

# Kontrollkarten auf Basis archimedischer Copulas

Dem Promotionsausschuss der Technischen Universität Dortmund  
vorgelegte Dissertation zur Erlangung des akademischen Grades

Doktorin der Naturwissenschaften (Dr. rer. nat.)

von

Dipl.-Stat. Editha Lockow

2013

1. Gutachter: Prof. Dr. Claus Weihs
2. Gutachter: Prof. Dr. Walter Krämer



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Abhängigkeits- und Abstandsmaße</b>	<b>5</b>
2.1	Kovarianz und Korrelation . . . . .	5
2.2	Maße auf Basis von Konkordanzen . . . . .	6
2.3	Abstandsmaße . . . . .	8
<b>3</b>	<b>Copulas</b>	<b>9</b>
3.1	Definition . . . . .	11
3.2	Sklar's Theorem . . . . .	13
3.3	Fréchet-Hoeffding-Schranken . . . . .	13
3.4	Ausgewählte Spezialfälle . . . . .	14
3.5	Beziehung zu den Korrelationskoeffizienten . . . . .	16
3.6	Parameterschätzung . . . . .	18
<b>4</b>	<b>Karten zur Prozessüberwachung</b>	<b>21</b>
4.1	Qualität und Statistik . . . . .	21
4.2	Qualitätsregelkarten . . . . .	23
4.3	Die mittlere Lauflänge ARL . . . . .	26
4.4	OC-Funktion . . . . .	27
4.5	Hotellings $T^2$ -Karte . . . . .	27
4.6	Steigers $Z$ -Test . . . . .	29
<b>5</b>	<b>Variante I: Überwachung von Prozesslage und -streuung</b>	<b>31</b>
5.1	Verwendete Kontrollkarten . . . . .	31
5.1.1	Adjustierte $T^2$ -Karte . . . . .	32
5.1.2	Alternative: Euklidische $E^2$ -Karte . . . . .	33

5.2	Effizienz der Kontrollkarten . . . . .	34
5.2.1	Bestimmung der Kontrollgrenzen . . . . .	34
5.2.2	Verhalten der Lauflänge bei Lageverschiebung . . . . .	36
5.3	Anwendungen . . . . .	42
5.3.1	Beispiel I: Ultraschall-Durchflussmesser . . . . .	42
5.3.2	Beispiel II: Inkrementelles Umformen . . . . .	47
5.4	Überwachung der Prozessstreuung . . . . .	53
<b>6</b>	<b>Variante II: Zusammenhangskontrolle</b>	<b>57</b>
6.1	Entwicklung eines geeigneten Testverfahrens . . . . .	57
6.1.1	$T^2$ -Karte für Abhängigkeitsänderungen . . . . .	57
6.1.2	Beschreibung der Teststatistik . . . . .	59
6.1.3	„Entwicklungsgeschichte“ des vorgestellten Tests . . . . .	64
6.2	Konstruktion der Grenzwerte $lcl$ und $ucl$ . . . . .	67
6.2.1	Herleitung der benötigten Wahrscheinlichkeiten . . . . .	67
6.2.2	Bestimmung der Wahrscheinlichkeitsbänder . . . . .	73
6.2.3	Beispiel Produktcopula . . . . .	74
6.3	Effizienz der Kontrollkarte . . . . .	76
6.3.1	Berechnung von $lcl$ und $ucl$ . . . . .	76
6.3.2	Ergebnisse der Lauflängensimulation . . . . .	78
6.3.3	Approximation der Grenzwerte am Beispiel der Claytoncopula	81
6.4	Vergleich mit Steigers $Z$ -Test . . . . .	85
6.5	Anwendungen . . . . .	87
6.5.1	Beispiel I: Aktienrenditen . . . . .	87
6.5.2	Beispiel II: Schrägseilbrücken . . . . .	90
6.6	Optimierung der Teststatistik . . . . .	98
6.6.1	Verteilung der Grenzwerte $lcl$ und $ucl$ . . . . .	99
6.6.2	Niveauausschöpfung verbessern . . . . .	101
6.6.3	ARL-Verhalten optimieren . . . . .	102
6.6.4	ARL-Verhalten und Niveauausschöpfung verbessern . . . . .	106
<b>7</b>	<b>Zusammenfassung und Ausblick</b>	<b>109</b>

<b>A</b>	<b>Ergänzende Formeln und Berechnungen</b>	<b>113</b>
A.1	Wahrscheinlichkeiten für Kapitel 6 . . . . .	113
A.1.1	Die Wahrscheinlichkeit $P(r \leq r^*, e_1^* \leq e \leq e_2^*)$ . . . . .	113
A.1.2	Die Wahrscheinlichkeit $P(e_1^* \leq e \leq e_2^*)$ . . . . .	114
A.1.3	Die Wahrscheinlichkeit $P(r \leq r^*   e_1^* \leq e \leq e_2^*)$ . . . . .	115
A.1.4	Beispiel Produktcopula - Herleitung der Integrale . . . . .	116
A.2	Copuladichten . . . . .	117
A.2.1	Dichte der Claytoncopula . . . . .	117
A.2.2	Dichte der Gumbelcopula . . . . .	118
A.2.3	Vergleich der Formeln 6.4 und 6.12 . . . . .	120
<b>B</b>	<b>Zusätzliche Tabellen und Abbildungen</b>	<b>123</b>



# Verzeichnis wichtiger Abkürzungen und mathematischer Ausdrücke

## Verteilungen

Term	Bedeutung
$\mathbb{B}()$	Binomialverteilung
$\mathbb{F}()$	F-Verteilung
$\mathbb{N}(0, 1)$	Standardnormalverteilung
$\mathbb{N}(\mu, \Sigma)$	multivariate Normalverteilung mit Erwartungswertvektor $\mu$ und Kovarianzmatrix $\Sigma$
$\mathbb{T}^2$	Hotellings T-Quadrat-Verteilung
$\mathbb{W}()$	Wishart-Verteilung

## Griechische Ausdrücke

Term	Bedeutung
$\alpha$	Testniveau beziehungsweise Niveau einer Kontrollkarte
$\eta$	Ausprägung des zu überwachenden Prozessparameters
$\eta_0$	Sollwert des zu überwachenden Prozessparameters
$\mu$	Erwartungswert(-vektor) einer Normalverteilung
$\Phi$	Verteilungsfunktion einer standardnormalverteilten Zufallsvariablen
$\varphi$	erzeugende Funktion einer archimedischen Copula
$\varphi^{[-1]}$	Pseudoinverse von $\varphi$
$\rho_{XY}$	Korrelation zwischen X und Y nach Pearson
$\rho_{XY}^S$	Rangkorrelation zwischen X und Y nach Spearman
$\sigma_{XY}$	Kovarianz zwischen X und Y
$\Sigma$	Kovarianzmatrix einer Normalverteilung
$\tau_{XY}$	Rangkorrelation zwischen X und Y nach Kendall
$\theta$	Copulaparameter (archimedische Copula)

## Sonstige Ausdrücke

Term	Bedeutung
$C(u_1, u_2, \theta)$	bivariate Copula mit dem Parameter $\theta$ am Punkt $(u_1, u_2)$
$c(u_1, u_2, \theta)$	Dichte von $C(u_1, u_2, \theta)$
$\text{Cov}(X, Y)$	Kovarianz von $X$ und $Y$
$d_M(X, Y)$	Mahalanobisdistanz zwischen $X$ und $Y$
$e$	euklidische Distanz eines Punktes auf der Winkelhalbierenden im $I^2$ zum Koordinatenursprung
$E^2$	Teststatistik der Euklidischen $E^2$ -Karte
$E(X)$	Erwartungswert von $X$
$f(\cdot)$	Fischertransformation
$F$	eindimensionale Verteilungsfunktion
$H_0$	Nullhypothese eines statistischen Tests
$H_1$	Alternativhypothese eines statistischen Tests
$H(x_1, x_2)$	bivariate Verteilungsfunktion am Punkt $(x_1, x_2)$
$I, I^2$	geschlossenes Intervall auf $[0, 1]$ , beziehungsweise $[0, 1] \times [0, 1]$
$I_p$	$p$ -dimensionale Einheitsmatrix
$i$	Zählvariable
$lcl$	Quantilsfunktion (obere Grenze) von $r$ in Abhängigkeit von $e$
$m$	Größe der Lernstichprobe
$n$	Stichprobengröße
$p$	Anzahl betrachteter Merkmale, Dimension der Verteilung
$P(\cdot)$	Wahrscheinlichkeit für ein Ereignis
$r$	orthogonaler Abstand zwischen einem Punkt und der Winkelhalbierenden im $I^2$
$R^*$	Korrelationsmatrix der multivariaten Normalverteilung
$rg(\cdot)$	Rang von $\cdot$
$R_i$	$\in \{0, 1, 2\}$ als Klasse für $r_i$ in Abhängigkeit von $e_i$ ( $i = 1, \dots, n$ )
$R_n$	Teststatistik der entwickelten R-Karte für $n$ Werte
$S_m$	Kovarianzmatrix einer Lernstichprobe der Größe $m$
$s_m^*$	Vektor der Standardabweichungen für alle Merkmale einer Lernstichprobe der Größe $m$
$s_{XY}$	empirische Kovarianz zwischen $X$ und $Y$



### Sonstige Ausdrücke (Fortsetzung)

Term	Bedeutung
$T$	Teststatistik eines statistischen Tests (allgemein)
$T^2$	Teststatistik zu Hotellings $T^2$ -Karte
$T^2_{adj}$	Teststatistik zur adjustierten $T^2$ -Karte
$U$	Zufallsvariable aus $I$ beziehungsweise $I^p$ , meist mit $p = 2$
$ucl$	Quantilsfunktion (untere Grenze) von $r$ in Abhängigkeit von $e$
$V$	Zufallsvariable aus $I$
$\text{Var}(X, Y)$	Varianz von $X$ und $Y$
$X$	Zufallsvariable aus dem $\mathbb{R}^p$
$Y$	Zufallsvariable aus dem $\mathbb{R}^p$
$\bar{Y}_m$	Mittelwertvektor einer Lernstichprobe der Größe $m$

### Abkürzungen

Abkürzung	Bedeutung
ARL	Average Run Length - mittlere Lauflänge
CL	Kontrollgrenze einer Kontrollkarte („Control Limit“)
CML	Canonical Maximum Likelihood (ML-Methode bei unbekanntem Rändern)
IFM	Inference for Margins (ML-Methode bei bekannten Rändern)
LCL	untere Kontrollgrenze einer Kontrollkarte („Lower Control Limit“)
ML	Maximum Likelihood
OC	Operationscharakteristik, misst den Fehler zweiter Art
QRK	Qualitätsregelkarte
RL	Run Length - Lauflänge
SDRL	Standard Deviation of the ARL - Standardabweichung der ARL
SPC	Statistical Process Control - Statistische Prozessüberwachung
UCL	obere Kontrollgrenze einer Kontrollkarte („Upper Control Limit“)
uiv	unabhängig identisch verteilt



# 1 Einleitung

Copulas waren bislang hauptsächlich ein Modellierungsinstrument der Ökonomen und Finanzstatistiker. Inzwischen erkennen auch vermehrt Naturwissenschaftler die Vorteile dieser multivariaten Datendarstellung, beispielsweise beim Modellieren von Umweltzusammenhängen und in der Ingenieurtechnik. Es werden immer mehr Bereiche gefunden, bei denen Copulamodelle bisher verwendete Verteilungsannahmen in ihrer Anpassungsgüte übertreffen. Die vorliegende Arbeit soll einen Beitrag zu dieser Entwicklung im Bereich der Qualitätskontrolle leisten.

Falsche oder „nicht optimale“ Verteilungsannahmen von Mess- und Qualitätsdaten führen unter Umständen zu rapiden Fehleinschätzungen der wahren Prozesslage. Wird der Prozess durch Copulamodelle beschrieben, sollten dies auch die verwendeten Instrumente zur Qualitätsüberwachung berücksichtigen. Eine genauere Kenntnis der Datenstruktur führt in der Praxis zur Einsparung verschiedener Ressourcen, etwa Material, Maschinen- oder Arbeiterstunden und ist damit finanziell und ökologisch interessant.

Die Idee, Copulamodelle auf den Bereich der Qualitätssicherung zu übertragen, insbesondere auf Qualitätsregelkarten, stammt von Yan (2007). Die vorliegende Dissertation greift diesen Gedanken auf. Betrachtet werden Qualitätsregelkarten für Prozesse, welche durch bivariate archimedische Copulas modellierbar sind.

Neben dem bekannten Problem, einen Prozess in Lage und Streuung der interessierenden Parameter konstant zu halten, wird die Abhängigkeitsstruktur zwischen den Merkmalen betrachtet. Hierauf liegt das Hauptaugenmerk der Arbeit.

Für Lage- und Streuungsänderungen werden Modifikationen der klassischen  $T^2$ -Karte verwendet. Zusätzlich wird die  $E^2$ -Karte entwickelt; diese verwendet ein alternatives Abstandsmaß. Zur Überwachung der Prozessstruktur wird eine neue Qualitätsregelkarte entwickelt und getestet. Diese Karte soll dabei sowohl eine Verstär-

kung als auch eine Abschwächung des Zusammenhangs zwischen den Daten erkennen. Alle Karten werden hinsichtlich ihres Güteverhaltens geprüft. Dazu wird die durchschnittliche Lauflänge des Prozesses betrachtet, ehe die Karte Alarm auslöst. Bei kontrolliert ablaufenden Prozessen sollte sie möglichst groß sein (in Abhängigkeit vom Testniveau), bei Prozessstörungen dagegen klein ausfallen.

Die Arbeit enthält fünf Hauptkapitel. Das nächste Kapitel stellt Abhängigkeits- und Abstandsmaße vor. Diese dienen zur Beschreibung der interessierenden Datenstrukturen und bilden die mathematische Grundlage der neu entwickelten Qualitätsregelkarte.

Das dritte Kapitel behandelt Copulas. Nach einer kurzen Darstellung der Anwendungsgebiete und der zugrunde liegenden Idee präsentiert dieses Kapitel ausgewählte archimedische Copulas. Neben der multivariaten Normalverteilung wird im Folgenden ausschließlich diese Gruppe betrachtet. Archimedische Copulas basieren auf einer Generatorfunktion; dabei ist die Auswahl beliebig groß. Sie beschreiben auf diese Weise sehr unterschiedliche Abhängigkeitsstrukturen und decken viele in der Praxis vorkommende Datenformationen ab. Ein weiterer Vorteil für diese Arbeit ist, dass Zufallszahlen auf Basis archimedischer Copulas relativ leicht zu erzeugen sind.

Das vierte Kapitel beschäftigt sich mit dem zweiten inhaltlich relevanten Theoriegebiet, den Qualitätsregelkarten. Neben Grundlagen und einigen Möglichkeiten, die Güte der Karte zu bewerten, werden Hotellings  $T^2$ -Karte und Steigers  $Z$ -Test vorgestellt.

Auf Basis der theoretischen Grundlagen beschäftigt sich Kapitel 5 mit den Möglichkeiten zur Überwachung von Prozesslage und -streuung. Zunächst wird eine Modifikation des klassischen  $T^2$ -Tests eingesetzt. Die verwendeten Copulas sind achsensymmetrisch, daher wird anschließend ein anderes Distanzmaß - der Euklidische Abstand - als Basis einer Regelkarte verwendet. Beide Karten werden für unterschiedliche Prozessszenarien vergleichend getestet. Zwei Anwendungen runden diese Betrachtung ab.

Copulas zielen darauf ab, die Datenstruktur gut abzubilden und Abhängigkeiten zwischen zwei oder mehr Merkmalen zu modellieren. Der interessanteste Aspekt dieser Arbeit ist daher die Überwachung der Abhängigkeitsstruktur. Diese soll für den betrachteten Prozess konstant bleiben. Die  $T^2$ -Karte erfasst Abhängigkeitsände-

rungen nicht zufriedenstellend. In Kapitel 6 wird daher eine neue Qualitätsregelkarte entwickelt. Sie soll Strukturänderungen eines Prozesses sichtbar machen. Dazu wird der Abstand der (Copula-)Datenpunkte zur Winkelhalbierenden auf dem Einheitsquadrat betrachtet. Dieser Abstand wird entsprechend aufbereitet und mit seiner Hilfe werden die Datenpunkte klassiert. Die Einordnung der Messpunkte erfolgt anhand entsprechender „Wahrscheinlichkeitsbänder“ eines kontrolliert ablaufenden Prozesses. Diese Bänder überdecken einen Bereich, in dem ein bestimmter Prozentsatz der Daten eines kontrolliert ablaufenden Prozesses liegt. Die Bänder decken den gesamten Wertebereich der modifizierten Messwerte ab. Damit ermöglichen sie eine eindeutige Dateneinordnung. Die Verteilung der klassifizierten Werte ist bestimmbar; sie bildet die Grundlage der Regelkarte.

Ein Vergleich mit Steigers  $Z$ -Test fällt zugunsten der neuen Qualitätsregelkarte aus, so dass deren praktischer Einsatz empfehlenswert scheint. Es folgen wiederum zwei Anwendungsbeispiele. Abschließend wird die Karte am Beispiel einer bestimmten Copula optimiert, indem die Wahrscheinlichkeitsschläuche variiert werden.

Das letzte Kapitel fasst die Ergebnisse zusammen und nennt Ansatzpunkte für weiterführende Arbeitsgebiete im Bereich der Qualitätsregelkarten auf Basis von Copulamodellen.



## 2 Abhängigkeits- und Abstandsmaße

Die deskriptive Statistik liefert verschiedene Möglichkeiten der Datenbeschreibung. Kennzahlen konzentrieren die Dateninformation auf wenige, relevante Maße, welche eine gute Übersicht über die Datengrundlage bieten. Einige Maße werden in diesem Kapitel erläutert, weiterführende Informationen gibt Greene (2012). Einen umfassenden Überblick zu „Standardmaßen“ liefern Hartung et al. (2005); die Definitionen aus Kapitel 2.1 stammen von ihm.

### 2.1 Kovarianz und Korrelation

Qualitätsmerkmale eines Objektes sind in der Regel nicht unabhängig, daher müssen Art und Ausmaß des Zusammenhangs für viele ingenieur- und naturwissenschaftliche Überlegungen bekannt sein. Für multivariat normalverteilte Daten sind Kovarianzen beziehungsweise Korrelationen geeignete Abhängigkeitsmaße. Die Kovarianz zweier Zufallsvariablen ist definiert als

$$\text{Cov}(X, Y) = \sigma_{XY} = E[(X - E(X))(Y - E(Y))]$$

Für eine Stichprobe  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , wird die Kovarianz geschätzt durch die **empirische Kovarianz**:

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (2.1)$$

Dabei bezeichnen  $\bar{x}$  und  $\bar{y}$  die arithmetischen Mittel der Beobachtungen  $x_1, \dots, x_n$  und  $y_1, \dots, y_n$ .

Die **Varianz** von  $X$  ( $\text{Var}(X)$ ) entspricht  $\sigma_{XX} = \sigma_X^2$  beziehungsweise  $s_{xx} = s_x^2$ . Die Kovarianz ist nicht beschränkt. Eine auf das Intervall  $[-1, 1]$  normierte Form der Kovarianz ist die **Korrelation nach Pearson**, die allgemein unter dem Begriff

„Korrelation“ verstanden wird. Sie ist gegeben durch:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2} \sqrt{\sigma_Y^2}} \quad (2.2)$$

Für die empirische Variante  $\hat{\rho}_{xy}$  oder auch  $r_{xy}$  werden die entsprechenden theoretischen Kovarianzen durch die empirischen ersetzt.

Für  $a, b \neq 0 \in \mathbb{R}$  und  $c, d \in \mathbb{R}$  ist  $\rho_{(aX+c, bY+d)} = \text{sign}(ab)\rho_{XY}$ . Die Korrelation ist also invariant gegenüber streng monoton wachsenden, linearen Transformationen der verwendeten Zufallsvariablen. Für streng monoton wachsende, nichtlineare Funktionen (hier:  $T$  und  $S$ ) gilt dieser Zusammenhang nicht, sondern allgemein ist

$$\rho_{T(X), S(Y)} \neq \rho_{XY}.$$

Betrachtet man beispielsweise die Vektoren

$$x = (0.2, 0.3, 0.4, 0.6, 0.2, 0.4)^\top \text{ und } y = (3.1, 4, 4.8, 6., 2.8, 5.2)^\top$$

und die Funktionen  $T(x) = x^2$  sowie  $S(y) = y^6$ , gilt  $\rho_{xy} = 0.97 \neq \rho_{T(x), S(y)} = 0.99$ .

Kovarianz und Korrelation sind meist einfach herzuleiten, bilden aber ausschließlich lineare Zusammenhänge ab. Aus der Unabhängigkeit zweier Zufallsvariablen folgt stets Unkorreliertheit, sofern die zweiten Momente existieren

$$(\sigma_{XY} = 0 \rightarrow \rho_{XY} = 0),$$

der Umkehrschluss gilt nur für multivariat normalverteilte Daten.

## 2.2 Maße auf Basis von Konkordanzen

Die Theorie zu Konkordanzen und darauf basierenden Abhängigkeitsmaßen sind erklärt Nelson (2006). Sowohl die Rangkorrelation von Kendall als auch Spearmans Rangkorrelation nutzen Konkordanzmessungen zur Beschreibung der Abhängigkeit zweier Variablen.



Betrachtet werden die Zufallsvariablen  $X$  und  $Y$ . Zwei Beobachtungspaare  $(x_i, y_i)$  und  $(x_j, y_j)$  ( $i, j \in 1, \dots, n$ ) heißen **konkordant**, falls

$$x_i < x_j \text{ und } y_i < y_j \quad \text{oder} \quad x_i > x_j \text{ und } y_i > y_j$$

gilt. beziehungsweise **diskonkordant**, falls gilt:

$$x_i < x_j \text{ und } y_i > y_j \quad \text{oder} \quad x_i > x_j \text{ und } y_i < y_j$$

Darauf aufbauend wird die **Rangkorrelation nach Kendall (Kendalls  $\tau$ )** von  $(X, Y)$  mit  $n$  unabhängigen Ziehungen  $(X_l, Y_l)$ ,  $l = 1, \dots, n$  folgendermaßen definiert:

$$\tau_{XY} = \mathbb{P}[(X_i - X_j)(Y_i - Y_j) > 0] - \mathbb{P}[(X_i - X_j)(Y_i - Y_j) < 0] \quad \forall i, j = 1, \dots, n$$

Sei  $c$  die Anzahl konkordanter Paare und  $d$  die Anzahl diskonkordanter Paare innerhalb der Stichprobe, so gilt

$$\hat{\tau}_{XY} = \frac{c - d}{c + d}. \quad (2.3)$$

Im Gegensatz zu Kendalls  $\tau$  nutzt die **Rangkorrelation von Spearman (Spearman's  $\rho$ )** die Differenz der Ränge und nicht nur deren Unterschied. Mit der Schreibweise von  $rg(X_i)$  als Rang von  $X_i$  gilt für  $n$  Wertepaare:

$$\hat{\rho}_{XY}^S = 1 - \frac{6 \sum_{i=1}^n (rg(X_i) - rg(Y_i))^2}{n \cdot (n^2 - 1)}. \quad (2.4)$$

Im Folgenden wird einheitlich Kendalls  $\tau$  zur Beschreibung von Abhängigkeiten verwendet. Das Anwachsen des Abhängigkeitsmaßes kann im direkten Vergleich zu Pearsons  $\rho$  nicht unbedingt linear interpretiert werden, siehe auch Abschnitt 3.5. Je nach betrachteter Grundgesamtheit kann beispielsweise eine Erhöhung von  $\tau$  um 0.2 einen wesentlich größeren Sprung von  $\rho$  bewirken.

## 2.3 Abstandsmaße

Viele multivariate Methoden basieren auf der Betrachtung von Abständen zwischen Objekten im multidimensionalen Raum. Unterschiedliche Methoden zur Abstandsberechnung können die Ergebnisse einer Methode stark beeinflussen. Für weiterführende Informationen sowie verschiedene konkrete Berechnungsmöglichkeiten für Abstände siehe etwa Fahrmeir et al. (1996).

Für eine endliche Menge  $M = \{X_1, X_2, \dots, X_N\}$  heißt die Funktion  $d : M \times M \rightarrow \mathbb{R}$  **Distanzmaß oder Distanzfunktion**, falls sie folgende Punkte erfüllt:

- $d(X_i, X_j) = d(X_j, X_i) \quad \forall X_i, X_j \in M$
- $d(X_i, X_j) \geq 0$  und  $d(X_i, X_i) = 0 \quad \forall X_i, X_j \in M$

Der Abstand zwischen zwei Punkten im  $n$ -dimensionalen Raum kann durch die  $L_q$ -Norm nach der **Gleichung von Minkowski**

$$d_q(x, y) = \left( \sum_{i=1}^p |x_i - y_i|^q \right)^{1/q} \quad \text{mit } q \geq 1 \quad (2.5)$$

berechnet werden. Dabei steht  $q$  für die Art des Abstands. Spezialfälle sind

- $q = 1$ : „City-Block-Metrik“ oder auch „Manhattan-Distanz“ und
- $q = 2$ : „Euklidische Distanz“ mit der Darstellung  $\|x - y\|$ .

Die Minkowski-Distanz berücksichtigt unterschiedliche Skalierungen der individuellen Koordinaten nicht. Wenn die Merkmale verschiedene Ausprägungsintervalle oder Maßeinheiten haben, dominiert die Koordinate mit dem größeren Bereich das Ergebnis. Dieses Problem kann durch Skalierung der Daten vor der Distanzberechnung gelöst werden.

Die **Mahalanobisdistanz** berücksichtigt die Korrelation und verschiedene Skalierungen durch Einbeziehen der inversen geschätzten Kovarianzmatrix  $S^{-1}$ ; für unkorrelierte, standardisierte Daten entspricht sie der Euklidischen Distanz. Sie berechnet sich nach der Formel

$$d_M(x, y) = (x - y)^T S^{-1} (x - y). \quad (2.6)$$

## 3 Copulas

Der Name „Copula“ stammt aus dem Lateinischen („Band, Koppel, Verbindung“), er betont die Eigenschaft der Funktion, Randverteilungen und Abhängigkeitsstruktur zu verknüpfen.

Copulas helfen, mehrdimensionale Zusammenhänge darzustellen und stochastische Abhängigkeit wesentlich flexibler modellieren als etwa mit Korrelationskoeffizienten. Klassische Beispiele hierfür stammen meist aus der Finanzwelt, siehe etwa Savu (2007). Weit verbreitet ist die Modellierung der Abhängigkeit mit Copulafunktionen in Anwendungen der finanziellen Risikobewertung und in versicherungsmathematischen Analysen. Das bekannteste Beispiel ist die Preisgestaltung von Wertpapieren, insbesondere von CDOs (Collateralized Debt Obligations), siehe etwa Cízek et al. (2005), Meucci (2011) oder Meneguzzo und Vecchiato (2003). CDOs bestehen aus einem Portfolio festverzinslicher Wertpapiere. Sie sind ein wichtiges Refinanzierungsmittel für Banken auf dem Kapitalmarkt und im Zuge der aktuellen Finanzkrise in die Kritik geraten. Mit den CDOs geriet auch die Anwendung der Gaußcopula auf Kreditderivate in Verruf (Salmon (2009)). Neue Veröffentlichungen im Bereich der Finanzstatistik dokumentieren die Beschränkungen der Gaußcopula. Wesentliche Punkte dabei sind die Unfähigkeit, Abhängigkeiten dynamisch zu modellieren, und die schlechte Wiedergabe extremer Ereignisse. Korrigierende Modelle stellen etwa Brigo et al. (2010) oder Donnelly und Embrechts (2010) vor.

Weniger umstritten ist die Anwendung von Copulas im Bereich anderer Anlagesektoren, beispielsweise bei der Preiskalkulation oder im Risikomanagement, für Devisenkurse und für Rentenfonds, siehe etwa Qu (2005).

Auch bei Versicherungsunternehmen spielt der Begriff „Risikomanagement“ eine immer größere Rolle, insbesondere die Risikoaggregation, also die Zusammenfassung von Einzelrisiken zu einem Gesamtrisiko. Abhängigkeiten von verschiedenen Einzelrisiken zu verstehen und zu bewerten, ist hier essentiell, um etwa die Gefahr

zu modellieren, dass verschiedene Ansprüche gleichzeitig geltend gemacht werden (Tran (2009)).

Salvadori und De Michele (2007) zeigen Anwendungsgebiete in der Hydrologie, wenn etwa Pegelstände eines Flusses an verschiedenen Messpunkten modelliert werden; auch im Bereich der Offshoretechnik gibt es diverse Einsatzgebiete (de Waal (2005)).

Bei technischen Produktionsprozessen interessiert das Auftreten von verbundenen Fehlern oder die zeitliche Verbundenheit von Misserfolgen. Copulafunktionen werden unter anderem erfolgreich im Bereich der Zuverlässigkeitsanalyse von Autobahnbrücken angewandt (Srinivas et al. (2006), Thompson und Kilgore (2011)).

Anwendungen im Bereich der Meteorologie und Klimaforschung (tägliche Niederschlagsmessungen, Temperatur, . . . ) stellen Schölzel und Friederichs (2008) vor. Im Bereich der Medizin können Copulafunktionen beispielsweise zur Modellierung von Aktionspotentialen (Zählimpulse von Spitzen im zeitlichen Ablauf) in der Neurologie angewendet werden, siehe etwa Onken et al. (2009).

In der Praxis werden oft multivariate statistische Verfahren zur Modellierung der genannten Beispiele verwendet. Diese basieren meist auf der Annahme, dass gemeinsam normalverteilte Daten vorliegen. Diese Bedingung ist oftmals nicht erfüllt, auch wenn die verbundenen Merkmale einzeln normalverteilt sind.

Hier sind andere multivariate Verteilungen geeigneter, in der Vergangenheit waren sie allerdings auch schwer zu konstruieren. Meist entstanden multivariate Verteilungen als Verallgemeinerung von univariaten Verteilungen, so dass nur gleiche Randverteilungen berücksichtigt werden konnten. Außerdem gab es oftmals starke Bedingungen an die Abhängigkeit der einzelnen Zufallsgrößen.

Copulas verfolgen einen anderen Ansatz. Mit ihnen lassen sich multivariate Verteilungsfunktionen in zwei Teile zerlegen. Dabei werden die univariaten Randverteilungen getrennt von der Copulafunktion betrachtet, welche die Abhängigkeitsstruktur der Zufallsvariablen beschreibt. Sowohl die Randverteilungen als auch die Copula sind dabei frei wählbar, wodurch die Modellierung der oben genannten Sachverhalte erheblich flexibler und einfacher wird.

Der Ansatz, Randverteilungen von der gemeinsamen Wahrscheinlichkeitsverteilung zu trennen, ist seit den 1940ern bekannt. (Enorm) gestiegen ist das Interesse

an Copulas aber erst um die Jahrtausendwende, wie man an der wachsenden Anzahl der Veröffentlichungen zu diesem Thema erkennen kann. Bislang gibt es daher nur wenige Standardwerke zu Copulas, etwa Nelson (2006) oder Joe (2001). Das Konzept der Copulas erscheint zunächst ungewöhnlich, als Einstiegslektüre ist das Werk von Genest und Favre (2007) geeignet.

Grundlage der folgenden mathematischen Erläuterungen ist die Arbeit von Nelson (2006). Hier werden alle angeführten Theoreme und Sätze für den zweidimensionalen Fall bewiesen, was für diese Arbeit ausreichend ist.

### 3.1 Definition

Copulas werden genutzt, um multivariate Verteilungen mit beliebigen Randverteilungen zu konstruieren. Sie veranschaulichen die Zusammenhangsstruktur verschiedener Zufallsvariablen. Gegeben sei dazu das geschlossene Intervall  $I = [0, 1]$  und analog  $I^2 = [0, 1] \times [0, 1]$ . Weiter sei  $(U_1, \dots, U_p)^T$  ein Zufallsvektor, wobei die Einträge  $U_i$ ,  $i = 1, \dots, p$ , gleichverteilt über dem Intervall  $I$  sind.

Die gemeinsame Verteilungsfunktion aller  $U_i$  heißt  **$p$ -dimensionale Copula ( $p$ -Copula)**  $C(u_1, \dots, u_p) = \mathbb{P}(U_1 \leq u_1, \dots, U_p \leq u_p)$ .

Eine **zweidimensionale Copula** (kurz: **Copula**) ist eine Funktion  $C : I^2 \rightarrow I$  mit  $C(u, v) = \mathbb{P}(U \leq u, V \leq v)$  für  $u, v \in I$ , für die gilt:

1.  $C(u, 0) = C(0, v) = 0 \quad \forall u, v \in I$
2.  $C(u, 1) = u$  und  $C(1, v) = v \quad \forall u, v \in I$
3.  $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$   
für alle Rechtecke  $(u_1, u_2] \times (v_1, v_2] \subset I^2$

Die obigen Bedingungen folgen direkt aus der Definition einer Verteilungsfunktion. Existiert die Dichtefunktion einer Copula, ergibt sie sich durch Ableiten nach den Randvariablen als

$$c(u, v) = \frac{\partial C(u, v)}{\partial u \partial v}. \quad (3.1)$$

**Elliptische Copulas** (am Beispiel der Gaußcopula) und **Archimedische Copulas** (Clayton-, Frank- und Gumbel-) werden hier vorgestellt. Informationen zu **Extremwertcopulas**, einer weiteren wichtigen Copulagruppe im Bereich der Ingenieurstatistik, liefert Joe (2001). Abbildung 3.1 zeigt einige Möglichkeiten, wie die gemeinsame Verteilung aussehen kann, wenn die Randverteilungen einzeln (identisch) normalverteilt sind.

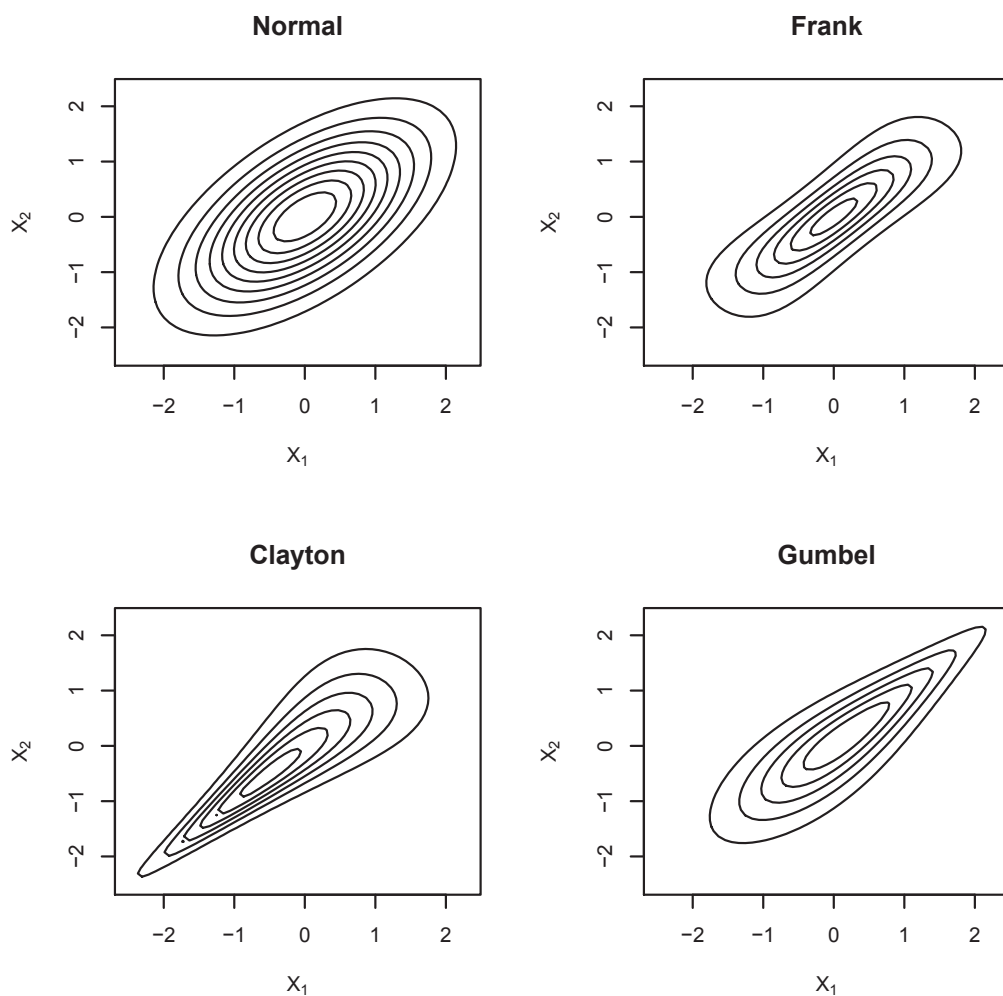


Abbildung 3.1: Dichtefunktionen von Verteilungen, die durch verschiedene Copulas modelliert wurden (Normal-, Frank-, Clayton- und Gumbel-). Die (beiden) Ränder sind standardnormalverteilt und Kendalls  $\tau = 0.6$ .

## 3.2 Sklar's Theorem

**Sklar's Theorem** besagt: Sei  $H$  eine  $p$ -dimensionale Verteilungsfunktion mit den Randverteilungen  $F_1, \dots, F_p$ . Dann existiert eine  $p$ -Copula  $C$  derart, dass  $\forall x \in \mathbb{R}^p$  gilt

$$H(x_1, \dots, x_p) = C(F_1(x_1), \dots, F_p(x_p)). \quad (3.2)$$

Auch der Umkehrschluss gilt: Es existiert für alle eindimensionalen Verteilungsfunktionen  $F_1, \dots, F_p$  sowie für jede Copula  $C$  eine mehrdimensionale Verteilungsfunktion  $H$  mit den Randverteilungen  $F_1, \dots, F_p$ . Für den zweidimensionalen Fall ist

$$H(x_1, x_2) = C(F_1(x_1), F_2(x_2))$$

mit  $F_1(x_1) = u$  und  $F_2(x_2) = v$ . Falls alle Randverteilungen stetig sind, ist  $C$  eindeutig bestimmt.

Stetige multivariate Verteilungen können also ohne Informationsverlust in zwei Komponenten zerlegt werden: Die univariaten Randverteilungen und die Copula. Die Abhängigkeitsstruktur multivariater Verteilungen kann so separat von den gegebenen Randverteilungen betrachtet werden - es reicht, sich auf die Copula zu beschränken. Diese Tatsache bietet völlig neue Möglichkeiten bei der Modellierung von Abhängigkeiten.

## 3.3 Fréchet-Hoeffding-Schranken

Für eine Copula  $C(u, v)$  gelten als obere beziehungsweise untere Grenze die Fréchet-Hoeffding-Schranken, die folgende Ungleichung bilden

$$\max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v) \quad \forall u, v \in I.$$

Die Schranken erfüllen selbst auch wieder die Copulabedingungen. Die untere Schranke ist nur für den bivariaten Fall eine Copula und wird als **Maximumscopula**  $W(u, v)$  bezeichnet; sie gibt perfekt negative Abhängigkeit an. Die obere Grenze, die **Minimumscopula**  $M(u, v)$ , entspricht perfekt positiver Abhängigkeit. Damit lautet die obige Ungleichung  $W(u, v) \leq C(u, v) \leq M(u, v) \quad \forall u, v \in I$ .

### 3.4 Ausgewählte Spezialfälle

Die einfachste Form der Copula ist die Unabhängigkeits- oder Produktcopula der Form  $C(u, v) = uv$ .

**Elliptische Copulas** ergeben sich aus elliptischen Verteilungen, beispielsweise der multivariaten Normalverteilung oder der multivariaten t-Verteilung. Sie sind punktsymmetrisch um den Erwartungswert der Verteilung, das erleichtert explizite Rechnungen beziehungsweise ermöglicht diese erst. Weiterhin können in dieser Klasse endlastige Verteilungen („heavy tails“) auftreten. Anschaulich besagt der Begriff, dass auf dem „Ende“ beziehungsweise den „Enden“ der Verteilung mehr Masse liegt als bei der Normalverteilung. Diese Verteilungen eignen sich unter anderem zur Modellierung von Großschäden und seltenen Ereignissen („Jahrhundertflut“, Radioaktivitätswerte (vor und) nach einer Reaktorexpllosion, ...).

Elliptische Copulas können genutzt werden, um mit Hilfe von Sklar's Theorem (Formel (3.2)) multivariate Verteilungen zu konstruieren. Zu beachten ist, dass zwar aus der Normalcopula eine multivariate Normalverteilung wird, aus der t-Copula in der Regel aber keine multivariate t-Verteilung (Ausnahme: Alle Randverteilungen haben gleichviele Freiheitsgrade). Die **Normal-** oder **Gaußcopula** gehört zur Klasse der elliptischen Verteilungen. Sie wird mit Hilfe der Normalverteilung definiert.

