} & BVar(X)BVar(Y) > 0; \\ 0 & BVar(X)BVar(Y) = 0; \end{cases}$$

die *Brownsche Distanz Korrelation*.

Bemerkung Wenn man für $(W_t)_{t \in \mathbb{N}}$ den Identitätsprozess bei BCor einsetzt, erhält man analog wie oben $BCor(X, Y) = |\rho(X, Y)|$.

Satz 3.17. Seien X und Y Zufallsvektoren mit $\mathbb{E}(\|X\|_2^2) + \mathbb{E}(\|Y\|_2^2) < \infty$, dann existiert BCov und es gilt

$$BCov^2(X, Y) = \mathbb{E}(\|X - X'\|_2 \|Y - Y'\|_2) + \mathbb{E}(\|X - X'\|_2) \mathbb{E}(\|Y - Y'\|_2) \\ - \mathbb{E}(\|X - X'\|_2 \|Y - Y''\|_2) - \mathbb{E}(\|X - X''\|_2 \|Y - Y'\|_2),$$

wobei $(X, Y), (X', Y')$ und (X'', Y'') unabhängig und identisch verteilt sind.

Beweis. Wir betrachten einfachheitshalber den Fall, dass $\dim(X) = \dim(Y) = 1$ ist. Nach Satz 3.11 gilt $\mathbb{E}(W(t)^4) = 3\mathbb{E}(W(t)^2)^2 = 12|t|^2$. Damit erhalten wir

$$\mathbb{E}(W(X)^4) = \mathbb{E}(\mathbb{E}(W(X)^4|X)) = \mathbb{E}(12|X|^2) < \infty.$$

Wegen $(x + y)^4 \leq 2^4(x^4 + y^4)$ und der Jensen Ungleichung gilt

$$\mathbb{E}(X_W)^4 = \mathbb{E}(W(X) - \mathbb{E}(W(X)|W))^4 \leq 2^4(\mathbb{E}(W(X)^4) + \mathbb{E}(\mathbb{E}(W(X)|W))^4) \\ \leq 2^5 \mathbb{E}(W(X)^4) = 2^5 12 \mathbb{E}(|X|^2) < \infty$$

Analog haben $X'_W, Y_{W'}$ und $Y'_{W'}$ alle ein existierendes viertes Moment. Mit der Ungleichung $abcd \leq a^4 + b^4 + c^4 + d^4$ folgt

$$BCov(X, Y) = \mathbb{E}(X_W X'_W Y_{W'} Y'_{W'}) \leq \mathbb{E}((X_W)^4 + (X'_W)^4 + (Y_{W'})^4 + (Y'_{W'})^4) < \infty.$$

Weiter gilt mit Satz 3.2

$$BCov(X, Y)^2 = \mathbb{E}(X_W X'_W Y_{W'} Y'_{W'}) = \mathbb{E}(\mathbb{E}(X_W X'_W Y_{W'} Y'_{W'} | X, X', Y, Y')) \\ = \mathbb{E}(\mathbb{E}(X_W X'_W | X, X', Y, Y') \mathbb{E}(Y_{W'} Y'_{W'} | X, X', Y, Y')). \quad (1)$$

Wegen

$$X_W X'_W = \left(W(X) - \int_{\mathbb{R}} W(t) dF_X(t) \right) \left(W(X') - \int_{\mathbb{R}} W(t) dF_X(t) \right) \\ = W(X)W(X') - \int_{\mathbb{R}} W(X)W(t) dF_X(t) - \int_{\mathbb{R}} W(X')W(t) dF_X(t) \\ + \int_{\mathbb{R}} \int_{\mathbb{R}} W(t)W(s) dF_X(t) dF_X(s)$$

und $\mathbb{E}(W(t)W(s)) = |t| + |s| - |t - s|$ gilt für die erste Größe in (1)