Dazu sei  $\Phi_{R^*}$  die Verteilungsfunktion einer multivariaten Normalverteilung mit der Korrelationsmatrix

$$R^* := \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{pmatrix}$$

und  $\Phi$  die Verteilungsfunktion einer standardnormalverteilten ( $\mathbb{N}(0, 1)$ ) Zufallsvariablen. Dann ist die Normalcopula mit Verteilungsparameter (auch: charakteristische Matrix)  $R^*$  gegeben durch

$$C^N(u_1, \dots, u_p, R^*) = \Phi_{R^*}[\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_p)].$$



Im bivariaten Fall gilt für den Copulaparameter  $\rho = \rho_{12}$  aus  $\mathbb{R}^*$  dementsprechend

$$C^N : I^2 \rightarrow I \text{ mit } C^N(u, v, \rho) = \Phi_\rho[\Phi^{-1}(u), \Phi^{-1}(v)]. \quad (3.3)$$

**Archimedische Copulas** werden - im Gegensatz zu den elliptischen Copulas - nicht mit Hilfe von Sklar's Theorem konstruiert. Sie leiten sich aus einer univariaten Funktion ab. Sie sind als geschlossener Ausdruck darstellbar und meist leicht zu konstruieren. Außerdem gibt es verschiedenste Copulafamilien aus der Klasse archimedischer Copulas. Daher findet man sie in der Praxis in vielen Bereichen. Neben Nelson (2006) bietet auch Joe (2001) einen ausführlichen Überblick über diverse archimedische Copulas.

Archimedische Copulas werden mit Hilfe einer **erzeugenden Funktion**  $\varphi$  (auch: **Erzeuger** oder **Generator**) konstruiert und sind durch diese eindeutig bestimmt. Dazu sei  $\varphi : I \rightarrow [0, \infty)$  eine stetige, streng monoton fallende Funktion mit  $\varphi(1) = 0$ . Die **Laplacetransformierte** oder **Pseudoinverse**  $\varphi^{[-1]}$  von  $\varphi$  ist definiert als:

$$\varphi^{[-1]} : [0, \infty] \rightarrow I \text{ mit } \varphi^{[-1]}(\nu) = \begin{cases} \varphi^{-1}(\nu) & 0 \leq \nu < \varphi(0), \\ 0 & \varphi(0) \leq \nu \leq \infty \end{cases}$$

Für  $\varphi(0) = \infty$  heißt  $\varphi$  **striker Erzeuger**. Dann gilt

$$\varphi^{[-1]}(t) = \varphi^{-1}(t) \quad \forall t$$

und die daraus resultierende Copula heißt **strikte (archimedische) Copula**. Mit der Laplacetransformierten ist die archimedische Copula  $C : I^p \rightarrow I$  mit Parameter  $\theta$  definiert als

$$C(u_1, \dots, u_p, \theta) = \varphi^{[-1]}[\varphi(u_1) + \dots + \varphi(u_p)].$$

Im bivariaten Fall vereinfacht sich die obige Formel zu

$$C : I^2 \rightarrow I \text{ mit } C(u, v, \theta) = \varphi^{[-1]}[\varphi(u) + \varphi(v)]. \quad (3.4)$$

Verschiedene Generatoren erzeugen unterschiedliche archimedische Copulafamilien, die durch geeignete Parameterwahl noch einmal unterscheidbar sind.

Exemplarisch wird hier die Claytoncopula aus ihrem Erzeuger hergeleitet. Es ist  $\varphi(t) = \frac{1}{\theta}(t^{-\theta} - 1)$  für  $\theta > 0$ . Damit ist  $\varphi(u) = \frac{1}{\theta}(u^{-\theta} - 1)$  und  $\varphi(v) = \frac{1}{\theta}(v^{-\theta} - 1)$  sowie  $\varphi^{[-1]}(t) = (\theta t + 1)^{-\frac{1}{\theta}}$ . Einsetzen in Formel (3.4) ergibt dann die gesuchte Darstellung:

$$\begin{aligned} C(u, v, \theta) &= \varphi^{[-1]}[\varphi(u) + \varphi(v)] = \varphi^{[-1]} \left[ \frac{1}{\theta}(u^{-\theta} + v^{-\theta} - 2) \right] \\ &= \left[ \theta \frac{1}{\theta}(u^{-\theta} + v^{-\theta} - 2) + 1 \right]^{-\frac{1}{\theta}} = \left( \frac{1}{u^\theta} + \frac{1}{v^\theta} - 1 \right)^{-\frac{1}{\theta}} \end{aligned}$$

Drei wichtige Copulafamilien zeigt Tabelle 3.1.

Familie	$\varphi(t)$	$\theta$	$C_\varphi(u, v, \theta)$
Clayton	$\frac{1}{\theta}(t^{-\theta} - 1)$	$\theta > 0$	$\left( \frac{1}{u^\theta} + \frac{1}{v^\theta} - 1 \right)^{-\frac{1}{\theta}}$
Gumbel	$(-\ln t)^\theta$	$\theta \geq 1$	$\exp \left( - \left[ \{-\ln(u)\}^\theta + \{-\ln(v)\}^\theta \right]^{\frac{1}{\theta}} \right)$
Frank	$\left( -\ln \frac{e^{-\theta t} - 1}{e^{-\theta} - 1} \right)$	$\theta \in \mathbb{R}$	$-\frac{1}{\theta} \ln \left( 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^\theta - 1} \right)$

Tabelle 3.1: Generator  $\varphi(t)$ , Parameter  $\theta$  und Verteilungsfunktion  $C(u, v, \theta)$  ausgewählter archimedischer Copulas

### 3.5 Beziehung zu den Korrelationskoeffizienten

Kendalls  $\tau$  und Spearmans  $\rho$  können als Funktionen bivariater Copulas dargestellt werden durch

$$\tau_{XY} = 4 \int \int_{[0,1]^2} C(u, v) dC(u, v) - 1 \quad \text{ sowie } \quad (3.5)$$

$$\rho_{XY}^S = 12 \int \int_{[0,1]^2} u \cdot v dC(u, v) - 3 = 12 \int \int_{[0,1]^2} C(u, v) dudv - 3, \quad (3.6)$$

wie beispielsweise Embrechts et al. (2010) zeigen.

Berechnet man Abhängigkeitsmaße für Datenpunkte, die durch verschiedene Copulastrukturen erklärbar sind, kann das Anwachsen von Kendalls  $\tau$  nicht im direkten Vergleich zu Pearsons  $\rho$  linear interpretiert werden, siehe auch Abbildung 3.2.

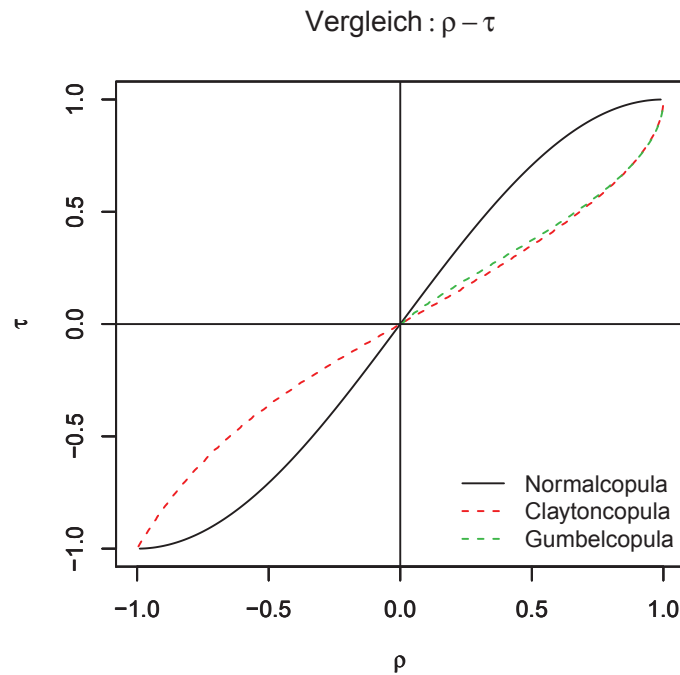


Abbildung 3.2: Zusammenhang zwischen Kendalls  $\tau$  und Pearsons  $\rho$  für verschiedene Datenstrukturen

Die Korrelationsmatrix  $R^*$  der **Gaußcopula**  $C^N$  ist eindeutig in die Rangkorrelationsmatrizen nach Kendall oder Spearman umformbar, siehe etwa Lindskog et al. (2001) und Lindskog (2000). Danach gilt für den bivariaten Fall

$$\tau = \frac{2}{\pi} \arcsin(\rho) \text{ und } \rho^S = \frac{6}{\pi} \arcsin\left(\frac{\rho}{2}\right). \quad (3.7)$$

Zwischen Kendalls  $\tau$  und dem Copulaparameter  $\theta$  der **Claytoncopula** lässt sich folgender Zusammenhang feststellen:

$$\theta = \frac{2\tau}{1 - \tau} \quad (3.8)$$

Für die **Gumbelcopula** gilt der Zusammenhang

$$\theta = \frac{1}{1 - \tau}. \quad (3.9)$$

Eine geschlossene Darstellung für den Parameter der Frankcopula existiert nicht.

### 3.6 Parameterschätzung

Um den Copulaparameter aus einem vorliegenden Datensatz zu schätzen, gibt es verschiedene Methoden. Einige Verfahren passen die direkt aus den Daten geschätzten Parameter noch weiter an und optimieren das Modell, indem abwechselnd Randverteilungen und Copulaparameter in mehreren Durchläufen aufeinander abgestimmt werden. Beispielhaft werden hier die Schätzung durch Rangkorrelation und per Maximum-Likelihood-Methode für den zweidimensionalen Fall dargestellt. Die Erläuterungen stammen aus Joe (2001) sowie Genest und MacKay (1986).

Die Copulafamilie (Gauß-, Clayton-, ...) wird meist als bekannt vorausgesetzt, da oft inhaltliche Gründe für eine bestimmte Verteilungsform der Daten sprechen. Ist dies nicht der Fall, gibt es eine Vielzahl mathematischer Modelle zur Bestimmung einer geeigneten Copula. Wichtig ist hier vor allem, verschiedene Anpassungen vergleichen zu können. In der vorliegenden Arbeit werden für praktische Beispiele Copulas angepasst. Dabei handelt es sich stets um den gleichen Datensatz. Es wird nur ein Copulaparameter geschätzt, da die Dissertation Qualitätsregelkarten auf Basis archimedischer Copulas betrachtet. Somit sind die Anforderungen an die Schätzmethode überschaubar und die Anpassungsgüte verschiedener Copulamodelle ist direkt vergleichbar.

Die **ML-Methode** („Maximum Likelihood“) liefert beispielsweise eine Möglichkeit, nach Parameterschätzung für mehrere Copulas diejenige mit der besten Anpassung auszuwählen. Dazu werden die ML-Werte verglichen. Alternativ können **Anpassungstests** helfen, eine geeignete Copula zu finden.

Bei der Anpassung einer Copula mit der ML-Methode tritt das Problem auf, dass

die  $u_i \in I^2$ ,  $i = 1 \dots n$ , meist nicht direkt beobachtet werden. Stattdessen liegen die Beobachtungen  $x_i \in \mathbb{R}^2$ ,  $i = 1 \dots n$ , vor, denen die gemeinsame Verteilung  $H$  zugrunde liegt. Es muss also zunächst eine Transformation erfolgen. Nach Sklar's Theorem ist  $u_i$  entsprechend der Copula verteilt, wenn  $u_i := (F_1(x_{i1}), F_2(x_{i2}))$  gewählt wird mit den Randverteilungen  $F_1$  und  $F_2$ . Diese Methode setzt bekannte Randverteilungen voraus. Hierbei spricht man auch von der **IFM-Methode** („Inference for Margins“).

Unbekannte Randverteilungen können beispielsweise durch die entsprechenden empirischen Verteilungsfunktionen geschätzt werden. Dann ergibt sich eine Stichprobe von Pseudorealisationen  $\hat{u}_i$ ,  $i = 1, \dots, n$ , mit

$$\hat{u}_i := (\hat{F}_1(x_{i1}), \hat{F}_2(x_{i2})).$$

Wird die log-Likelihood-Funktion der Pseudorealisationen betrachtet und für  $\theta$  maximiert, spricht man von der **CML-Methode** („Canonical Maximum Likelihood“).

Die Parameterschätzung per **Rangkorrelation** setzt voraus, dass eine Beziehung zwischen dem Copulaparameter und einem Rangkorrelationskoeffizienten (Spearman, Kendall) existiert. Dabei wird auf die Rangkorrelation zurückgegriffen, weil sie nur von der Copula abhängt, während die Korrelation nach Pearson auch durch die Randverteilungen beeinflusst wird. Dieser Zusammenhang wurde in Abschnitt 3.5 für Gauß-, Clayton- und Gumbelcopula gezeigt und wird in Tabelle 3.2 genutzt, um aus der geschätzten Korrelation der Daten einen Schätzer für den Copulaparameter abzuleiten.

Copula	<b>Gaußcopula</b>	<b>Claytoncopula</b>	<b>Gumbelcopula</b>
Parameterschätzer	$\hat{\rho} = \sin(\frac{\pi\hat{\tau}}{2})$	$\hat{\theta} = \frac{2\hat{\tau}}{1-\hat{\tau}}$	$\hat{\theta} = \frac{1}{1-\hat{\tau}}$

Tabelle 3.2: Parameterschätzer für verschiedene Copula, dargestellt als Funktion eines Abhängigkeitsmaßes



# 4 Karten zur Prozessüberwachung

## 4.1 Qualität und Statistik

Der Erfolg eines produzierenden Unternehmens hängt unter anderem von dessen Fähigkeit ab, Kunden termingerecht, kostengünstig und deren Qualitätserwartungen entsprechend beliefern zu können. Damit beeinflusst die Qualitätsstrategie eines Unternehmens dessen langfristige Überlebenschancen stark. Einen Überblick über die Anwendungsmöglichkeiten der Statistik in der Qualitätssicherung liefern Montgomery (2009) sowie Weihs und Jessenberger (1999).

Die Qualität eines Produktes ist die Gesamtheit von Eigenschaften und Merkmalen, die sich auf dessen Eignung zur Erfüllung festgelegter oder vorausgesetzter Erfordernisse beziehen. Montgomery (2009) definiert Qualität auf zwei Arten, einmal durch „Qualität bedeutet Gebrauchstauglichkeit“ und zum anderen mit „Qualität ist umgekehrt proportional zur Variabilität“. Die Statistik liefert wesentliche Werkzeuge, um Qualitätsvereinbarungen von Kunden und Lieferanten übersichtlich zu machen. Statistische Methoden der Qualitätskontrolle dienen zur Überwachung und Verbesserung eines Prozesses, beispielsweise verschiedene Regelkarten, statistische Versuchsplanung, die Fehler-, Möglichkeits- und Einflussanalyse (FMEA), Prozessfähigkeitsindizes oder die Methodensammlung „Six-Sigma“. Mehr Informationen zu den genannten Methoden bieten auch Yang und El-Haik (2003). Weihs und Jessenberger (1999) teilen den Qualitätssicherungsprozess in drei Bereiche ein: Schwachstellenanalyse, Optimierung und Stabilisierung.

Die **statistische Prozessüberwachung SPC** („Statistical Process Control“) soll Qualitätsdefizite bereits während der Produktion entdecken und ihnen durch steuernde Eingriffe entgegenwirken. Außerdem wird eine geringere Variabilität der interessierenden Kennwerte angestrebt, so dass der Prozess stabiler läuft.

Die SPC ist dabei nicht geeignet, die Qualität von Produkten zu erhöhen. Entsprechend der Definition von Weihs und Jessenberger (1999) wird sie in der Phase der „Stabilisierung des Prozesses“ eingesetzt und dient lediglich dazu, ein festgelegtes Maß an Qualität möglichst einzuhalten. Eine über das benötigte Maß hinausgehende Qualität hätte zusätzliche Kosten zur Folge und wird daher nicht angestrebt.

Als Pionierwerk der SPC gilt das Buch von Shewhart (1931). Deming (1975) übertrug die Erkenntnisse und Werkzeuge der SPC auf andere Prozesse, etwa Verwaltungs- oder Geschäftsprozesse und erreichte auch dort Verbesserungen. Er nutzte die Methoden in den Vereinigten Staaten während des Zweiten Weltkrieges im Bereich der Munitionsproduktion. Nach dem Krieg stellte er die Techniken der japanischen Industrie vor, wo sie unter anderem innerhalb des Toyota-Produktionssystems weiterentwickelt wurden.

Laut Shewhart (1931) ist jeder Produktionsprozess Störkomplexen ausgesetzt. Dabei ist es egal, wie gut er durchgeführt wird: Auf jede Produktion wirken zufällige und nicht zufällige („systematische“) Ursachen ein. Zufällige Einwirkungen sind durch die Herstellung selbst bestimmt und in ihrer Wirkung nicht zu beseitigen. Ein Beispiel dafür sind Vibrationen in den Maschinen. Dem gegenüber sind systematische Ursachen lokalisierbar und beeinflussbar. Sie führen zu allmählichen oder plötzlichen Veränderungen (beispielsweise Werkzeugverschleiß, Temperaturanstieg oder Ermüdung des Personals). Schwankungen der Merkmalsausprägungen sind also unvermeidlich und sollten innerhalb eines Toleranzbereichs um den Sollwert zugelassen werden. Ein Prozess, der nur mit zufälligen Schwankungen läuft, wird als **statistisch unter Kontrolle** und ein Prozess mit systematischen Abweichungen wird als **statistisch außer Kontrolle** bezeichnet.

Mit Hilfe der SPC soll das Auftreten von systematischen Ursachen möglichst schnell erkannt werden, so dass Nachforschungen unternommen und korrigierende Maßnahmen eingeleitet werden können. Die **Qualitätsregelkarte (QRK)** ist ein effektives Werkzeug zur Prozessüberwachung und wird im folgenden Abschnitt vorgestellt.



## 4.2 Qualitätsregelkarten

**Prozess- beziehungsweise Qualitätsregelkarten (QRK)** wurden von Shewhart (1931) entwickelt und sind ein Frühwarnsystem, um Fehler und Ausschuss zu vermeiden.

Kontrollkarten werden zur Auswertung von Prüfdaten eingesetzt. Beispiele für überwachbare Merkmale sind Länge oder Durchmesser in Zentimeter, Gewicht in Gramm, Wartezeit in Minuten oder die Anzahl von Verkaufsgesprächen. Die Karten sind also nahezu universell einsetzbar, „lohnenswert“ sich aber meist erst bei der Anwendung auf große Stückzahlen. Klassische Einsatzfelder liegen in der Industrie, im Bereich der (Intensiv-)Medizin oder auch im Verwaltungs- oder Dienstleistungssektor.

QRK betrachten die Realisationen von Zufallsvariablen eines Prozesses. Für Prozesse unter statistischer Kontrolle ist die Verteilung der interessierenden Merkmale konstant und das Prozessverhalten ist prognostizierbar. Dies entspricht dem  $H_0$ -Szenario. Bei einem Prozess außer Kontrolle kann sich beispielsweise der Erwartungswert oder Erwartungswertvektor der Zufallsvariablen verschieben oder die Variabilität der Messwerte ändern. Es kann auch passieren, dass die Beobachtungen im Außer-Kontrolle-Fall autokorreliert werden.

Gerät der Prozess außer Kontrolle, sollte das natürlich schnell entdeckt werden, um eingreifen und die Ursache beheben zu können. Dies geschieht mit Hilfe von QRK. Einige Karten beziehen dabei in ihre aktuellen Teststatistiken Messungen aus der Vergangenheit ein, um Trends im Prozessverhalten zu berücksichtigen (siehe etwa Roberts (1959) oder Page (1954)). Im Folgenden werden aber nur Karten „ohne Gedächtnis“ betrachtet.

Meist werden zur **Konstruktion einer Kontrollkarte** unabhängige metrische Qualitätsdaten vorausgesetzt. Der Produktion wird eine Stichprobe entnommen und das interessierende Merkmal (beziehungsweise die interessierenden Merkmale) als Realisierung einer (oder mehrerer) Zufallsvariablen betrachtet. Nach Mittag (1993) müssen dabei folgende Annahmen für den Prozess unter Kontrolle gelten:

1. Die sukzessiven Messwerte jeder gezogenen Einzelstichprobe sind unabhängig identisch verteilt (iiv) mit endlicher Varianz.

2. Stichprobenvariablen von unterschiedlichen Zeitpunkten sind uiv.
3. Die genaue Verteilung oder wenigstens die Verteilungsfamilie der Stichprobenvariablen ist bekannt.
4. Messfehler der Stichprobenvariablen sind vernachlässigbar.

Der dritte Punkt entfällt, wenn statt eines verteilungsbasierten Tests eine nicht-parametrische Alternative verwendet wird. Die Messwerte des zu überwachenden Qualitätsmerkmals werden entweder direkt als Einzelwerte weiterverwendet oder zu einer Funktion verdichtet. Diese Funktion ist im einfachsten Fall der Mittelwert, der Median oder die Varianz der Werte.

Die daraus abgeleitete Teststatistik wird gegen die Zeit beziehungsweise die Anzahl entnommener Proben abgetragen. Untersucht wird dann mit Hilfe eines statistischen Signifikanztests:

$H_0$ : Der Prozess ist unter Kontrolle

gegen

$H_1$ : Der Prozess ist außer Kontrolle

Widersprechen die erhobenen Messwerte der Nullhypothese  $H_0$  nicht, ist keine Aktion vom Anwender nötig. Eine Testentscheidung zugunsten der Alternativhypothese  $H_1$  wird getroffen, wenn der Stichprobenbefund den unter  $H_0$  erwarteten Resultaten widerspricht. Die Teststatistik liegt dann außerhalb der Eingriffsgrenzen. Der Prozess ist damit als nicht mehr unter statistischer Kontrolle anzusehen und ein korrigierender Eingriff in den Fertigungsablauf ist nötig.

Zusätzlich gibt es noch die Möglichkeit „unter  $H_0$  unwahrscheinlichere“ Ereignisse zu betrachten, welche aber  $H_0$  noch nicht widersprechen. In diesem Fall liegt der Stichprobenbefund zwischen den sogenannten Warn- und Eingriffsgrenzen und ein Verdacht auf Eintritt einer Störung liegt vor. Dieser Verdacht versetzt den Prüfer in „Alarmbereitschaft“: Häufigere Probenentnahme und/oder eine Erhöhung der Stichprobengröße zur schnellen Entdeckung eventueller Fehler sind denkbare Reaktionen.

Die Warn- und Kontrollgrenzen der QRK bestimmen sich durch die Grenzen des  $(1-\alpha)$ -Konfidenzintervalls der Teststatistik unter  $H_0$ . Abbildung 4.1 stellt diesen Zu-

sammenhang dar. Üblich ist, als Warngrenze  $\alpha = 0.10$  (oder  $\alpha = 0.05$ ) und als Kontrollgrenze  $\alpha = 0.05$  (oder  $\alpha = 0.01$ ) zu setzen.

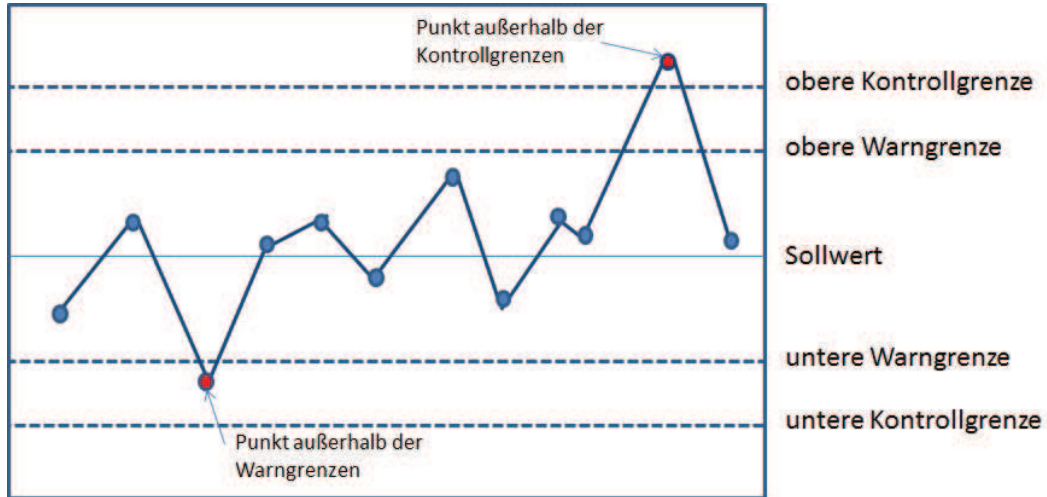


Abbildung 4.1: Beispielhafte Kontrollkarte

Multivariate Karten werden für mehrdimensionale Daten verwendet, deren Objekteigenschaften voneinander abhängen, etwa Füllgewicht und Volumen von Mehltüten. Einen guten Einblick in die Arbeitsweise der verschiedenen multivariaten Karten bietet Ryan (2011). Das bekannteste Beispiel ist Hotellings  $T^2$ -Karte (Hotelling (1947)). Sie ist optimal geeignet zur Überwachung von Lage- und Streuungsänderungen multivariat normalverteilter Daten. Diese ursprüngliche  $T^2$ -Karte wurde zudem vielfach neuen Bedürfnissen angepasst, so dass inzwischen verschiedene Varianten und Unterformen existieren (siehe Oyeyemi (2011)).

Bei den meisten multivariaten Karten handelt es sich um Kontrollkarten bezüglich der Prozesslage, nur wenige überwachen die Streuung eines Prozesses. Beispiele dafür sind die  $S^2$ -Karte oder die Influenzfunktions-Varianz-Karte (siehe etwa Montgomery (2009), Aparasi et al. (1999) oder Jaupi (2002)).

Die meisten statistischen Tests können als Kontrollkarten aufbereitet werden. Daher sind die Anwendungsfelder der Copulamodelle aus Abschnitt 3 prinzipiell auf Kontrollkarten übertragbar. Anstelle der gemeinsamen Modellierung von Pegelständen durch eine Copulafunktion könnte beispielsweise die Frage betrachtet

werden, ab welcher Höhe der Pegelstände von „Flutgefahr“ gesprochen werden muss. Da Copulas Zusammenhänge modellieren, sind hier allerdings nur multivariate Qualitätsregelkarten sinnvoll.

### 4.3 Die mittlere Lauflänge ARL

Die **Lauflänge RL** („Run Length“) ist die Zeitspanne bis zu einem Eingriff in den Prozess. Die RL ist eine Zufallsvariable mit Erwartungswert, Varianz und einer Verteilung. Sie existiert für jede Qualitätsregelkarte. Der Erwartungswert der Lauflänge ist die **mittlere Lauflänge ARL** („Average RL“), also die mittlere Zeitspanne bis zum Eingreifen in den Prozess. Sie ist darstellbar als

$$\begin{aligned} \text{ARL}(\eta) &:= \text{E}(\text{RL}(\eta) \mid \text{Verteilungsannahme unter } H_0) \\ &= \text{P}(\text{T liegt außerhalb der Kontrollgrenzen} \mid \eta)^{-1}. \end{aligned}$$

Dabei kann  $\eta$  beliebige Werte aus dem Wertebereich des interessierenden Prozessparameters annehmen (vergleiche Weihs und Jessenberger (1999)). Die Wahrscheinlichkeit eines Ereignisses ist durch  $\text{P}(\cdot)$  gegeben, T steht für die gewählte Teststatistik. Die Kontrollgrenzen ergeben sich aus der Verteilungsannahme unter  $H_0$ .

Ist der Prozess unter Kontrolle ( $\eta$  entspricht dem **Sollwert**  $\eta_0$ ), hängt die ARL direkt von  $\alpha$  ab. Hier gilt

$$\text{ARL}(\eta_0) = 1/\alpha.$$

Mit zunehmender Abweichung der Stichprobe von der Sollverteilung nimmt die ARL ab. Dieser Erwartungswert kann nicht direkt berechnet werden, das Verhalten der ARL-Funktion wird daher durch Simulationen ermittelt. Neben der ARL ist auch die Standardabweichung der Lauflängen interessant, die ebenfalls mittels Simulation über eine große Anzahl von Stichproben approximiert werden muss. Allgemein gilt für die **Standardabweichung der RL (SDRL)**:

$$\text{SDRL}(\eta) := \sqrt{\text{Var}(\text{RL}(\eta) \mid \text{Verteilungsannahme unter } H_0)}.$$

## 4.4 OC-Funktion

Durch die Wahl des  $\alpha$  wird die Wahrscheinlichkeit für einen Fehler erster Art eines Tests bestimmt. Der Fehler zweiter Art (auch  $\beta$ -Fehler, das heißt der Prozess wird irrtümlich als unter Kontrolle angesehen) wird durch die **Operationscharakteristik**, die sogenannte **OC-Funktion**, ausgedrückt. Unter Beibehaltung der Notation aus Abschnitt 4.3 gilt:

$$OC(\eta) := P(T \text{ liegt innerhalb der Kontrollgrenzen} \mid \eta)$$

$OC(\eta)$  gibt demzufolge die Wahrscheinlichkeit an, dass der Prozess als unter Kontrolle gilt, wenn der interessierende Prozessparameter den Wert  $\eta$  annimmt. Unter Idealbedingungen hätte die OC-Funktion an der Stelle des Sollwertes  $\eta_0$  den Wert 1 und sonst den Wert 0. In der Realität ist die OC-Funktion meist glockenförmig. Dabei gilt: Je näher die Funktion dem Ideal kommt, umso besser ist die Güte der zugehörigen Kontrollkarte oder je größer ihr Wert abseits des Sollwertes ist, umso ineffizienter ist die Kontrollkarte.

Für unabhängige Beobachtungen gilt bei Kontrollkarten ohne Gedächtnis

$$OC(\eta) = 1 - \frac{1}{ARL(\eta)}. \quad (4.1)$$

## 4.5 Hotellings $T^2$ -Karte

Die Hotelling  $T^2$ -Karte ist eine Karte für multivariat normalverteilte Einzelbeobachtungen, um den Prozess auf Abweichungen von den Sollwerten für das Prozessmittel sowie die Prozessstreuung (durch Betrachtung der Kovarianzmatrix) zu überwachen, siehe Hotelling (1947). Dabei ist sie unter den geforderten Voraussetzungen die optimale Kontrollkarte; hier sinkt die ARL im Außer-Kontrolle-Fall also am schnellsten, während das Niveau im In-Kontrolle-Fall eingehalten wird.

Im Außer-Kontrolle-Fall muss anschließend noch das alarmlösende Merkmal gefunden werden. Dies kann beispielsweise durch gleichzeitige, separate Verwendung univariater Kontrollkarten geschehen. Es gibt verschiedene Versionen der Karte, die vorgestellte nutzt Schätzwerte für das Prozessmittel und die Prozessvarianz.

Zunächst wird eine Zufallsstichprobe des Umfangs  $m$  des Prozesses unter statis-

tischer Kontrolle gezogen. Die Beobachtungen  $Y_1, \dots, Y_m \in \mathbb{R}^p$  seien unabhängig identisch  $\mathbb{N}(\mu, \Sigma)$ -verteilt. Sie werden zur Berechnung des Mittelwertvektors ( $\bar{Y}_m$ ) und der Stichprobenkovarianzmatrix ( $S_m$ ) genutzt.

Die Teststatistik  $T^2$  basiert auf der Mahalanobisdistanz. Für jede neue unabhängige Beobachtung  $X \in \mathbb{R}^p$  gilt

$$T^2(X) := (X - \bar{Y}_m)^T S_m^{-1} (X - \bar{Y}_m). \quad (4.2)$$

Unter der Nullhypothese ist die Teststatistik  $T^2$  nach entsprechender Gewichtung  $\mathbb{F}_{p, m-p}$ -verteilt. Die F-Verteilung ergibt sich aus der Tatsache, dass

$$X \sim \mathbb{N}(\mu, \Sigma) \quad \text{und} \quad \bar{Y}_m \sim \mathbb{N}(\mu, m^{-1}\Sigma) \quad \text{sowie} \quad mS_m \sim \mathbb{W}(\Sigma, m-1),$$

wobei  $\mathbb{W}$  für die Wishart-Verteilung steht, also die multivariate Form der  $\chi^2$ -Verteilung. Die Differenz von  $X$  und  $\bar{Y}_m$  ist normalverteilt mit dem Mittelwert null und der Kovarianzmatrix  $\frac{m+1}{m}\Sigma$ . Da  $X$  und  $\bar{Y}_m$  jeweils unabhängig von  $S_m$  sind, ist auch ihre Differenz ( $X - \bar{Y}_m$ ) unabhängig von  $S_m$ . Damit gilt (siehe etwa Mardia et al. (1995)), dass

$$m(X - \bar{Y}_m)^T \left( \frac{m+1}{m} mS_m \right)^{-1} (X - \bar{Y}_m) \sim \mathbb{T}_{p, m-1}^2$$

wobei  $\mathbb{T}_{p, m-1}^2$  für Hotellings T-Quadrat-Verteilung mit  $p$  und  $m-1$  Freiheitsgraden steht. Zwischen ihr und der F-Verteilung gilt der Zusammenhang

$$\mathbb{T}_{k, n}^2 = \frac{nk}{n-k+1} \mathbb{F}_{k, n-k-1} \quad \text{für beliebige } k < n \in \mathbb{R}.$$

Damit ist

$$m(X - \bar{Y}_m)^T \left( \frac{m+1}{m} mS_m \right)^{-1} (X - \bar{Y}_m) \sim \frac{(m-1)p}{m-p} \mathbb{F}_{p, m-p}$$

und schließlich

$$(X - \bar{Y}_m)^T S_m^{-1} (X - \bar{Y}_m) \sim \frac{(m+1)(m-1)p}{m(m-p)} \mathbb{F}_{p, m-p}.$$

Die obere Kontrollgrenze  $\text{UCL}_\alpha$  zum Niveau  $\alpha$  ist dann das entsprechende  $(1 - \alpha)$ -

Quantil der hergeleiteten F-Verteilung, also

$$\text{UCL}_\alpha := \frac{p(m+1)(m-1)}{m^2 - mp} \mathbb{F}_{1-\alpha, p, m-p}.$$

Simulationen zur Lauflänge zeigen, wie sich Änderungen der Prozessverteilung unter  $H_0$  auf die Karte auswirken. In Kapitel 6.3 wird so beispielsweise untersucht, wie die  $T^2$ -Karte sich für Werte verhält, deren Verteilung durch verschiedene Copulastrukturen erklärbar ist.

## 4.6 Steigers $Z$ -Test

Einen Ansatz, Änderungen in der Abhängigkeitsstruktur von Daten zu testen, erläutert Steiger (1980). Steigers  $Z$ -Test nutzt den Korrelationskoeffizienten nach Pearson  $\rho$  (Formel 2.2). Die **Fishertransformation**  $f(\cdot)$  korrigiert die Tatsache, dass für bivariat normalverteilte Daten der aus  $n$  Datenpunkten resultierende Korrelationskoeffizient  $\rho$  unimodal rechtssteil verteilt ist. Die (fischer-)transformierte Variable

$$f(\rho) = 0.5 \ln \left( \frac{1 + \rho}{1 - \rho} \right)$$

ist dann annähernd normalverteilt. Daher setzt der  $Z$ -Test näherungsweise bivariat normalverteilte Daten voraus und testet im zweiseitigen Fall:

$$H_0 : \rho = \rho_0 \quad \text{vs.} \quad H_1 : \rho \neq \rho_0$$

Die Teststatistik ist gegeben durch

$$T(r) = \frac{f(r) - f(\rho_0) - \rho_0/(n-2)}{1/\sqrt{n-3}},$$

sie ist unter  $H_0$  annähernd  $N(0, 1)$ -verteilt.

Eine Qualitätsregelkarte, welche auf diesem Test basiert, verwendet also den zu  $T(r)$  modifizierten empirischen Korrelationskoeffizienten der Stichprobe und vergleicht diesen mit dem entsprechenden oberen und unteren Quantil der Standardnormalverteilung.





# 5 Variante I: Überwachung von Prozesslage und -streuung

Yan (2007) erläutert, dass Copulastrukturen neben den typischen Anwendungsfeldern in der Finanzwelt auch zur Beschreibung von Produktionsprozessen denkbar sind. Darauf bezugnehmend werden in diesem Abschnitt Karten zur Lage- und Streuungsüberwachung konstruiert. Die interessierenden Prozessdaten sollen dazu in ihrer gemeinsamen Verteilung einer gegebenen bivariaten Copula mit normalverteilten Rändern folgen.

Im Abschnitt 5.1 werden die Kontrollkarten beschrieben. Ihr Güteverhalten für verschiedene Prozessszenarien wird durch Simulationsstudien in Kapitel 5.2 untersucht. Dabei zeigt sich, dass die adjustierten  $T^2$ -Karten Abweichungen im Prozessmittel für alle untersuchten Copulas gut erkennen. Die  $E^2$ -Karte basiert auf einem alternativen Distanzmaß. Sie übertrifft die adjustierten  $T^2$ -Karten in einigen Fällen der Prozessverschiebung. Schließlich werden die entwickelten Karten in den Kapiteln 5.3.1 und 5.3.2 auf praktische Beispiele angewendet.

In Abschnitt 5.4 wird abschließend untersucht, inwiefern die klassische  $T^2$ -Karte Abweichungen in der Streuung einzelner Ränder aufdeckt, wenn Copulastrukturen für die Daten verwendet werden.

## 5.1 Verwendete Kontrollkarten

Traditionell wird zur Lageüberwachung korrelierter Daten Hotellings  $T^2$ -Karte genutzt. Für verschiedene Copulatypen aus Kapitel 3, Tabelle 3.1, werden im Folgenden Kontrollkarten konstruiert. Die vorgestellten Karten basieren auf verschiedenen Distanzmessungen einer Beobachtung zum erwarteten beziehungsweise geschätzten Mittel der Grundgesamtheit, siehe auch Abschnitt 2.3.

### 5.1.1 Adjustierte $T^2$ -Karte

Sind die betrachteten Daten unter  $H_0$  nicht multivariat normalverteilt, verliert die Qualitätsregelkarte nach Hotelling aus Kapitel 4.5 ihre Optimalitätseigenschaften, etwa die Einhaltung des Niveaus sowie das schnelle Absinken der ARL unter  $H_1$ . Verglichen mit Konkurrenzkarten sinkt die ARL also unter Umständen nicht mehr am schnellsten. Auch das Maximum der ARL muss nicht mehr bei  $\eta_0$  erreicht werden. Die Verteilung der  $T^2$ -Statistik ändert sich in diesem Fall und damit auch die Kontrollgrenzen. Verteilungsaussagen sind für nicht multivariat normalverteilte Daten schwierig, siehe Abschnitt 4.5. Daher werden hier adjustierte, also angepasste, Karten auf Basis der  $T^2$ -Statistik erzeugt.

Um möglichst viele Datenstrukturen abbilden zu können, werden die Frank-, Clayton- und Gumbelcopula verwendet, siehe auch Abbildung 3.1. Die Gaußcopula (also die multivariate Normalverteilung) als vierte Möglichkeit erfordert keine Adjustierung der traditionellen  $T^2$ -Karte nach Hotelling. Die Kontroll- und Warngrenzen der klassischen  $T^2$ -Karte hängen von der Anzahl betrachteter Merkmale  $p$  und der Größe der Lernstichprobe  $m$  ab. Die Grenzen der adjustierten Karten variieren außerdem mit unterschiedlichem Maß der Abhängigkeit zwischen den betrachteten Variablen.

Die „neuen“ Kontrollgrenzen werden mit Hilfe einer Trainingsstichprobe aus einem Prozess unter Kontrolle geschätzt, dabei werden per Monte-Carlo-Simulation Stichproben erzeugt und daraus die benötigten Kennzahlen berechnet. Mit wachsender Stichprobengröße wird die Verteilung der Teststatistik immer besser angenähert und die empirischen Quantile sind gute Schätzer der theoretischen Quantile. Abbildung B.1 im Anhang zeigt das Flussdiagramm der Grenzwertsimulation. Die Simulation zur Bestimmung der Grenzwerte verläuft folgendermaßen:

1. Kendalls  $\tau$  auswählen, Copulafamilie festlegen mit standardnormalverteilten Rändern ( $Z_i \sim \mathcal{N}(0, 1)$ ,  $i = 1, 2$ ) und den Parameter  $\theta$  entsprechend  $\tau$  wählen. Damit gilt

$$H_0 : H(Z_1, Z_2) = C(F(Z_1), F(Z_2), \theta)$$

2.  $m$  bivariate Zufallvektoren  $(Y_1, \dots, Y_m)$  aus  $H(Z_1, Z_2)$  generieren und daraus  $S_m$  und  $\bar{Y}_m$  berechnen

3. Weitere  $B$  Zufallsvektoren  $X_i$ ,  $i = 1, \dots, B$  aus  $H(Z_1, Z_2)$  erzeugen

4. Berechnung von  $T^2(X_i) := (X_i - \bar{Y}_m)^T S_m^{-1} (X_i - \bar{Y}_m)$

Schritt 2. bis 4. wird  $K$ -mal wiederholt. Die entsprechenden Quantile der  $B \cdot K$  simulierten  $T^2$ -Werte sind die ermittelten (empirischen) Kontrollgrenzen.

Einige Beispielergebnisse für verschiedene Verteilungen unter  $H_0$  und Niveaus  $\alpha$  zeigt Tabelle 5.1 in Abschnitt 5.2.

### 5.1.2 Alternative: Euklidische $E^2$ -Karte

Die  $T^2$ -Statistik repräsentiert den normierten Abstand von  $X$  zum Mittelwertvektor einer elliptischen Datenverteilung. Dazu wird die Mahalanobisdistanz verwendet.

Die genannten Copulas modellieren aber nichtelliptische Datenstrukturen. Daher wird ein alternatives Distanzmaß zur Konstruktion einer Prozessregelkarte verwendet. Als Teststatistik bietet sich die Euklidische Distanz der (einzeln) normierten Ursprungsdaten an. Dieses Maß setzt unter  $H_0$  keine elliptisch um den Sollwert verteilten Daten voraus, so dass es für die unter  $H_0$  achsensymmetrisch angeordneten Punkte gute Resultate liefert.

Analog zu Hotellings  $T^2$ -Karte wird zunächst eine Zufallstichprobe vom Umfang  $m$   $Y_1, \dots, Y_m \in \mathbb{R}^p$  gezogen und daraus der  $p$ -dimensionale Mittelwertvektor  $\bar{Y}_m$  bestimmt. Zusätzlich wird der Vektor der für jedes Merkmal separat berechneten Standardabweichungen  $s_m^* = (s_{m,1}^*, \dots, s_{m,p}^*)^T$  nach Formel (2.1) bestimmt. Für eine neue unabhängige Beobachtung  $X \in \mathbb{R}^p$  gilt dann:

$$E^2 := \left( \text{diag}(s_{m,1}^{-1}, \dots, s_{m,p}^{-1})(X - \bar{Y}_m) \right)^T \left( \text{diag}(s_{m,1}^{-1}, \dots, s_{m,p}^{-1})(X - \bar{Y}_m) \right) \quad (5.1)$$

Dabei steht  $\text{diag}$  für die Diagonalmatrix. Unterstellt man den Daten unter  $H_0$  Normalverteilung (wie beispielsweise bei Hotellings  $T^2$ -Karte in Abschnitt 4.5), ist die Verteilungsfunktion der Teststatistik  $E^2$  nicht geschlossen darstellbar. Es gilt zunächst analog zur  $T^2$ -Statistik, dass

$$X \sim \mathbb{N}(\mu, \Sigma) \quad \text{und} \quad \bar{Y}_m \sim \mathbb{N}(\mu, m^{-1}\Sigma),$$

beziehungsweise je Vektoreintrag  $X_i \sim \mathbb{N}(\mu_i, \sigma_i^2)$  und  $\bar{Y}_{m,i} \sim \mathbb{N}(\mu_i, m^{-1}\sigma_i^2)$ .

Außerdem ist

$$s_{m,i}^* \sim \frac{\sigma_i^2}{m-1} \chi_{m-1}^2 \quad \text{und} \quad (X_i - \bar{Y}_{m,i}) \sim \mathbb{N}(0, (\frac{m+1}{m})\sigma_i^2).$$

Weiter sind  $s_{m,i}^*$  und  $(X_i - \bar{Y}_{m,i})$  unabhängig. Demzufolge ist

$$\frac{X_i - \bar{Y}_{m,i}}{s_{m,i}^*} \sim \sqrt{\frac{m+1}{m}} \mathbb{t}_{m-1} \quad \text{für alle } i = 1, \dots, p.$$

Das Quadrat einer  $\mathbb{t}_{m-1}$ -verteilten Zufallsvariable ist  $\mathbb{F}_{1,m-1}$ -verteilt. Es gilt daher komponentenweise für  $i = 1, \dots, p$ :

$$\left( s_{m,i}^{*-1} (X_i - \bar{Y}_{m,i}) \right) \left( s_{m,i}^{*-1} (X_i - \bar{Y}_{m,i}) \right) \sim \frac{m+1}{m} \mathbb{F}_{1,m-1}$$

Die Summe der  $p$  Werte (also die Teststatistik  $E^2$ ) ist eine Faltungspotenz dieser F-Verteilung und damit nicht durch Standardverteilungen darstellbar. Die Grenzwerte für alle betrachteten Copulastrukturen werden simuliert. Wie bei den adjustierten  $T^2$ -Karten hängen die Kontrollgrenzen von  $p$ ,  $m$ , der zugrundeliegenden Datenstruktur und dem Abhängigkeitsmaß der Variablen ab. Die Simulation der Grenzwerte erfolgt analog zu Abschnitt 5.1.1. Die gesuchten Grenzen für verschiedene Copulas und  $\alpha$  zeigt Tabelle 5.2 in Abschnitt 5.2.

## 5.2 Effizienz der Kontrollkarten

Der Vergleich der Güteeigenschaften der Kontrollkarten aus Kapitel 4.6 und 5.1.1 erfolgt durch eine Simulationsstudie. Dazu muss der betrachtete Prozess einige Anforderungen erfüllen. Die Anzahl beobachteter Merkmale wird auf  $p = 2$  beschränkt, die Stichprobengröße zur Parameterschätzung sei  $n = 30$  und die einzelnen Merkmale seien univariat normalverteilt, also  $X = (X_1, X_2)^T$  sowie  $Y = (Y_1, Y_2)^T$  mit  $X_j, Y_j \sim \mathbb{N}(\mu_j, \sigma_j)$  für  $j = 1, 2$ .

### 5.2.1 Bestimmung der Kontrollgrenzen

Die Simulationsergebnisse gelten für beliebige normalverteilte Ränder, also für alle Werte von  $\mu_j$  und  $\sigma_j$ , weil die gewählten Methoden skaleninvariant sind. Prozessabweichungen werden daher nur für den Bereich von  $\mu_1 = -3$  bis  $\mu_1 = +3$  betrachtet.

Zunächst werden Kontrollgrenzen nach der Routine aus Abschnitt 5.1.1 simuliert mit  $B = 100$  und  $K = 200.000$ . Verschiedene Grade der Abhängigkeit werden durch unterschiedliche  $\tau$  nach Kendall beschrieben.

Für die adjustierten Karten ( $T_{adj}^2$ ) sind die Werte der oberen Kontrollgrenzen in Tabelle 5.1 aufgelistet und für die Euklidische Kontrollkarte ( $E^2$ ) in Tabelle 5.2.

Niveau $\alpha =$	10%	5%	1%	0.27%
$T^2(N)$	5.357	7.150	11.671	15.753
$T_{adj}^2(F, \tau = 0.2)$	5.381	7.223	11.981	16.388
$T_{adj}^2(F, \tau = 0.4)$	5.455	7.449	12.900	18.321
$T_{adj}^2(F, \tau = 0.6)$	5.637	7.970	14.873	22.175
$T_{adj}^2(F, \tau = 0.8)$	5.948	9.178	20.682	34.471
$T_{adj}^2(C, \tau = 0.2)$	5.381	7.260	12.181	16.818
$T_{adj}^2(C, \tau = 0.4)$	5.422	7.457	13.072	18.578
$T_{adj}^2(C, \tau = 0.6)$	5.582	8.069	15.776	24.162
$T_{adj}^2(C, \tau = 0.8)$	5.849	9.376	23.510	41.485
$T_{adj}^2(G, \tau = 0.2)$	5.405	7.305	12.244	16.873
$T_{adj}^2(G, \tau = 0.4)$	5.435	7.408	12.682	17.778
$T_{adj}^2(G, \tau = 0.6)$	5.474	7.584	13.595	19.853
$T_{adj}^2(G, \tau = 0.8)$	5.516	7.841	15.072	23.378

Tabelle 5.1: Simulierte Kontrollgrenzen der klassischen und adjustierten  $T^2$ -Karten für verschiedene zugrundeliegende Copulastrukturen (N=Normal, F=Frank, C=Clayton, G=Gumbel) und verschiedene Kendalls  $\tau$

Die erste Zeile in Tabelle 5.1 zeigt die Grenzwerte der traditionellen  $T^2$ -Karte, die anschließenden Zeilen enthalten die Resultate für verschiedene Copulas. Betrachtet werden verschiedene Wahrscheinlichkeiten für den Fehler erster Art. Das  $\alpha$  von 0.0027 entspricht dem gebräuchlichen  $3\sigma$ -Kontrollniveau bei einer  $T^2$ -Karte.

Für  $\tau = 0$  sind die Grenzen der  $T^2$ -Karte und der Copula-basierten Karten gleich. Die Grenzwerte der adjustierten Karten wachsen mit ansteigendem  $\tau$  und offensichtlich mit sinkendem  $\alpha$ . Die simulierten Grenzwerte weichen demzufolge umso weiter von den Werten der klassischen  $T^2$ -Karte ab, je stärker die Abhängigkeit der beiden Randverteilungen wird.

Tabelle 5.2 zeigt die UCL-Werte der  $E^2$ -Statistik für unabhängige Variablen, diese Werte sind für alle Copulas identisch. Anschließend sind die Grenzen für verschiedene Copulas gegeben. Wie bei Tabelle 5.1 steigen auch hier die UCL-Werte mit wachsendem  $\tau$  und sinkendem  $\alpha$ .

Niveau $\alpha =$	10%	5%	1%	0.27%
$E^2 (\tau = 0)$	2.268	2.612	3.311	3.822
$E^2 (N, \tau = 0.2)$	2.271	2.624	3.349	3.882
$E^2 (N, \tau = 0.4)$	2.288	2.667	3.456	4.036
$E^2 (N, \tau = 0.6)$	2.320	2.738	3.605	4.240
$E^2 (N, \tau = 0.8)$	2.375	2.835	3.781	4.472
$E^2 (F, \tau = 0.2)$	2.282	2.633	3.343	3.863
$E^2 (F, \tau = 0.4)$	2.322	2.692	3.427	3.956
$E^2 (F, \tau = 0.6)$	2.385	2.781	3.551	4.097
$E^2 (F, \tau = 0.8)$	2.449	2.886	3.711	4.290
$E^2 (C, \tau = 0.2)$	2.276	2.654	3.465	4.098
$E^2 (C, \tau = 0.4)$	2.312	2.727	3.613	4.290
$E^2 (C, \tau = 0.6)$	2.367	2.796	3.694	4.375
$E^2 (C, \tau = 0.8)$	2.424	2.870	3.778	4.460
$E^2 (G, \tau = 0.2)$	2.281	2.660	3.470	4.093
$E^2 (G, \tau = 0.4)$	2.318	2.742	3.638	4.310
$E^2 (G, \tau = 0.6)$	2.376	2.836	3.783	4.482
$E^2 (G, \tau = 0.8)$	2.424	2.910	3.905	4.634

Tabelle 5.2: Simulierte Grenzwerte der  $E^2$ -Karte für verschiedene Copulas (N=Normal, F=Frank, C=Clayton, G=Gumbel) und Kendalls  $\tau$

### 5.2.2 Verhalten der Lauflänge bei Lageverschiebung

Die Güte der erzeugten Karten wird mit Hilfe simulierter ARL und SDRL (siehe Abschnitt 4.3) für die Grenzwerte der Tabellen 5.1 und 5.2 verglichen.

Um Verschiebungen bezüglich der Mittelwerte aufzudecken, werden zwei verschiedene Szenarien betrachtet. Bei Szenario (I) wird der Lageparameter  $\mu_1$  (nur) einer Randverteilung verschoben.

Berücksichtigt man die Abhängigkeit der beiden Ränder, ist auch eine Änderung von  $\mu_2$  sinnvoll. Diese kann direkt aus der Verschiebung von  $\mu_1$  abgeleitet werden. Perfekt positiv linear abhängige Daten liegen auf einer Geraden mit der Steigung  $\rho = 1$ . Damit gilt auch  $\mu_2^* = \mu_1^*$ . Allgemein streuen abhängige Daten um eine Gerade mit der Steigung  $\rho$ . Für eine realitätsnahe Verschiebung des Datenschwerpunktes sollte der Punkt  $(\mu_1, \mu_2)$  also entlang dieser Geraden verschoben werden. Genutzt wird dazu der in Tabelle 3.2 dargestellte Zusammenhang von  $\rho$  und  $\tau$ . Danach ist  $\mu_2^* = \rho \cdot \mu_1^* = \sin(\frac{\pi}{2}\tau) \cdot \mu_1^*$ .

Abbildung 5.1 zeigt beide Varianten, den Mittelwert zu verschieben. Dabei ist  $\mu_1$  beide Male gleich weit vom Sollwert entfernt.

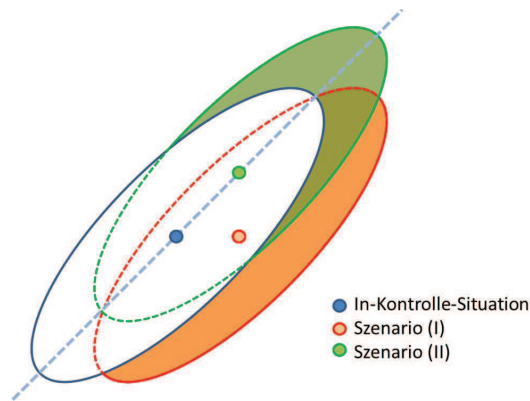


Abbildung 5.1: Veranschaulichung von Szenario (I) und (II) im Vergleich zur In-Kontrolle-Lage des Prozesses

Beide Untersuchungen verlaufen wie im Flussdiagramm aus Abbildung 5.2 beschrieben:

Zunächst wird ein  $N$ -dimensionaler Nullenvektor  $K$  erzeugt, der im Folgenden mit simulierten Lauflängen gefüllt wird. Anschließend wird die Copula festgelegt: Der Copulaparameter  $\theta$  entspricht einem festen Kendalls  $\tau$ , die Ränder sind standardnormalverteilt ( $X_i \sim \mathbb{N}(\mu_i = 0, 1)$ ,  $i = 1, 2$ ), es gilt:

$$H_0 : H(X_1, X_2) = C(F(X_1), F(X_2), \theta)$$

Zur Simulation der ARL und SDRL wird die Prozesslage „künstlich“ verschoben, entweder durch

- (I) Änderung von  $\mu_1$  zu  $\mu_1^*$  in der ersten Randverteilung oder durch
- (II) Wie (I), gleichzeitige Änderung von  $\mu_2$  auf  $\mu_2^* = \sin(\frac{\pi}{2}\tau) \cdot \mu_1^*$  im zweiten Rand.

Die so konstruierte Verteilung wird mit  $H^*(X_1, X_2)$  gekennzeichnet. Für die neuen Verteilungen werden  $N$  Lauflängen simuliert. Die Zählvariable  $j = 1, \dots, N$  gibt an, die wievielte RL gerade betrachtet wird. Für jede Lauflänge wird folgende Schleife durchlaufen:

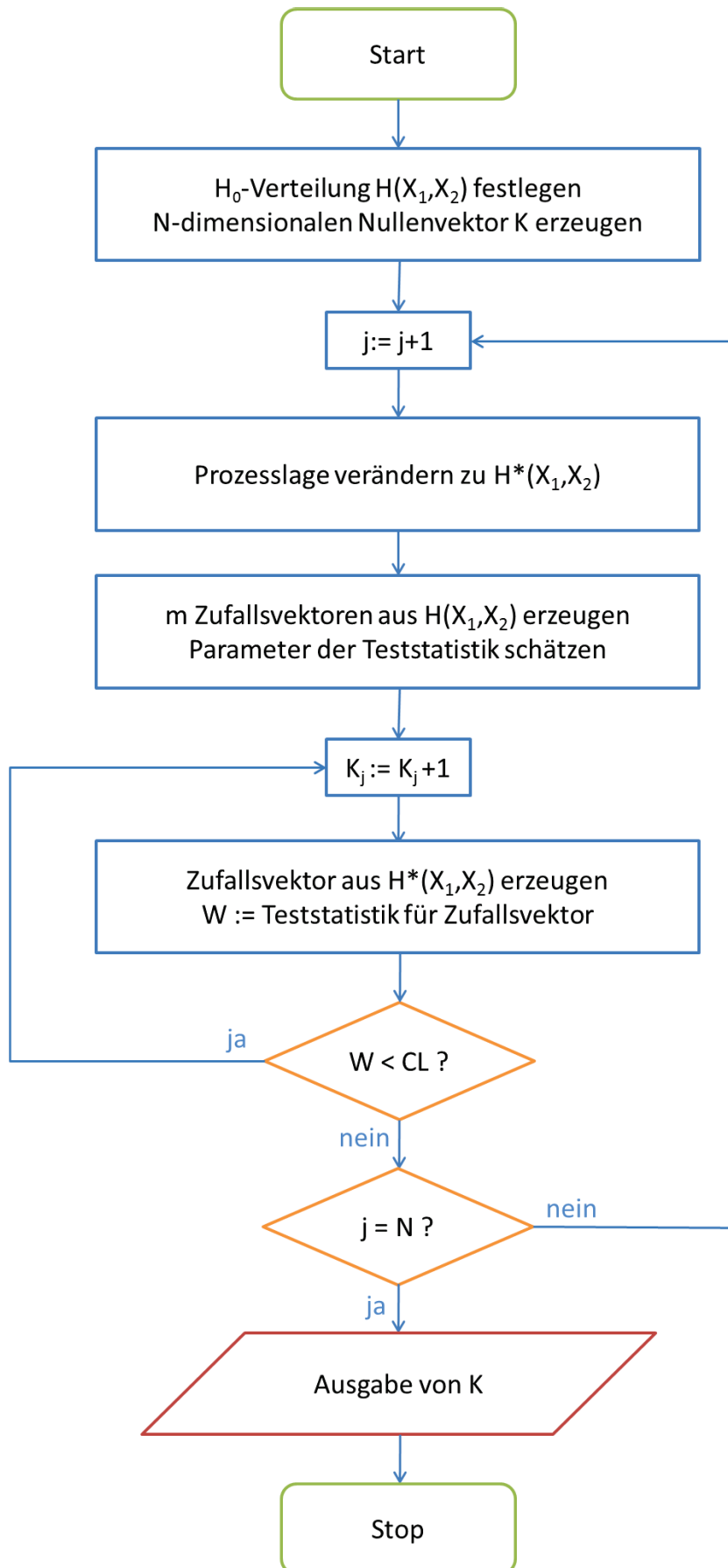


Abbildung 5.2: Flussdiagramm: Simulationsstudie zur Lageverschiebung des Prozesses



Erst werden  $m$  Zufallvektoren  $(Y_1, \dots, Y_m)$  aus der Verteilung des kontrolliert laufenden Prozesses  $H(X_1, X_2)$  generiert und daraus  $S_m$  beziehungsweise  $s_m^*$  und  $\bar{Y}_m$  berechnet. Diese Schätzer werden zur Berechnung der Teststatistik benötigt und bis zum Abbruch der aktuellen  $j$ -ten Lauflängensimulation beibehalten.

In einer zweiten Schleife werden solange Zufallsvektoren (Messwerte) aus der Verteilung  $H^*(X_1, X_2)$  erzeugt, bis die daraus abgeleitete Teststatistik größer als die zugehörige Kontrollgrenze wird. Als Teststatistik kann entweder  $T^2(X_i)$  nach Formel (4.2) verwendet werden oder  $E^2(X_i)$  laut Formel (5.1).

Für jeden generierten Messwert wird der Eintrag von  $K_j$  um eins erhöht. Nach dem Abbruch der Schleife entspricht  $K_j$  der  $j$ -ten simulierten Lauflänge des Prozesses.

Wurden alle  $j$  betrachtet, also  $N$  Lauflängen simuliert, wird der Lauflängenvektor  $K$  ausgegeben und die Simulation beendet.

Die ARL entspricht dann  $\bar{K}$  und die SDRL ist  $\sqrt{\text{Var}(K_1, \dots, K_N)}$ . In der folgenden Simulationsstudie erfolgen die Änderungen nach Szenario (I) und (II) schrittweise in 0.1-Intervallen von  $\mu_1$ .

Die Abbildungen 5.3 und 5.4 zeigen die Ergebnisse der Simulationsstudie in Form von ARL- und die SDRL-Kurven für die Karten aus Abschnitt 5.1 ( $\alpha = 0.05$ ,  $m=30$ ,  $B=100$ ,  $N=100.000$ ). Grafik 5.3 visualisiert dabei die Ergebnisse für die erste beschriebene Möglichkeit, wie ein Prozess außer Kontrolle geraten kann (Szenario I), Abbildung 5.4 zeigt die Kurven für das zweite beschriebene Szenario.

Die ARL wird (bei beiden Szenarien) durch die zugrundeliegende Datenbasis beeinflusst. Die ARL von Hotellings  $T^2$ -Karte ist bei der Anwendung bivariat normalverteilter Daten unabhängig vom gewählten  $\tau$  (jeweils oberes linkes Bild). Das ist durch die Konstruktion der  $T^2$ -Statistik begründet. Weiterhin gilt, dass die Ergebnisse für  $\tau = 0$  für alle Karten gleich ausfallen, da hier Unabhängigkeit zwischen den Variablen vorliegt.

Für beide Karten und Szenarien gilt, dass die ARL umso schneller abfällt, je größer die Abhängigkeit zwischen den Rändern wird, bei den Abbildungen wurde

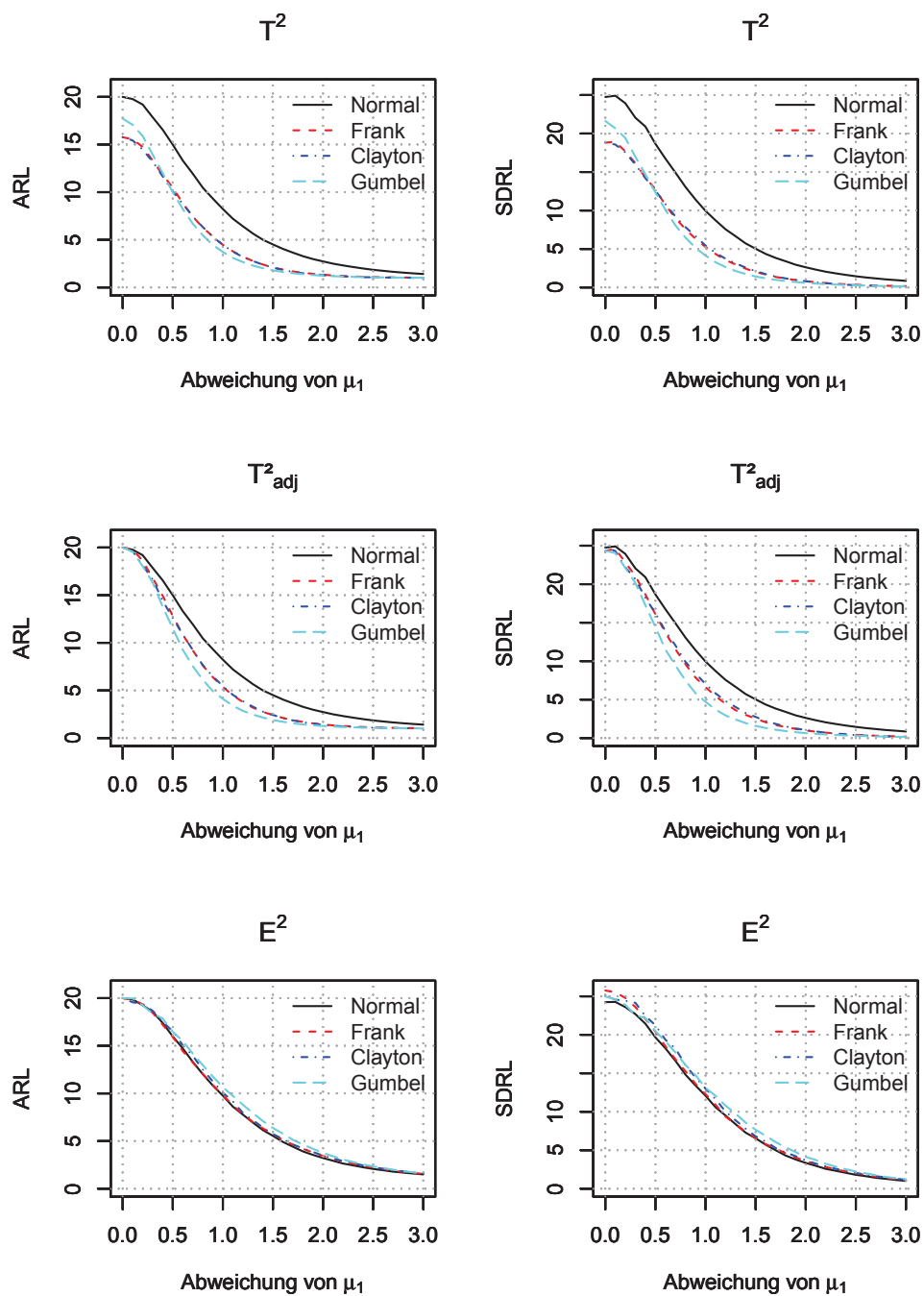


Abbildung 5.3: ARL und SDRL für verschiedene Kontrollkarten (klassische und adjustierte Hotellings  $T^2$ -Karten ( $T^2$  und  $T^2_{adj}$ ) sowie die Euklidische Karte ( $E^2$ ));  $\tau = 0.6$ , Szenario (I)

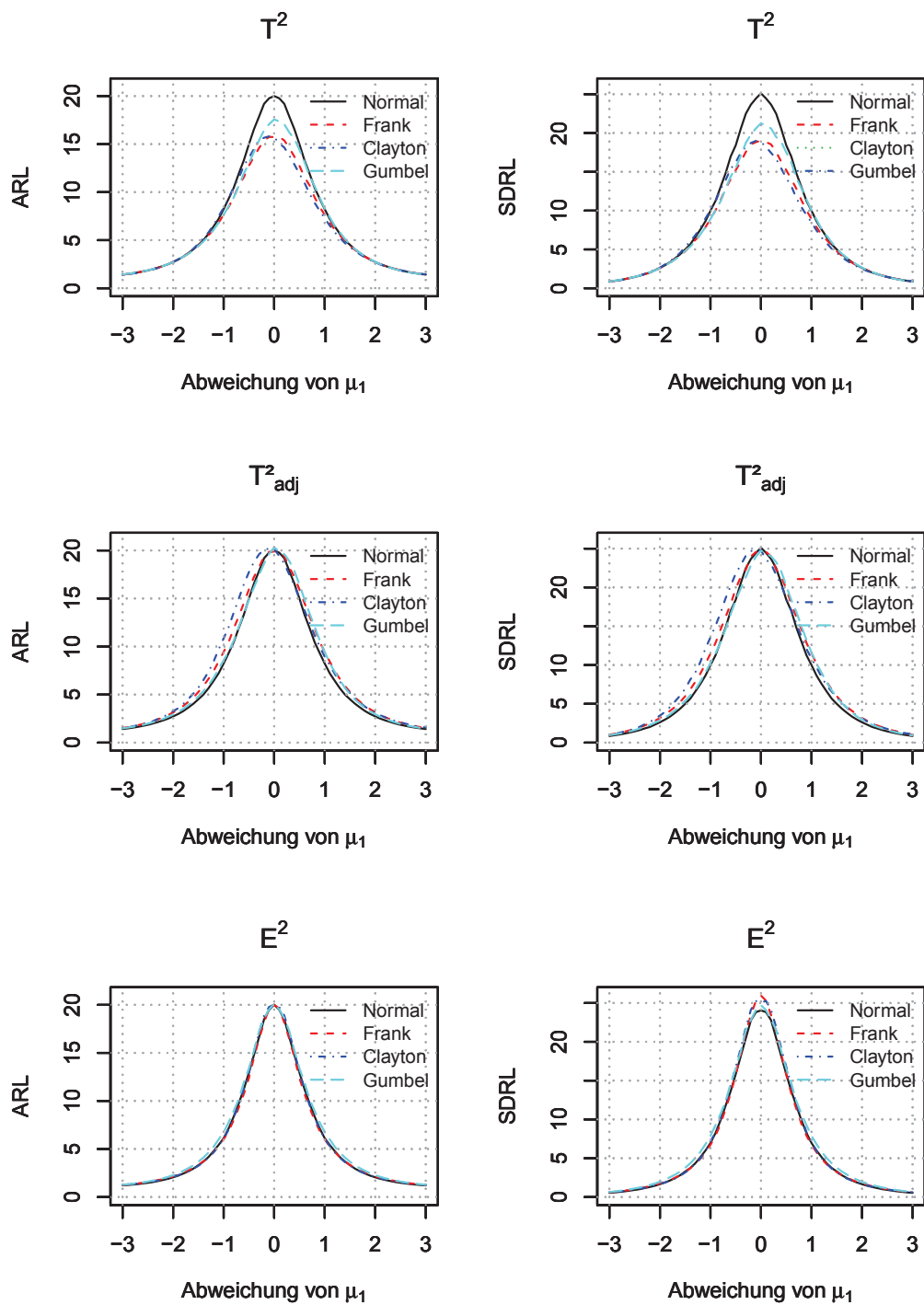


Abbildung 5.4: ARL und SDRL für verschiedene Kontrollkarten,  $\tau = 0.6$ , Szenario (II)

daher eine mittlere Abhängigkeit zwischen den Daten gewählt ( $\tau = 0.6$ ).

Für die klassische  $T^2$ -Karte sind die Verläufe der ARL nicht zufriedenstellend: Im In-Kontrolle-Fall erreicht die ARL für  $\alpha = 0.05$  nur für normalverteilte Daten den theoretisch erwarteten Wert 20 und liegt sonst darunter. In der Praxis gäbe es also zu viele Ablehnungen der Karte und damit Prozessunterbrechungen. Diesen Nachteil zeigen die adjustierte  $T^2$ -Karte und die  $E^2$ -Karte nicht.

Für die Clayton- und Gumbelcopula zeigt die ARL der (adjustierten)  $T^2$ -Karte mit stärker werdender Abhängigkeit außerdem eine wachsende Abweichung von der Symmetrie um den Sollwert, wie sie bei unabhängigen oder punktsymmetrisch angeordneten Daten auftritt. Das ARL-Verhalten der  $E^2$ -Statistik ist dagegen symmetrisch um den Sollwert.

Für die  $T_{adj}^2$ -Karte fällt die ARL für Abweichung vom Mittelwert in Szenario (I) schneller ab als bei der  $E^2$ -Karte. Das ist damit zu erklären, dass die  $T^2$ -Statistik Abstände zum gemeinsamen Schwerpunkt der Verteilung unter Berücksichtigung der gemeinsamen Kovarianzstruktur misst. Bei der  $E^2$ -Karte werden dagegen beide Merkmale unabhängig voneinander standardisiert und anschließend die Summe der einzelnen Abstände betrachtet. Die „Hälfte der Abstandswerte“ (also die eines der Merkmale) entspricht bei Szenario (I) den Anforderungen unter  $H_0$ . Daher sinkt hier die ARL nicht so schnell wie bei der  $T^2$ -Statistik. Für Szenario (I) ist also die adjustierte Hotellings  $T^2$ -Karte geeignet.

Bei Szenario (II) sinkt die ARL für die  $T_{adj}^2$ -Karte langsamer als für die  $E^2$ -Karte. Hier erkennt die  $E^2$ -Statistik die Abweichungen beider Mittelwerte separat und übertrifft damit die  $T^2$ -Statistik. Daher ist die  $E^2$ -Karte hier zu bevorzugen.

Das Verhalten der Standardabweichung der ARL ähnelt dem der ARL, die SDRL liefert also keinen Grund, die oben empfohlene Kartenwahl zu revidieren.

## 5.3 Anwendungen

### 5.3.1 Beispiel I: Ultraschall-Durchflussmesser

Eine mögliche Anwendung der Karten stammt aus dem Bereich der Messtechnik. Ultraschall-Durchflussmesser erfassen die Geschwindigkeit von Flüssigkeiten, Ga-

sen oder Dämpfen mit Hilfe akustischer Wellen. Die Messmethode ist größtenteils unabhängig von den Eigenschaften der verwendeten Durchflussmedien (elektrische Leitfähigkeit, Dichte, Temperatur und Viskosität). Sie deckt einen großen Bereich ab und ist einfach zu warten. Dadurch ist sie in der Praxis beliebt. Die Deutsche Industrienorm 1319-1 (1995) beinhaltet Anforderungen an Ultraschall-Durchflussmesser; weiterführende Informationen bietet Hofmann (2000).

Der Bereich Petrochemie, also die Herstellung chemischer Produkte aus Erdgas und geeigneten Teilen des Erdöls, stellt mit Temperaturen bis über  $500^{\circ}\text{C}$  oder Drücken bis 1500 bar hohe Anforderungen an die Messgeräte. Im Folgenden wird der Durchfluss von Benzol betrachtet. Diese flüssige Kohlenwasserstoffverbindung ist die Grundlage einiger Industriechemikalien (etwa Ethylbenzol).

Im Bereich der Hydrologie gibt es Copulamodelle von Flüssigkeiten. Salvadori und De Michele (2007) erläutert beispielsweise, wie Pegelstände eines Flusses an verschiedenen Erfassungspunkten zusammenhängen. Diese Idee wird analog auf das vorliegende Problem übertragen.

Für eichpflichtige Bereiche in der Produktion ist es inzwischen Standard, dass der Massestrom durch mehrere Sensoren erhoben wird. Diese sogenannte Redundanz sichert zuverlässige Ergebnisse auch nach Ausfall eines Messkanals und in extremer Umgebung.

Unter  $H_0$  ist ein konstanter Massestrom des Benzols gewünscht. Betrachtet werden im Folgenden die Zufallsvariablen

$X_1$  : Messwert des ersten Durchflussmessers und

$X_2$  : Messwert des zweiten Durchflussmessers.

$X_1$  und  $X_2$  sind einzeln durch Normalverteilungen modellierbar. Erwartungswert ist der Sollwert des Benzolstroms. Die Variation der Ergebnisse zwischen den Geräten ist bei größeren Messwerten tendenziell höher als bei kleineren. Die Messunsicherheit nimmt mit wachsenden Merkmalswerten zu.

Die gemeinsame Verteilung von  $X_1$  und  $X_2$  wird daher durch eine Claytoncopula mit hohem Abhängigkeitsgrad ( $\tau = 0.8$ ) und normalverteilten Rändern modelliert. Der  $\tau$ -Wert muss in der praktischen Anwendung auf die jeweils verwendeten Geräte abgestimmt werden.

Nachfolgend werden 200 simulierte Messwerte  $(x_1, x_2)$  des statistisch kontrollierten Prozesses genutzt. Ihnen werden mit Hilfe einer CML-Schätzung (siehe Abschnitt 3.6) Copulas angepasst. Der Maximum-Likelihood-Wert der Claytoncopula liegt mit 306.06 für  $\theta = 8.269$  etwas über dem für die Normalcopula (297.88 für  $\rho = 0.98$ ), so dass die Claytoncopula verwendet wird. Abbildung 5.5 zeigt die normierten Ursprungsdaten sowie die Konturplots der angepassten Verteilungen.

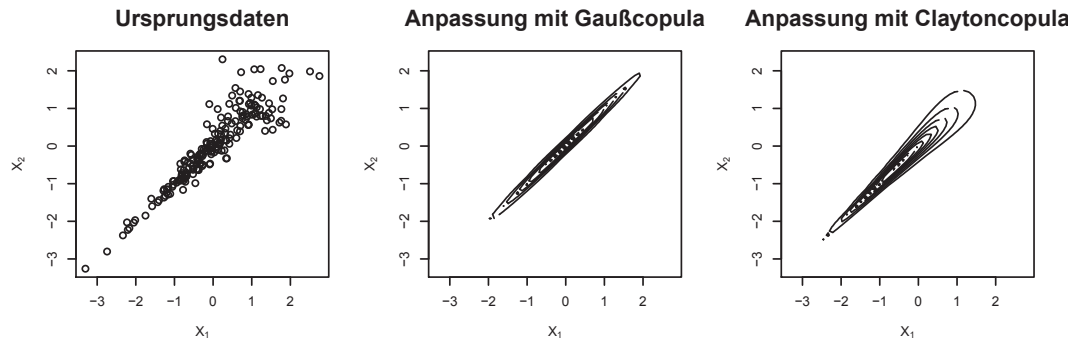


Abbildung 5.5: Ursprungsdaten und Konturlinien der angepassten Verteilungen

Weil jede normalverteilte Variable linear in eine standardnormalverteilte Variable umgeformt werden kann, werden die Kontrollgrenzen aus Abschnitt 5.2 verwendet. Das geschätzte Kendalls  $\tau$  beträgt etwa 0.79, so dass die UCL-Werte für  $\tau = 0.8$  verwendet werden. Die ersten 30 Datenpunkte werden zur Bestimmung der Werte  $\bar{Y}_m$ ,  $S_m$  und  $s_m^*$  genutzt. Daraus wird die Teststatistik für die anderen 170 Datenpunkte bestimmt.

Szenario I und II aus Kapitel 5.2 zeigen, wie der Prozess außer Kontrolle geraten kann. Sie können wie folgt gedeutet werden:

- (I) Ein Messgerät ist außer Kontrolle, zum Beispiel beschädigt.
- (II) Der zugrundeliegende Prozess ist außer Kontrolle, das Benzol strömt zu schnell oder zu langsam.

Für Szenario I wird eine Unterstichprobe aus Zufallszahlen konstruiert, bei der einer der Erwartungswerte der Randverteilungen von 0 auf 2 wechselt; der andere Erwartungswert bleibt konstant null. Für Szenario II werden Daten simuliert,

bei welcher beide Erwartungswerte der Ränder passend geändert werden. Die erste Randverteilung streut nun um 2, die zweite um 1.902.

Abbildung 5.6 stellt jeweils 15 Datenpunkte aus Szenario I (rot) und Szenario II (grün) vergleichend zum In-Kontrolle-Datensatz dar. Das linke Bild zeigt, wie sich

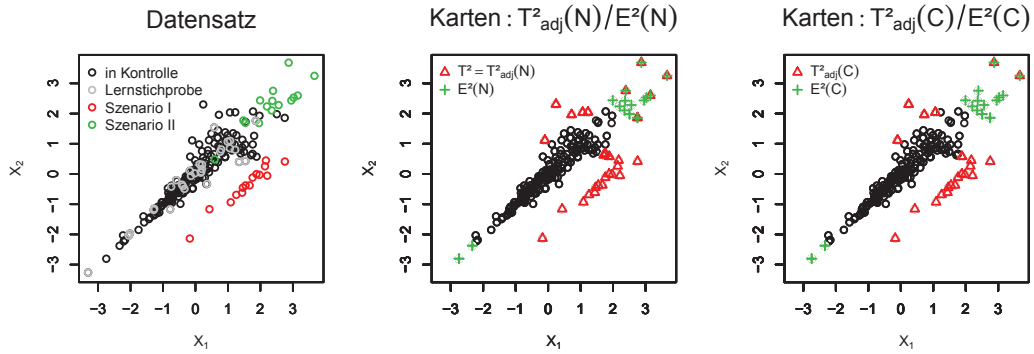


Abbildung 5.6: Datensatz mit Außer-Kontrolle-Werten für verschiedene Szenarien der Mittelwertänderung (links) und durch die verschiedenen Karten als außer Kontrolle eingeordnete Punkte bei Unterstellung einer Normal- (Mitte) sowie Claytoncopula (rechts)

die simulierten Daten in die Ursprungswerte einfügen. Diese Datenpunkte werden genutzt, um die Karte zu testen. Das mittlere und rechte Bild visualisieren, welche Datenpunkte bei Anwendung der verschiedenen Karten Teststatistiken außerhalb der Kontrollgrenzen (Niveau  $\alpha = 0.05$ ) erzeugen.

Das arithmetische Mittel der Lernstichprobe liegt bei  $(-0.0356, -0.0805)$ . Die  $E^2$ -Karte erkennt hauptsächlich weit von diesem geschätzten Erwartungswert gelegene Datenpunkte, während die  $T^2$ -Statistik Werte aufspürt, die weit von der geschätzten Ellipse um den Mittelwert entfernt liegen. Wie schon in Abschnitt 5.2 beschrieben, ist die  $E^2$ -Karte für Szenario II geeigneter, die  $T^2$ -Karte zeigt bessere Ergebnisse für Szenario I.

Fügt man die konstruierten Außer-Kontrolle-Daten ans Ende der 170 verwendeten Messpunkte, ergeben sich die Karten aus den Abbildungen 5.7 und 5.8. Die Bilder zeigen die beschriebenen Kontrollkarten für beide Szenarien der Mittelwertänderung zum Niveau  $\alpha = 0.05$ .

Abbildung 5.7 betrachtet für Szenario I das Verhalten der  $T^2$ - und der  $T^2_{adj}(C)$ -Karte im linken Bild und das der  $E^2$ -Karte im rechten Bild. Die vertikalen Linien

(schwarz gepunktet) zeigen den Punkt an, zu dem der Prozess außer Kontrolle gerät, also den 171. Datenpunkt.