$$\begin{aligned}
& \mathbb{E}(X_W X'_W | X, X', Y, Y') \\
&= |X| + |X'| - |X - X'| - \int_{\mathbb{R}} |X| + |t| - |X - t| dF_X(t) \\
&\quad - \int_{\mathbb{R}} |X'| + |t| - |X' - t| dF_X(t) + \int_{\mathbb{R}} \int_{\mathbb{R}} |t| + |s| - |t - s| dF_X(t) dF_X(s) \\
&= |X| + |X'| - |X - X'| - (|X| + \mathbb{E}(|X|) - \mathbb{E}'(|X - X'|)) \\
&\quad - (|X'| + \mathbb{E}(|X|) - \mathbb{E}''(|X' - X''|)) + (\mathbb{E}(|X|) + \mathbb{E}(|X'|) - \mathbb{E}(|X - X'|)) \\
&= \underbrace{\mathbb{E}'(|X - X'|) + \mathbb{E}''(|X' - X''|) - |X - X'| - \mathbb{E}(|X - X'|)}_{=:A},
\end{aligned}$$

wobei \mathbb{E}' den Erwartungswert bezüglich X' und \mathbb{E}'' den Erwartungswert bezüglich X'' bezeichnet und X', X'' unabhängig identisch verteilte Kopien von X sind. Analog gilt für die zweite Größe in (1)

$$\mathbb{E}(Y_W Y'_W | X, X', Y, Y') = \underbrace{\mathbb{E}'(|Y - Y'|) + \mathbb{E}''(|Y' - Y''|) - |Y - Y'| - \mathbb{E}(|Y - Y'|)}_{=:B},$$

wobei \mathbb{E}' den Erwartungswert bezüglich Y' und \mathbb{E}'' den Erwartungswert bezüglich Y'' bezeichnet und Y', Y'' unabhängig identisch verteilte Kopien von Y sind.

Damit haben wir in (1)

$$\begin{aligned}
BCov^2(X, Y) &= \mathbb{E}(A \cdot B) = \mathbb{E}(|X - X'| |Y - Y'|) + \mathbb{E}(|X - X'|) \mathbb{E}(|Y - Y'|) \\
&\quad - \mathbb{E}(|X - X'| |Y - Y''|) - \mathbb{E}(|X - X''| |Y - Y'|).
\end{aligned}$$

□

Wir kommen nun zu einem überraschenden Ergebnis: Die Distanz Kovarianz und die Brownsche Distanz Kovarianz stimmen miteinander überein. Das Ergebnis ist insofern überraschend, da beide Maßzahlen völlig verschiedene Definitionsansätze haben und bei der Brownschen Distanz Kovarianz die Existenz des zweiten Moments der Zufallsvariablen gefordert wurde, damit die Brownsche Distanz Kovarianz wohldefiniert ist und bei der Distanz Kovarianz schon die Existenz des ersten Moments ausreichte, damit diese wohldefiniert war.

Satz 3.18. Seien X und Y Zufallsvariablen. Dann gilt $DCov(X, Y) = BCov(X, Y)$.

Beweis. Es genügt zu zeigen, dass

$$\begin{aligned}
DCov(X, Y)^2 &= \mathbb{E}(\|X - X'\|_2 \|Y - Y'\|_2) + \mathbb{E}(\|X - X'\|_2) \mathbb{E}(\|Y - Y'\|_2) \\
&\quad - \mathbb{E}(\|X - X'\|_2 \|Y - Y''\|_2) - \mathbb{E}(\|X - X''\|_2 \|Y - Y'\|_2),
\end{aligned}$$

wobei $(X, Y), (X', Y')$ und (X'', Y'') unabhängig und identisch verteilt sind. Die Behauptung folgt dann mit Satz 3.17, da $DCov \geq 0, BCov \geq 0$. Der Zähler im Integral von $DCov^2(X, Y)$ kann wie folgt umgeschrieben werden:

$$\begin{aligned} |\varphi_{X,Y}(t, s) - \varphi_X(t)\varphi_Y(s)|^2 &= |\varphi_{X,Y}(t, s)|^2 - \overline{\varphi_{X,Y}(t, s)}\varphi_X(t)\varphi_Y(s) \\ &\quad - \varphi_{X,Y}(t, s)\overline{\varphi_X(t)}\overline{\varphi_Y(s)} + |\varphi_X(t)|^2 |\varphi_Y(s)|^2 := A. \end{aligned}$$