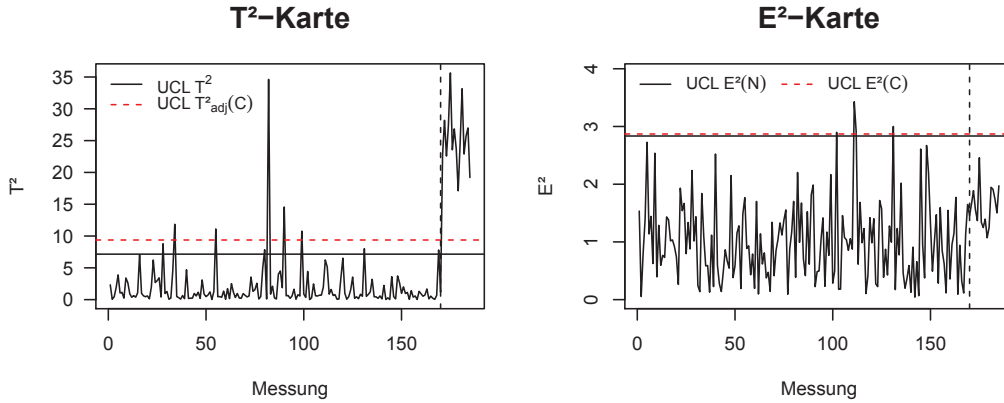


Abbildung 5.7: Anwendung der Kontrollkarten auf Szenario I

Die Grenzwerte für eine unterstellte gemeinsame Normalverteilung der Variablen sind für beide Teststatistiken kleiner als bei einer unterstellten Claytoncopula. Dieser Effekt ist bei der  $T^2$ -Statistik deutlicher ausgeprägt. Hier liegen 24 Werte für die traditionelle  $T^2$ -Karte oberhalb der Grenzen, darunter alle 15 Außer-Kontrolle-Werte.

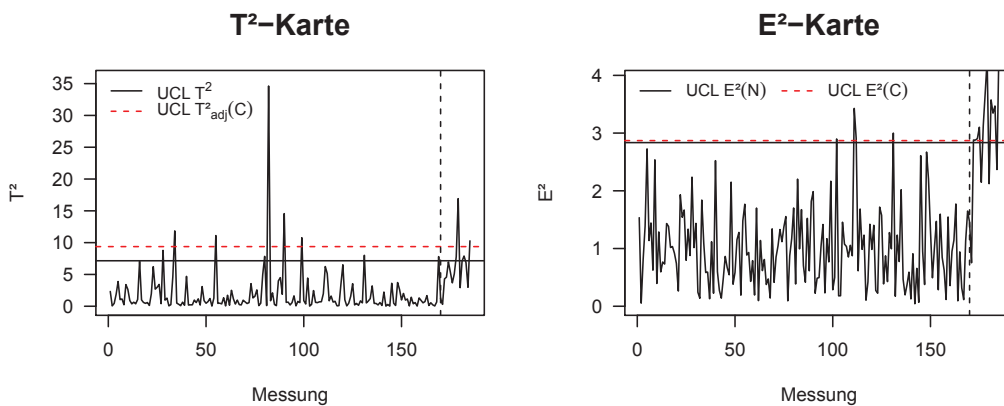


Abbildung 5.8: Anwendung der Kontrollkarten auf Szenario II

Für die  $T^2_{adj}(C)$  liegen nur 20 Werte über der oberen Kontrollgrenze, darunter die 15 anders verteilten. Die Trefferquote ist bei der adjustierten Karte also besser. Die



$E^2$ -Statistik entdeckt keinen Wechsel für Szenario I. Für beide Grenzwerte werden vier Werte als Ausreißer markiert, davon stammt keiner aus der Außer-Kontrolle-Stichprobe.

Abbildung 5.8 bildet Szenario II ab. Auch hier sind für beide Teststatistiken die Grenzwerte bei unterstellter Normalverteilung kleiner als die der Claytoncopula und wiederum ist der Unterschied bei der  $T^2$ -Statistik deutlicher als bei der  $E^2$ -Statistik. Bei der klassischen  $T^2$ -Karte liegen 14 Werte über der Kontrollgrenze, vier davon stammen aus der abweichenden Grundgesamtheit. Für  $T_{adj}^2(C)$  sind von sieben Datenpunkten oberhalb der Grenzen zwei aus der Außer-Kontrolle-Stichprobe. Bei der  $E^2$ -Statistik liegen für beide Möglichkeiten 15 Werte außerhalb der Grenzen. Elf Punkte davon sind Außer-Kontrolle-Werte.

Wechsel in nur einer Randverteilung wurden mit hoher Wahrscheinlichkeit durch die (adjustierte)  $T^2$ -Karte erkannt, Änderungen in beiden Rändern sind schwerer nachzuweisen, da die Außer-Kontrolle-Werte stärker mit den In-Kontrolle-Daten überlappen (siehe auch Abbildung 5.6). Hier war die  $E^2$ -Karte besser geeignet.

### 5.3.2 Beispiel II: Inkrementelles Umformen

Der vorliegende Fall betrachtet Zylinder, die per ISF („Incremental Sheet Forming - inkrementelles Umformen“) gefertigt wurden. Diese Fertigungstechnik ist den konventionellen Umformtechniken überlegen. Sie liefert nahtlose Zylinder, die leicht zu reinigen sind. Eingesetzt werden sie beispielsweise in der Baubranche, bei Landmaschinen oder in der Lebensmittel- und Pharmaindustrie.

Die Daten wurden im Rahmen des Sonderforschungsbereichs 823 „Statistical modelling of nonlinear dynamic processes“ der Technischen Universität Dortmund, Projekt B2 „Charakterisierung des dynamischen Prozessverhaltens bei der inkrementellen Blechumformung“ erhoben und sind bei Melsheimer et al. (2011) detailliert beschrieben.

Die betrachteten Zylinder haben ein langes, schmales Endstück. Durch kleinste Materialunterschiede oder minimale Unregelmäßigkeiten im Umformprozess kann das Metall im Bereich des Endstückes reißen oder Falten schlagen. Der interessierende Parameter ist also die Blechdicke des Zylinders im Bereich der Umformung. Die

Dicke im nichtbearbeiteten Teil ist nahezu konstant und braucht nicht überwacht werden. Insgesamt haben die Blechscheiben einen Radius von etwa 63mm, die herausgedrückte Spitze erreicht eine Höhe von 31.5mm bei einem Radius von etwa 9mm (siehe Abbildung 5.9).

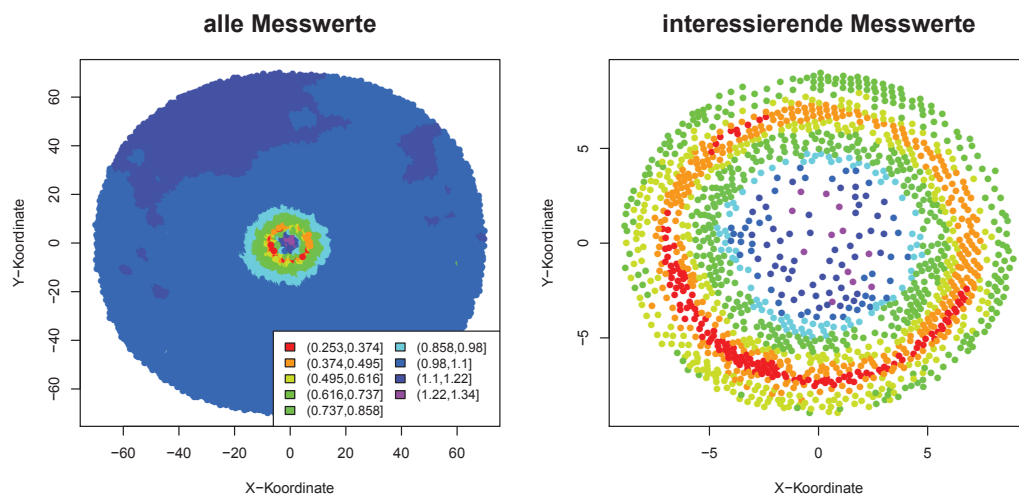


Abbildung 5.9: Blechdicke in mm für alle Messpunkte (links) und für den untersuchten Ausschnitt (rechts)

Zur Messung der Blechdicke wurde der analysierte Bereich in horizontale Abschnitte mit äquidistanter Höhe (sogenannte „Bänder“) unterteilt. Als optimal zur Erkennung von Außer-Kontrolle-Daten erwiesen sich Abstände von 3.6mm. Zur Datenerfassung werden die Messwerte der Blechdicken des ersten Bandes von links nach rechts sortiert. Orientierung bietet die  $X$ -Koordinate. Für das zweite Band werden die Werte dann von rechts nach links angeordnet und so weiter. Der interessierende Bereich wird also alternierend abgetastet. Die Folge dieser Werte wird analysiert. Die logarithmierten Werte der Blechdicke in Millimeter können durch eine Normalverteilung beschrieben werden, wie Abbildung 5.10 zeigt. Auch formale Tests auf Normalverteilung lehnen die Normalverteilungsannahme nicht ab.

Im Folgenden werden die Daten der rechten Grafik aus Abbildung 5.9 analysiert. Das Beispiel ist eine Ex-Post-Studie; das heißt, erst nach Fertigstellung des Zylinders werden alle relevanten Werte zusammenhängend erhoben. Daher können alle verfügbaren Daten verwendet werden, um die zugrundeliegende Datenstruktur und die erforderlichen Werte von  $\bar{Y}_m$ ,  $S_m$  und  $s_m^*$  zu berechnen.

Unter  $H_0$  zeigt die Blechdicke keine „Auffälligkeiten“. Dazu zählen beispielsweise Risse, Einschlüsse im Material, zu dicke oder zu dünne Bereiche. Diese können durch die Überwachung räumlich benachbarter Blechdickenmessungen gefunden werden.

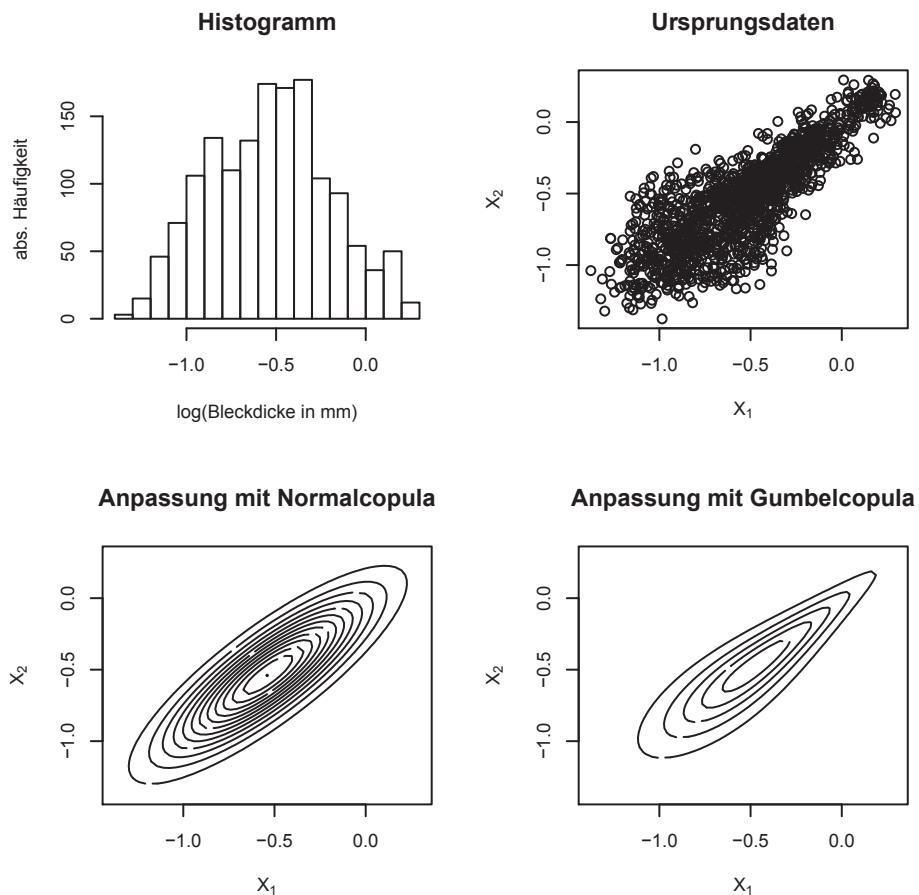


Abbildung 5.10: Histogramm der logarithmierten Messwerte (oben links), Ausgangsdaten (oben rechts) und Höhenlinien der angepassten gemeinsamen Verteilungen (unten)

Überwacht werden daher die Paare  $(X_1, X_2)$  der Zufallsvariablen

$X_1$  : erste Blechdickenmessung sowie

$X_2$  : Nachbarmessung zu  $X_1$  laut des erstellten Bandes.

Durch Verwendung der logarithmierten Messungen sind  $X_1$  und  $X_2$  normalverteilt. Die benachbarten Dichtemessungen sind abhängig.

Abbildung 5.10 zeigt die Ursprungsdaten und die Höhenlinien der per CML-Methode (siehe Abschnitt 3.6) angepassten Copulas. Der ML-Wert für die Gumbelcopula beträgt ungefähr 915.5 für den Parameter  $\theta = 2.76$  und für die Normalcopula etwa 816.5 für  $\rho = 0.84$ . Daher wird die Gumbelcopula bevorzugt.

Da jede normalverteilte Zufallsvariable leicht in eine standardnormalverteilte Zufallsvariable umgewandelt werden kann, wird das Verfahren aus Abschnitt 5.2 genutzt, um passende Grenzwerte zu erstellen. Das geschätzte Kendalls  $\tau$  beträgt ungefähr 0.637. Die entsprechenden UCL-Werte für einige Niveaus  $\alpha$  zeigt Tabelle 5.3.

Niveau $\alpha =$	10%	5%	1%	0.27%
$T^2$	4.611	6.001	9.230	11.860
$T^2(G)$	4.643	6.299	10.801	15.227
$E^2(N)$	2.145	2.449	3.040	3.442
$E^2(G)$	2.269	2.690	3.524	4.099

Tabelle 5.3: Simulierte obere Kontrollgrenzen für  $n = 2974$ ,  $\tau = 0.637$

Die zugehörigen Karten sind in Abbildung 5.11 dargestellt. Bei Anwendung der traditionellen  $T^2$ -Statistik liegen 62 der knapp 3000 Werte außerhalb, bei der  $T^2(G)$ -Statistik sind es 52 Werte. Die  $E^2$ -Karte liefert unter vorausgesetzter bivariater Normalverteilung 112 Werte oberhalb der Grenze, bei unterstellter Gumbelcopula 80 Werte. Sowohl Szenario I als auch II aus Abschnitt 5.2 sind in diesem Prozess vorstellbar:

- (I) Messung 1 und 2 unterscheiden sich stark. Das kann auf lokale Beulen, Materialfehler, Microrisse oder kleine Wellen hindeuten.
- (II) Messung 1 und 2 sind ähnlich, aber ungewöhnlich klein oder groß und damit Anzeichen für sehr dicke oder dünne Regionen des Blechs.

Wo sich die außenliegenden Messwerte im Vergleich zu den anderen Messwerten beziehungsweise im umgeformten Blech befinden, zeigt Abbildung 5.12.

Punkte außerhalb der  $T^2$ -Grenzen liegen ohne erkennbare Muster über das betrachteten Gebiet verstreut (obere Grafik). Die Außer-Kontrolle-Werte der  $E^2$ -Statistik konzentrieren sich dagegen auf zwei Bereiche (untere Grafik).

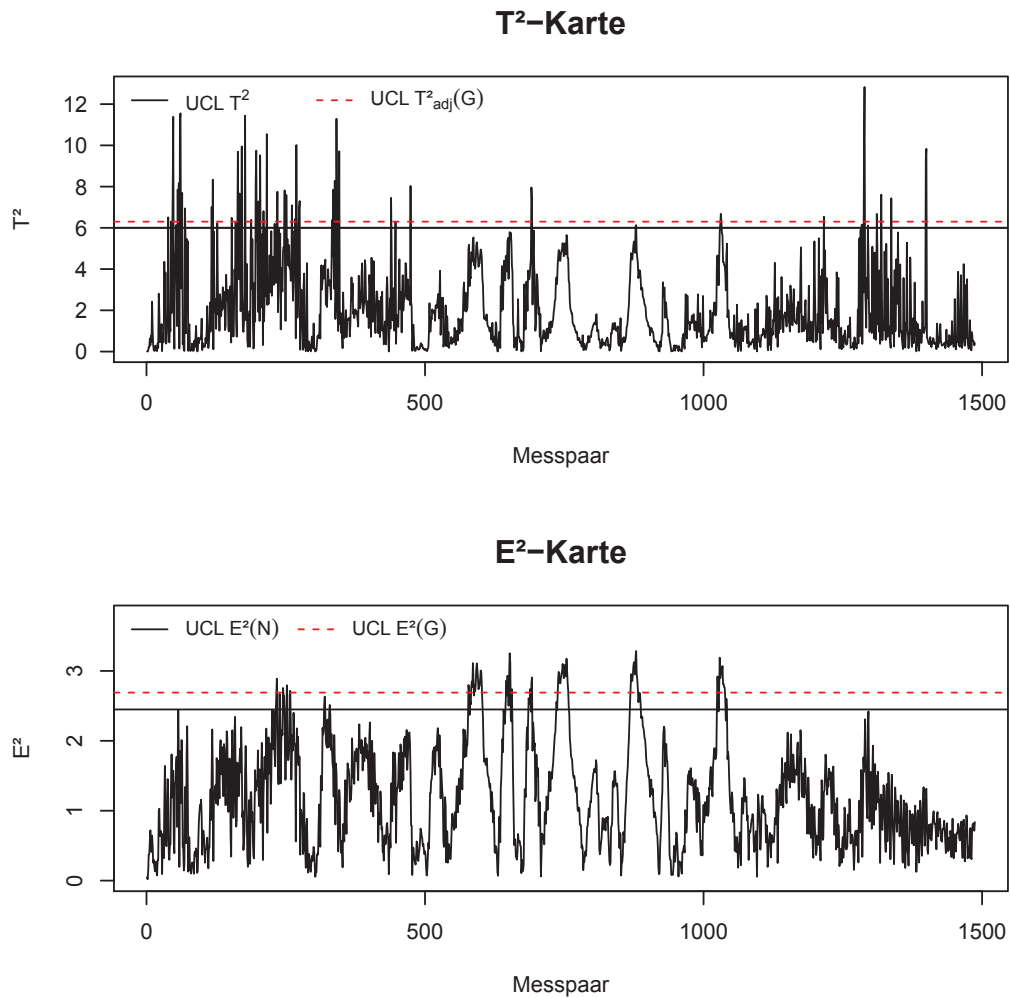


Abbildung 5.11: Kontrollkarten für die Messungen der Wanddicke, basierend auf der  $T^2$ - (oben) und der  $E^2$ -Statistik (unten)

Bekanntermaßen reagiert die  $T^2$ -Karte sensibler auf Szenario I, die  $E^2$ -Karte erkennt eher Wechsel, wie sie Szenario II beschreibt.

Ein Vergleich dieser Grafik mit der mittleren Darstellung aus Abbildung 5.9 erlaubt eine gute Interpretation der Ergebnisse der  $E^2$ -Statistik: Der rote Kreis aus Abbildung 5.9 markiert Regionen, wo das Blech besonders dünn gedrückt wurde. Insbesondere im linken unteren Bereich ist eine sehr dünne Region zu finden. Dieser Bereich wurde auch durch die  $E^2$ -Karte gefunden. Das Blech wird instabil und kann leicht reißen, wenn es zu dünn wird. Daher sollte eine Mindestblechdicke eingehalten werden. Sehr dicke Bereiche befinden sich dagegen in der Kreismitte, wo die

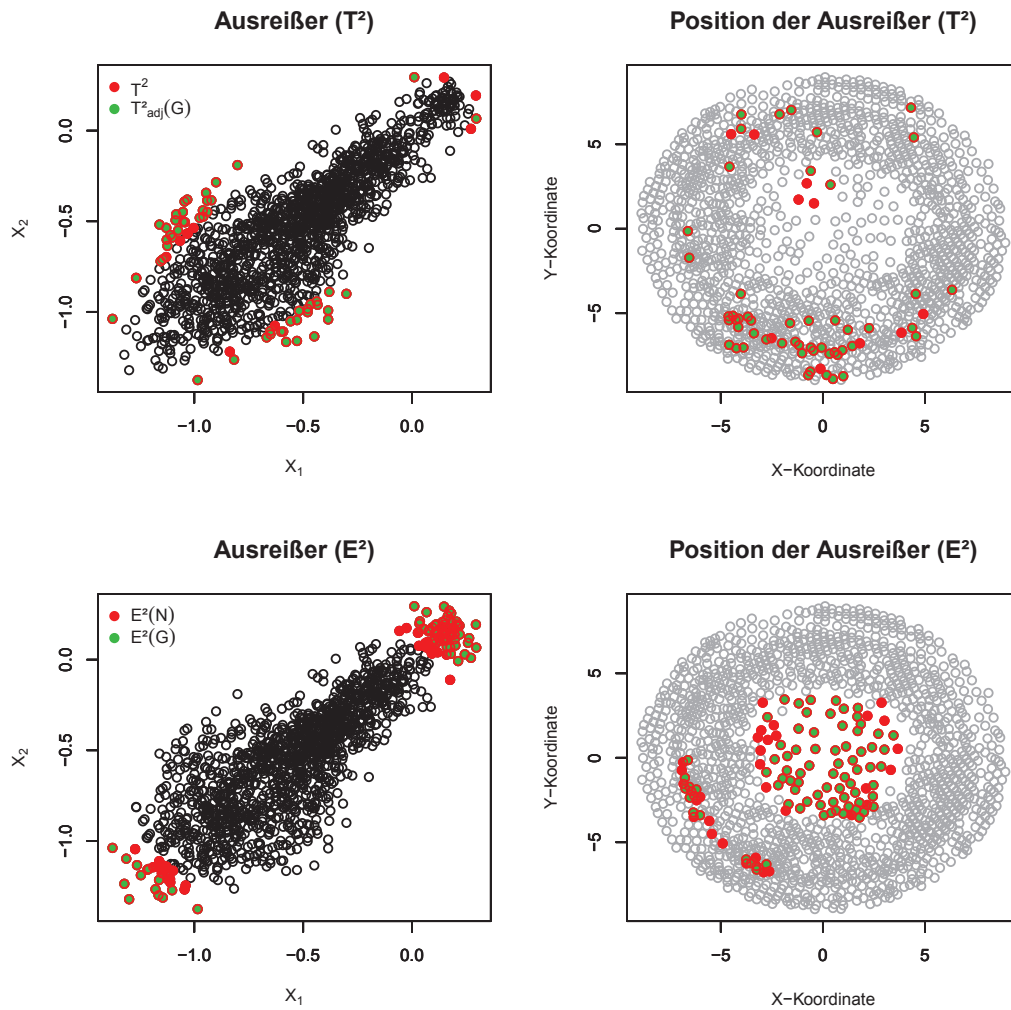


Abbildung 5.12: Kritische Werte, gekennzeichnet in der Datenwolke (links) und in der Blechplatte (rechts)

Platte im Drückprozess durch den Dorn erfasst wurde. Eine Minimierung dieser Region würde mehr Material für die restliche Zylinderspitze bedeuten, die damit tendenziell dicker würde. Zu dicke Bereiche sollten also auch vermieden werden. Die durch die  $T^2$ -Statistik als kritische Werte gefundenen Datenpunkte sind inhaltlich gut so offensichtlich nachvollziehbar. Dass es sich um räumlich meist weiter gestreute Einzelpunkte handelt, passt allerdings gut zu der obigen Interpretation von lokalen Beulen, Microrissen und ähnlichen Phänomenen.

## 5.4 Überwachung der Prozessstreuung

Die bisherigen Abschnitte untersuchen Lageänderungen. Die  $T^2$ -Karte zeigt aber auch wachsende Prozessstreuung an. Ob Lage- oder Streuungsänderungen im Prozess den Alarm ausgelöst haben, kann dabei nicht unterschieden werden.

Eine Simulationsstudie mit verschiedenen Copulamodellen zeigt das Verhalten der  $T^2$ -Statistik mit sinkender beziehungsweise wachsender Streuung in den Randverteilungen. Es werden Normal-, Clayton- und Gumbelcopula für verschiedene Zusammenhangsmaße  $\tau$  betrachtet. Die Copulas verknüpfen jeweils zwei  $\mathbb{N}(0, 1)$ -Verteilungen. Untersucht wurde, wie sich ARL- und SDRL-Werte verändern, wenn die Standardabweichung einer Randverteilung zwischen 0.5 und 1.5 variiert.

Nach dem in Abschnitt 5.2 beschriebenen Verfahren wurden je 10000 ARL- und SDRL-Werte simuliert und in Abbildung 5.13 zusammengestellt.

Bei unterstellter Normalverteilung realisiert sich die ARL bei einem Wert von 20 unter der Nullhypothese (für  $\alpha = 0.05$ ) für alle getesteten  $\tau$ . Bei der Clayton- und Gumbelcopula erreicht die ARL den gewünschten Wert nur für  $\tau = 0$ . Anschließend sinkt die ARL unter  $H_0$  mit wachsendem  $\tau$ . Dieser Effekt ist von den Untersuchungen zur Lageverschiebungen bekannt, die Erklärung ist hier die selbe. Für Claytoncopulas ist der Effekt der ARL-Absenkung stärker ausgeprägt als für Gumbelcopulas. Dieses Phänomen ist durch die Form beider Funktionen begründet.

Die Teststatistik  $T^2$  erkennt eine Vergrößerung der Streuung, die ARL- und SDRL-Werte rechts des Sollwertes sinken monoton. In der Praxis tritt eine wachsende Streuung bei den gemessenen Merkmalen beispielsweise auf, wenn sich bei einer Maschine im Laufe des Fertigungsprozesses Schrauben an den Gelenken durch Bewegung lockern. Auch die Abnutzung von Schleifelementen, Bohrern oder sonstigen Verschleißteilen an der Maschine können diesen Effekt bewirken. Dadurch werden die angestrebten Sollwerte nicht mehr präzise erreicht und die gefertigten Produkte weichen qualitativ stärker voneinander ab. Ein Anschlagen der Karte weist dann darauf hin, Verschleißteile auszutauschen oder die entsprechenden Maschinenbereiche zu warten.

Eine geringe Streuung um den Sollwert ist in der Regel gewünscht, um ein gesetztes Qualitätsziel möglichst gut zu erreichen. Schwächere Streuung ist interessant,

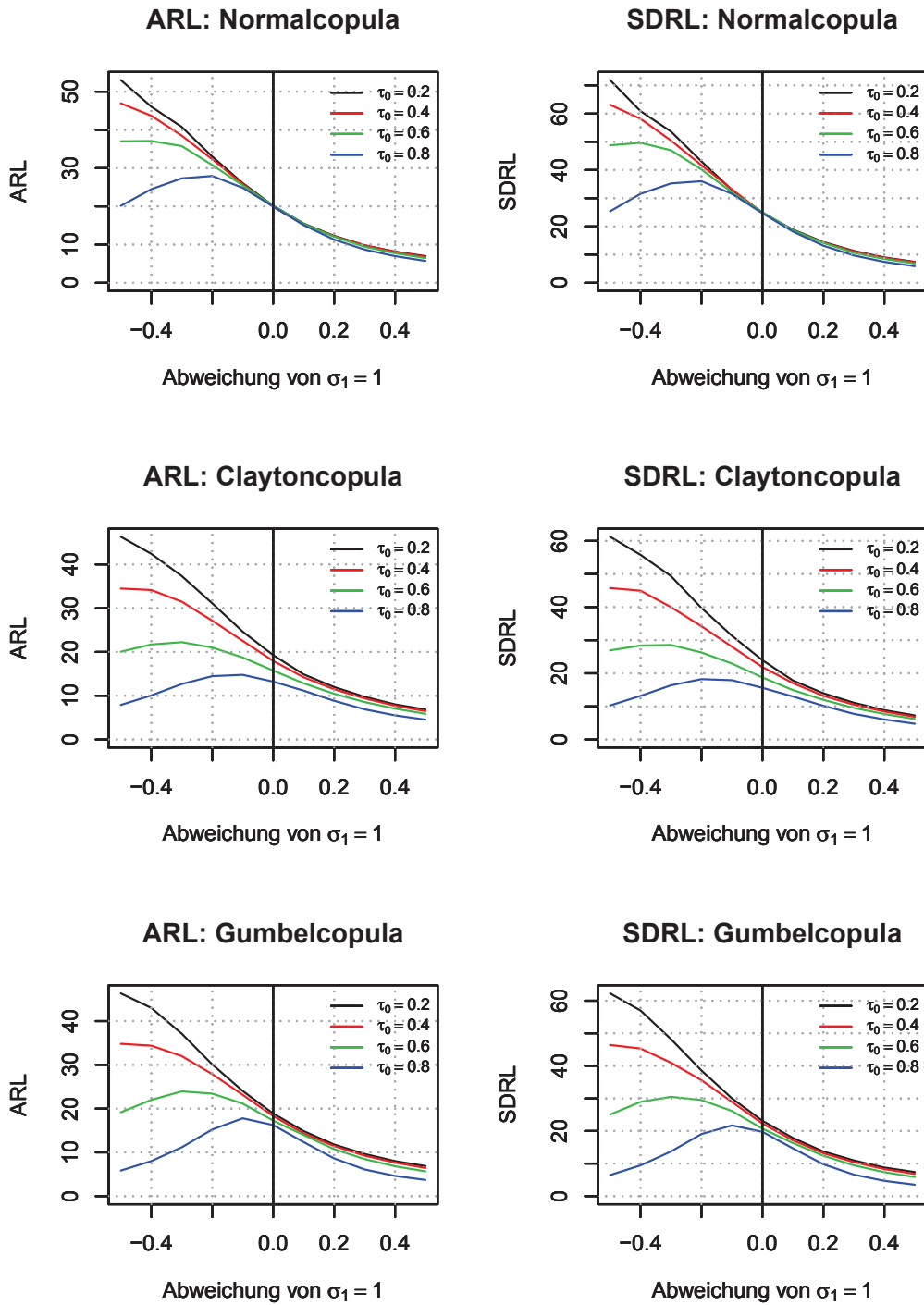


Abbildung 5.13: ARL und SDRL der  $T^2$ -Statistik bei Abweichungen in der Streuung einer Randverteilung bei Verwendung verschiedener Copulas und einzeln normalverteilten Rändern



wenn der Prozess verändert wurde. So kann zum Beispiel untersucht werden, ob eine bestimmte Maschineneinstellung den Prozess besser kontrollierbar macht. Dieser Fall interessiert also hauptsächlich, um eine Qualitätsverbesserung nachzuweisen. Die vorliegende Teststatistik erkennt eine sinkende Datenstreuung kaum. Je größer die Abhängigkeit zwischen zwei Merkmalen ist, umso eher führt das Absinken von  $\sigma_1$  zu einer fallenden ARL. Für alle Werte gilt allerdings zunächst, dass die ARL ansteigt. Für den Bereich der Qualitätssicherung ist dieser Effekt durchaus positiv zu bewerten, da eine sinkende Varianz in der Regel keine Minderung der Qualität impliziert. Hier ist der Fall wachsender Streuung interessanter.

Wie bei der Überwachung der Lage ist es möglich, durch Monte-Carlo-Simulation adjustierte Karten zu erstellen. Die Grenzwertberechnung erfolgt analog zu der in Abschnitt 5.2 beschriebenen. Die adjustierten  $T^2$ -Karten aus Kapitel 5.1.1 halten auch in diesem Fall unter  $H_0$  das Niveau ein. Das ARL-Verhalten ähnelt dann dem aus Abbildung 5.13 für die Normalcopula.



# 6 Variante II: Zusammenhangskontrolle

## 6.1 Entwicklung eines geeigneten Testverfahrens

### 6.1.1 $T^2$ -Karte für Abhängigkeitsänderungen

Abwandlungen der klassischen  $T^2$ -Karte erkennen Abweichungen in den Lage- und Streuungsparametern der Randverteilung gut. Nun soll eine Teststatistik gefunden werden, die auf Änderungen in der Abhängigkeitsstruktur zweier Messgrößen reagiert.

Zunächst wird die  $T^2$ -Karte aus Abschnitt 4.5 auf die vorliegende Problemstellung angewandt. Kapitel 5 zeigt, dass die Teststatistik unter Anwendung modifizierter Grenzwerte Prozessabweichungen aufspürt. Um eine geeignete Kontrollkarte zur Überwachung der Prozessstruktur zu finden, wird zunächst das Verhalten der  $T^2$ -Karte bei Strukturänderungen untersucht.

Abbildung 6.1 zeigt, wie sich Änderungen in der Zusammenhangsstruktur von Datenpunkten aus verschiedenen Copulastrukturen auf Hotellings  $T^2$ -Karte auswirken. Die Zusammenhangsstruktur wird zur besseren Vergleichbarkeit durch Kendalls  $\tau$  ausgedrückt. Es wurden analog zur Beschreibung in Abschnitt 5.2 jeweils 10000 ARL- und SDRL-Werte simuliert.

Ein Abschwächen des Zusammenhangs wird erkannt, eine Zunahme der Zusammenhänge lässt die ARL-Werte dagegen sogar ansteigen. Das ist dadurch zu erklären, dass sich ein abschwächender Zusammenhang in weiter streuenden Punkten zeigt, deren Abstand zum Mittelpunkt tendenziell wächst und daher die (einseitigen) Grenzwerte relativ schnell überschreitet. Andererseits bewirkt ein Ansteigen der Abhängigkeit, dass die Datenpunkte tendenziell konzentrierter um die gemeinsame Schwerpunktsachse auftreten. Die Distanz zum Mittelpunkt sinkt also und die Grenzwerte werden seltener überschritten, folglich steigt die ARL.

Die  $T^2$ -Statistik misst die standardisierte Abweichung eines Punktes von dem

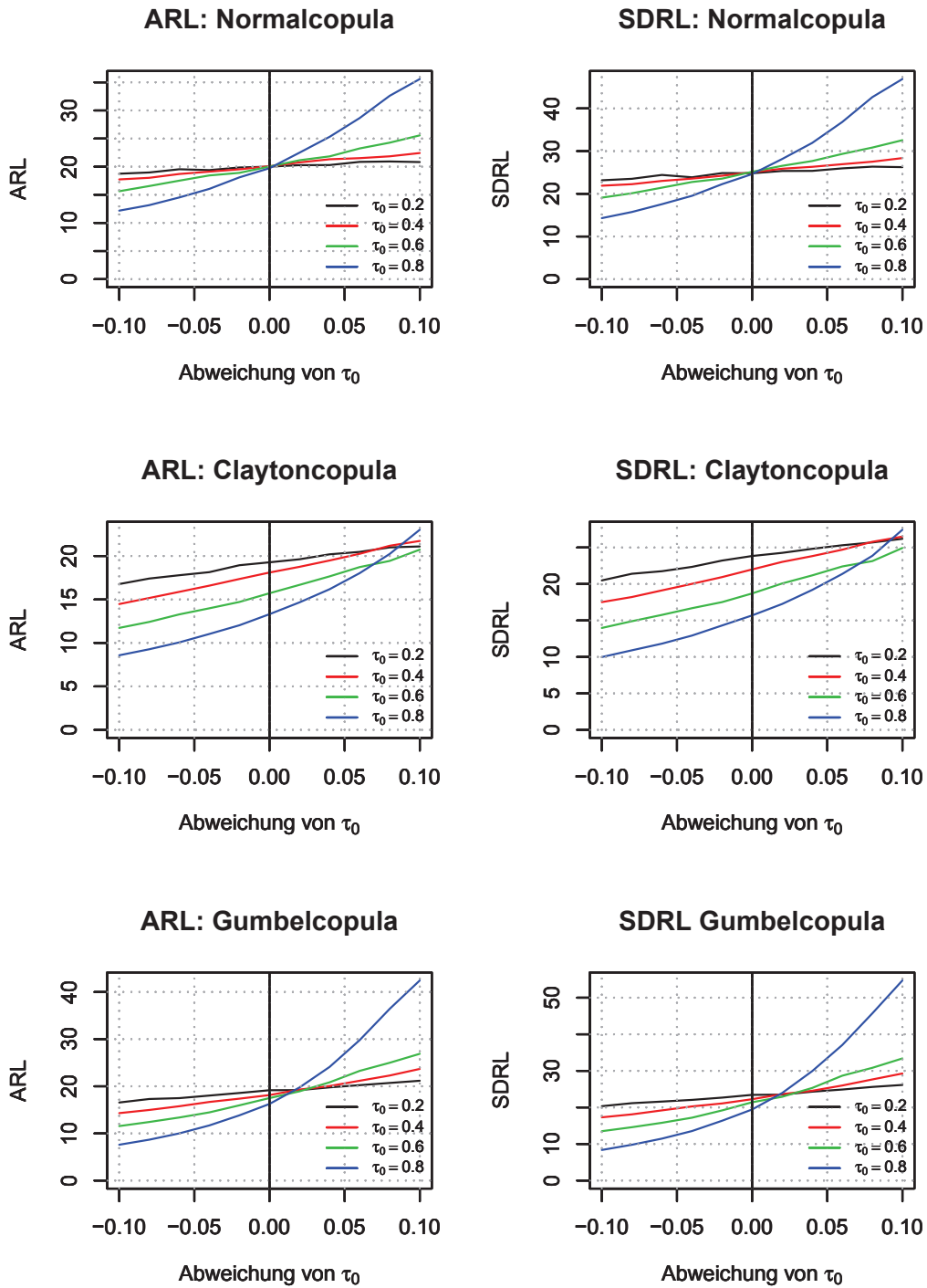


Abbildung 6.1: ARL und SDRL der  $T^2$ -Statistik bei Abweichungen in der Abhängigkeitsstärke unter Zugrundelegung verschiedener Copulas bei einzeln standardnormalverteilten Messdaten  $X_1$  und  $X_2$

Wahrscheinlichkeitsschwerpunkt einer multivariaten Normalverteilung. Es handelt sich demzufolge um ein ideales Abstandsmaß für die punktsymmetrische Normalverteilung. Die betrachteten Copulas sind hingegen achsensymmetrisch. Auf dieser Grundlage soll eine Teststatistik zum Erkennen von Strukturbrüchen (also Änderungen der Abhängigkeitsstruktur) entwickelt werden.

Ähnliche Testverfahren sind so konstruiert, dass beispielsweise Datenpunkte aus der Verteilung unter  $H_0$  simuliert und mit der erhobenen Stichprobe verglichen werden. Fügen sich diese Stichprobenwerte „plausibel“ in die erzeugten Daten ein, geht man davon aus, dass beide derselben Grundgesamtheit entstammen.

Tests zu Strukturbrüchen auf Basis von Copulamodellen wurden beispielsweise Krämer und van Kampen (2011), Wied et al. (2012), Wied und Galeano (2012) sowie Brodski et al. (2012) veröffentlicht.

### 6.1.2 Beschreibung der Teststatistik

Begründet auf den bisherigen Überlegungen wird nun ein Testverfahren zur Überwachung von Strukturänderungen entwickelt. Bei der  $T^2$ -Statistik werden die Euklidischen Distanzen der Datenpunkte zum Datenschwerpunkt unter  $H_0$  auf Plausibilität geprüft. Um die Achsensymmetrie der verwendeten Copulas auszunutzen, verwendet das vorgestellte Verfahren ein Distanzmaß, das den kürzesten Abstand zwischen Punkt und Achse misst. Gering voneinander abhängende Daten bedingen eher große Abweichungen von der betrachteten Achse, stark abhängige Daten zeigen meist kleinere Abweichungen. In diesem Fall entspricht die betrachtete Achse der Winkelhalbierenden des Einheitsquadrates, die Datenpunkte müssen dazu passend auf den Wertebereich  $[0, 1]^2$  normiert sein beziehungsweise aus einer Copulastruktur stammen.

Je nach betrachteter Copula streuen die Datenpunkte unterschiedlich um die Winkelhalbierende. Diese Besonderheit wird bei den vorgestellten Testverfahren berücksichtigt. Abbildung 6.2 veranschaulicht die vorgestellte Testidee. Anstelle eines fixen Abschnittes (linke Grafik), wird ein interessierender Bereich mit relativ hoher Datendichte betrachtet. Die mittlere Grafik zeigt die Idee, einen an der Winkelhalbierenden ausgerichteten Streifen zu betrachten. Der Vollständigkeit halber ist in der rechten Abbildung die später definierte Teststatistik dargestellt. Diese konzentriert sich auf den „aussagekräftigsten“ Abschnitt zwischen den grau gestrichelten

Linien. Die an der Winkelhalbierenden gespiegelten Datenpunkte in diesem Bereich werden für Aussagen über Änderungen der Struktur verwendet.