Mit einem ähnlichen Vorgehen wie in (2.5) erhält man:

$$\begin{aligned} |\varphi_{X,Y}(t, s)|^2 &= \mathbb{E}(\cos((X - X')^T t)\cos((Y - Y')^T s)); \\ |\varphi_X(t)|^2 |\varphi_Y(s)|^2 &= \mathbb{E}(\cos((X - X')^T t))\mathbb{E}(\cos((Y - Y')^T s)); \\ \overline{\varphi_{X,Y}(t, s)}\varphi_X(t)\varphi_Y(s) &= \mathbb{E}(\cos((X - X'')^T t)\cos((Y - Y')^T s)); \\ \varphi_{X,Y}(t, s)\overline{\varphi_X(t)}\overline{\varphi_Y(s)} &= \mathbb{E}(\cos((X - X')^T t)\cos((Y - Y'')^T s)). \end{aligned}$$

Mit Satz 2.3 (vi) folgt

$$\begin{aligned} A &= \mathbb{E}(\cos((X - X')^T t)\cos((Y - Y')^T s)) + \mathbb{E}(\cos((X - X')^T t))\mathbb{E}(\cos((Y - Y')^T s)) \\ &\quad - \mathbb{E}(\cos((X - X'')^T t)\cos((Y - Y')^T s)) - \mathbb{E}(\cos((X - X')^T t)\cos((Y - Y'')^T s)) \\ &= 1 - (1 - \mathbb{E}(\cos((X - X')^T t))) - (1 - \mathbb{E}(\cos((Y - Y')^T s))) \\ &\quad + \mathbb{E}[(1 - \cos((X - X')^T t))(1 - \cos((Y - Y')^T s))] + \mathbb{E}(\cos((X - X')^T t))\mathbb{E}(\cos((Y - Y')^T s)) \\ &\quad - 1 + (1 - \mathbb{E}(\cos((X - X')^T t))) + (1 - \mathbb{E}(\cos((Y - Y'')^T s))) \\ &\quad - \mathbb{E}[(1 - \cos((X - X')^T t))(1 - \cos((Y - Y'')^T s))] - 1 + (1 - \mathbb{E}(\cos((X - X'')^T t))) \\ &\quad + (1 - \mathbb{E}(\cos((Y - Y')^T s))) - \mathbb{E}[(1 - \cos((X - X'')^T t))(1 - \cos((Y - Y')^T s))] \\ &= \mathbb{E}[(1 - \cos((X - X')^T t))(1 - \cos((Y - Y')^T s))] \\ &\quad + \mathbb{E}(1 - \cos((X - X')^T t))\mathbb{E}(1 - \cos((Y - Y')^T s)) \\ &\quad - \mathbb{E}[(1 - \cos((X - X')^T t))(1 - \cos((Y - Y'')^T s))] \\ &\quad - \mathbb{E}[(1 - \cos((X - X'')^T t))(1 - \cos((Y - Y')^T s))]. \end{aligned}$$

Mit dem Satz von Fubini und Lemma 2.2 gilt

$$\begin{aligned} &\int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \frac{\mathbb{E}[(1 - \cos((X - X')^T t))(1 - \cos((Y - Y')^T s))] dt ds}{c_n c_m \|t\|_2^{1+n} \|s\|_2^{1+m}} \\ &= \mathbb{E} \left[\int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \frac{(1 - \cos((X - X')^T t))(1 - \cos((Y - Y')^T s)) dt ds}{c_n c_m \|t\|_2^{1+n} \|s\|_2^{1+m}} \right] \\ &= \mathbb{E} \left[\int_{\mathbb{R}^n} \frac{(1 - \cos((X - X')^T t)) dt}{c_n \|t\|_2^{1+n}} \int_{\mathbb{R}^m} \frac{(1 - \cos((Y - Y')^T s)) ds}{c_m \|s\|_2^{1+m}} \right] \end{aligned}$$

$$= \mathbb{E}(\|X - X'\|_2 \|Y - Y'\|_2)$$

und mit ähnlichen Schritten folgt

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \mathbb{E}(1 - \cos((X - X')^T t)) \mathbb{E}(1 - \cos((Y - Y')^T s)) d\omega = \mathbb{E}(\|X - X'\|_2) \mathbb{E}(\|Y - Y'\|_2).$$