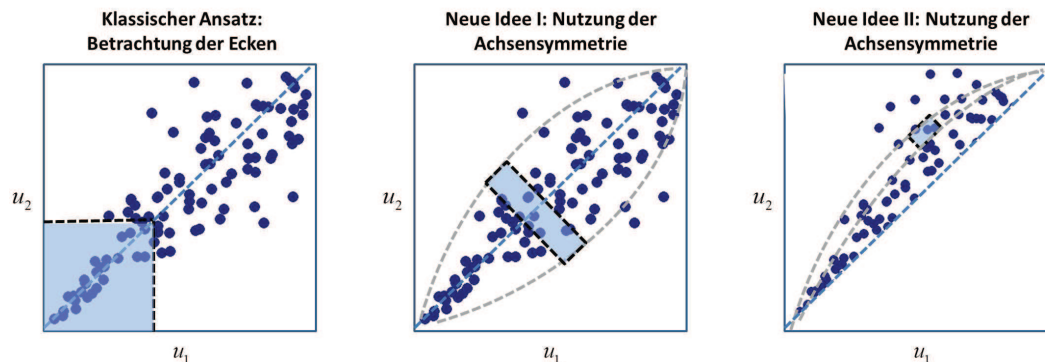


Abbildung 6.2: Motivation der Teststatistik

Das vorgestellte Verfahren ist zweistufig. Zuerst werden einige Distanzen bestimmt, etwa die Entfernung des Punktes zur Winkelhalbierenden oder ein spezieller Abstand zum Koordinatenursprung, der sich wiederum an der Winkelhalbierenden orientiert. Anschließend werden die umgeformten Daten entsprechend der zugrundeliegenden Copula als zu  $H_0$  passend, zu weit streuend oder zu konzentriert klassifiziert. Zeigt die Stichprobe also Unregelmäßigkeiten in der Abhängigkeitsstruktur der Daten, führen sowohl eine Verstärkung als auch eine Abschwächung des Zusammenhangs zur Alarmauslösung. Wie die Karte genau arbeitet, wird im Folgenden beschrieben.

### Schritt 1: Datentransformation

Zunächst wird der orthogonale Abstand eines zweidimensionalen Zufallsvektors  $U := (U_1, U_2)$ , also eines Punktes im  $I^2$  zur Symmetrieachse bestimmt. Die folgenden Überlegungen gelten für beliebige Copulas. Beim Punkt  $U$  handelt es sich um einen Vektor der Verteilungsfunktionen an den Stellen der Einträge einer zugrundeliegenden Ausgangsvariablen  $X := (X_1, X_2)$ , demzufolge ist  $U = (F_1(X_1), F_2(X_2))$ . Anstelle der gemeinsamen Verteilungsfunktion wird nur der Zusammenhang der Messwerte berücksichtigt, ausgedrückt durch die zugrundeliegende Copula. Abbildung

6.3 zeigt ein Beispiel mit normalverteilten Rändern ( $\mathbb{N}(0,1)$  und  $\mathbb{N}(8,1)$ ), deren gemeinsame Struktur einer Claytoncopula folgt. Die unterschiedlichen Normalverteilungen als Ränder zeigen erneut, dass die Copulastruktur davon unberührt bleibt. Die Abhängigkeit beträgt im ersten Fall  $\tau = 0$  und im zweiten  $\tau \approx 0.79$ .

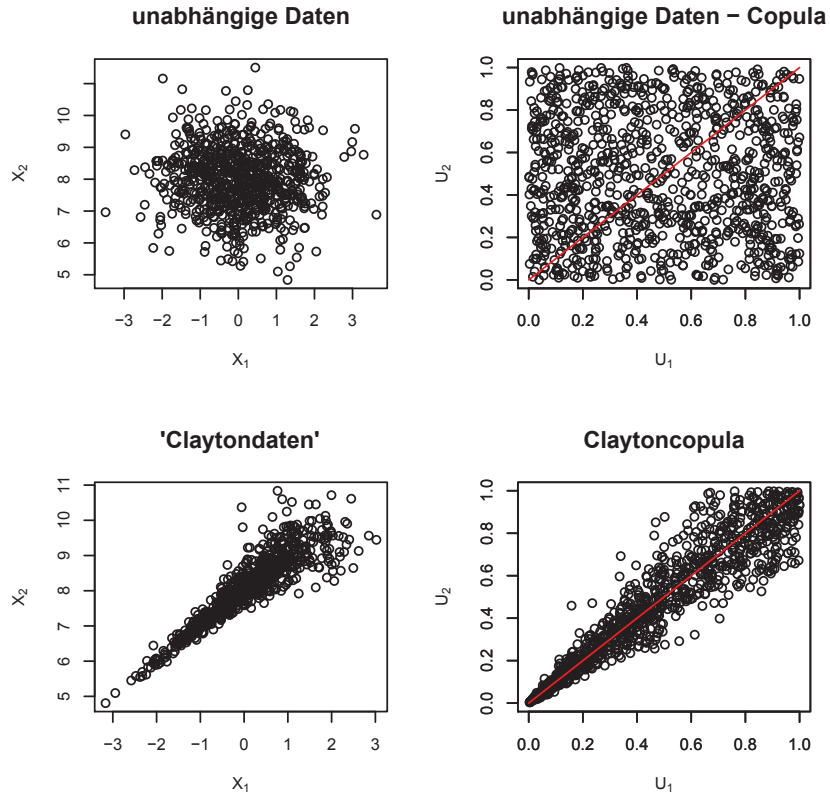


Abbildung 6.3: Streu- und Copula-Diagramme für einzeln normalverteilte Daten  $X_1$  und  $X_2$  (einmal unabhängig, einmal auf Basis einer Claytoncopulastruktur)

Die Karte betrachtet den orthogonalen Abstand eines Punktes  $(U_1, U_2)$  von der Winkelhalbierenden  $U_2^* = U_1^*$ . Der Punkt auf der Winkelhalbierenden, den die dazu orthogonale Gerade durch  $(U_1, U_2)$  schneidet, liegt bei

$$\left( \frac{U_2 + U_1}{2}, \frac{U_2 + U_1}{2} \right).$$

Der orthogonale Abstand zwischen diesem Schnittpunkt und  $(U_1, U_2)$  ist dann

$$r := r(U_1, U_2) := \left[ \left( \frac{U_1 + U_2}{2} - U_1 \right)^2 + \left( \frac{U_1 + U_2}{2} - U_2 \right)^2 \right]^{\frac{1}{2}} \quad (6.1)$$

mit  $r \in [0, \frac{1}{\sqrt{2}}]$ .

Berechnet wird außerdem  $e$ . Diese Variable gibt die auf den Wertebereich  $[0, \sqrt{2}]$  normierte Euklidische Distanz des errechneten Lotes von  $(U_1, U_2)$  zum Koordinatenursprung an. Es gilt:

$$e := e(U_1, U_2) := \frac{U_1 + U_2}{\sqrt{2}} \quad (6.2)$$

Abbildung 6.4 verdeutlicht den Zusammenhang der Maße  $r$  nach Formel (6.1) und  $e$  nach Formel (6.2). Die Verteilung von  $r$  hängt von der verwendeten Copula ab. Für feste Copulas und  $\tau_0$  können die Quantile von  $r$  in Abhängigkeit von  $e$  ausgedrückt werden.

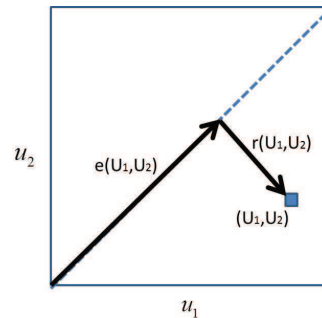


Abbildung 6.4: Die Abstandsmaße  $r$  und  $e$  zum Punkt  $U = (U_1, U_2)$

## Schritt 2: Anwendung

Um sowohl Abhängigkeitsverstärkungen als auch nachlassende Zusammenhänge der Merkmale erfassen zu können, werden das erste ( $lcl$ ) und das dritte ( $ucl$ ) Quartil der Statistik  $r$  in Abhängigkeit von  $e$  betrachtet. Dadurch ergeben sich „Wahrscheinlichkeitsbänder“, in denen sich die transformierte Beobachtung  $(r(U_1, U_2), e(U_1, U_2))$  mit einer Wahrscheinlichkeit von 0.5 realisiert. Zur Aussage, ob eine Stichprobe der



unter  $H_0$  erwarteten Copulastruktur entstammt, wird die Summe der Zufallsvariablen

$$R_i := \begin{cases} 0 & \text{falls } r_i \leq lcl(e_i) \\ 1 & \text{falls } lcl(e_i) < r_i \leq ucl(e_i) \\ 2 & \text{falls } r_i > ucl(e_i) \end{cases} \quad (6.3)$$

gebildet. Gleichung (6.3) wird durch Abbildung 6.5 im (um  $45^\circ$ ) gekippten Einheitswürfel veranschaulicht.

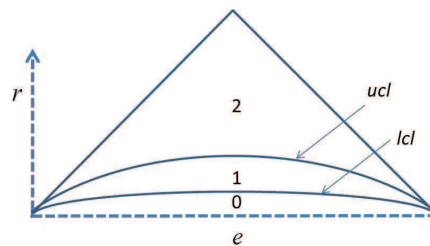


Abbildung 6.5: Gleichung (6.3), veranschaulicht im gekippten Einheitswürfel (um  $45^\circ$ )

Die einzelnen  $R_i$  verhalten sich unter  $H_0$  unabhängig identisch binomialverteilt nach  $\mathbb{B}(2, 0.5)$ . Sie beschreiben die Tatsache, dass (für unendlich große Stichproben) ein Viertel der Werte oberhalb und ein Viertel der Werte unterhalb des Wahrscheinlichkeitsbandes liegen.

Die Summe

$$R_n = \sum_{i=1}^n R_i \quad (6.4)$$

verhält sich wiederum binomialverteilt ( $\mathbb{B}(2n, 0.5)$ ). Für die entsprechende Verteilungsfunktion gilt also

$$F_{R_n}(x) = P(R_n \leq x) = \sum_{k=0}^x \binom{2n}{k} 0.5^k (1 - 0.5)^{2n-k},$$

wobei  $k$  die Anzahl der insgesamt  $2n$  betrachteten Objekte ist, die mit Wahrscheinlichkeit  $0.5$  realisiert werden. Prinzipiell ist dieser Test auch einseitig durchführbar. Die  $R_n$ -Werte können direkt in eine Kontrollkarte abgetragen werden, welche die

entsprechenden Quantile der Binomialverteilung als Grenzwerte nutzt. Damit gilt

$$\text{LCL}(R_n) = P(R_n \leq x)_{\alpha/2} \text{ und}$$

$$\text{UCL}(R_n) = P(R_n \leq x)_{1-\alpha/2}.$$

### Zusammenfassung der Testroutine

Der Test erfolgt zweistufig. Gegeben sei eine Stichprobe aus Zufallszahlen  $X_i = (X_{i1}, X_{i2})$  mit  $i = 1, \dots, n$ . Für die einzelnen Stichprobenpunkte werden in Stufe 1 zunächst die  $X_i$  umgeformt zu  $U_i = (U_{i1}, U_{i2})$  und daraus wird nach den Formeln (6.1) und (6.2)  $r(U_{i1}, U_{i2})$  sowie  $e(U_{i1}, U_{i2})$  berechnet. Schritt 1 ist wichtig, um die Abweichung in beide Richtungen nachweisen zu können. Das wird durch die Spiegelung an der Winkelhalbierenden erreicht. Dadurch werden sowohl Abweichungen nach unten als auch nach oben in der Abhängigkeitsstruktur der Daten sichtbar.

Im zweiten Schritt werden die  $R_i$  nach Formel (6.3) berechnet und summiert (Formel (6.4)). Eine Kurzfassung beider Schritte liefert Tabelle B.1 im Anhang.

### 6.1.3 „Entwicklungsgeschichte“ des vorgestellten Tests

Das vorgestellte Testverfahren ist das Ergebnis einer mehrstufigen Entwicklung. Ziel war es, einen Test zu konstruieren, der auf Grundlage einer relativ kleinen Stichprobe Veränderungen in der Abhängigkeitsstruktur zweier Merkmale findet.

Möglich wäre, nur einen abgeschlossenen (rechteckigen) Teilbereich des Einheitsquadrates zu betrachten. Die Messwerte aus diesem Bereich müssten dann bestimmten Anforderungen genügen. Dadurch bleiben viele Informationen ungenutzt. Bei großen Datenmengen ist das weniger wichtig, da trotzdem noch „genügend viele“ Werte für genaue Aussagen verfügbar sind.

Die hier benötigte Stichprobe soll aber nicht zu groß werden. Daher betrachtet die entwickelte Teststatistik den gesamten Datenraum.

Erst wurde das Abstandsmaß  $r$  entwickelt und verwendet. Die Abhängigkeit des Abstandes von der Distanz des Lotes eines Punktes zum Koordinatenursprung  $e$

blieb unberücksichtigt. Diese Idee hat den großen Nachteil, dass alle „Eckwerte“ - also Werte mit kleinem und großen  $e$ -Wert - nach Konstruktion einer Copula nur kleinere  $r$ -Werte erreichen können als Daten mit mittleren  $e$ -Werten. Die Testentscheidung basiert also hauptsächlich auf den mittleren Datenpunkten. Zunächst wurde berücksichtigt, auf welcher Seite der Winkelhalbierenden der Datenpunkt liegt. Die Werte rechts der Geraden erhielten ein negatives Vorzeichen. Das sollte helfen, alle Informationen der Messwerte zu nutzen. Als mögliche Teststatistiken wurden zunächst das arithmetische Mittel und die Summe der  $r$ -Werte in Betracht gezogen. Unter  $H_0$  müssten diese etwa null betragen. Durch die Achsensymmetrie der betrachteten Copulas sind Veränderungen in der Struktur so nicht nachzuweisen. Strukturänderungen zeigen sich in enger oder weiter um die Winkelhalbierende streuenden Messwerten, deren Erwartungswert wieder null beträgt.

Alternative Teststatistiken sind das Maximum oder Minimum aller  $r$ -Werte oder die Anzahl der Werte, die über oder unter einer bestimmten Grenze liegen. Hierbei sind zwei Grenzwerte nötig; gewählt wurde das  $\pm(1 - \frac{\alpha}{2})$ -Quantil der Verteilung. Nachweisbar ist damit ein schwächer werdender Zusammenhang zwischen den Messwerten. Die Abnahme der ARL-Werte mit zunehmender Abweichung von  $H_0$  erfolgt relativ langsam, so dass die Karte keinen praktischen Nutzen verspricht.

Als zweiter Schritt wurde der Absolutwert von  $r$  betrachtet. Damit werden das arithmetische Mittel oder die Summe der  $r$  sinnvolle Teststatistiken. Wählt man als Grenzwerte das  $\frac{\alpha}{2}$  und das  $(1 - \frac{\alpha}{2})$ -Quantil der Verteilung der Teststatistik, ist sowohl ein verstärkter als auch ein abgeschwächter Zusammenhang nachweisbar. Die Güte der Karte, gemessen an der ARL des Außer-Kontrolle-Prozesses, ist dabei nicht zufriedenstellend. Weiterhin beeinflusst nur der Bereich der mittleren  $e$ -Werte die Testentscheidung.

In einem dritten Versuch wurde der  $e$ -Wert mit aufgenommen und die Abhängigkeit zwischen  $r$  und  $e$  modelliert.

Der Mittelwert oder die Summe der  $r$ -Werte als Teststatistik erweisen sich als ungeeignet, da der zugehörige  $e$ -Wert nicht beachtet wird. Versuche,  $e$  hier mit einzubeziehen, schlugen fehl. Eine Idee war beispielsweise, den absoluten Abstand des realisierten  $e$ -Wertes zur Intervallmitte ( $\frac{1}{\sqrt{2}}$ ) zu nutzen. Bis dahin steigen die

möglichen Ausprägungen der  $r$ -Werte, anschließend sinken sie. Das arithmetische Mittel der  $r$  sollte abhängig vom Mittelwert der  $(\frac{1}{\sqrt{2}} - e)$ -Werte betrachtet werden. Hier kam es teilweise sogar zu wachsenden ARL-Werten außerhalb von  $H_0$ . Das lässt sich folgendermaßen erklären: Die  $r$  können theoretisch bei gleichem Abstand rechts und links von  $e = \frac{1}{\sqrt{2}}$  Ausprägungen bis zum gleichen Wert erreichen, ehe sie an die „Grenzen des Einheitsquadrates“ stoßen. Je nach Form der zugrundeliegenden Copula wird der zur Verfügung stehende Platz allerdings unterschiedlich genutzt, wie beispielsweise Abbildung 6.3 zeigt.

Schließlich entstand die Idee, die Datenpunkte  $(r, e)$  vor Berechnung der Teststatistik zu klassieren.

Zunächst wurde eine Grenze auf Höhe des Erwartungswertes von  $r$  in Abhängigkeit von  $e$  betrachtet. Unter  $H_0$  müsste etwa die Hälfte der Realisationen über dieser Grenze liegen. Werden es (deutlich) mehr, kann das auf einen abschwächenden Zusammenhang der Messwerte hindeuten und umgekehrt. Als Teststatistik bietet sich die Anzahl der  $r$  an, die den Erwartungswert übertreffen. Binär kodiert ist die Wahrscheinlichkeit der einzelnen Messwerte durch eine Bernoulli-Verteilung modellierbar und ein entsprechender Test leicht herzuleiten. Die ARL sinkt für Außer-Kontrolle-Fälle. Der Test ist geeignet, Strukturänderungen nachzuweisen. Allerdings werden die Änderungen teilweise erst relativ spät erkannt.

Daher wurde schließlich versucht, mehrere Klassen zu bilden. Wie bereits erwähnt, betrachten bestehende Tests meist nur einen Ausschnitt des Einheitsquadrates. Meist wird die linke untere Ecke gewählt, teilweise auch die obere rechte Ecke. Änderungen in der Struktur sollen sich am ehesten in diesen Abschnitten zeigen. In diesen Ecken realisieren sich umso mehr Daten, je stärker der Zusammenhang zwischen den Merkmalen wird. Die beiden anderen Ecken werden dagegen bei nachlassendem Zusammenhang „voller“. Betrachtet man das komplette Einheitsquadrat, realisieren sich die Datenpunkte bei starker Abhängigkeit tendenziell dichter an der Winkelhalbierenden. Bei schwacher Abhängigkeit kommen vermehrt auch Messwerte nahe der Außengrenzen des Einheitsquadrates vor. Es sollten also vor allem Bereiche nahe und fern der Winkelhalbierenden betrachtet werden. Damit ergeben sich zwei Grenzwerte, um die Messungen in die drei Klassen „nah“, „mittel“ und „fern“ einzuteilen. Mögliche Einteilungen waren das  $\frac{1}{3}$ - und das  $\frac{2}{3}$ -Quantil der Verteilung von  $r$

in Abhängigkeit von  $e$  zu wählen oder die beschriebene Quartileinteilung. Die Entscheidung zugunsten der Quartile ergab sich aus der einfachen Modellierbarkeit der Datenpunkte unter  $H_0$  mit Hilfe einer Binomialverteilung. Spätere Untersuchungen bestätigen diese Entscheidung auch inhaltlich. Die so entstehende Karte zeigt ein adäquates Güteverhalten.

## 6.2 Konstruktion der Grenzwerte $lcl$ und $ucl$

### 6.2.1 Herleitung der benötigten Wahrscheinlichkeiten

Der Prozess unter Kontrolle folge einer bestimmten Copulastruktur. Gegeben sei der Zufallsvektor  $U := (U_1, U_2)$  mit gleichverteilten Einträgen auf dem Intervall  $[0, 1]$  ( $U_k \sim \mathbb{C}[0, 1]$ ,  $k = 1, 2$ ) sowie die transformierten Punkte  $e$  und  $r$  nach den Formeln (6.1) und (6.2).

Es bezeichne  $(u_1, u_2)$  eine (beliebige) Realisation von  $(U_1, U_2)$  mit  $r = r(u_1, u_2)$  und  $e = e(u_1, u_2)$ . Durch die Beschränkung auf achsensymmetrische Copulamodelle ändert eine Vertauschung von  $u_1$  und  $u_2$  die Werte von  $(r, e)$  nicht, daher gelte ohne Beschränkung der Allgemeinheit  $u_1 \geq u_2$ .

Die bedingte Wahrscheinlichkeit  $P(r \leq r^* | e_1^* \leq e \leq e_2^*)$  wird mit zunächst beliebigen, aber festen Punkten  $r^* \in [0, \frac{1}{\sqrt{2}}]$  und  $e_1^*, e_2^* \in [0, \sqrt{2}]$ , verwendet. Dieser Wert wird mit Hilfe der Wahrscheinlichkeit

$$P(e_1^* \leq e \leq e_2^*) = P(e \leq e_2^*) - P(e \leq e_1^*) \quad (6.5)$$

ermittelt. Außerdem wird die Wahrscheinlichkeit dafür benötigt, dass  $e$  sich innerhalb eines Intervalls  $e_1^*$  und  $e_2^*$  realisiert, also  $P(r \leq r^*, e_1^* \leq e \leq e_2^*)$ . Dieses Intervall wird in Grafik 6.6 (links) verdeutlicht.

Es gilt

$$P(r \leq r^*, e_1^* \leq e \leq e_2^*) = P(r \leq r^*, e \leq e_2^*) - P(r \leq r^*, e \leq e_1^*). \quad (6.6)$$

**Die Wahrscheinlichkeit  $P(r \leq r^*, e \leq e^*)$**

Das Dreieck  $B$  enthalte alle Punkte  $(v_1, v_2) \in [0, 1]^2$  innerhalb einer Fläche mit den Eckpunkten  $\{(0, u_1 - u_2), (0, u_1 + u_2), (u_1, u_2)\}$ , wie die rechte Grafik aus Abbildung 6.6 zeigt. Dann ist

$$P(r \leq r^*, e \leq e^*) = P(e \leq e^*) - 2 P(B). \quad (6.7)$$

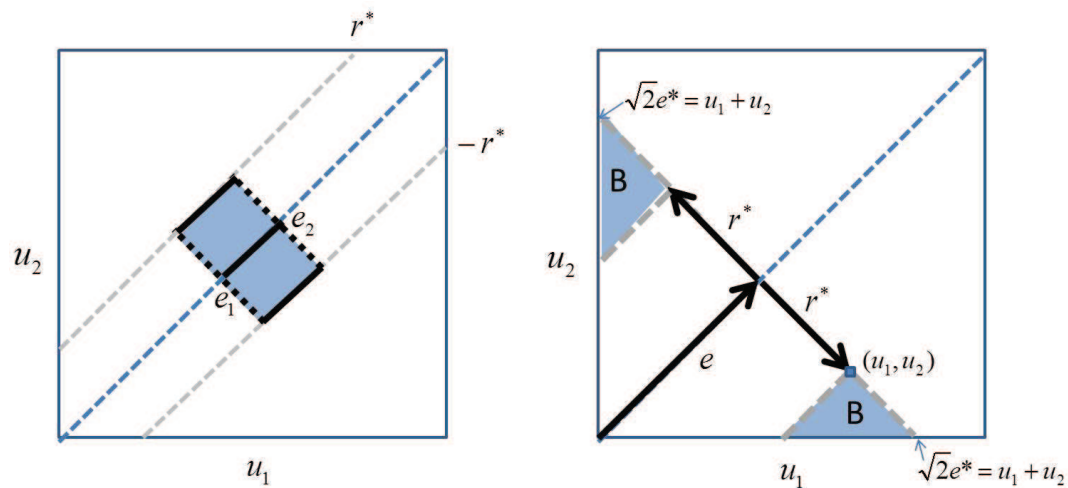


Abbildung 6.6: Veranschaulichung von  $P(r \leq r^*, e_1^* \leq e \leq e_2^*)$  (links) und des Dreiecks  $B$  zur Berechnung von  $P(r \leq r^*, e \leq e^*)$  (rechts)

Für die Wahrscheinlichkeit  $P(e \leq e^*)$  gilt unter Verwendung von  $c(\tau_1, \tau_2)$  als Dichte der Copula an der Stelle  $(\tau_1, \tau_2)$ :

$$P(e \leq e^*) = P\left(\frac{u_1 + u_2}{\sqrt{2}} \leq e^*\right) = \int \int_{\tau_1 + \tau_2 \leq \sqrt{2}e^*} c(\tau_1, \tau_2) d\tau_1 d\tau_2$$

Dabei liegt  $\sqrt{2}e^*$  im Intervall  $[0, 2]$ . Rechnet man ab der Hälfte des Intervalls (das entspricht der Mitte der Winkelhalbierenden) mit der Gegenwahrscheinlichkeit, so gilt

$$P(u_1 + u_2 \leq \sqrt{2}e^*) =$$

$$\begin{cases} 0 & \text{falls } \sqrt{2}e^* < 0 \\ \int_0^{\sqrt{2}e^*} \int_0^{\sqrt{2}e^* - \tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 & \text{falls } 0 \leq e^* < \frac{1}{\sqrt{2}} \\ 1 - \int_{\sqrt{2}e^* - 1}^1 \int_{\sqrt{2}e^* - \tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 & \text{falls } \frac{1}{\sqrt{2}} \leq e^* < \sqrt{2} \\ 1 & \text{falls } e^* \geq \sqrt{2} \end{cases}$$

Daraus folgt mit dem Zusammenhang  $\sqrt{2}e^* = u_1 + u_2$ :

$$P(e \leq e^*) = \begin{cases} 0 & \text{falls } e^* < 0 \\ \int_0^{u_1+u_2} \int_0^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 & \text{falls } 0 \leq e^* < \frac{1}{\sqrt{2}} \\ 1 - \int_{u_1+u_2-1}^1 \int_{u_1+u_2-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 & \text{falls } \frac{1}{\sqrt{2}} \leq e^* < \sqrt{2} \\ 1 & \text{falls } e^* \geq \sqrt{2} \end{cases}$$

Und für die Dreiecksfläche  $B$  gilt:

$$P(B) = \begin{cases} 0 & \text{falls } e^* \leq 0 \text{ oder } r^* \leq 0 \text{ oder } e^* \leq r^* \\ \int_0^{u_2} \int_{u_1-u_2+\tau_2}^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 & \text{falls } 0 \leq e^* \leq \frac{1}{\sqrt{2}}, 0 \leq r^* \leq e^*, e^* + r^* < \sqrt{2} \\ \int_0^{u_1+u_2-1} \int_{u_1-u_2+\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 \\ + \int_{u_1+u_2-1}^{u_2} \int_{u_1-u_2+\tau_2}^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 & \text{falls } \frac{1}{\sqrt{2}} \leq e^* < \sqrt{2}, 0 \leq r^* \leq e^*, e^* + r^* < \sqrt{2} \\ \int_0^{1-u_1+u_2} \int_{u_1-u_2+\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 & \text{falls } e^* > 0, 0 \leq r^* \leq e^*, e^* + r^* \geq \sqrt{2} \end{cases}$$

Diese vier Möglichkeiten für  $P(B)$  sind durch geometrische Überlegungen gut erklärbar, wie Abbildung 6.7 zeigt.

Im ersten Fall liegt der Punkt  $(u_1, u_2)$  auf der linken oder unteren Grenze des Einheitsquadrates oder sogar schon links außerhalb oder darunter. Damit ist  $B$  leer

und  $P(B) = 0$ .

Bei der zweiten Variante liegt  $(u_1, u_2)$  im Einheitsquadrat unterhalb der Linie  $u_1 + u_2 = 1$ , also der Lotrechten zur Winkelhalbierenden durch die obere linke und die untere rechte Ecke des Einheitsquadrates.

Als dritte Möglichkeit kann sich der Punkt oberhalb dieser Linie, aber innerhalb des Quadrates realisieren. Durch die Beschränkung des Einheitsquadrates wird die interessierende Fläche  $B$  viereckig. Daher wird zunächst der Streifen bis zum äußeren Rand des  $I^2$  berechnet (erster Summand - in der Grafik „ $B_1$ “) und anschließend das fehlende Dreieck (zweiter Summand - in der Grafik „ $B_2$ “).

Zuletzt wird die Möglichkeit betrachtet, dass der Punkt  $(u_1, u_2)$  rechts außerhalb oder auf der oberen Grenze des Einheitsquadrates liegt,  $B$  allerdings trotzdem Punkte innerhalb des Einheitsquadrates enthält (angeschnittene Ecken). Hier muss (wie bei Fall 3) die Beschränkung des Einheitsquadrates bei den Grenzen berücksichtigt werden.

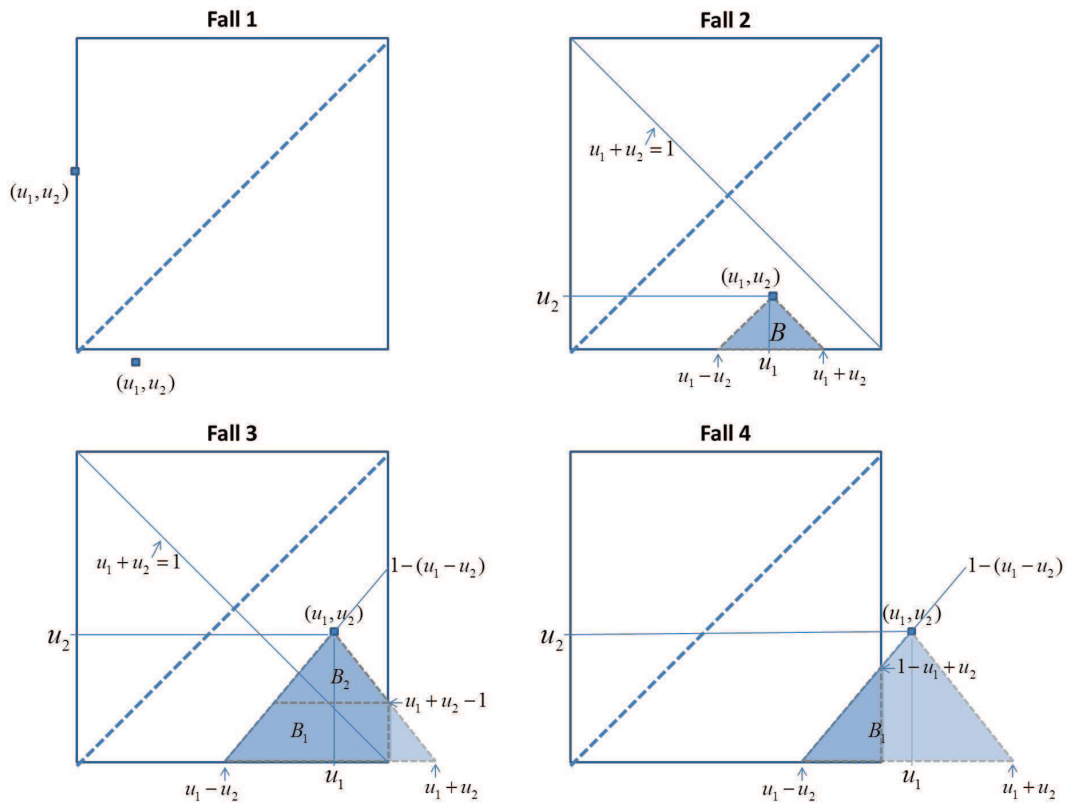


Abbildung 6.7: Veranschaulichung der verschiedenen Möglichkeiten für  $P(B)$



Die Wahrscheinlichkeiten  $P(e \leq e^*)$  und  $P(B)$  können durch Zusatzinformationen vereinfacht werden. Die Zufallsvariable  $(U_1, U_2)$  hat laut Konstruktionsvorschrift (siehe Abschnitt 6.1.2, Schritt 1) nur Realisierungen innerhalb oder auf den Grenzen des Einheitsquadrates. Weiterhin sei gefordert, dass die untere Intervallgrenze  $e_1^*$  kleiner als die obere  $e_2^*$  ist, also dass  $0 \leq e_1^* < e_2^* \leq \sqrt{2}$ . Weiter gilt:

$$r^* \leq \begin{cases} e_1^* & \text{falls } e_2^* \leq \frac{1}{\sqrt{2}} \\ \min(e_1^*, 1 - e_2^*) & \text{falls } e_1^* \leq \frac{1}{\sqrt{2}} \text{ und } e_2^* \geq \frac{1}{\sqrt{2}} \\ 1 - e_2^* & \text{falls } e_1^* \geq \frac{1}{\sqrt{2}} \end{cases}$$

Hieraus folgt

$$P(e \leq e^*) =$$

$$\begin{cases} \int_0^{u_1+u_2} \int_0^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 & \text{falls } 0 \leq e^* < \frac{1}{\sqrt{2}} \\ 1 - \int_{u_1+u_2-1}^1 \int_{u_1+u_2-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 & \text{falls } \frac{1}{\sqrt{2}} \leq e^* < \sqrt{2} \\ 1 & \text{falls } e^* = \sqrt{2} \end{cases} \quad (6.8)$$

und

$$P(B) =$$

$$\begin{cases} \int_0^{u_2} \int_{u_1-u_2+\tau_2}^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 & \text{falls } 0 \leq e^* < \frac{1}{\sqrt{2}} \\ \int_0^{u_1+u_2-1} \int_{u_1-u_2+\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 \\ + \int_{u_1+u_2-1}^{u_2} \int_{u_1-u_2+\tau_2}^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 & \text{falls } \frac{1}{\sqrt{2}} \leq e^* < \sqrt{2} \\ 0 & \text{falls } e^* = \sqrt{2} \end{cases} \quad (6.9)$$

Die Punkte  $(u_1, u_2)$  und  $(u_3, u_4)$  seien Eckpunkte des interessierenden Quadrates. Beide Punkte hängen von  $e_1^*$  und  $e_2^*$  sowie von  $r^*$  ab, wie Abbildung 6.8 zeigt.

Dann liefert die Anwendung der Formeln (6.8), (6.9) und (6.7) auf Gleichung (6.6) die gesuchte Wahrscheinlichkeit  $P(r \leq r^*, e_1^* \leq e \leq e_2^*)$ . Eine Aufstellung für die verschiedenen Möglichkeiten der Grenzen  $e_1^*$  und  $e_2^*$  zeigt Anhang A.1.1. Betrachtet

werden die folgenden Fälle:

- $e_2^* < \frac{1}{\sqrt{2}}$
- $e_1^* < \frac{1}{\sqrt{2}}$  und  $\frac{1}{\sqrt{2}} \leq e_2^* < \sqrt{2}$
- $\frac{1}{\sqrt{2}} \leq e_1^* < e_2^* < \sqrt{2}$
- $\frac{1}{\sqrt{2}} \leq e_1^* < \sqrt{2}$  und  $e_2^* = \sqrt{2}$

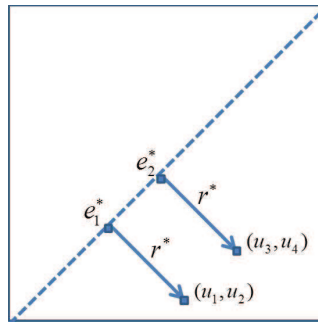


Abbildung 6.8: Skizze der Punkte  $(u_1, u_2)$  und  $(u_3, u_4)$

Auch der Fall  $e_1^* \leq \frac{1}{\sqrt{2}}$  und  $e_2^* = \sqrt{2}$  wäre denkbar, da er für die Konstruktion der Grenzwerte aber nicht verwendet wird, ist er nicht aufgeführt.

**Die Wahrscheinlichkeit**  $P(e_1^* \leq e \leq e_2^*)$

Die Wahrscheinlichkeit  $P(e_1^* \leq e \leq e_2^*)$  ergibt sich durch Einsetzen von Formel (6.8) in Gleichung (6.5). Die möglichen Ausprägungen sind im Anhang A.1.2 zu finden.

**Die Wahrscheinlichkeit**  $P(r \leq r^* | e_1^* \leq e \leq e_2^*)$

Die bedingte Wahrscheinlichkeit

$$P(r \leq r^* | e_1^* \leq e \leq e_2^*) = \frac{P(r \leq r^*, e_1^* \leq e \leq e_2^*)}{P(e_1^* \leq e \leq e_2^*)}$$

ergibt sich als Quotient der Formeln aus dem Anhang A.1.1 und A.1.2.

Alternativ bewirkt das Einsetzen von Formel (6.7) in (6.6) die Gleichung

$$P(r \leq r^*, e_1^* \leq e \leq e_2^*) = P(e \leq e_2^*) - 2 P(B_2) - [P(e \leq e_1^*) - 2 P(B_1)],$$

wobei  $B_1$  das zu  $e_1^*$  gehörige Dreieck bezeichnet und  $B_2$  das zu  $e_2^*$  gehörende. Der obige Ausdruck dividiert durch  $[P(e \leq e_2^*) - P(e \leq e_1^*)]$  (siehe Formel (6.5)) ergibt die gesuchte bedingte Wahrscheinlichkeit. Damit gilt:

$$P(r \leq r^* | e_1^* \leq e \leq e_2^*) = 1 - \frac{2P(B_2) - 2P(B_1)}{P(e \leq e_2^*) - P(e \leq e_1^*)} \quad (6.10)$$

Die entsprechende Fallunterscheidung ist im Anhang A.1.3 angegeben.

## 6.2.2 Bestimmung der Wahrscheinlichkeitsbänder

Als Grenzwerte ( $lcl(e), ucl(e)$ ) werden die Quartile der Statistik  $r$  in Abhängigkeit von  $e$  genutzt. Wie die Einzelwahrscheinlichkeiten aus Abschnitt 6.2.1 hängen auch die Grenzen von der unterstellten Copulastruktur der Daten ab. Je nach Copula sind die Grenzwerte geschlossen darstellbar oder müssen numerisch berechnet werden. Allgemein sind folgende Gleichungen zu erfüllen:

$$\begin{aligned} lcl &= \arg \max_{r^*} (P(r \leq r^* | e_1^* \leq e < e_2^*) = 0.25) \\ ucl &= \arg \max_{r^*} (P(r \leq r^* | e_1^* \leq e < e_2^*) = 0.75) \end{aligned} \quad (6.11)$$

In der Anwendung werden die Grenzen so bestimmt, dass die Winkelhalbierende des Einheitsquadrates zunächst in gleichmäßige Abschnitte der Länge  $2\varepsilon$  eingeteilt wird. Formel (6.11) wird dann auf alle Intervalle angewendet. Insgesamt gibt es  $K := \frac{\sqrt{2}}{(2\varepsilon)}$  Intervalle mit den Grenzen ( $e_1^* = 2k\varepsilon, e_2^* = 2(k+1)\varepsilon$ ).

Damit stehen für alle Wertepaare  $(r_i, e_i), i = 1, \dots, n$ , einer Stichprobe Referenzwerte zur Verfügung, indem  $e_i$  in das  $\lfloor \frac{e_i}{2\varepsilon} + 1 \rfloor$ -te Intervall eingeordnet wird, wobei  $\lfloor \cdot \rfloor$  für die **Abrundungsfunktion** oder **Gaußklammer** steht, also die nächstkleinere ganze Zahl des innenliegenden Wertes verwendet wird. Die Grenzwerte  $lcl(e_i)$  und  $ucl(e_i)$  können dann entsprechend Formel (6.3) mit  $r_i$  verglichen werden.

### 6.2.3 Beispiel Produktcopula

Die Produktcopula bildet unabhängige Daten ab (siehe Abschnitt 3.4). Mit der Formel (3.1) ergibt sich die Dichte  $c(u_1, u_2) = 1$ .

Die für Gleichung (6.10) nötigen Wahrscheinlichkeiten  $P(B_1)$ ,  $P(B_2)$ ,  $P(e \leq e_1^*)$  und  $P(e \leq e_2^*)$  ergeben sich mit Hilfe der Formeln (6.8) und (6.9) für  $p = 1, 2$  als

$$P(e \leq e_p^*) = \begin{cases} 0.5(u_1 + u_2)^2 & \text{falls } 0 \leq e_p^* < \frac{1}{\sqrt{2}} \\ -1 + 2(u_1 + u_2) - 0.5(u_1 + u_2)^2 & \text{falls } \frac{1}{\sqrt{2}} \leq e_p^* < \sqrt{2} \\ 1 & \text{falls } e_p^* = \sqrt{2} \end{cases}$$

und

$$P(B_p) = \begin{cases} u_2^2 & \text{falls } 0 \leq e_p^* < \frac{1}{\sqrt{2}} \\ u_2^2 - 0.5(u_1 + u_2 - 1)^2 & \text{falls } \frac{1}{\sqrt{2}} \leq e_p^* < \sqrt{2} \\ 0 & \text{falls } e_p^* = \sqrt{2} \end{cases}$$

Eine schrittweise Herleitung der Gleichungen ist im Anhang A.1.4 aufgelistet.

Einfügen der obigen Wahrscheinlichkeiten in Formel (6.10) ergibt

$$P(r \leq r^* | e_1^* \leq e \leq e_2^*) = \begin{cases} 1 - \frac{4(u_4^2 - u_2^2)}{(u_3 + u_4)^2 - (u_1 + u_2)^2} & \text{falls } 0 \leq e_1^* < e_2^* < \frac{1}{\sqrt{2}} \\ 1 - \frac{2u_4^2 - 2u_2^2 - (u_3 + u_4 - 1)^2}{-1 + 2(u_3 + u_4) - 0.5(u_3 + u_4)^2 - 0.5(u_1 + u_2)^2} & \text{falls } 0 \leq e_1^* < \frac{1}{\sqrt{2}} \\ & \text{und } \frac{1}{\sqrt{2}} \leq e_2^* < \sqrt{2} \\ 1 - \frac{2u_4^2 - 2u_2^2 - (u_3 + u_4 - 1)^2 + (u_1 + u_2 - 1)^2}{0.5(u_1 + u_2)^2 - 0.5(u_3 + u_4)^2 + 2[(u_3 + u_4) - (u_1 + u_2)]} & \text{falls } \frac{1}{\sqrt{2}} \leq e_1^* < e_2^* < \sqrt{2} \\ 1 - \frac{-2u_2^2 + (u_1 + u_2 - 1)^2}{2 + 0.5(u_1 + u_2)^2 - 2(u_1 + u_2)} & \text{falls } \frac{1}{\sqrt{2}} \leq e_1^* < \sqrt{2} \\ & \text{und } e_2^* = \sqrt{2} \end{cases}$$

Die Produktcopula verbindet unabhängige Zufallsvariablen und kommt daher ohne zusätzlichen Parameter zur Abbildung von Abhängigkeiten aus. Die Grenzwerte sind somit für alle  $\tau$  identisch und symmetrisch um  $\frac{1}{\sqrt{2}}$ .