Insgesamt also

$$\begin{aligned} DCov(X, Y)^2 &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} |\varphi_{X,Y}(t, s) - \varphi_X(t) \varphi_Y(s)|^2 d\omega \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \mathbb{E}[(1 - \cos((X - X')^T t))(1 - \cos((Y - Y')^T s))] d\omega \\ &\quad + \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \mathbb{E}(1 - \cos((X - X')^T t)) \mathbb{E}(1 - \cos((Y - Y')^T s)) d\omega \\ &\quad - \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \mathbb{E}[(1 - \cos((X - X')^T t))(1 - \cos((Y - Y'')^T s))] d\omega \\ &\quad - \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \mathbb{E}[(1 - \cos((X - X'')^T t))(1 - \cos((Y - Y')^T s))] d\omega \\ &= \mathbb{E}(\|X - X'\|_2 \|Y - Y'\|_2) + \mathbb{E}(\|X - X'\|_2) \mathbb{E}(\|Y - Y'\|_2) \\ &\quad - \mathbb{E}(\|X - X'\|_2 \|Y - Y''\|_2) - \mathbb{E}(\|X - X''\|_2 \|Y - Y'\|_2). \end{aligned}$$

□

Bemerkung Für die Distanz Varianz gilt

$$DVar(X)^2 = \mathbb{E}(\|X - X'\|_2^2) + \mathbb{E}(\|X - X''\|_2^2) - 2\mathbb{E}(\|X - X'\|_2 \|X - X''\|_2),$$

wobei X, X' und X'' unabhängig und identisch verteilt sind.

Wegen Satz 3.18 können nun alle Eigenschaften der Distanz Kovarianz\Korrelation, die wir in Kapitel 2 gezeigt haben, auf die Brownsche Distanz Kovarianz\Korrelation übertragen werden. Daher ist die empirische Distanz Kovarianz auch ein Schätzer für die Brownsche Distanz Kovarianz.

Nun kann man mittels obigen Sätzen einen leichten Beweis für Satz 2.9 (iii) angeben, denn mittels Cauchy-Schwarzscher Ungleichung gilt $\mathbb{E}(VW) \leq \sqrt{\mathbb{E}(V^2)\mathbb{E}(W^2)}$. Wählen wir nun wie im Beweis von Satz 3.17

$$V = \mathbb{E}'(\|X - X'\|_2) + \mathbb{E}''(\|X' - X''\|_2) - \|X - X'\|_2 - \mathbb{E}(\|X - X'\|_2)$$

und

$$W = \mathbb{E}'(\|Y - Y'\|_2) + \mathbb{E}''(\|Y' - Y''\|_2) - \|Y - Y'\|_2 - \mathbb{E}(\|Y - Y'\|_2),$$

dann ist $DCov(X, Y)^2 = BCov(X, Y)^2 = \mathbb{E}(VW)$, $DVar(X)^2 = BVar(X)^2 = \mathbb{E}(V^2)$ und $DVar(Y)^2 = BVar(Y)^2 = \mathbb{E}(W^2)$. Insgesamt also

$$\begin{aligned} DCov(X, Y)^2 &\leq \sqrt{DVar(X)^2 DVar(Y)^2} \Leftrightarrow \frac{DCov(X, Y)^2}{\sqrt{DVar(X)^2 DVar(Y)^2}} \leq 1 \\ &\Leftrightarrow DCor(X, Y) \leq 1 \Leftrightarrow BCor(X, Y) \leq 1. \end{aligned}$$

Außerdem können wir nun einen Zusammenhang zwischen Kovarianz und Distanz Kovarianz angeben, es ist nämlich

$$\begin{aligned} DCov(X, Y)^2 &= \mathbb{E}(\|X - X'\|_2 \|Y - Y'\|_2) + \mathbb{E}(\|X - X'\|_2) \mathbb{E}(\|Y - Y'\|_2) \\ &\quad - 2\mathbb{E}(\|X - X'\|_2 \|Y - Y''\|_2) \\ &= Cov(\|X - X'\|_2, \|Y - Y'\|_2) - 2Cov(\|X - X'\|_2, \|Y - Y''\|_2). \end{aligned}$$

Für die Distanz Varianz folgt somit

$$\begin{aligned} DVar(X)^2 &= \mathbb{E}(\|X - X'\|_2^2) + \mathbb{E}(\|X - X'\|_2)^2 - 2\mathbb{E}(\|X - X'\|_2 \|X - X''\|_2) \\ &= Var(\|X - X'\|_2) - 2Cov(\|X - X'\|_2, \|X - X''\|_2). \end{aligned}$$

Kapitel 4

Simulation

In diesem Kapitel beschäftigen wir uns mit der Anwendung der Distanz Korrelation und der klassischen Korrelation auf reelle Datenmengen.

Anwendung 1 Als Datenmenge wurden die Aktienpreise und die dazugehörigen Log>Returns der RWE AG und E.ON im Zeitraum vom 01.01.2009 bis 14.09.2012 gewählt. Insgesamt sind dies 956 Preise bzw. 955 Log>Returns. In Abbildung 4.1

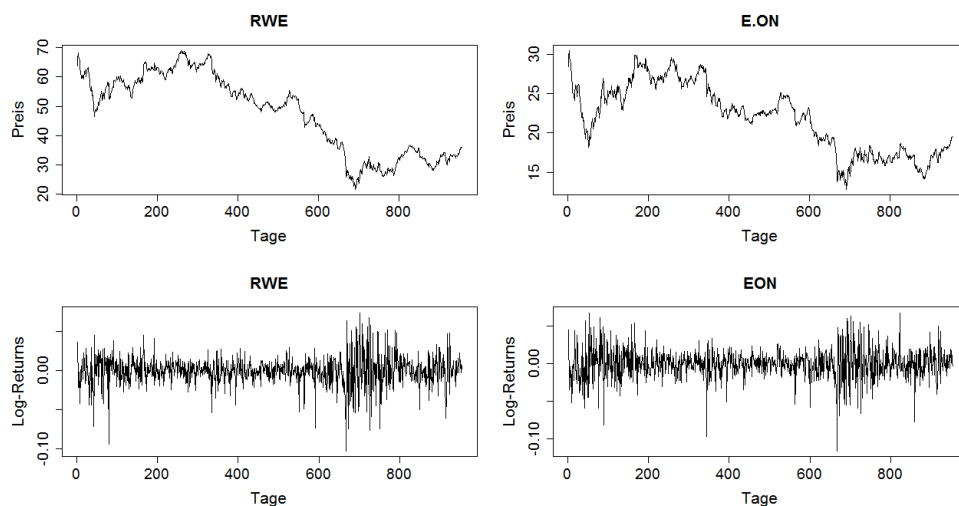


Abbildung 4.1: Preisentwicklungen(oben) und Log>Returns(unten) der RWE AG und E.ON im Zeitraum vom 01.01.2009 bis 14.09.2012

3.2. Brownsche Distanz Korrelation

sehen wir die Preisentwicklung und die Log>Returns der beiden Aktien im Betrachtungszeitraum. Abbildung 4.2 stellt ein Streudiagramm der beiden Preise und beider Log>Returns dar, auf dem man sofort einen linearen Zusammenhang erkennt. Diesen Zusammenhang hätte man von vornherein vermutet, da es sich um Aktien aus derselben Branche handelt und zudem beide im DAX vertreten sind.