Auf Basis der hergeleiteten Wahrscheinlichkeiten werden die Grenzen  $lcl$  und  $ucl$  nach Formel (6.10) numerisch bestimmt. Die Intervalllänge  $2\epsilon$  betrage dabei 0.005. Kleinere Distanzen liefern genauere Ergebnisse. Dies ist vor allem in der Region  $e = \frac{1}{\sqrt{2}}$  interessant.

Durch die Unabhängigkeit der Daten bei der Produktcopula gilt hier

$$lcl = \begin{cases} \frac{1}{4}e & \text{für } e \leq \frac{1}{\sqrt{2}} \\ \frac{\sqrt{2}}{4} - \frac{1}{4}e & \text{für } \frac{1}{\sqrt{2}} \leq e \leq \sqrt{2} \end{cases}$$

und

$$ucl = \begin{cases} \frac{3}{4}e & \text{für } e \leq \frac{1}{\sqrt{2}} \\ \frac{3\sqrt{2}}{4} - \frac{3}{4}e & \text{für } \frac{1}{\sqrt{2}} \leq e \leq \sqrt{2}. \end{cases}$$

Dabei handelt es sich um Umrissse von Dreiecken, die  $\frac{1}{4}$  beziehungsweise  $\frac{3}{4}$  der gesamten Fläche (ein Dreieck mit den Eckpunkten  $(0, 0)$ ,  $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$  und  $(\sqrt{2}, 0)$ ) erfassen, in denen sich die Werte realisieren können. Die Grenzen sind in Abbildung 6.9 dargestellt.

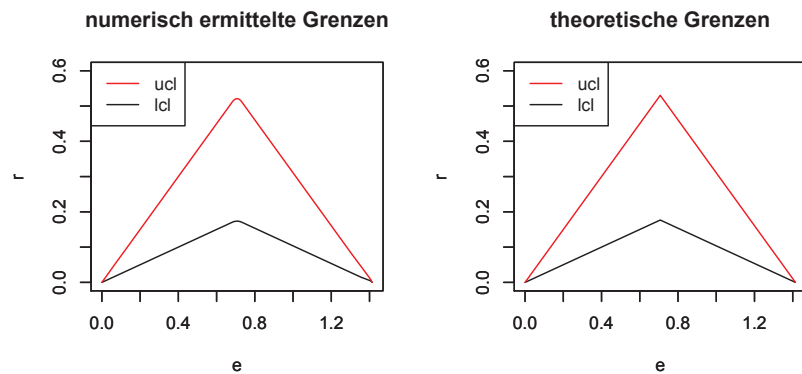


Abbildung 6.9: Kontrollgrenzen beim Verwenden einer Produktcopula

Eine analoge Herleitung der Grenzwerte für die Clayton- oder Gumbelcopula ist nach dieser Herangehensweise nicht möglich, da die Integrale über beide Copuladichten nicht geschlossen darstellbar sind.

## 6.3 Effizienz der Kontrollkarte

Das Verhalten der Kontrollkarte mit der in Kapitel 6.1 beschriebenen Testroutine wird durch Simulationsstudien geprüft. Die Simulationsstudie behandelt eine Karte mit der Anfangsannahme, dass der Prozess unter Kontrolle und die  $H_0$ -Verteilung bekannt ist. Der Prozess wird in Echtzeit, also „online“, überwacht. Die Außer-Kontrolle-Situationen werden durch Verlagerungen des Copulaparameters ausgedrückt. Diese Verschiebungen werden durch Kendalls  $\tau$  quantifiziert.

Die Kontrollgrenzen werden anhand der unterstellten Prozessverteilung festgelegt, unter  $H_0$  also eine bestimmte archimedische Copula. Die Grenzwerte für unabhängige Daten ( $\tau = 0$ ) sind für alle Copulas identisch (siehe Beispiel 6.2.3). Eine Überwachung der (nicht vorhandenen) Abhängigkeitsstruktur ist in diesem Fall nicht sinnvoll.

Es werden zunächst die Kontrollgrenzen  $lcl$  und  $ucl$  sowie LCL und UCL hergeleitet, dann zufällig Stichproben verschiedener Prozesssituationen erzeugt und die daraus resultierenden Lauflängen zur Gütebeurteilung des Tests herangezogen. In Kapitel 4.3 wird die ARL als mittlere Lauflänge des Prozesses definiert. Das Verhalten der ARL-Funktion sowie der Verlauf der zugehörigen Standardabweichung SDRL wird in diesem Kapitel durch Monte-Carlo-Simulationen ermittelt. Zusätzlich wird auch die OC-Funktion angegeben. Diese wurde aus den ARL-Werten berechnet (siehe Abschnitt 4.4).

### 6.3.1 Berechnung von $lcl$ und $ucl$

Die Kontrollgrenzen  $lcl(e)$  und  $ucl(e)$  werden durch Integration aus der Dichtefunktion der zugrunde gelegten Copula bestimmt. Dazu werden die in Kapitel 6.2 hergeleiteten Formeln genutzt. Der Wertebereich  $E$  von  $e$  wurde dazu in 148 Abschnitte unterschiedlicher Länge eingeteilt, die zur Mitte ( $\frac{1}{\sqrt{2}}$ ) kleiner wurden.

Abbildung 6.10 zeigt die entsprechenden Grenzwerte für  $r$  unter  $H_0$  für eine Claytoncopula mit bekanntem Parameter (oben) und für eine festgelegte Gumbelcopula (unten). Diese Grenzen gelten unter der Nullhypothese, dass die untersuchte neue Stichprobe aus der erwarteten Verteilung stammt.

Die Grenzwerte zeigen dabei eine Struktur, die sich entsprechend der Copula er-

klären lässt. Ein Vergleich der Grenzwerte mit der Dichtestruktur aus Abbildung 3.1 zeigt einen deutlichen Zusammenhang beider Maße. Das ist durch die alternative Darstellungsmöglichkeit der Grenzwerte als Integrale aus der Dichtefunktion erklärbar, siehe Formel (6.11).

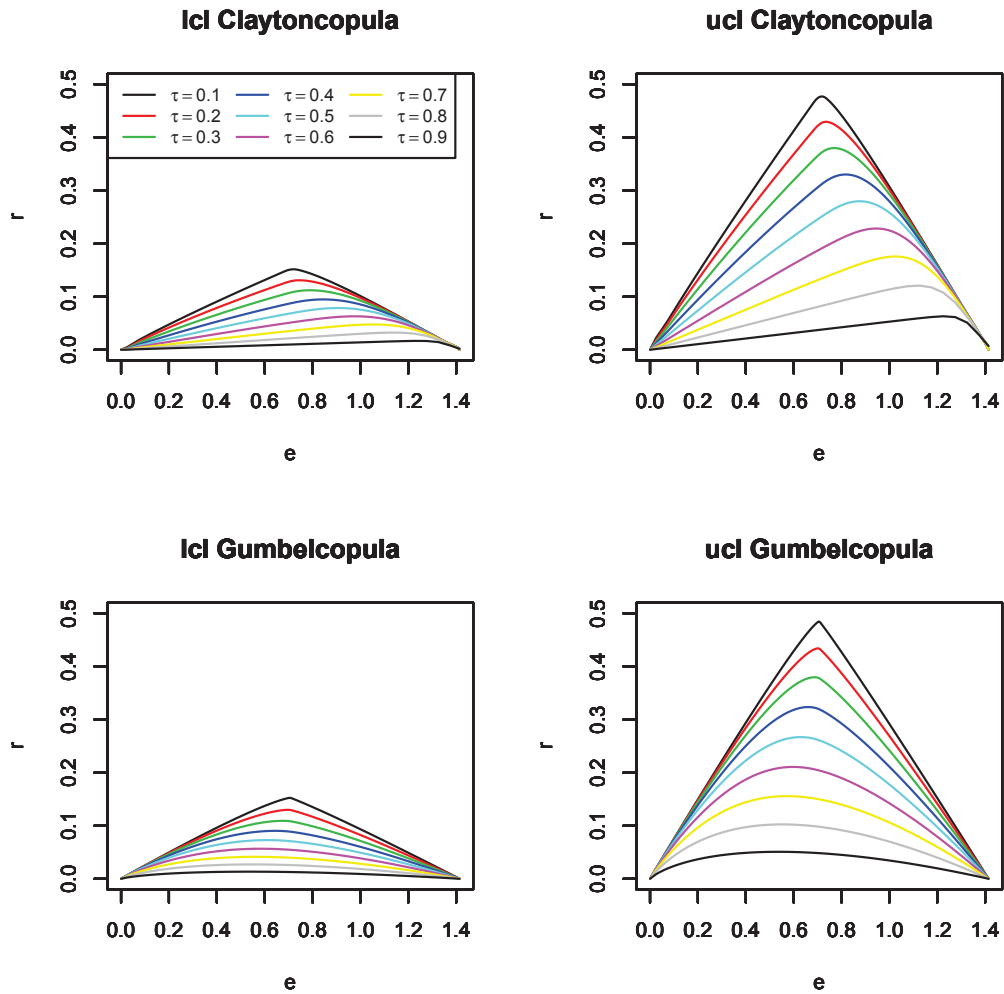


Abbildung 6.10: Kontrollgrenzen bei Unterstellung einer festgelegten Copulastruktur für verschiedene  $\tau$

Bei den Grenzen handelt es sich um Quantilsfunktionen - in dem vorliegenden Fall werden das erste und das dritte Quartil betrachtet. Eine allgemeinere Übersicht über die Quantile von  $r$ , abhängig von  $e$  für feste  $\tau$  liefert Kapitel 6.6. Bei einer

Claytoncopula konzentriert sich die Datenmasse im unteren linken Bereich des Einheitsquadrates, die Grenzwerte sind hier dementsprechend linksschief beziehungsweise rechtssteil. Bei den Gumbelcopulas konzentriert sich die Datenmasse eher im oberen rechten Quadranten, die Grenzwerte weisen analog eine linkssteile, rechtschiefe Struktur auf. Die Graphen der Grenzwerte zeigen einen ähnlichen Verlauf. Das lässt vermuten, dass eine Approximation der Grenzen durch stetige Funktionen möglich ist. Diese Anpassung und eine Untersuchung, inwieweit die Verwendung angenäherter Grenzwerte die Qualität der Regelkarten beeinflusst, erfolgt in Abschnitt 6.3.3.

### 6.3.2 Ergebnisse der Lauflängensimulation

Die entwickelten Kontrollkarten nutzen die im letzten Abschnitt hergeleiteten Grenzen in Kombination mit der Teststatistik  $R_i$  aus Abschnitt 6.1. Die Lauflängen werden durch Monte-Carlo-Simulationen auf Basis von Zufallsstichproben aus verschiedenen Prozesssituationen ermittelt. Neben den ARL- und SDRL-Werten wird die OC-Funktion zur Gütebeurteilung des Tests herangezogen.

Für kontrollierte Prozesse ( $\tau = \tau_0$ ) ergeben sich Werte für die ARL entsprechend der Quantile der Binomialverteilung, abhängig von der gewählten Stichprobengröße. Nachfolgend werden verschiedene Prozesssituationen für die im letzten Abschnitt festgelegten Regeln betrachtet (unter Kontrolle oder außer Kontrolle). Die verschiedenen Prozesszustände werden durch Variieren des Copulaparameters beziehungsweise des zugehörigen  $\tau$  erzeugt. Ein Intervall von  $[\tau_0 - 0.1, \tau_0 + 0.1]$  erweist sich als ausreichend, um Abweichungen von der In-Kontrolle-Situation nachzuweisen.

Abweichungen des wahren  $\tau$  von  $\tau_0$  werden in 0.02-er Schritten betrachtet,  $\tau_0$  wurde in 0.1-er Abständen im Bereich von  $\tau_0 = 0.1$  bis  $\tau_0 = 0.9$  gewählt.

Insgesamt wurden  $N = 20.000$  Simulationsläufe durchgeführt, also 20.000 Lauflängen simuliert, die auf Einzelstichproben von je  $n = 30$  berücksichtigten Messpunkten basieren. Diese Zahlen ergaben sich durch Vorabuntersuchungen als Kompromiss zwischen Rechenzeit und Streuung der Ergebnisse. Vor allem die Erzeugung der Zufallszahlen auf Basis von Copulas erwies sich dabei als rechenaufwendig.

Für alle Stichprobenpunkte wurde zunächst die Teststatistik  $R_i$ ,  $i = 1, \dots, 30$ , mit Hilfe der Grenzen  $lcl$  und  $ucl$  aus Abschnitt 6.3.1 berechnet. Die kumulierten  $R_i$  entsprechen der in Gleichung (6.4) beschriebenen Teststatistik  $R_n$ . Zu einem



Signifikanzniveau  $\alpha \approx 0.02734$  ergibt sich als untere Grenze von  $R_n$  LCL= 22 und als obere Grenze UCL= 38 (siehe Abschnitt 6.1.2, Schritt 2). Daraus resultiert eine In-Kontrolle-ARL von etwa 36.58. Das unübliche Signifikanzniveau  $\alpha$  erklärt sich aus der Vorgabe, das nächstmögliche Niveau  $\alpha \leq 0.05$  zu wählen, für das (nach Konstruktion des Tests notwendigerweise ganzzahlige) Grenzwerte existieren.

Abbildung 6.11 zeigt die Ergebnisse der Simulation unter Annahme einer Claytoncopulastruktur in den Daten, einige Datenpunkte sind zusätzlich in Tabelle B.2 in Anhang B aufgelistet.

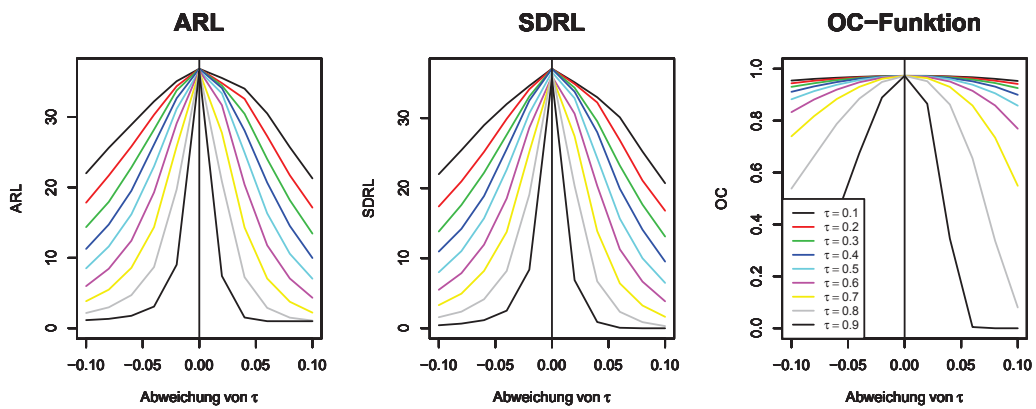


Abbildung 6.11: Simulationsergebnisse zur Lauflänge bei Unterstellung einer Claytoncopulastruktur mit verschiedenen Kendalls  $\tau$

Die Resultate für die Gumbelcopula zeigen Abbildung 6.12 und Tabelle B.3 (Anhang B).

Die Simulation verdeutlicht, dass der Test für die untersuchte Fragestellung geeignet ist. Für beide Abhängigkeitsstrukturen erreicht die ARL-Funktion ihr Maximum - etwa entsprechend der theoretisch erwarteten ARL - unter der Nullhypothese, also bei Übereinstimmung von  $\tau$  und  $\tau_0$ . Dieser ARL-Wert wird für alle betrachteten Fälle von  $\tau_0 = 0.1$  bis  $\tau_0 = 0.9$  erreicht.

Mit wachsendem Abstand vom wahren  $\tau$ -Wert des Prozesses zu dem  $\tau_0$  unter Kontrolle verringert sich die ARL. Je deutlicher man also in den Bereich der Alternativhypothese kommt, umso früher erkennt der Test, dass die Sollwerte nicht eingehalten wurden.

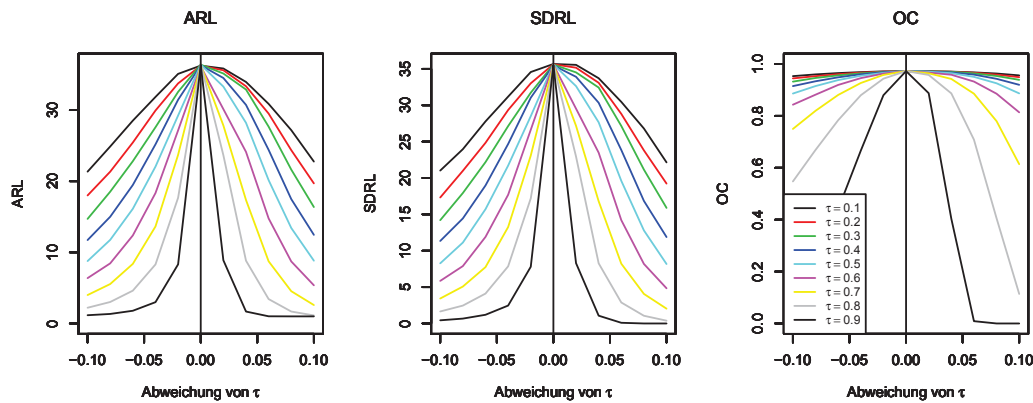


Abbildung 6.12: Simulationsergebnisse zur Lauflänge bei Unterstellung einer Gumbelcopula mit verschiedenen Kendalls  $\tau$

Die Kontrollkarte erkennt sowohl eine Verstärkung als auch eine Abschwächung des Zusammenhangs der Datenpunkte. Kendalls  $\tau$  ist nichtsymmetrisch, siehe auch Abbildung 3.2. Es gilt mit steigenden Abhängigkeiten zwischen den Variablen, dass für gleichbleibende Zuwächse der  $\tau$ -Werte kleinere Zuwächse der  $\rho$  denselben Zusammenhang widerspiegeln. Das heißt, eine Erhöhung von  $\tau$  um einen festen Wert bewirkt ein geringeres Wachstum von  $\rho$  als eine Erniedrigung von  $\tau$  um denselben Wert. Dies könnte erklären, warum die ARL und die SDRL einen Bogen hin zu den größeren  $\tau$ -Werten (verzögertes Absinken der Werte rechts) zeigen. Dass die ARL mit wachsendem  $\tau_0$  trotzdem so schnell sinkt, spricht für die Qualität der Teststatistik. Die SDRL zeigt keine Auffälligkeit; sie erscheint in sich konsistent und zur jeweiligen ARL passend. Die Karte ist damit also offenbar geeignet, die Strukturkonstanz der Daten zu überwachen.

Obwohl die ARL bei der Claytoncopula rechts von  $\tau_0$  zunächst schwächer fällt, sind die ARL für  $\tau_0 + 0.1$  durchweg geringer als für  $\tau_0 - 0.1$ : Ansteigende Zusammenhänge werden also - nach einer kurzen Übergangszeit - schneller erkannt als ein abschwächender Zusammenhang zwischen Merkmalen. In der Praxis dürfte dieser Fall auch der interessantere sein, da beispielsweise in der Finanzökonomie ein verstärkter Zusammenhang zwischen Aktienrenditen auf Krisen hinweisen kann.

Für die Gumbelcopula zeigen die Simulationen ähnliche Resultate, Abweichungen in der Zusammenhangsstruktur werden sicher erkannt und auch hier sinken die ARL

und die SDRL schneller mit wachsendem Zusammenhang zwischen den Datenpunkten. Dabei ist ein schnelleres Absinken links von  $\tau_0$  als rechts davon zu erkennen, die SDRL zeigt also dieselbe Tendenz wie die ARL. Auch die Beobachtung ist durch die unsymmetrischen  $\tau$  erklärbar.

Die OC-Funktion liefert weitere Hinweise zur Güte der Karte: Je steiler sie abfällt, desto besser ist die zugrundeliegende Kontrollkarte. Auffällig sind die durchgängig relativ hohen OC-Werte für Merkmale, die nur einen schwachen Zusammenhang haben. Für kleine  $\tau$  ist es relativ wahrscheinlich, den Prozess irrtümlich als unter Kontrolle anzusehen. Diese Wahrscheinlichkeit liegt für  $\tau_0 \leq 0.4$  bei 90%, wenn die Abweichung  $\pm 0.1$  betrachtet wird (siehe Tabelle B.2 und B.3). Ein direkter Vergleich mit der ARL zeigt allerdings auch, dass die betrachteten Kontrollkarten für  $\tau_0 = 0.4$  und einem wahren  $\tau = 0.3$  beziehungsweise  $\tau = 0.5$  schon dreimal so häufig ablehnen wie für  $\tau = 0.4$ .

Werden dagegen stärker zusammenhängende Daten betrachtet, so fällt die OC-Funktion relativ schnell.

Insgesamt zeigt die Simulationsstudie die Fähigkeit der Kontrollkarten, Daten anhand ihrer Zusammenhangsstruktur zu überwachen. Für den praktischen Gebrauch ist zu erwarten, dass hauptsächlich Daten mit offensichtlichem Zusammenhang überwacht werden, die also ein vergleichsweise großes Abhängigkeitsmaß  $\tau_0$  besitzen. Gerade für diese Fälle ist die vorgestellte Kontrollkarte geeignet.

### 6.3.3 Approximation der Grenzwerte am Beispiel der Claytoncopula

Die verwendeten Integrale des Abschnitts 6.2.2 liefern die exakten Ergebnisse für  $ucl(e)$  und  $lcl(e)$  in Abhängigkeit von der zugrundeliegenden Copula. Eine Betrachtung dieser Grenzwerte lässt vermuten, dass sie durch Funktionen angenähert werden können (siehe beispielsweise Abbildung 6.10). Während bei der Integration stückweise Grenzwerte für Intervalle von  $E$  berechnet werden, bietet eine gute Approximation den Vorteil stetiger Grenzen - damit ist im besten Fall eine genauere Einschätzung der untersuchten Stichprobe möglich. Ein weiterer Vorteil angepasster Grenzwerte wäre die Übertragbarkeit auf Zwischenwerte des Zusammenhangsmaßes  $\tau$ . So lägen in der praktischen Anwendung der Karte für beliebige  $\tau$  direkt die

entsprechenden Grenzwerte vor.

Es werden Funktionen für die Clayton-Grenzen konstruiert, eine analoge Übertragung des Vorgehens auf jede andere Copulastruktur ist problemlos möglich.

### Stückweise Anpassung mit Hilfe von Polynomen

Die Grenzwerte können nicht ausreichend gut durch nur eine stetige Funktion über den gesamten Wertebereich von  $e$  approximiert werden. Eine Aufteilung in drei Teilschnitte liefert die besten Resultate.

Zur Bestimmung der Abschnitte wird das Maximum der  $lcl$ - beziehungsweise  $ucl$ -Kurve betrachtet. In einem Intervall von  $\pm 0.05$  um dieses Maximum wird eine Parabel angepasst, rechts und links davon Polynome dritten Grades.

Es wird somit zunächst das  $e$  bestimmt, für das der Grenzwert  $cl = ucl$  oder  $cl = lcl$  sein Maximum erreicht, also

$$e_{\max} := \operatorname{argmax}_e cl(\tau).$$

Sowohl  $\max cl(\tau)$  als auch das dazugehörige  $e_{\max}$  hängen nach Konstruktion vom Zusammenhang der ursprünglich betrachteten Variablen ab. Die gesuchten Werte lassen sich mit Hilfe der Formeln

$$e_{\max} = 0.5326801 + 0.7549989\tau + 0.0011910\tau^2 \text{ für die } lcl$$

$$e_{\max} = 0.5161385 + 0.7474091\tau + 0.0014055\tau^2 \text{ für die } ucl$$

annähern. Die Güte der Annäherung beträgt für die  $lcl$ -Funktion etwa 0.94 und für die  $ucl$ -Funktion rund 0.92.

Anschließend werden die Grenzwerte  $cl$  in drei Abschnitten angepasst:

$$r = \begin{cases} \text{a) } \beta_{a,1}e + \beta_{a,2}e^2 + \beta_{a,3}e^3 & \text{für } e < \operatorname{argmax}_e cl(\tau) - 0.05 \\ \text{b) } \beta_{b,0} + \beta_{b,1}e + \beta_{b,2}e^2 & \text{sonst} \\ \text{c) } \beta_{c,0} + \beta_{c,1}e + \beta_{c,2}e^2 + \beta_{c,3}e^3 & \text{für } e > \operatorname{argmax}_e cl(\tau) + 0.05 \end{cases}$$

Tabelle 6.1 beinhaltet die Modellparameter für ausgewählte  $\tau$  für den Fall  $cl = lcl$ , Abbildung 6.13 zeigt die Approximation der Grenzen mit Hilfe der Funktionen a)

bis c) und den Parametern aus Tabelle 6.1. Die Anpassungsgüte ist dabei sehr hoch, sie liegt für alle betrachteten Fälle über 0.98.

$\tau$	$e_{\max}$	Abschnitt a)			Abschnitt b)			Abschnitt c)			
		$\beta_{a,1}$	$\beta_{a,2}$	$\beta_{a,3}$	$\beta_{b,0}$	$\beta_{b,1}$	$\beta_{b,2}$	$\beta_{c,0}$	$\beta_{c,1}$	$\beta_{c,2}$	$\beta_{c,3}$
0.1	0.61	0.25	-0.07	0.03	-0.99	3.14	-2.16	0.03	0.58	-0.71	0.20
0.2	0.68	0.22	-0.06	0.01	-0.59	1.93	-1.29	-0.19	1.08	-1.09	0.30
0.3	0.76	0.18	-0.03	-0.01	-0.28	0.99	-0.63	-0.38	1.48	-1.37	0.36
0.4	0.83	0.14	0.00	-0.03	-0.20	0.69	-0.41	-0.59	1.90	-1.65	0.43
0.5	0.91	0.10	0.03	-0.04	-0.20	0.61	-0.33	-1.04	2.87	-2.36	0.60
0.6	0.99	0.07	0.02	-0.02	-0.19	0.52	-0.27	-1.47	3.75	-2.96	0.74
0.7	1.06	0.05	0.01	-0.01	-0.25	0.57	-0.27	-3.19	7.59	-5.82	1.45
0.8	1.14	0.03	0.00	0.00	-0.19	0.39	-0.17	-4.64	10.57	-7.87	1.92
0.9	1.21	0.01	0.00	0.00	-0.16	0.29	-0.12	-16.08	35.46	-25.94	6.30

Tabelle 6.1: Parameterschätzer und  $e_{\max}$  für verschiedene Zusammenhangsmaße  $\tau$  bei Unterstellung einer Claytoncopula für die untere Kontrollgrenze  $lcl$

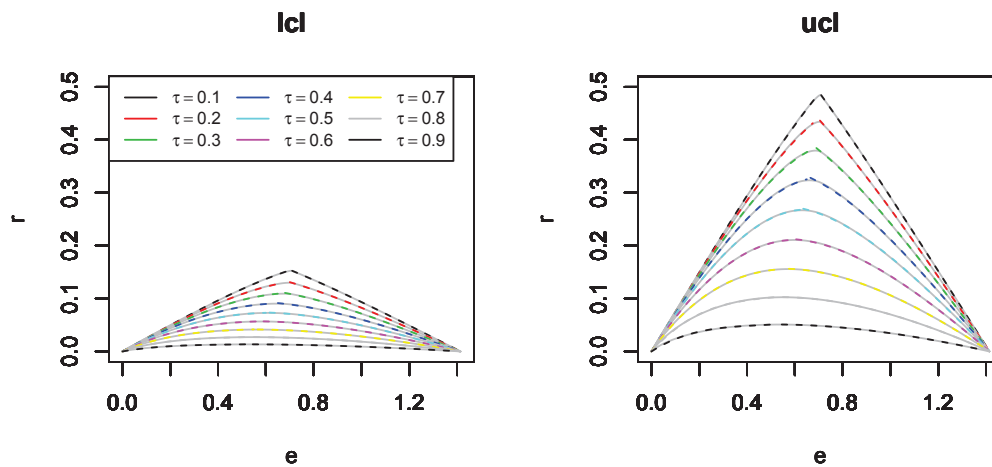


Abbildung 6.13: Approximierte Kontrollgrenzen (gestrichelt) verglichen mit den wahren Kontrollgrenzen (grau, durchgehend)

Die Parameterschätzer aus Tabelle 6.1 gelten für die aufgeführten  $\tau$ . In einem zweiten Schritt soll eine Verallgemeinerung der Formeln auf  $\tau$ -Zwischenwerte gefunden werden. Es zeigt sich, dass für die Parameterschätzer kein funktionaler Zusammenhang in Abhängigkeit von  $\tau$  existiert, der ausreichend gute Resultate liefert. Die kleinen Abweichungen in den  $\beta$ -Werten führen relativ schnell zu Abweichungen in der Anpassung. Dadurch werden die Funktionen als Grenzwerte zu ungenau. Für

gute Approximationen sollten demzufolge zunächst einige Grenzwerte für das interessierende  $\tau$  simuliert und dann die Modelle entsprechend der gezeigten Formeln angepasst werden.

### **Anpassung mit Hilfe von Splines**

Die Anpassung durch Polynome aus dem vorherigen Abschnitt hat den großen Nachteil, dass zwischen den stückweise angepassten Funktionen in der Regel Sprungstellen auftreten.

Eine Alternative bietet die Anpassung durch **Splines**. Ein kubischer Spline ist eine Funktion, die stückweise aus Polynomen maximal dritten Grades zusammengesetzt ist. Stellen, an denen zwei Polynomstücke zusammenstoßen („Knoten“), müssen bestimmte Bedingungen erfüllen, etwa dass der Spline zweimal stetig differenzierbar ist. Durch diese Forderung entsteht eine glatte Kurve ohne Sprungstellen.

Für die vorliegenden Daten zeigt sich, dass im Intervall  $[0, \sqrt{2}]$  mindestens 12 Knoten nötig sind, um die *lcl*- und *ucl*-Daten gut anzunähern. Pro Grenzfunktion werden also mindestens 11 Polynome betrachtet. Verglichen mit den 4 Knoten und 3 Polynomen des letzten Abschnittes ist dieser Wert zu hoch. Die große Diskrepanz in der Anzahl benötigter Teilstücke ist durch die Forderung der zweifachen stetigen Differenzierbarkeit erklärbar.

Die Verallgemeinerung der Funktionen auf Zwischenwerte von  $\tau$  beziehungsweise die Darstellung in Abhängigkeit von  $\tau$  gelingt hier nicht.

### **Fazit**

Die gefundenen Approximationen der Grenzwerte scheinen beide für die Praxis nicht geeignet. Die exakte Berechnung der Grenzwerte für festes  $\tau$  stellt einen überschaubaren Aufwand dar. Außerdem entfallen bei der Verwendung genauer Werte mögliche Rundungs Nachteile. Daher wird dieser Ansatz nicht weiter verfolgt und nachfolgend die Integraldarstellung der Grenzfunktionen verwendet.

## 6.4 Vergleich mit Steigers $Z$ -Test

Im Folgenden wird die Güte der vorgestellten Kontrollkarte mit Steigers  $Z$ -Test aus Abschnitt 4.6 verglichen. Dieser setzt annähernd bivariat normalverteilte Daten voraus.

Wie mehrfach erwähnt, ist der Schluss von univariat normalverteilten Daten auf eine bivariate Normalverteilung als gemeinsame Verteilung ein häufiger Fehler. Daher soll das Verhalten der Lauflänge für den Fall untersucht werden, dass eine auf Steigers  $Z$ -Test basierende Kontrollkarte irrtümlich auf Copuladaten angewandt wird.

Grafik 6.14 zeigt die Simulationsergebnisse für Datenpunkte einer Claytoncopula unter Nutzung verschiedener Abhängigkeitsstärken der Datenpunkte. Die Steiger-Grenzen wurden derart gewählt, dass die ARL unter Kontrolle das gleiche Niveau wie die neu konstruierte Karte erreicht. Die ARL unter Kontrolle ist im Mittel 36.63 ( $\alpha = 0.0273$ ). Wie im vorherigen Kapitel wurde  $n = 30$  und  $N = 20000$  gewählt.

Die Kontrollkarte nach Steiger erweist sich als deutlich schlechter als die neu entwickelte Karte. Schon für vergleichsweise kleine  $\tau$  unterschreitet sie das angestrebte Niveau massiv. Das Maximum der ARL stimmt nicht mit dem gewählten  $\tau_0$  überein, sondern verschiebt sich immer weiter nach rechts. Dieses Verhalten ist durch die Form der Claytoncopula erklärbar, deren Datengewicht unsymmetrisch um den Datenschwerpunkt verteilt ist, wohingegen normalverteilte Werte punktsymmetrisch um den Datenursprung streuen.

Tabelle 6.2 und Abbildung 6.14 verdeutlichen noch einmal, dass der Steigertest im Vergleich zur  $R$ -Karte versagt. Verglichen wurden nur mittlere Abhängigkeiten zwischen den betrachteten Merkmalen ( $\tau = 0.3$  bis  $\tau = 0.7$ ). Dadurch wird der Effekt besonders deutlich.

$\tau =$	0.3	0.4	0.5	0.6	0.7
ARL $R$ -Karte	36.84	36.84	36.83	36.54	36.60
ARL Steigers $Z$ -Test	23.67	21.86	18.91	13.76	8.13

Tabelle 6.2: ARL-Simulationsergebnisse bei Unterstellung einer Claytoncopulastruktur für die  $R$ -Karte im Vergleich zu Steiger's  $Z$ -Karte für verschiedene  $\tau$

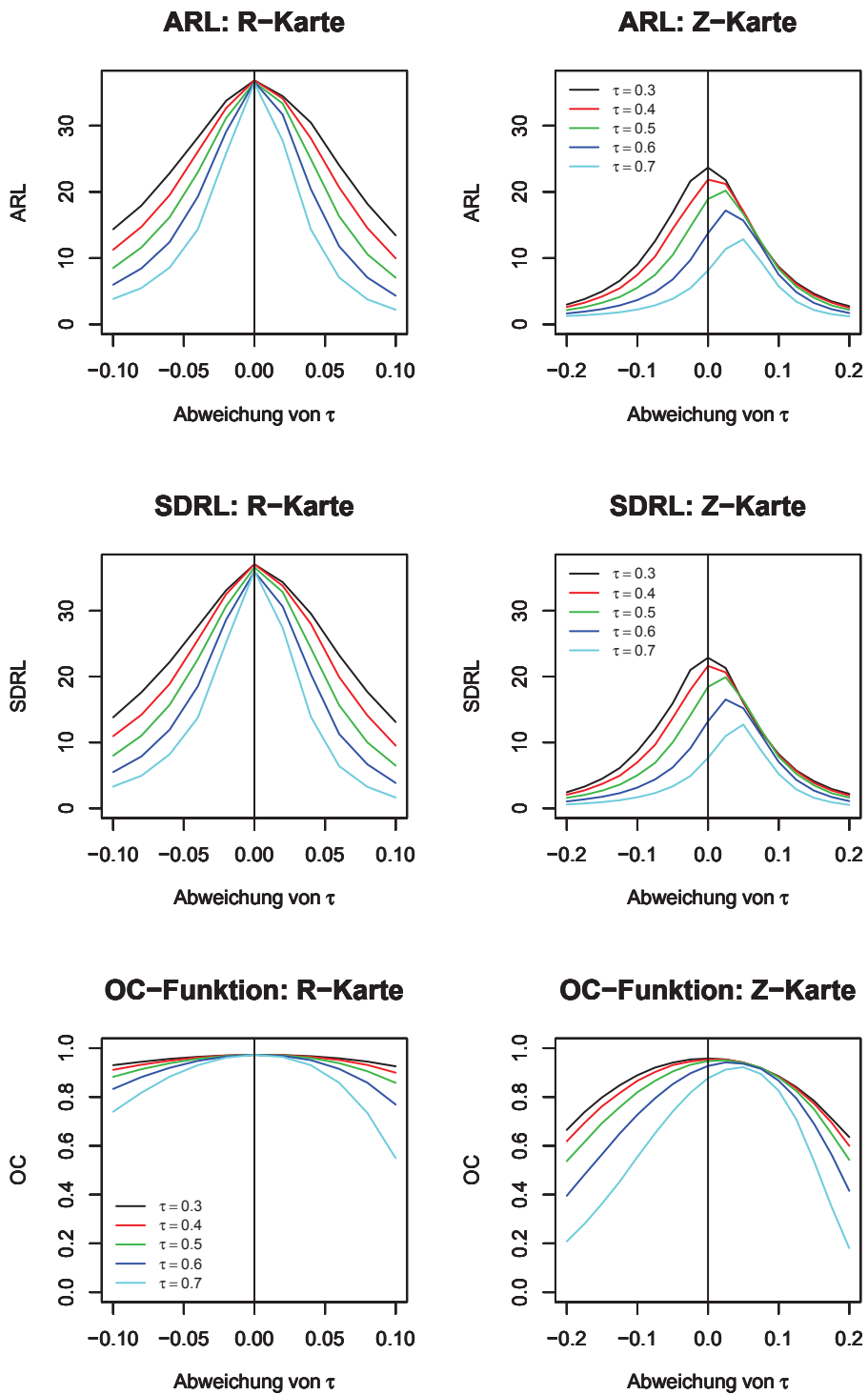


Abbildung 6.14: Simulationsergebnisse zu ARL, SDRL und OC-Kurve bei Unterstellung verschiedener Claytoncopulas zum Vergleich Steiger's Z-Karte und der R-Karte



## 6.5 Anwendungen

### 6.5.1 Beispiel I: Aktienrenditen

Ein Beispiel zeigt die Anwendung der Karten aus Kapitel 6 auf Finanzmarktdaten. Renditedaten werden miteinander verglichen und in Beziehung zum DAX gesetzt. Die Theorie des „Diversification Meltdowns“ („Zusammenbruch der Risikostreuung“) besagt, dass sich die Abhängigkeit von Aktienkursen in Zeiten wirtschaftlicher Abschwünge verstärkt (siehe etwa Campbell et al. (2008)). Eine Veränderung des gemeinsamen Verhaltens kann daher ein Anzeichen für eine bevorstehende Unruhe im Finanzmarkt sein, so dass es sich anbietet, die gemeinsame Abhängigkeitsstruktur verschiedener „Indikatoraktien“ zu überwachen.

Nachfolgend werden Finanzmarktdaten für den Zeitraum Januar 2003 bis Februar 2010 ausgewertet; die Entwicklung des Zusammenhangs zwischen Renditen zweier geeigneter Aktien wird vor dem Hintergrund der weltweiten Banken- und Finanzmarktkrise betrachtet, die 2007 begann und bis in die Gegenwart reicht. Auslöser der Unruhe war die Subprime-Krise in den USA. Viele Staaten versuchten durch spezielle Rettungspakete entgegenzusteuern. Die Betrachtung in diesem Beispiel erfolgt auf Grundlage der DAX-Firmen BASF und BAYER und wird daher vor allem mit der deutschen Entwicklung verglichen. Beide Firmen sind weltweit agierende Chemiekonzerne. Ein ähnliches Verhalten beider Aktienkurse kann also angenommen werden. Der Bereich Chemie ist zudem ein guter Indikator für gesamtwirtschaftliche Tendenzen, da es sich um typische Zuliefererbranchen handelt.

Interessierendes Merkmal ist in diesem Fall die (zeitdiskret) Rendite einer Aktie (tageweise gemessen). Betrachtet werden

$X_{i1}$  = Rendite von BASF am  $i$ -ten Tag und

$X_{i2}$  = Rendite von BAYER am Tag  $i$ .

Das Streudiagramm der Renditen beider Aktien (Abbildung 6.15, links) zeigt einen deutlichen Zusammenhang; Abbildung 6.15 zeigt außerdem die Höhenlinien von zwei möglichen Verteilungen. Die Claytoncopula passt die Daten besser an. Der ML-Schätzer für den Parameter der Claytoncopula ist  $\theta = 1.746$  (ML-Wert von 285.3)

und für die Normalcopula gilt  $\rho = 0.629$  (ML-Wert von 267.5). Zur Modellierung der gemeinsamen Verteilung der BASF- und BAYER-Renditen wird im Folgenden somit eine Claytoncopula mit  $\theta = 1.746$  verwendet.

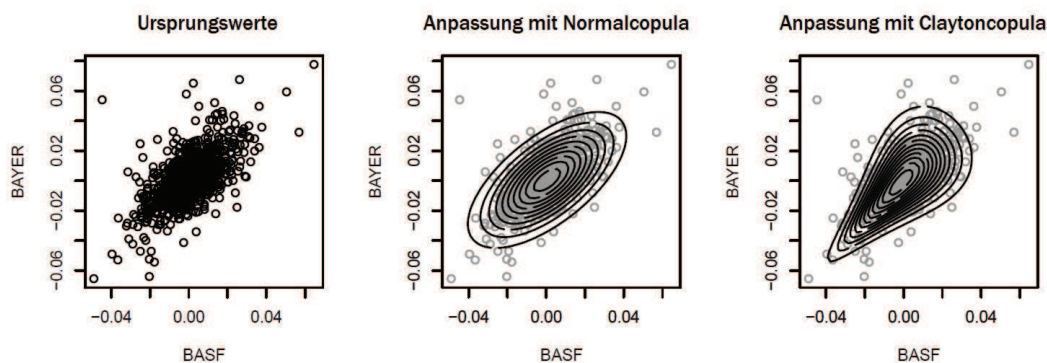


Abbildung 6.15: Ursprungsdaten und Konturlinien der angepassten gemeinsamen Verteilung

Zur Erstellung der Karte werden statt der Renditenpaare  $(X_{i1}, X_{i2})$  die daraus gebildeten Wertepaare  $r_i$  (Formel (6.1)) und  $e_i$  (Formel (6.2)) verwendet. Für das vorliegende  $\hat{\tau} = 0.47$  sind die Grenzwerte  $lcl$  und  $ucl$  für  $\tau=0.5$  aus Abbildung 6.10 geeignet, um die  $R_i$  anhand von  $(r_i, e_i)$  zu bestimmen.

Die  $R$ -Werte, also die nach Gleichung (6.4) kumulierten  $R_i$ , bilden schließlich die Entscheidungsgrundlage, ob zu einem Zeitpunkt die Kurse der betrachteten Aktien „zu starke“ oder „zu schwache“ Abhängigkeit voneinander zeigen. Die  $R$ -Statistik basiert auf den letzten 30 Werten. Die Reaktionszeit der Karte hat daher eine Verzögerung von bis zu 30 Tagen, die berücksichtigt werden muss. Mit der unteren Grenze  $LCL=22$  und der oberen Grenze  $UCL=38$  ergibt sich für die Karte das Niveau  $\alpha \approx 0.02734$ .

Abbildung 6.16 visualisiert die Ergebnisse der beschriebenen Karte, die grauen vertikalen Linien markieren Werte, für die der Prozess als außer Kontrolle klassifiziert wird.

Von den 788 untersuchten Zeitfenstern liegen 68 außerhalb der Grenzwerte, 34 davon über der oberen Grenze. Diese Abweichung nach oben spricht nach der Theorie des „Diversication Meltdowns“ für eine Anspannung des Finanzmarktes zu diesem Zeitpunkt. Nach unten abweichende Werte würden demzufolge für eine un-

gewöhnliche Entspannung an den Börsen sprechen und sind für unsere Überlegungen zunächst uninteressant. Grob können diese Daten in zwei Zeitabschnitte eingeteilt werden: Anfang 2008 (hauptsächlich Mitte Februar) und Anfang 2009 (konzentriert auf den Zeitraum von Ende Februar bis Mitte April). Um diese Spitzen zu erklären, werden die genannten Zeiträume mit dem Verlauf der Finanzmarktkrise in Verbindung gebracht.

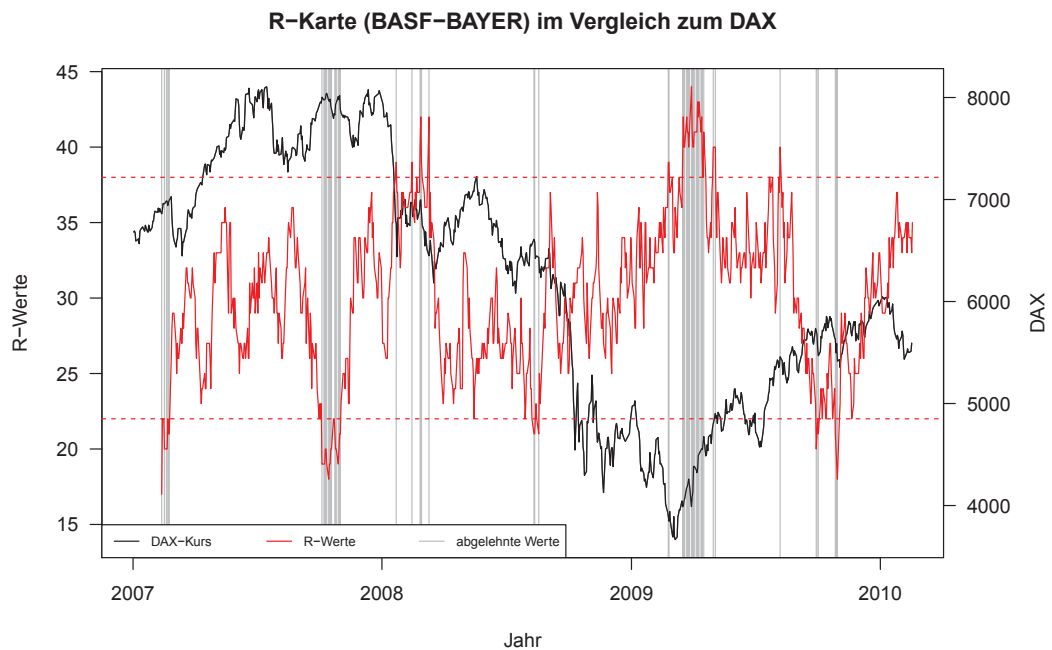


Abbildung 6.16: *R*-Karte für die gegebenen BASF- und BAYER-Renditen im Vergleich zum DAX-Kurs mit markierten Außer-Kontrolle-Punkten)

Die erste Zeitspanne lässt sich mit einer „Achterbahnfahrt“ der Aktienpreise im Januar 2008 erklären, die durch die Angst der Anleger vor einer US-Rezession und den daraus resultierenden Folgen für den Weltmarkt entstand. Am 25.01.2008 verabschiedet der US-Kongress ein 150-Milliarden-Dollar-Konjunkturprogramm, um eine Rezession zu verhindern. In Deutschland kündigte die Regierung am 13.02.2008 staatliche Unterstützung für die IKB Deutsche Industriebank an, was als Stoppsignal des Diversication Meltdowns interpretiert werden kann.

Im Zeitbereich der zweiten Phase auffälliger Werte erreichte die Finanzkrise ih-

ren Höhepunkt: Die Weltwirtschaft rutscht immer stärker in die Krise und auch in Deutschland häufen sich Negativmeldungen. Im Februar 2009 wurde die Hypo Real Estate von der deutschen Regierung mit 102 Milliarden Euro gestützt, was zu Forderungen führt, die Bank zu verstaatlichen. Im März ist die Frage bestimmend, ob der deutsche Staat die General-Motors-Tochter Opel retten sollte. Rettungspläne der Konzernspitze beinhalten meist massive Stellenstreichungen und stoßen auf Kritik. Als Wendepunkt kann der 31. März angenommen werden: Hier verspricht Bundeskanzlerin Dr. Angela Merkel staatliche Hilfe bei der Investorensuche für Opel und zeigt sich zu möglichen Bürgschaften bereit. Am 7. April beschließt die Bundesregierung außerdem eine massive finanzielle Aufstockung der sogenannten „Abwrackprämie“. Eine Entspannung des Marktes folgt und im Mai gibt es wieder gute Nachrichten: Der Staat besitzt nun 47.31% der Hypo Real Estate und Opel gilt aufgrund der angekündigten Übernahme durch den Zulieferer Magna als vorerst gerettet.

### **6.5.2 Beispiel II: Schrägseilbrücken**

Geraten Bauwerke oder technische Einrichtungen in Schwingungen, ist das Resultat unter Umständen verheerend. Durch gleichmäßig wiederkehrende Anregung von außen „schaukelt“ sich die Bewegung auf und es kommt im schlimmsten Fall zur Überlastung des Materials. Beispiele dafür sind der Einsturz der Broughton Suspension Kettenbrücke (England) von 1831. Dieser Einsturz wurde durch marschierende Soldaten verursacht, die mit ihrer gleichförmigen Trittfrequenz die Eigenschwingung der Brücke „aufpeitschten“. Detailliertere Informationen dazu liefern Taylor und Phillips (1881).

Um diese sogenannten Resonanzkatastrophen zu verhindern, gibt es verschiedene Ansätze. Maßnahmen zur günstigen Beeinflussung des Verhaltens erklären Dreßig und Holzweißig (2011) für verschiedene dynamische Effekte. Ein Ansatz sind Schwingungsdämpfer; das weltgrößte Tilgerpendel (im Taipei Financial Center) erstreckt sich beispielsweise über mehrere Stockwerke. Ferner wird versucht, so zu bauen, dass die Eigenfrequenz des Systems, also das Schwingungsverhalten nach einmaliger Anregung, nicht von der erwarteten äußeren Anregung verstärkt wird. So werden in Erdbebengebieten die lokal typischen Schwingungsfrequenzen der Erderschütterungen bei Bau- und Trägerkonstruktionen berücksichtigt. Im Flugzeug-

bau, bei Eisenbahntunneln, Autobahnen oder Unterführungen treten ebenfalls hohe Schwingungsbelastungen auf, die bedacht werden müssen.

Auch die Schwingungsüberwachung von Maschinen ist sinnvoll: Beginnende Schäden sind oft frühzeitig durch Veränderungen im Schwingungsspektrum feststellbar.

In diesem Kapitel wird eine Schrägseilbrücke betrachtet. Abbildung 6.17 zeigt schematisch den Aufbau einiger Schrägseilbrücken. Die Grafik wurde von Mehlhorn (2007) veröffentlicht. Er liefert in der angegebenen Quelle außerdem weiterführende physikalische Erläuterungen zu verschiedenen Brückenarten. Schrägseilbrücken sind

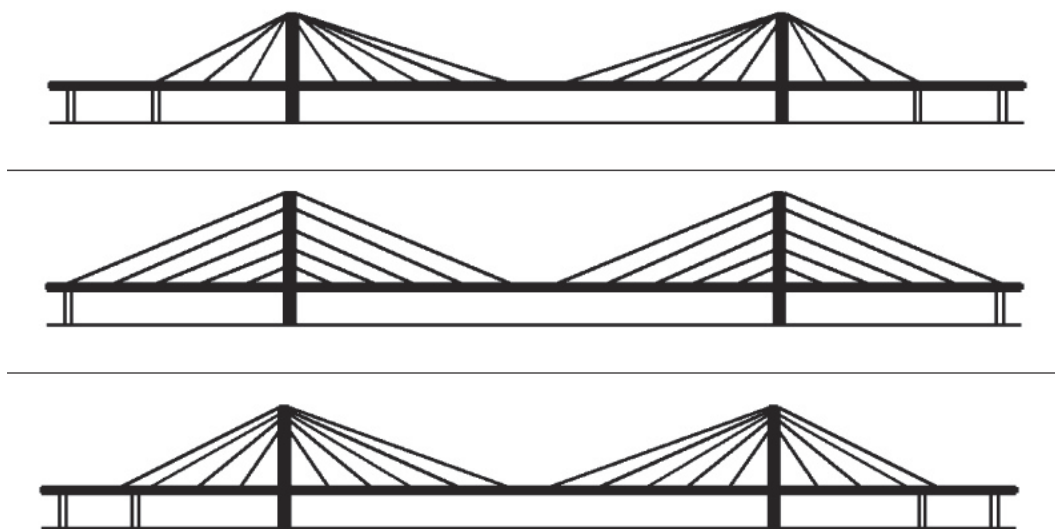


Abbildung 6.17: Skizze verschiedener Schrägseilbrücken: Bündelsystem (oben), Harfensystem (Mitte) und Fächersystem (unten), ursprünglich Abbildung 5.5.1 aus Mehlhorn (2007)

dadurch gekennzeichnet, dass der Überbau an Seilen aufgehängt ist, die schräg von einem Pfeiler gespannt sind. Beispiele für diese Brückenkonstruktion sind die momentan im Bau befindliche Russki-Brücke (Vladivostok, 3100 Meter Länge), die Fleher Brücke (Düsseldorf, 1166 Meter), die Rheinbrücke Neuwied - Raiffeisenbrücke (485 Meter) oder die Severinbrücke (Köln, 691 Meter).

Schrägseilbrücken sind empfindlich gegen Windkräfte und können durch große bewegliche Massen - etwa Güterzüge oder LKW-Korsos - in Bewegung gebracht werden. Dabei ist es insbesondere gefährlich, wenn die einzelnen Fahrbahn Teile in eine gleichgerichtete Bewegung geraten oder genau entgegengesetzt schwingen.

Sensoren erfassen die Schwingamplitude in der Mitte des Tragseils, also dessen maximale Auslenkung. Um Materialschäden oder schlimmstenfalls einem Versagen der Tragfähigkeit vorzubeugen, werden Brücken für den Verkehr gesperrt, sobald die Umgebungsbedingungen zu schlecht werden. Kriterien können hohe Windstärken, Hagel oder durch Eisbildung veränderte Tragseileigenschaften sein. Die Belastbarkeitsgrenze ist dabei brückenspezifisch verschieden.

Ferngesteuerte Seildämpfungssysteme mit aktiv beeinflussbarem Widerstand können dazu beitragen, Sperrungen zu vermeiden. Diese werden bisher aber nur vereinzelt eingesetzt. Ein Regelalgorithmus reagiert dabei auf aktuelle Seilschwingungen und vergrößert die Dämpfungskraft für stärker schwingende Seile. Bei einem Stromausfall arbeiten die Systeme mit einer Grunddämpfungskraft; daher ist das System relativ sicher. Ein aktuelles Forschungsprojekt dazu betreibt die Firma EMPA - Material Science & Technology, ein Forschungs- und Dienstleistungsinstitut für Materialwissenschaften und Technologieentwicklung (siehe EMPA (2012)).

Nachfolgend werden zwei Tragseile betrachtet. Gefährlich sind eine starke Ähnlichkeit oder ein stark entgegengesetztes Verhalten im Schwingungsverhalten beider Seile. Solche Schwingungen sind ein Indikator für Resonanzverstärkung. In diesem Fall sollte Alarm ausgelöst werden. Nachfolgend wird eine zweispurige Schrägseilautobahnbrücke von etwa 500 Metern Länge mit einem Stützpfeiler betrachtet. Im Idealzustand ist das Seil gespannt und wird nur durch normale Verkehrsbelastung bewegt.

Zur Entwicklung einer geeigneten Qualitätsregelkarte muss die Schwingbewegung der Seile erfasst werden. Die dazugehörige Formel gibt die Auslenkung  $y(t)$  zum Zeitpunkt  $t$  an. Eine ungedämpfte Schwingung lässt sich beschreiben durch

$$y(t) = y_0 \cdot \sin(2\pi ft + \varphi_0),$$

wobei  $y_0$  den Wert der Amplitude,  $t$  die Zeit,  $f = 1/T$  die Frequenz und  $\varphi_0$  die

Anfangsphase der Schwingung darstellt. Abbildung 6.18 zeigt die Kenngrößen zur Charakterisierung einer Schwingbewegung, vergleiche auch Dresig und Holzweißig (2011).

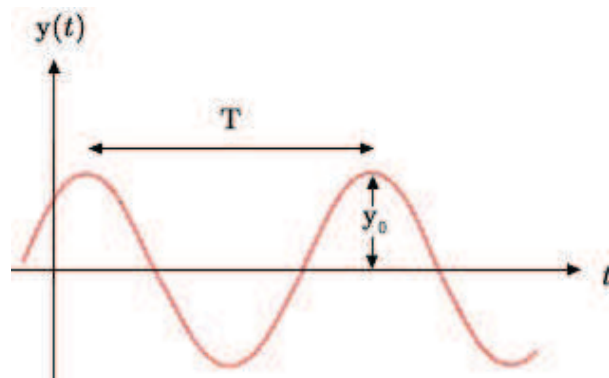


Abbildung 6.18: Schaubild der verschiedenen Schwingungsvariablen

Physikalische Systeme sind dagegen fast immer gedämpft. Im Falle von Trageil-schwingung beispielsweise durch Luftreibung, die genannte Schwingungsformel wird dann entsprechend erweitert, um das Abklingen der Funktion darzustellen. Diese lautet dann

$$y(t) = y_0 \cdot \delta e^{-\frac{t}{\beta}} \cdot \sin(2\pi ft + \varphi_0).$$

Die Abklingkonstante  $\delta$  und die Abklingdauer  $\beta$  beschreiben dabei die Form des Abklingens der Schwingung.

Im normalen Verkehrsbetrieb überlagern sich viele Reize an den Trageilen. Etwas Bewegung ist immer zu verzeichnen, eine totale Ruheposition der Trageile ist in der Praxis nicht zu erwarten. Die Messwerte der absoluten Amplituden sind normalverteilt für jedes einzelne Seil.

Durch die Konstruktion der Brücke ist eine Unabhängigkeit in dem Verhalten der Trägerseile nicht realistisch. Das gemeinsam betrachtete Schwingverhalten von zwei und mehreren Seilen unterliegt einer bestimmten Abhängigkeitsstruktur. Für kleinere Stöße gibt es kaum „Interaktion“ zwischen den Seilen, sie wirken lokal sehr begrenzt. Die Trageile verhalten sich dadurch im Ruhezustand der Brücke relativ unabhängig zueinander. Stärkere Anreize bewirken höhere Schwingamplituden bei den

einzelnen Seilen, ihre Wirkung ist außerdem lokal weiter ausgedehnt bis hin zu globalen Tendenzen, die die gesamte Brückenkonstruktion betreffen. Die Abhängigkeit der Schwingungen untereinander wächst also für stärkere Auslenkungen.

Anhand dieser theoretischen Überlegungen lässt sich die gemeinsame Brückenschwingung zweier Tragseile modellieren.

Methoden der Zeitreihenanalyse werden dafür bewusst nicht verwendet. Beispielsweise wäre eine (mathematisch sinnvolle) Saisonbereinigung kontraproduktiv, da diese Kräfte effektiv an den Seilen auftreten und daher zu berücksichtigen sind. Ein auftretender Trend sollte auch nicht eliminiert werden. Im Optimalfall sollte kein Trend auftreten. Werden trotz eingespannter Tragseile die Ausschläge tendenziell größer, so kann das beispielsweise auf Materialermüdung in den Seilen oder Aufhängungen hinweisen. Am sinnvollsten erscheint daher eine gemeinsame Darstellung ohne vorherige Komponentenzerlegung des Schwingverhaltens.

Die Amplituden beider Seile werden in Abhängigkeit von der Maximalauslenkung betrachtet. Je größer die Übereinstimmung beider Werte wird, umso eher besteht auch die Gefahr einer gleichförmigen Schaukelbewegung der Brücke.

Möglichkeiten zur Modellierung extremer Wellenhöhen und -perioden durch Copulas erläutert de Waal (2005). Sriramula et al. (2005) und Srinivas et al. (2006) stellen die Simulation und Abhängigkeitsmodellierung von Achsenbelastungen an Fahrzeugen dar, unter anderem werden auch dort Copulas genutzt. Diese Modelle sind in ähnlicher Form auf Autobahnbrücken übertragbar (siehe etwa Srinivas (2006)). Caprani et al. (2012) beschreiben das dynamische Verhältnis von statischer zur Gesamtbelastung auf Brücken mit Hilfe von (Gumbel-)Copulas.

Die gemeinsame Verteilung der Amplituden zweier Tragseile in Abhängigkeit von der Maximalauslenkung  $(y_1(t), y_2(t))^T$  kann dementsprechend durch eine Gumbelcopula mit dem Parameter  $\theta$  und zwei normalverteilten Rändern modelliert werden. Stellt man den Grad der Abhängigkeit in Beziehung zu Kendalls  $\tau$ , so fällt dieses unter Ruhebelastung eher klein aus.

Inhaltlich ist die Wahl der Verteilung durch die Entstehung der Schwingbewegung begründet. Zugrunde gelegt wird eine gedämpfte Schwingung, die sich aus jedem „Schock“, also jeder plötzlichen Krafteinwirkung, auf das Tragseil ergibt (siehe



Abbildung 6.19, links).

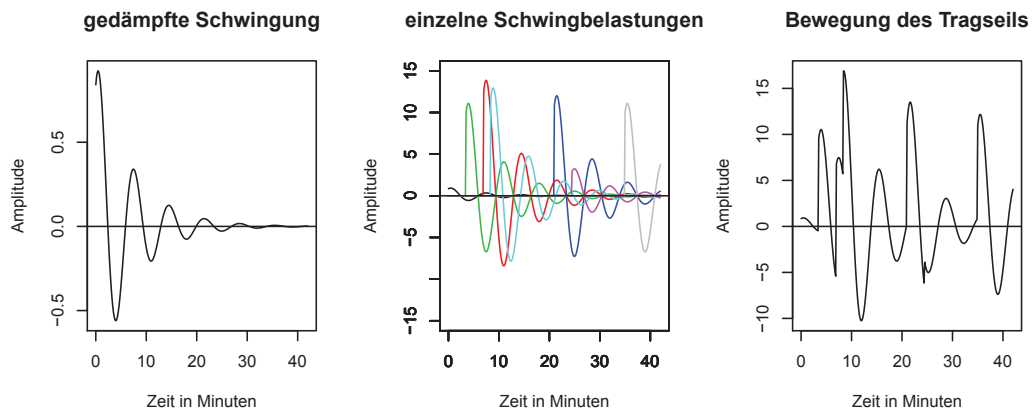


Abbildung 6.19: Darstellung gedämpfter Schwingungen sowie von Trageislschwingungen als Summe gedämpfter Schwingungen

Diese Schocks treten - mehr oder weniger - zufällig verteilt über die Zeit auf, beispielsweise bei jedem vorbeifahrenden Auto. Die Summe aller Einzelamplituden ergibt dann die Seilbewegung unter geringer Belastung, von einem Zusammenspiel mit anderen Seilen ist hier zunächst nicht auszugehen. Größere einwirkende Kräfte haben weitreichendere Auswirkungen. In der Praxis ist außerdem die Aufnahme eines Störterms in das Modell sinnvoll. Dieser ist in Abbildung 6.19 noch nicht berücksichtigt.

Für Abbildung 6.20 wurden zwei Datenreihen parallel zur dargestellten Schwingung in Abbildung 6.19 (rechts) konstruiert. Diese dienen als exemplarische Grundschwingung zweier Trageisile. Die Schwingungen wurden jeweils mit einem standardnormalverteiltem Fehlerterm verrauscht.

In der 17.5ten und der 28ten Minute wurden globale Schocks simuliert, die auf beide Trageisile gleichzeitig einwirken. Dabei handelt es sich wiederum um abklingende Schwingungen, die auf die Grundschwingung summiert wurden. Zur Vereinfachung werden die Schocks als „mittig auftretend“ modelliert; sie wirken also bei beiden Trageisilen gleichzeitig mit gleicher Kraft.

Das rechte Bild der Grafik zeigt die gegeneinander im Streudiagramm abgetragenen absoluten Amplituden. Dabei wird noch einmal deutlich, dass die Daten nicht

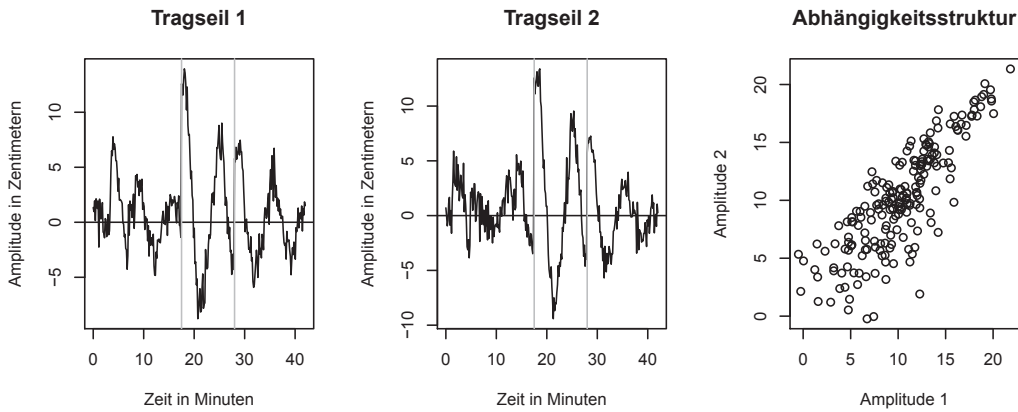


Abbildung 6.20: Vergleich zweier gekoppelter Belastungsdiagramme

normalverteilt sind. Die abgebildeten Daten haben ML-Werte von 142.9 bei Anpassung einer Claytoncopula, 146.0 bei der Wahl einer bivariaten Normalverteilung und 164.4 bei Unterstellung einer Gumbelcopula. Testläufe mit analog konstruierten Schwingungen lieferten vergleichbare Resultate. Die Gumbelcopula ist also ein geeignetes mathematisches Instrument zur Modellierung der Abhängigkeit beider Amplitudenreihen.

Die entwickelte Qualitätsregelkarte ist auf diesen Sachverhalt anwendbar. Sie erkennt, ob die Brücke in gleichförmige Schwingungen gerät. Untersucht wird für verschiedene Zeitintervalle die These

$H_0$ : Das gemeinsame Schwingverhalten der Tragseile folgt einer bestimmten festen Copulastruktur. Diese Struktur ergibt sich entweder aus Messungen in der Vergangenheit oder aus baustatischen Überlegungen.

$H_1$ : Die Abhängigkeitsstruktur entspricht nicht der unter  $H_0$  beschriebenen.

Im Folgenden werden Punkte der gemeinsamen Verteilung der Amplituden zweier Tragseile in Abhängigkeit von der Maximalauslenkung  $(y_1(t), y_2(t))^T$  simuliert. Als Basis der Simulation dient eine Gumbelcopula. Der Copulaparameter wird entsprechend der zu modellierenden Abhängigkeit beider Seile (dargestellt durch Kendalls  $\tau$ ) gewählt.

Abbildung 6.21 zeigt die  $R$ -Karte für vier verschiedene Szenarien. Oben links

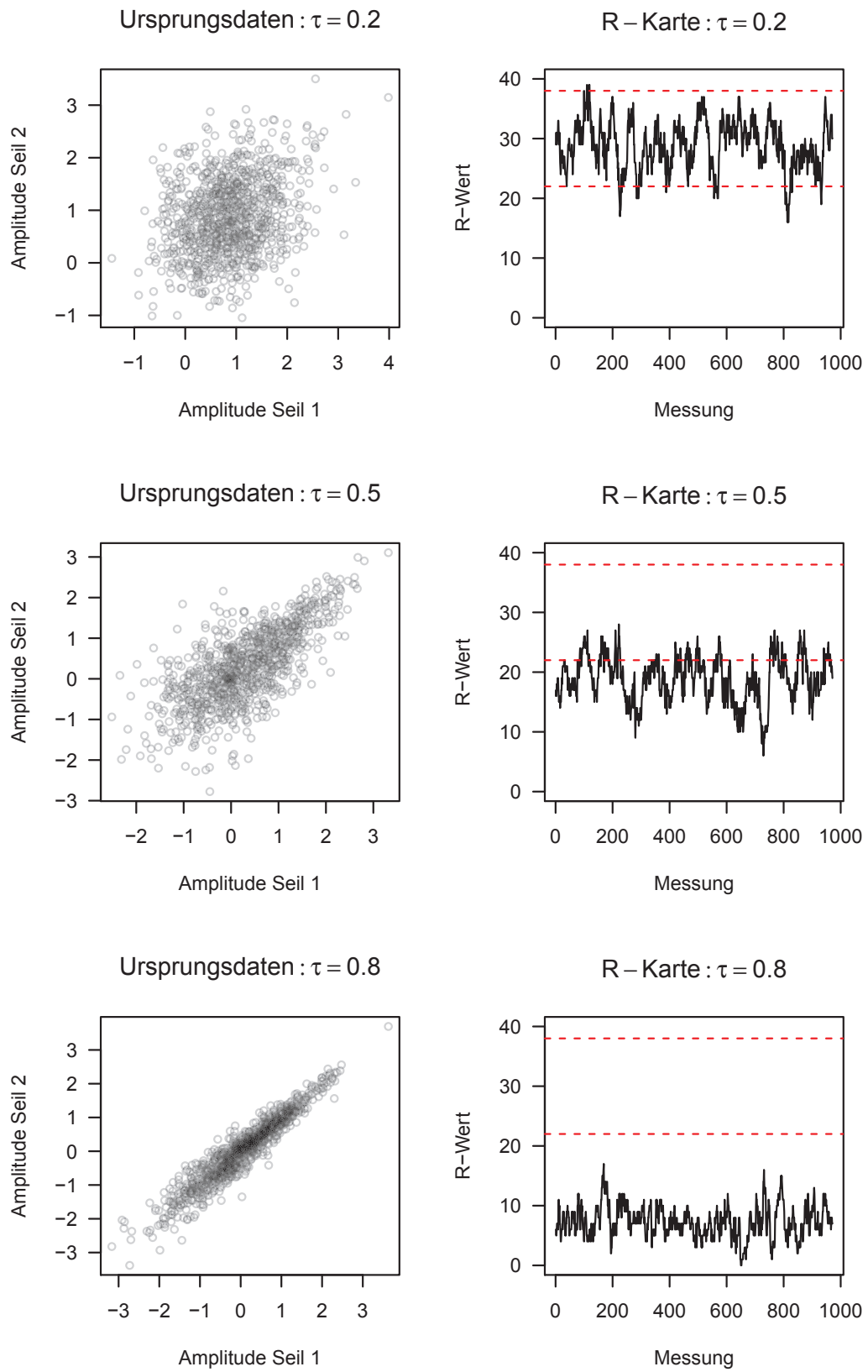


Abbildung 6.21: Anwendung der  $R$ -Karte auf verschiedene Szenarien unter Nutzung der Gumbelcopula und  $\tau_0 = 0.2$

ist der In-Kontrolle-Fall ( $\tau = 0.2$ ) abgebildet. In der Bildkomposition ist links ein Streudiagramm des Datensatzes dargestellt; rechts ist die zugehörige Teststatistik in der Qualitätsregelkarte abgetragen, die Grenzwerte sind gekennzeichnet. Die anderen drei Grafiken zeigen die Auswirkung einer erhöhten Gleichschwingung der Tragseile auf die Daten und die Qualitätsregelkarte. Die Bilder sind analog zum In-Kontrolle-Fall aufgebaut. Es wird deutlich, dass eine stark gleichmäßig gerichtete Schwingung gut erkannt wird. Von 971 betrachteten (zufälligen) Testwerten liegen bei diesem Beispiel für  $\tau = 0.2$  42 außerhalb der Grenzen, für  $\tau = 0.4$  sind es 374, für  $\tau = 0.6$  liegen 831 außerhalb und für  $\tau = 0.8$  alle Datenpunkte. Auf weitergehende Simulationen wird verzichtet; die Lauflängenversuche aus Abschnitt 6.3.2 sind hier übertragbar.

## 6.6 Optimierung der Teststatistik

Die bisherigen Kapitel zeigen, dass die gefundene Kontrollkarte geeignet ist, Abhängigkeitsänderungen in den Daten nachzuweisen. Dieses Kapitel untersucht, inwieweit eine Änderung der Grenzwerte  $lcl$  und  $ucl$  und eine entsprechende Anpassung von LCL und UCL zu besseren Resultaten führt.

Ein verbesserter Test könnte beispielsweise das gegebene Testniveau besser ausschöpfen oder die ARL im Außer-Kontrolle-Fall schneller absinken lassen als der Test aus Abschnitt 6.1. Dort wurde als  $lcl$  das 25%-Quantil der Verteilung von  $r$  für ein gegebenes  $e$  gewählt. Für die  $ucl$  wurde analog das 75%-Quantil verwendet. So wird das „mittlere Datenfeld“ mit  $R_i = 1$  gleichmäßig erfasst. Darüber ( $R_i = 2$ ) und darunter ( $R_i = 0$ ) liegende Datenpunkte treten im kontrollierten Prozess zu etwa gleichen Teilen auf. Damit gilt im Erwartungswert  $R_i = 1$  für jeden Punkt der Datenmenge. Die Grenzwerte LCL und UCL können einfach als die entsprechenden Quantile einer Binomialverteilung bestimmt werden.

Um zu erkennen, wie sich eine Verschiebung der Grenzwerte  $lcl$  und  $ucl$  auf die Güte des Tests auswirkt, wird zunächst die Verteilung der Summe der Zufallsvariablen  $R_i$  (Formel (6.3)) neu bestimmt. Dementsprechend sind die Grenzwerte LCL und UCL zu ändern. Auf dieser Basis werden dann die Optimalitätskriterien „ARL“ und „Niveau“ betrachtet.

### 6.6.1 Verteilung der Grenzwerte $lcl$ und $ucl$

Die Gleichung

$$P(r_i \leq lcl(e_i)) = P(r_i > ucl(e_i)) = 0.25$$

aus Kapitel 6.1.2 ist jetzt nicht mehr erfüllt, stattdessen seien die einzelnen  $R_i$  unter  $H_0$  unabhängig und identisch verteilt mit

- $P(R_i = 0) = p_1$ ,
- $P(R_i = 1) = 1 - p_1 - p_2$  und
- $P(R_i = 2) = p_2$ .

Die Wahrscheinlichkeit, dass die Summe der  $R_i$  einen bestimmten Wert  $x$  annimmt, lässt sich im Gegensatz zu Abschnitt 6.1.2, Formel (6.4) nicht durch eine bekannte Verteilung ausdrücken. Die einzelnen Wahrscheinlichkeiten für Ausprägungen der Summe über die  $R_i$  sind jedoch geschlossen darstellbar durch

$$P\left(\sum_{i=1}^n R_i = x\right) =$$

$$\sum_{j=0}^{\lfloor x/2 \rfloor} n! \frac{1}{(n-x+j)!(x-2j)!j!} p_1^{(n-x+j)} (1-p_1-p_2)^{(x-2j)} p_2^j. \quad (6.12)$$

Der Term  $\lfloor x/2 \rfloor$  (Gaußklammerausdruck) steht dabei für den auf ganze Zahlen abgerundeten Wert von  $\frac{x}{2}$ .

Die angegebene Wahrscheinlichkeit ergibt sich durch folgende Überlegung: Es ist

- $(n-x+j)$  die Anzahl der  $R_i$  mit  $R_i = 0$ ,
- $(x-2j)$  die Zahl der  $R_i$  mit der Ausprägung 1 und
- $j$  die Anzahl der  $R_i$  mit der Ausprägung 2.

Insgesamt werden  $n$   $R_i$  mit der gemeinsamen Summe  $x$  betrachtet. Dabei ist

$$x = 0 \cdot (n-x+j) + 1 \cdot (x-2j) + 2 \cdot j = (x-2j) + 2j.$$

Die Wahrscheinlichkeit, diese Summe mit  $(n - x + j)$  Nullen,  $(x - 2j)$  Einsen und  $j$  Zweien zu erreichen, setzt sich zusammen aus der Anzahl möglicher Anordnungen

$$n! \cdot [(n - x + j)!(x - 2j)!j!]^{-1}$$

multipliziert mit den Wahrscheinlichkeiten für die einzelnen Ergebnisse.

Für  $p_1 = p_2 = 0.25$  entsprechen die Resultate denen einer Binomialverteilung (siehe Formel 6.4), wie im Anhang A.2.3 exemplarisch für  $x = 0$ ,  $x = 1$  und  $x = 2$  gezeigt wird.

Zur Berechnung der Verteilung der  $R_n$  wird Formel 6.12 für alle in Frage kommenden  $x$  bestimmt. Die summierten Wahrscheinlichkeiten entsprechen den Werten der Verteilungsfunktion. Da es sich um eine diskrete Verteilung handelt, wird das gewünschte Niveau durch die gewählten Quantile in der Regel nicht exakt getroffen.

Einige Wahrscheinlichkeitswerte  $p_1$  und  $p_2$  zeigt Tabelle 6.3.

$p_1 \backslash p_2$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	25 35	27 39	30 42	32 46	35 49	38 52	40 55	44 57	46 60
$\alpha$	0.024	0.026	0.046	0.031	0.037	0.039	0.025	0.042	0.008
0.2	21 33	23 37	25 40	28 44	30 47	33 50	36 53	38 56	
$\alpha$	0.027	0.030	0.038	0.037	0.035	0.039	0.040	0.020	
0.3	18 30	20 35	22 38	24 42	27 45	29 48	32 52		
$\alpha$	0.047	0.038	0.044	0.036	0.046	0.042	0.026		
0.4	14 28	16 32	18 36	20 40	23 43	26 46			
$\alpha$	0.032	0.037	0.036	0.031	0.041	0.038			
0.5	11 25	13 30	15 33	17 37	20 40				
$\alpha$	0.038	0.035	0.046	0.041	0.043				
0.6	8 22	10 27	12 31	14 34					
$\alpha$	0.040	0.039	0.042	0.038					
0.7	5 20	7 24	8 28						
$\alpha$	0.026	0.040	0.026						
0.8	3 16	4 22							
$\alpha$	0.043	0.020							
0.9	0 14								
$\alpha$	-								

Tabelle 6.3: Kontrollgrenzen UCL und LCL bei verschiedenen Wahrscheinlichkeiten  $p_1$  und  $p_2$  mit Angabe des zugehörigen Niveaus  $\alpha$

Als Rahmenbedingungen wurden  $n = 30$  sowie  $\alpha \leq 0.05$  gewählt. Das jeweils

tatsächlich erreichte Niveau  $\alpha$  wird mit angegeben.

Für  $p_1 = (1 - p_2)$  ist  $lcl = ucl$ , da beide Grenzen dem gleichen Quantil der Verteilung entsprechen, beispielsweise gilt für den oberen rechten Tabelleneintrag  $p_1 = (1 - p_2) = 0.1$ . Damit ist  $R_i = 0$  für Werte unterhalb dieser Grenze, oberhalb gilt  $R_i = 2$ , die Ausprägung  $R_i = 1$  kann nicht auftreten. Wird die Summe der  $R_i$  kleiner als 46 oder größer als 60, ist der Prozess zum Niveau  $\alpha = 0.008$  als außer Kontrolle anzunehmen.

Die Grenzwerte LCL und UCL hängen von  $p_1$  und  $p_2$  ab. Eine Erhöhung des Wertes  $p_1$  bewirkt sinkende Grenzen, da tendenziell eher kleinere  $R_i$ -Werte vergeben werden. Die Kombination  $p_1 = 0.7$  und  $p_2 = 0.1$  sorgt beispielsweise dafür, dass die meisten  $R_i$  (theoretisch 70%) null werden. Eine Erhöhung von  $p_2$  bewirkt dagegen einen Anstieg der Grenzen, da hierbei der Bereich von  $R_i = 2$  größer ist. Die bisher verwendeten Grenzen für  $p_1 = p_2 = 0.25$  passen sich mit LCL= 22 und UCL= 38 gut in die Tabelle ein.

Zwei Optimierungsansätze werden im Folgenden betrachtet:

- $n$  oder  $p_1$  und  $p_2$  variieren, um das Niveau zu optimieren (Abschnitt 6.6.2)
- $p_1$  und  $p_2$  variieren, um die ARL außer Kontrolle zu verbessern (Kapitel 6.6.3)

## 6.6.2 Niveauausschöpfung verbessern

In Tabelle 6.3 wird die Kontrollkarte dahingehend optimiert, dass sie für  $n = 30$  das 5%-Niveau möglichst ausschöpft. Mit Beschränkung auf eine Nachkommastelle liegt das Maximum bei den zu  $p_1 = 0.3$  sowie  $p_2 = 0.1$  gehörenden  $lcl$  beziehungsweise  $ucl$ , LCL=18 und UCL=30 ( $\alpha = 0.0469$ ).

Um das Niveau  $\alpha$  besser auszureizen, ohne den Stichprobenumfang zu erhöhen, werden  $p_1$  und  $p_2$  feiner gerastert. So ist es möglich, das Niveau beliebig genau anzunähern. Für zwei Nachkommastellen wird  $\alpha \leq 0.05$  beispielsweise gut mit den Werten  $p_1 = 0.42$  und  $p_2 = 0.3$  und den Grenzen 18 und 35 realisiert ( $\alpha = 0.04954$ ), für drei Nachkommastellen wird das 5%-Niveau dagegen mit den Werten  $p_1 = 0.235$  und  $p_2 = 0.63$  und den Grenzen 33 und 50 bestmöglich angenähert ( $\alpha = 0.04985$ ). Der Rechenaufwand wächst mit jeder zusätzlich betrachteten Nachkommastelle deutlich, daher sollte die Unterteilung nicht zu fein gewählt werden.