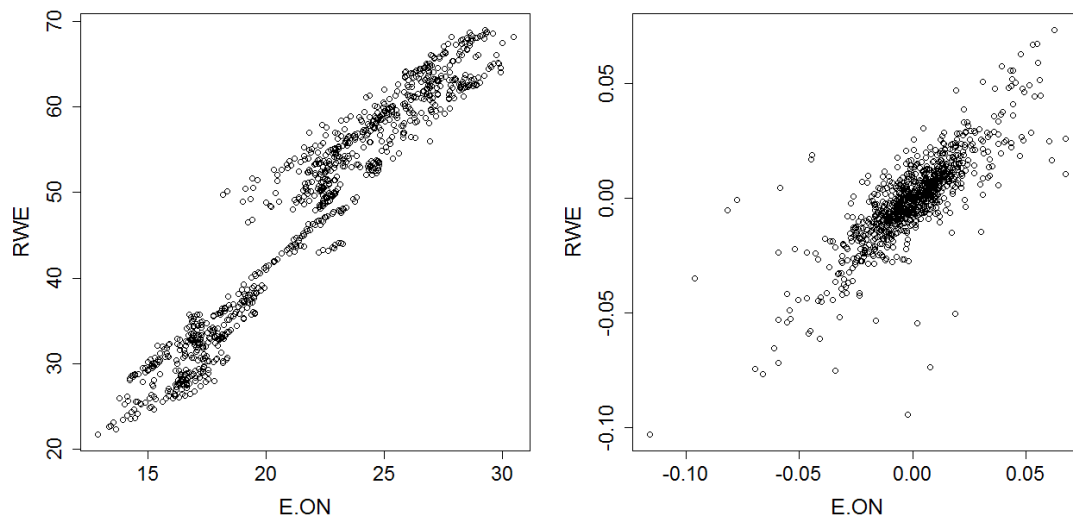


Abbildung 4.2: Streudiagramm der Aktienpreise(links) und Log>Returns(rechts) von der RWE AG und E.ON

Dieser lineare Zusammenhang wird zusätzlich durch die Pearsonsche Korrelation untermalt, denn es ist für die Aktienpreise $\tilde{\rho} = 0.9671158$. Zudem ist der Pearsonsche Korrelationstest signifikant (P-Wert $< 2.2e - 16$).

Die Distanz Korrelation ermittelt ebenso einen Zusammenhang der beiden Stichproben: $DCor = 0.9647255$ und P-Wert = 0.005, ist also auch signifikant.

Für die Log>Returns ist $\tilde{\rho} = 0.8020525$ mit P-Wert $< 2.2e - 16$ und $DCor = 0.7940915$ mit P-Wert = 0.005, also geht der lineare Zusammenhang etwas zurück.

Anwendung 2 Als Datenmenge verwenden wir nun die Aktienpreise und die dazugehörigen Log>Returns von BMW und von Adidas im Zeitraum vom 01.01.2003 bis 26.01.2009. Bei Adidas ist zu erwähnen, dass am 06.06.2006 ein 4:1 Aktiensplit durchgeführt wurde. Insgesamt sind es 1568 Preise bzw. 1566 Log>Returns, da der Log-Return bei dem Aktiensplit rausgenommen wurde. In Abbildung 4.3 sind die Preisentwicklungen und Log>Returns im Betrachtungszeitraum der beiden DAX-Unternehmen abgebildet.

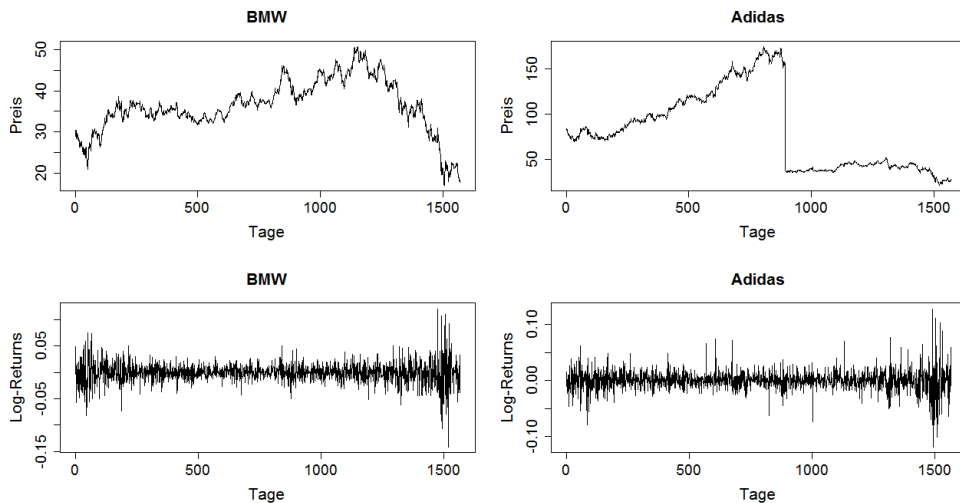


Abbildung 4.3: Aktienpreisentwicklung(oben) und Log>Returns(unten) bei BMW und Adidas im Zeitraum vom 01.01.2003 bis 26.01.2009

In Abbildung 4.4, welche ein Streudiagramm der beiden Preise und der Log>Returns darstellt, sieht man, dass ein nichtlinearer Zusammenhang zwischen beiden Preisen besteht und ein leichter linearer Zusammenhang zwischen den Log>Returns besteht. Da beide Unternehmen im DAX enthalten sind, aber in völlig verschiedenen Branchen tätig sind, hätte man einen Zusammenhang vermutet.

Mit dem Pearsonschen Korrelationstest erhalten wir für die Aktienpreise $\tilde{\rho} = 0.01801265$ mit P-Wert = 0.476. Also ist kein linearer Zusammenhang vorhanden. Der Distanz Korrelationstest ermittelt $\text{DCor} = 0.4147271$, welcher dazu noch signifikant ist (P-Wert = 0.005). Also handelt es sich bei den Aktienpreisen um zwei Datenmengen, die nahezu unkorreliert sind, aber nach dem Distanz Korrelationstest abhängig sind. Für die Log>Returns ist $\tilde{\rho} = 0.496137$ mit P-Wert $< 2.2e - 16$ und $\text{DCor} = 0.3991172$ mit P-Wert = 0.005. Somit verändert sich die Distanz Korrelation im Vergleich zur Distanz Korrelation bei den Aktienpreisen kaum, aber die Korrelation ermittelt nun einen stärkeren linearen Zusammenhang.

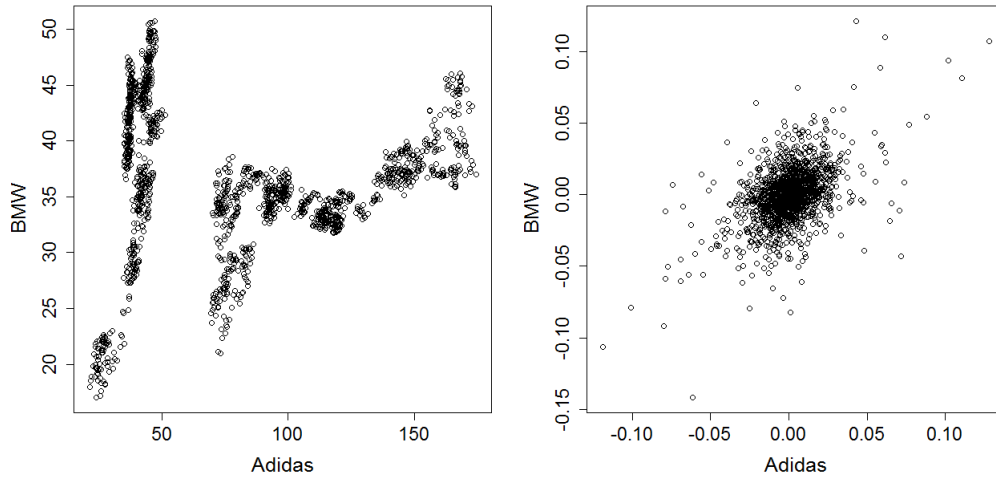


Abbildung 4.4: Streudiagramm der Aktienpreise(links) und der Log>Returns(rechts) von BMW und Adidas

Anwendung 3 Die Abhängigkeit zwischen zwei Zufallsvariablen kann eingeteilt werden in einen linearen und einen nicht-linearen Teil. Hier zeigen wir nun wie man den nicht-linearen Teil der Abhängigkeit erfassen kann. Da weder die Distanz Kovarianz noch die Distanz Korrelation in lineare und nicht-lineare Komponenten zerlegt werden kann, muss zuerst die klassische Korrelation herausgezogen werden.

Um die lineare Komponente der Abhängigkeit heraus zu ziehen, wird auf die Stichprobe (X^d, Y^d) ein lineares Modell der Form $Y = X\beta + \varepsilon$ mittels der Kleinsten-Quadrate-Methode angepasst. Die Fehler $\hat{\varepsilon}_i = X_i\hat{\beta} - Y_i$ sind mit den Prädiktoren \mathbf{X} unkorreliert. Um die nicht-lineare Komponente zu erhalten, wird der Distanz Korrelationstest auf die Fehler ε und die Prädiktoren \mathbf{X} angewandt.

Beispiel 4.1. (Aus [5], Aufgabe 10.2.1)

In einem Experiment wurden Meeresbakterien einer Röntgenstrahlung in insgesamt 15 aufeinander folgenden sechsminütigen Strahlungsintervallen ausgesetzt. Man erhielt die folgenden Ergebnisse:

Bakt.	355	211	197	166	142	106	104	60	56	38	36	32	21	19	15
Intervall	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Passen wir für die Daten ein Regressionsmodell der Form $Y = \beta X + \varepsilon$ an, wobei die Kleinst-Quadrate-Methode für die Schätzung des Regressionskoeffizienten β verwendet wird, so erhalten wir die Fehler

3.2. Brownsche Distanz Korrelation

1	2	3	4	5	6	7
114.883333	-9.652381	-4.188095	-15.723810	-20.259524	-36.795238	-19.330952
8	9	10	11	12	13	14
-43.866667	-28.402381	-26.938095	-9.473810	5.990476	14.454762	31.919048
15						
47.383333						

Wenden wir nun den Distanz Korrelationstest auf die Fehler ε und die Prädikatoren \mathbf{X} an, erhalten wir für $DCor = 0.5864254$. Dieser Test ist signifikant (P-Wert = 0.05).

Literaturverzeichnis

- [1] Aaronson, J., Burton, R., Dehling, H., Gilat, D., Hill, T., Weiss, B.: *Strong law for L- and U-Statistics*. Transactions of the American Mathematical Society Volume 348, Number 7(1996).
- [2] Bauer, H.: *Wahrscheinlichkeitstheorie*. Walter de Gruyter, 5.Auflage, Berlin, New York(2002).
- [3] Billingsley, P.: *Probability and Measures*. Wiley, 1.Auflage, New York(1979).
- [4] Hoeffding, W.: *The Strong law of large numbers for U-Statistics*. Institute of Statistics Mimeo Series 302(1961).
- [5] Holzmann, H.: *Skript zum Modul "Praktikum zur Stochastik", WS 10/11*.
- [6] Holzmann, H.: *Skript zum Modul "Elementare Stochastik", WS 09/10*.
- [7] Klenke, A.: *Wahrscheinlichkeitstheorie*. Springer Berlin Heidelberg, 1.Auflage(2006).
- [8] Kuo, H.H.: *Gaussian Measures in Banach Spaces*. Lecture Notes in Math. 463. Springer, Berlin(1975).
- [9] Mammitzsch, V.: *Skript zum Modul "Wahrscheinlichkeitstheorie", WS 11/12*.
- [10] Prudnikov, A.P., Brychkov, A., Marichev, O.I.: *Integrals and Series*. New York: Gordon and Breach Science Publishers(1986).
- [11] Székely, G. J. and Bakirov, N. K.: *Brownian Covariance and central limit theorem for stationary sequences*. Technical Report Nr. 08-01. Dept. Mathematics and Statistics, Bowling Green State University (2008).
- [12] Székely, G. J. and Bakirov, N. K.: *Extremal probabilities for Gaussian quadratic forms*. Probability Theory Related Fields 126, Seiten 184-202(2003).
- [13] Székely, G. J. and Rizzo M. L.: *Brownian distance Covariance*. The Annals of Applied Statistics, Vol 3 (2009).

- [14] Székely, G. J. and Rizzo M. L.: *Measuring and Testing Dependence by Correlation of Distances*. The Annals of Statistics, Vol 35 (2007).
- [15] Székely, G. J. and Rizzo M. L.: *Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method*. Journal of Classification 22: 151-183 (2005).

Erklärung

Hiermit versichere ich, dass ich diese Bachelorarbeit eigenständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Diese Arbeit hat in dieser oder einer ähnlichen Form noch nicht im Rahmen einer anderen Prüfung vorgelegen.

Marburg, den 26. September 2012

(Viktor Bengs, Verfasser)