Tabelle B.4 im Anhang B zeigt zusätzlich, wie sich eine Erhöhung der Stichprobengröße  $n$  auf die Entwicklung des Niveaus auswirkt. Zur Verdeutlichung des Einflusses der Stichprobengröße werden für  $p_1$  und  $p_2$  konstant zwei Nachkommastellen verwendet. Die gewählte Niveaugrenze von 0.05 wird für größere Stichproben erwartungsgemäß eher erreicht als für kleinere. Durch die diskrete Verteilung erfolgt die Verbesserung der Niveauschöpfung nicht stetig, sondern es treten Sprünge und „Rückschritte“ für ungünstige Stichprobengrößen auf. Um die praktische Umsetzung zu erleichtern, sollte  $n$  nicht zu groß werden. Für die Niveaueinhaltung ist der Vorteil größerer Stichproben allerdings eher unbedeutend.

In diesem Abschnitt wird nur betrachtet, wie gut das Testniveau ausgeschöpft wird. Durch die diskrete Verteilungsfunktion können die Werte von  $p_1$  und  $p_2$  je nach Anzahl der Nachkommastellen völlig verschieden ausfallen. Die Testgüte bleibt bei diesem Verfahren unberücksichtigt. Dadurch können beispielsweise die ARL-Werte im Außer-Kontrolle-Fall sehr schlecht werden. Das Verhalten der ARL-Werte wird im nächsten Absatz untersucht.

### 6.6.3 ARL-Verhalten optimieren

Im Folgenden wird untersucht, wie sich die Wahl der Grenzen  $lcl$  und  $ucl$  beziehungsweise LCL und UCL auf das Verhalten der ARL auswirkt. Zur besseren Vergleichbarkeit der Resultate werden bei der Gütestudie weitestgehend konstante Bedingungen gewählt. Getestet werden verschiedene Grenzwerte unter den Voraussetzungen  $n = 30$  und  $\alpha \leq 0.05$ . Beispielhaft wird das Abhängigkeitsmaß  $\tau_0 = 0.5$  betrachtet. Dafür werden drei Prozesssituationen untersucht:

- a) deutlich außer Kontrolle:  $\tau = 0.4$  beziehungsweise  $\tau = 0.6$
- b) außer Kontrolle:  $\tau = 0.45$  beziehungsweise  $\tau = 0.55$
- c) unter Kontrolle:  $\tau = 0.5$

Fall c) dient der Plausibilitätsprüfung, da sich die ARL unter Kontrolle durch das Niveau des jeweiligen Tests ergibt. Diese fünf Stützwerte reichen aus, um die ARL der verschiedenen Grenzpaare zu vergleichen. Abbildungen 6.11 und 6.12 zeigen, dass die Funktionen global stetig sind. Sie steigen monoton bis zum Punkt  $\tau = \tau_0$  und fallen anschließend monoton.



Die Simulation erfolgt analog zu der aus Abschnitt 6.3.2. Jeweils  $N = 20000$  Lauflängen werden durch Monte-Carlo-Simulationen erzeugt und daraus die ARL-Werte geschätzt. Um trotz unterschiedlicher Ausgangsniveaus  $\alpha$  (siehe Tabelle 6.3) die ARL-Werte vergleichbar zu machen, wird der Wert unter  $H_0$  auf 1 gesetzt und die restlichen Werte diesem proportional angepasst. Somit liegen für die verschiedenen Szenarien durchschnittliche Lauflängen zwischen 0 (außer Kontrolle) und 1 (unter Kontrolle) vor.

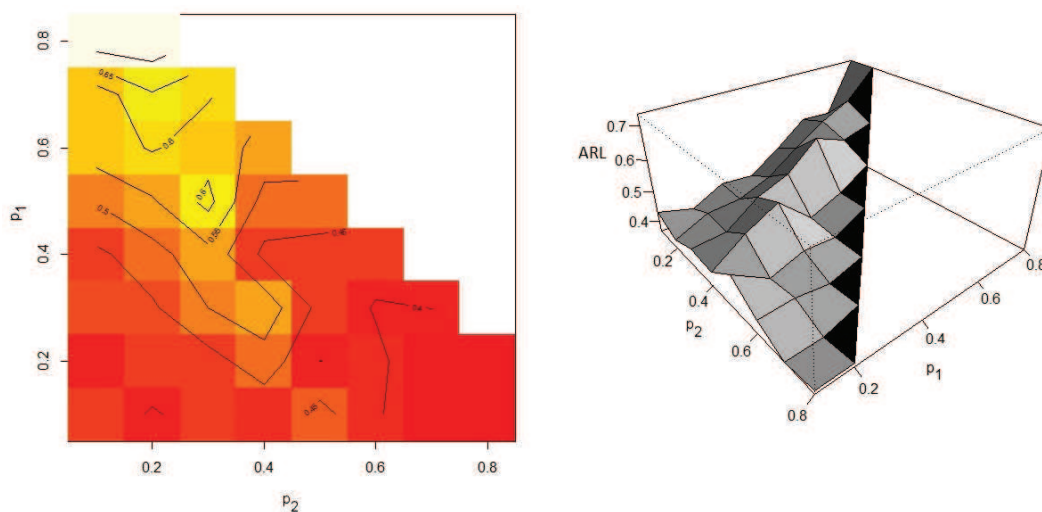


Abbildung 6.22: Mittlere ARL-Werte für verschiedene  $p_1$ - $p_2$ -Kombinationen

Abbildung 6.22 zeigt die Ergebnisse der Simulation unter Annahme einer Claytoncopula für verschiedene Kombinationen von  $p_1$  und  $p_2$  (orientiert an Tabelle 6.3). Dargestellt wird für jede  $p_1$ - $p_2$ -Kombination das arithmetische Mittel ihrer vier ARL-Werte aus Situation a) und b). Wenn nachfolgend also von „durchschnittlicher ARL im Außer-Kontrolle-Fall“ gesprochen wird, ist das Mittel dieser (beliebig gewählten) Punkte gemeint. Je kleiner dieser Wert ist, umso schneller werden Außer-Kontrolle-Situationen erkannt und umso besser ist die zugrundeliegende Kontrollkarte zur Problemlösung geeignet.

Die rechte Grafik visualisiert die Punkte in einem dreidimensionalen Flächendiagramm. Das Diagramm zeigt einen „Gebirgskamm“. Er beginnt in der hinteren Ecke bei  $p_1 = 0.8$  und  $p_2 = 0.1$  beziehungsweise  $p_2 = 0.2$ . Die durchschnittliche ARL beträgt hier etwa 73.3% der In-Kontrolle-ARL. Nicht erwünschte Prozesssituationen

werden also erst spät erkannt. Der Höhenzug verläuft diagonal bis zu  $p_1 = 0.1$  und  $p_2 = 0.5$  mit durchschnittlich 46.8% der ursprünglichen ARL. Die Karte lehnt demzufolge im Schnitt etwa doppelt so oft ab wie im kontrolliert ablaufenden Prozess. Für die Bereiche oberhalb und unterhalb dieser Diagonale sinkt die ARL im Außer-Kontrolle-Fall schneller.

Im linken Bild werden Höhenlinien und Farbmarkierungen genutzt, um die interessierenden Werte darzustellen. Beide Bilder ergänzen sich in ihren Informationen. Es wird deutlich, dass sowohl kleine als auch große Werte von  $p_2$  in Kombination mit kleinen  $p_1$ -Werten das ARL-Verhalten optimieren. Diese Kombinationen liefern also die Karten mit den besten Güteeigenschaften.

Die bereits erläuterte Abbildung 6.5 veranschaulicht den Zusammenhang zwischen den Grenzwerten  $ucl$  und  $lcl$ . Ein größeres  $p_1$  verschiebt die untere Grenze  $lcl$  nach oben, so dass für mehr Datenpunkte  $R_i = 0$  gilt. Ein wachsendes  $p_2$  bewirkt dagegen, dass sich die obere Grenze  $ucl$  der Winkelhalbierenden nähert und dadurch mehr  $R_i$ -Werte den Wert 2 annehmen. Dieser Trend ist durch die Bedingung  $lcl < ucl$  begrenzt.

Die vorliegende Simulationsstudie betrachtet sowohl sehr schmale als auch breite Bereiche mit  $R_i = 1$ : Zwischen 10% (für  $p_1 = 0.1$  und  $p_2 = 0.8$ ) und 80% (für  $p_1 = 0.1$  und  $p_2 = 0.1$ ) der Werte werden mit  $R_i = 1$  klassifiziert.

Nach der Interpretation von Abbildung 6.22 zeigen  $lcl$ -Funktionen aus kleinen  $p_1$ -Werten, kombiniert mit  $ucl$ -Funktionen aus entweder kleinen oder großen  $p_2$ -Werten das gewünschte ARL-Verhalten. Am effektivsten scheint daher ein Streifen für  $R_i = 1$  zu sein, der entweder

- vergleichsweise schmal ist und dicht an der Winkelhalbierenden anliegt  
( $0.1 \leq p_1 \leq 0.3, 0.6 \leq p_2 \leq 0.8$ )
- oder relativ breit ( $0.1 \leq p_1 \leq 0.3, 0.1 \leq p_2 \leq 0.4$ ) ist.

Bei der ersten Variante erhalten die meisten Punkte den Wert  $R_i = 2$  (60%-80%). Bei der zweiten Möglichkeit ist  $R_i = 1$  am häufigsten (30%-80%). Die Grenzwerte UCL und LCL fallen daher bei der ersten Variante größer aus als bei der zweiten.

Abbildung 6.23 zeigt geeignete  $p_1$ - $p_2$ -Kombinationen, die sich anhand von Abbildung 6.22 ergeben. Der Plot oben links stellt die bisher verwendeten Quartile als

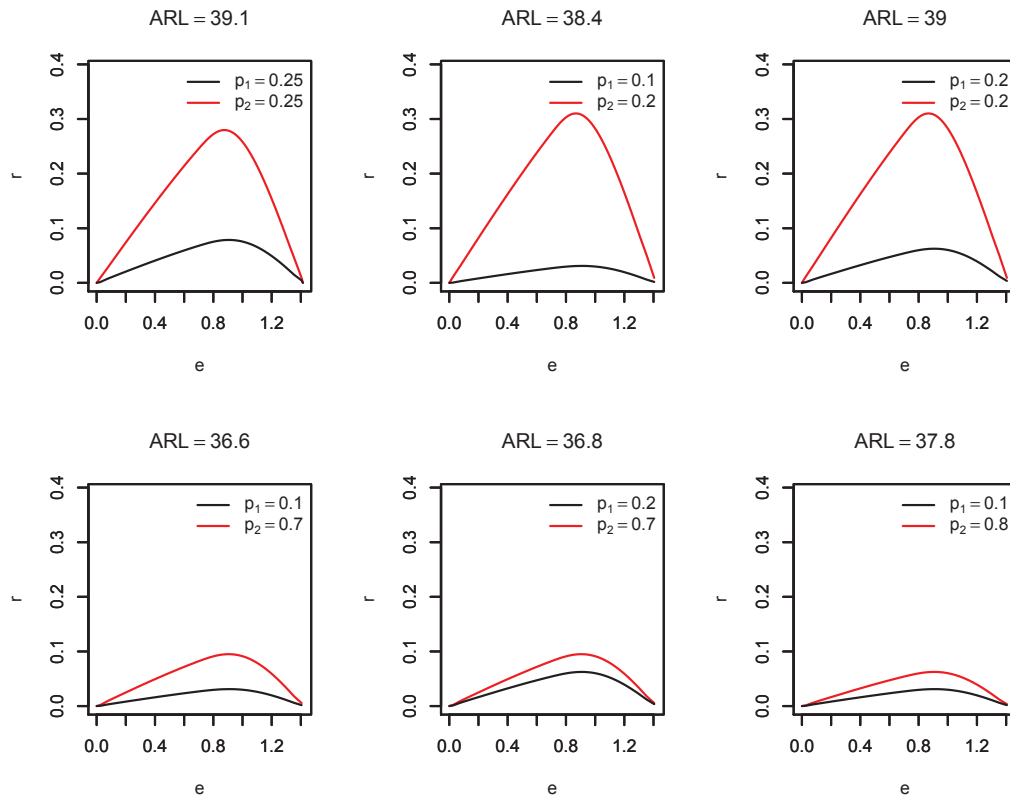


Abbildung 6.23: *lcl* (schwarz) und *ucl* (rot) für ARL-optimierte Kontrollkarten

Grenzwerte ( $p_1 = p_2 = 0.25$ ) dar. Die Ergebnisse hierfür liegen im akzeptablen Bereich, die durchschnittliche ARL außerhalb von  $H_0$  beträgt etwa 39% der ARL unter Kontrolle.

Die obere mittlere und rechte Grafik zeigen Grenzwertpaare, für welche die ARL im Bereich „links“ des Bergkamms optimal wird. Das Minimum in der Simulationsstudie mit einer Nachkommastelle liegt bei  $p_1 = 0.1$  und  $p_2 = 0.7$  mit einem Wert von rund 36.6% (Abbildung 6.23, unten links). Die nächstkleineren Werte liefern die Kombination  $p_1 = 0.2$  und  $p_2 = 0.7$  ( $\sim 36.8\%$ , Abbildung 6.23, unten Mitte),  $p_1 = 0.1$  und  $p_2 = 0.8$  ( $\sim 37.7\%$ , Abbildung 6.23, unten rechts) sowie  $p_1 = 0.2$  und  $p_2 = 0.8$  ( $\sim 37.8\%$ , ohne Abbildung). Die nächsten Optimierungsschritte erfolgen daher im Bereich von  $0.1 \leq p_1 \leq 0.3$  und  $0.6 \leq p_2 \leq 0.8$ .

#### 6.6.4 ARL-Verhalten und Niveaue Ausschöpfung verbessern

Bislang wurden das ARL-Verhalten und die Niveaueausnutzung unabhängig voneinander betrachtet. Jetzt werden beide Eigenschaften kombiniert untersucht. Zwei der in Abschnitt 6.6.2 gefundenen Grenzen zur Niveaueoptimierung ( $p_1 = 0.42$ ,  $p_2 = 0.3$ ;  $p_1 = 0.235$ ,  $p_2 = 0.63$ ) liegen in den Bereichen mit gutem ARL-Verhalten (siehe Abbildung 6.22) und sind daher als mögliche Optima zu berücksichtigen.

Allerdings wird ein schnelles Absinken der ARL im Außer-Kontrolle-Fall als wesentlich wichtiger angesehen als eine optimale Niveaueinhaltung, so dass die Optimierung hier startet. Dazu wird zunächst der Bereich von  $p_1$  und  $p_2$  eingegrenzt, für den die ARL am deutlichsten sinkt - unabhängig von dem realisierten Niveau. Anschließend werden in der Umgebung der gefundenen  $p_1$  und  $p_2$  die Grenzwerte mit der maximalen Niveaueausnutzung gewählt.

Abbildung 6.22 zeigt, dass die ARL am schnellsten im Bereich  $p_1 = 0.1$  bis  $p_1 = 0.3$  und  $p_2 = 0.6$  bis  $p_2 = 0.8$  sinkt. Kombinationen mit kleinem  $p_1$  und kleinem  $p_2$  zeigen ebenfalls gute Resultate. Der letztgenannte Abschnitt ist allerdings kleiner und die durchschnittliche ARL im Außer-Kontrolle-Fall ist höher als in der ersten Region. Daher erfolgt die Suche nach dem globalen Minimum in der Umgebung von  $0.1 \leq p_1 \leq 0.3$  und  $0.6 \leq p_2 \leq 0.8$ .

Abbildung 6.24 zeigt für den interessierenden Ausschnitt von  $p_1$ - $p_2$ -Kombinationen die durchschnittliche ARL im Außer-Kontrolle-Fall. Die mittlere ARL ist im gewählten Ausschnitt erwartungsgemäß kleiner als bei der ersten Simulation. Ist der Prozess außer Kontrolle, nimmt die ARL im Bereich von  $0.1 \leq p_1 \leq 0.14$  und  $0.66 \leq p_2 \leq 0.70$  am schnellsten ab. Das Minimum der durchschnittlichen ARL beträgt 0.341 für  $p_1 = 0.1$  und  $p_2 = 0.68$ .

In einem zweiten Schritt wird das Niveau maximiert. Die  $p_1$ - $p_2$ -Intervalle für optimale Grenzwerte  $lcl$  und  $ucl$  sind relativ klein; daher ist es wahrscheinlich, dass das gewünschte Niveau nicht exakt erreicht wird. Für  $\alpha \leq 5\%$  liefert die Kombination  $p_1 = 0.122$ ,  $p_2 = 0.67$ , LCL= 39, UCL= 53 und  $\alpha = 0.039$  die bestmögliche Annäherung. Die Optimalwerte für  $\alpha \leq 10\%$  liegen bei  $p_1 = 0.137$ ,  $p_2 = 0.679$ , LCL= 40, UCL= 52 und  $\alpha = 0.08$ .

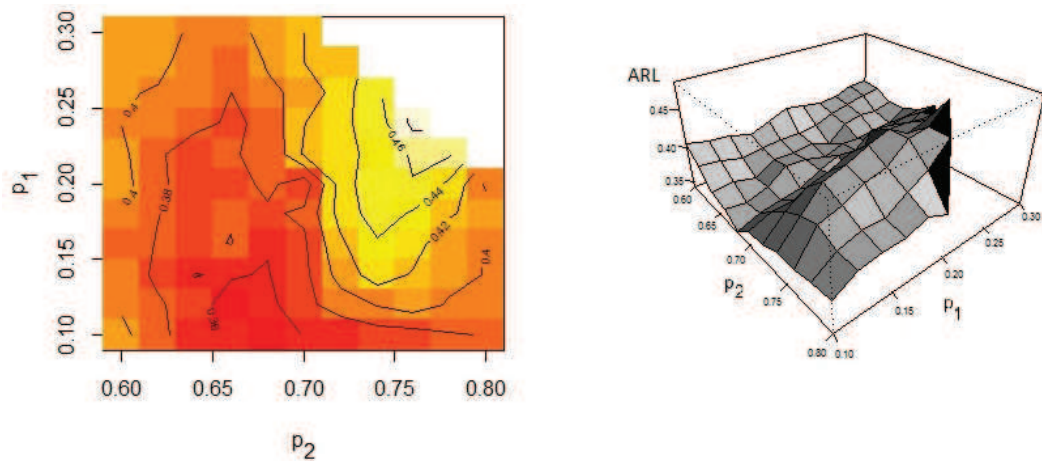


Abbildung 6.24: Interessierender Ausschnitt: mittlere ARL-Werte für verschiedene  $p_1$ - $p_2$ -Kombinationen

Um auszuschließen, dass der (zunächst unberücksichtigte) Bereich mit geringen  $p_1$ - und  $p_2$ -Werten unter Umständen die Niveauschöpfung maximiert, wird auch hierfür die optimale Kombination für  $\alpha \leq 5\%$  bestimmt. Sie liegt bei  $p_1 = 0.165$ ,  $p_2 = 0.2$ , LCL= 25, UCL= 37 und  $\alpha = 0.03575$ . Damit bietet sie gegenüber dem getesteten Bereich mit hohen  $p_2$ -Werten keinen Vorteil.

Das obige Verfahren optimiert die Grenzwerte nach zwei Gesichtspunkten. Im direkten ARL-Vergleich übertrifft die optimierte Karte die Ursprungskarte ( $p_1 = p_2 = 0.25$ ) leicht. Das Niveau  $\alpha$  wird um 1.2 Prozentpunkte besser ( $\alpha = 0.039$  statt  $\alpha = 0.027$ ), gleichzeitig sinkt die normierte ARL etwas schneller. Tabelle 6.4 zeigt die ARL für die betrachteten Außer-Kontrolle-Situationen. Ob der Rechenaufwand insgesamt gerechtfertigt ist, soll hier nicht beurteilt werden.

	$\tau = 0.4$	$\tau = 0.45$	$\tau = 0.55$	$\tau = 0.6$
$p_1 = 0.25, p_2 = 0.25$ ( $\alpha = 0.027$ )	0.250	0.534	0.566	0.213
$p_1 = 0.122, p_2 = 0.67$ ( $\alpha = 0.039$ )	0.179	0.507	0.539	0.160

Tabelle 6.4: Normierte ARL für optimale Grenzwerte und „Standardgrenzen“ zum Niveau  $\alpha \leq 0.05$  für Daten einer Claytoncopula mit  $\tau_0 = 0.5$



## 7 Zusammenfassung und Ausblick

Die Vorteile einer Datenmodellierung mit Hilfe von Copulas werden zunehmend auch von Ingenieuren und Naturwissenschaftlern erkannt. Ziel dieser Dissertation war es daher, Qualitätsregelkarten für (bivariate) Messgrößen zu entwickeln, denen eine Copulastruktur zugrunde liegt. Damit soll der Schritt gelingen, die neuen Kenntnisse über das Verteilungsverhalten der Prozessdaten in die Qualitätssicherung und -verbesserung zu übertragen.

Um das Thema überschaubar zu halten, wurden lediglich bivariate archimedische Prozesse betrachtet. Als Randverteilungen der Einzeldaten dienten Normalverteilungen. Das verdeutlichte, dass univariat normalverteilte Merkmale nicht zwingend gemeinsam durch eine multivariate Normalverteilung modellierbar sind. Unter diesen Voraussetzungen wurden Qualitätsregelkarten für mehrere Fragestellungen entwickelt.

Die Güte aller Karten wurde anhand der durchschnittlichen Lauflänge des Prozesses bewertet. Bei kontrolliert ablaufenden Prozessen sollte sie möglichst groß sein (in Abhängigkeit vom Testniveau), bei Prozessstörungen dagegen klein ausfallen.

Nach der Einleitung und drei Grundlagenkapiteln widmete sich Kapitel 5 zunächst der Lage und Streuung des Prozesses. Weicht die gemeinsame Verteilung zweier Merkmale von der multivariaten Normalverteilung (stark) ab, wird das Verhalten der klassischen  $T^2$ -Karte (extrem) schlecht. Die ARL liegt dann oft deutlich unter den erwarteten Werten, das Testniveau wird in der Regel nicht eingehalten.

Diesen Nachteil beheben die vorgestellten modifizierten Karten. Die adjustierten  $T^2$ -Karten verwenden Grenzwerte aus Monte-Carlo-Simulationen. Eine Simulationsstudie für Verschiebungen der Prozesslage zeigt, dass die Karten für das vorliegende Problem geeignet sind.

Je nach Art der erwarteten Lageverschiebung des Prozesses erwies sich auch die Wahl eines alternativen Abstandsmaßes als sinnvoll. Die daraus resultierende Karte wurde analog zu Shewarts  $T^2$ -Karte als Euklidische  $E^2$ -Karte bezeichnet. Beide Karten liefern plausible Resultate für Beispiele aus der Messtechnik.

Eine Studie zu Änderungen in der Prozessstreuung zeigte, dass Shewarts  $T^2$ -Karte ein ähnliches Verhalten aufweist wie für Verschiebungen in der Lage. Die Entwicklung verbesserter Karten auf Grundlage von Monte-Carlo-Simulationen wäre problemlos analog zum Vorgehen bei Lageüberwachung möglich. Darauf wurde in dieser Arbeit verzichtet.

Die Karten aus Abschnitt 5 sind zur Überwachung der Prozessstruktur nicht geeignet. Daher wurde in Kapitel 6 eine neue Karte entwickelt, die sowohl eine Verstärkung als auch eine Abschwächung des Zusammenhangs zwischen den Daten erkennt.

Eine gemeinsame Verteilung von Datenpunkten kann eindeutig in Randverteilungen und Copula zerlegt werden. Die entsprechenden theoretischen Grundlagen vermittelt Kapitel 3. Bei der Lage- und Streuungsüberwachung beeinflussen die Randverteilungen die Teststatistik. Für die Überwachung der Struktur reicht es, nur die Copula zu betrachten. Die Beschränkung auf einzeln normalverteilte Daten war demzufolge in diesem Kapitel nicht nötig, wurde allerdings trotzdem beibehalten.

Entwickelt wurde ein zweistufiges Verfahren: Zur Anwendung der Karte werden zunächst anhand von Wahrscheinlichkeitsbändern die modifizierten Datenpunkte klassiert; die Testgröße  $R_n$  ist dann die Summe aller  $n$  Klassenausprägungen einer Stichprobe. Sie lässt sich mit Hilfe einer Binomialverteilung beschreiben, so dass die Grenzwerte einfach herzuleiten sind.

Die Wahrscheinlichkeitsbänder werden durch die untere Grenze ( $lcl$ ) und die obere Grenze ( $ucl$ ) festgelegt. Die Funktionen von  $lcl$  und  $ucl$  ergeben sich als Integraldarstellung. Eine ausreichend genaue geschlossene Anpassung der Funktionen konnte nicht gefunden werden, die benötigten Grenzwerte sind aber durch numerische Integration bestimmbar.

Ein Vergleich mit Steigers  $Z$ -Test zeigte die Überlegenheit der gefundenen Teststatistik.

Ein klassisches Anwendungsgebiet der Copulas ist die gemeinsame Modellierung



von Aktienkursen. Um die inhaltliche Plausibilität der Karte zu untersuchen, bot sich daher eine Untersuchung der Kurse zur Zeit der Finanzkrise an. Insbesondere im asiatischen Raum gibt es derzeit Tendenzen, den Verkehrsfluss und dessen Einwirkungen auf Straßen, Schienen und Brücken mit Hilfe von Copulas darzustellen. Dieser Trend wurde hier berücksichtigt, indem als zweites Beispiel Tragseile von Verkehrsbrücken betrachtet wurden, deren gemeinsame Schwingung nicht zu stark voneinander abhängig werden sollte.

Eine Optimierung der Grenzwerte  $lcl$  und  $ucl$  komplettierte das Kapitel. Betrachtet wurde eine feste Copula (Claytoncopula mit dem Parameter  $\theta = 2$ ,  $\tau = 0.5$ ). Die gefundenen Optimalgrenzen nutzen das 5%–Niveau 1.2 Prozentpunkte besser aus als die ursprünglich gewählten Grenzen ( $\alpha = 0.039$  statt  $\alpha = 0.027$ ), gleichzeitig sinkt die ARL etwas schneller. Da die Optimierung allerdings sehr rechenaufwendig ist, müssen in der Praxis Aufwand und Nutzen im Einzelfall gegeneinander abgewogen werden.

Aus dem letzten Abschnitt ergibt sich auch der erste Ansatz für weitere Forschungen auf dem Gebiet copulabasierter Qualitätsregelkarten. So könnte die vorgestellte Teststatistik für verschiedene Fälle optimiert werden. Dazu müssten allerdings konkrete Einsatzgebiete betrachtet werden; eine „allumfassende“ Verbesserung erscheint unrealistisch aufgrund des hohen Arbeitsaufwandes.

Eine Ausweitung der betrachteten Grundgesamtheiten ist ebenfalls denkbar. Archimedische Copulas decken schon viele Datenstrukturen ab, eine Ausweitung auf weitere Copulas würde die Flexibilität allerdings noch erhöhen. Extremwertcopulas könnten ein lohnendes Ziel sein, um seltene Ereignisse (Reaktorausfälle, bestimmte Emissionswerte, ...) zu berücksichtigen. Auch Kombinationen von Copulas ergeben wieder Copulas und erweitern die Einsatzmöglichkeiten noch einmal.

Interessanter erscheint zunächst jedoch die Ausweitung der gefundenen Teststatistik auf mehr als zwei abhängige Merkmale.

Dazu müssen zunächst das betrachtete Lot auf die Winkelhalbierende des  $I^2$ , also  $r$ , und die normierte Entfernung zum Ursprung  $e$  auf höhere Dimensionen verallgemeinert werden. Anschließend sind die  $lcl$  und  $ucl$  als Quantile der Verteilung von  $r$  in Abhängigkeit von  $e$  analog zum Vorgehen hier bestimmbar. Für  $p$  betrachtete Merkmale handelt es sich dabei um gekrümmte Ebenen im Raum  $[0, \sqrt{2}]^{p-1}$ .

Die Bereitstellung der gefundenen Resultate für die Praxis ist ebenfalls wünschenswert. Dies kann beispielsweise durch die Erstellung einer geeigneten Software erfolgen. Noch besser wäre es, die Karte direkt in eine bestehende Statistiksoftware für Ingenieure einzubinden, beispielsweise Minitab oder qs-STAT.

# A Ergänzende Formeln und Berechnungen

## A.1 Wahrscheinlichkeiten für Kapitel 6

### A.1.1 Die Wahrscheinlichkeit $P(r \leq r^*, e_1^* \leq e \leq e_2^*)$

Anwenden der Formeln (6.8) und (6.9) auf Gleichung (6.6) führt zu folgenden Wahrscheinlichkeiten:

$$P(r \leq r^*, e_1^* \leq e \leq e_2^*) = \tag{A.1}$$

für  $e_2^* < 1/\sqrt{2}$  :

$$\begin{aligned} & \int_0^{u_3+u_4} \int_0^{u_3+u_4-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 - \int_0^{u_1+u_2} \int_0^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 \\ & - \left[ \int_0^{u_4} \int_{u_3-u_4+\tau_2}^{u_3+u_4-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 - \int_0^{u_2} \int_{u_1-u_2+\tau_2}^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 \right] \end{aligned}$$

für  $e_1^* < 1/\sqrt{2}$  und  $1/\sqrt{2} \leq e_2^* < \sqrt{2}$  :

$$\begin{aligned} & 1 - \int_{u_3+u_4-1}^1 \int_{u_3+u_4-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 - \int_0^{u_1+u_2} \int_0^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 \\ & - \left[ \int_0^{u_3+u_4-1} \int_{u_3-u_4+\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 + \int_{u_3+u_4-1}^{u_4} \int_{u_3-u_4+\tau_2}^{u_3+u_4-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 \right. \\ & \left. - \int_0^{u_2} \int_{u_1-u_2+\tau_2}^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 \right] \end{aligned}$$

für  $1/\sqrt{2} \leq e_1^* < e_2^* < \sqrt{2}$  :

$$\begin{aligned} & - \int_{u_3+u_4-1}^1 \int_{u_3+u_4-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 + \int_{u_1+u_2-1}^1 \int_{u_1+u_2-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 \\ & - \left[ \int_0^{u_3+u_4-1} \int_{u_3-u_4+\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 + \int_{u_3+u_4-1}^{u_4} \int_{u_3-u_4+\tau_2}^{u_3+u_4-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 \right. \\ & \left. - \int_0^{u_1+u_2-1} \int_{u_1-u_2+\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 - \int_{u_1+u_2-1}^{u_2} \int_{u_1-u_2+\tau_2}^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 \right] \end{aligned}$$

und für  $1/\sqrt{2} \leq e_1^* < \sqrt{2}$  und  $e_2^* = \sqrt{2}$  :

$$\begin{aligned} & \int_{u_1+u_2-1}^1 \int_{u_1+u_2-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 \\ & - \left[ \int_0^{u_1+u_2-1} \int_{u_1-u_2+\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 + \int_{u_1+u_2-1}^{u_2} \int_{u_1-u_2+\tau_2}^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 \right] \end{aligned}$$

### A.1.2 Die Wahrscheinlichkeit $P(e_1^* \leq e \leq e_2^*)$

Einsetzen der Formel (6.8) in die Gleichung (6.5) führt zu:

$$P(e_1^* \leq e \leq e_2^*) = P(e \leq e_2^*) - P(e \leq e_1^*) = \tag{A.2}$$

für  $e_2^* < 1/\sqrt{2}$  :

$$\int_0^{u_3+u_4} \int_0^{u_3+u_4-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 - \int_0^{u_1+u_2} \int_0^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2$$

für  $e_1^* < 1/\sqrt{2}$  und  $1/\sqrt{2} \leq e_2^* < \sqrt{2}$  :

$$1 - \int_{u_3+u_4-1}^1 \int_{u_3+u_4-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 - \int_0^{u_1+u_2} \int_0^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2$$

für  $1/\sqrt{2} \leq e_1^* < e_2^* < \sqrt{2}$  :

$$- \int_{u_3+u_4-1}^1 \int_{u_3+u_4-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 + \int_{u_1+u_2-1}^1 \int_{u_1+u_2-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2$$

und für  $1/\sqrt{2} \leq e_1^* < \sqrt{2}$  und  $e_2^* = \sqrt{2}$ :

$$\int_{u_1+u_2-1}^1 \int_{u_1+u_2-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2$$

### A.1.3 Die Wahrscheinlichkeit $P(r \leq r^* | e_1^* \leq e \leq e_2^*)$

Die zur Berechnung der Grenzwerte *lcl* und *ucl* nötige bedingte Wahrscheinlichkeit ergibt sich als Quotient der Gleichungen aus A.1.1 und A.1.2. Hierbei gelten:

$$P(r \leq r^* | e_1^* \leq e \leq e_2^*) = \frac{P(r \leq r^*, e_1^* \leq e \leq e_2^*)}{P(e_1^* \leq e \leq e_2^*)} =$$

für  $e_2^* < 1/\sqrt{2}$ :

$$1 - \frac{\int_0^{u_4} \int_{u_3-u_4+\tau_2}^{u_3+u_4-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 - \int_0^{u_2} \int_{u_1-u_2+\tau_2}^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2}{\int_0^{u_3+u_4} \int_0^{u_3+u_4-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2 - \int_0^{u_1+u_2} \int_0^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2}$$

für  $e_1^* < 1/\sqrt{2}$  und  $1/\sqrt{2} \leq e_2^* < \sqrt{2}$ :

$$1 - \frac{\int_0^{u_3+u_4-1} \int_{u_3-u_4+\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 + \int_{u_3+u_4-1}^{u_4} \int_{u_3-u_4+\tau_2}^{u_3+u_4-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2}{1 - \int_{u_3+u_4-1}^1 \int_{u_3+u_4-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 - \int_0^{u_1+u_2} \int_0^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2} + \frac{\int_0^{u_2} \int_{u_1-u_2+\tau_2}^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2}{1 - \int_{u_3+u_4-1}^1 \int_{u_3+u_4-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 - \int_0^{u_1+u_2} \int_0^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2}$$

für  $1/\sqrt{2} \leq e_1^* < e_2^* < \sqrt{2}$ :

$$1 - \frac{\int_0^{u_3+u_4-1} \int_{u_3-u_4+\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 + \int_{u_3+u_4-1}^{u_4} \int_{u_3-u_4+\tau_2}^{u_3+u_4-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2}{\int_{u_1+u_2-1}^1 \int_{u_1+u_2-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 - \int_{u_3+u_4-1}^1 \int_{u_3+u_4-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2} + \frac{\int_0^{u_1+u_2-1} \int_{u_1-u_2+\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 + \int_{u_1+u_2-1}^{u_2} \int_{u_1-u_2+\tau_2}^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2}{\int_{u_1+u_2-1}^1 \int_{u_1+u_2-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 - \int_{u_3+u_4-1}^1 \int_{u_3+u_4-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2}$$

und für  $1/\sqrt{2} \leq e_1^* < \sqrt{2}$  und  $e_2^* = \sqrt{2}$ :

$$1 - \frac{\int_0^{u_1+u_2-1} \int_{u_1-u_2+\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2 + \int_{u_1+u_2-1}^{u_2} \int_{u_1-u_2+\tau_2}^{u_1+u_2-\tau_2} c(\tau_1, \tau_2) d\tau_1 d\tau_2}{\int_{u_1+u_2-1}^1 \int_{u_1+u_2-\tau_2}^1 c(\tau_1, \tau_2) d\tau_1 d\tau_2}$$

### A.1.4 Beispiel Produktcopula - Herleitung der Integrale

Fall 1:  $0 \leq e^* < 1/\sqrt{2}$

$$\begin{aligned}
 P(B) &= \int_0^{u_2} \int_{u_1-u_2+\tau_2}^{u_1+u_2-\tau_2} 1 \partial\tau_1 \partial\tau_2 \\
 &= \int_0^{u_2} [u_1 + u_2 - \tau_2 - u_1 + u_2 - \tau_2] \partial\tau_2 = \int_0^{u_2} [2u_2 - 2\tau_2] \partial\tau_2 \\
 &= [(2u_2 - \tau_2^2)]_0^{u_2} = 2u_2^2 - u_2^2 = u_2^2
 \end{aligned}$$

$$\begin{aligned}
 P(e \leq e^*) &= \int_0^{u_1+u_2} \int_0^{u_1+u_2-\tau_2} 1 \partial\tau_1 \partial\tau_2 \\
 &= \int_0^{u_1+u_2} (u_1 + u_2 - \tau_2) \partial\tau_2 \\
 &= [\tau_2(u_1 + u_2 - 0.5\tau_2)]_0^{u_1+u_2} \\
 &= ((u_1 + u_2) - 0.5(u_1 + u_2))(u_1 + u_2) \\
 &= 0.5(u_1 + u_2)^2
 \end{aligned}$$

Fall 2:  $1/\sqrt{2} \leq e^* < \sqrt{2}$

$$\begin{aligned}
 P(B) &= \int_0^{u_1+u_2-1} \int_{u_1-u_2+\tau_2}^1 1 \partial\tau_1 \partial\tau_2 + \int_{u_1+u_2-1}^{u_2} \int_{u_1-u_2+\tau_2}^{u_1+u_2-\tau_2} 1 \partial\tau_1 \partial\tau_2 \\
 &= \int_0^{u_1+u_2-1} (1 - u_1 + u_2 - \tau_2) \partial\tau_2 \\
 &\quad + \int_{u_1+u_2-1}^{u_2} (u_1 + u_2 - \tau_2 - u_1 + u_2 - \tau_2) \partial\tau_2 \\
 &= \int_0^{u_1+u_2-1} (1 - u_1 + u_2) \partial\tau_2 - \int_0^{u_1+u_2-1} (\tau_2) \partial\tau_2 \\
 &\quad + \int_{u_1+u_2-1}^{u_2} 2(u_2 - \tau_2) \partial\tau_2 \\
 &= (u_1 + u_2 - 1)(1 - u_1 + u_2) - [0.5\tau_2^2]_0^{u_1+u_2-1} \\
 &\quad + 2u_2(u_2 - u_1 - u_2 + 1) - [\tau_2^2]_{u_1+u_2-1}^{u_2} \\
 &= (u_1 + u_2 - 1)(1 - u_1 + u_2) - 0.5(u_1 + u_2 - 1)^2 \\
 &\quad + 2u_2(1 - u_1) - u_2^2 + (u_1 + u_2 - 1)^2 \\
 &= (u_1 + u_2 - 1)(1 - u_1 + u_2 - 0.5u_1 - 0.5u_2 + 0.5) \\
 &\quad + 2u_2 - 2u_1u_2 - u_2^2 + u_1^2 + u_2^2 + 1 - 2u_1 - 2u_2 + 2u_1u_2 \\
 &= (u_1 + u_2 - 1)0.5(3 - 3u_1 + u_2) + u_1^2 - 2u_1 + 1
 \end{aligned}$$

$$\begin{aligned}
&= 0.5(3u_1 - 3u_1^2 + u_1u_2 + 3u_2 - 3u_1u_2 + u_2^2 - 3 + 3u_1 - u_2) \\
&\quad + u_1^2 - 2u_1 + 1 \\
&= -0.5u_1^2 + 0.5u_2^2 - u_1u_2 + u_1 + u_2 - 0.5 \\
&= u_2^2 - 0.5(u_1 + u_2 - 1)^2
\end{aligned}$$

Zur Berechnung von  $P(e \leq e^*)$  wird zunächst umparametrisiert, es sei  $u := u_1 + u_2$ .

$$\begin{aligned}
P(e \leq e^*) &= 1 - \left( \int_{u-1}^1 \int_{u-\tau_2}^1 1 \, \partial\tau_1 \partial\tau_2 \right) \\
&= 1 - \left( \int_{u-1}^1 (1 - u + \tau_2) \, \partial\tau_2 \right) \\
&= 1 - \left( \int_{u-1}^1 (1 - u) \, \partial\tau_2 + \int_{u-1}^1 \tau_2 \, \partial\tau_2 \right) \\
&= 1 - \left( (1 - u)(1 - u + 1) + [0.5\tau_2^2]_{u-1}^1 \right) \\
&= 1 - \left( (1 - u)(2 - u) + 0.5 - 0.5(u - 1)^2 \right) \\
&= 1 - \left( (2 - 3u + u^2 + 0.5 - 0.5(u^2 - 2u + 1)) \right) \\
&= 1 - \left( (2 - 2u + 0.5u^2) \right) = -1 + 2u - 0.5u^2 \\
&= -1 + 2(u_1 + u_2) - 0.5(u_1 + u_2)^2.
\end{aligned}$$

## A.2 Copuladichten

Zur Berechnung der Grenzen ist es nötig, die Dichtefunktionen der zugrundeliegenden Copula zu kennen. Mit Hilfe der Generatoren und der daraus resultierenden Copulaverteilung (siehe Tabelle 3.1) werden im Folgenden die bivariaten Dichten durch Differenzieren nach beiden Variablen abgeleitet.

### A.2.1 Dichte der Claytoncopula

Für die Claytoncopula gilt mit der Formel (3.1)

$$c(u, v, \theta) = \frac{\partial C(u, v, \theta)}{\partial u \partial v} = \frac{\partial}{\partial v} \frac{\partial}{\partial u} \left( u^{-\theta} + v^{-\theta} - 1 \right)^{-\frac{1}{\theta}}.$$

Zunächst wird die partielle Ableitung nach  $u$  gebildet.

$$\begin{aligned}
\frac{\partial}{\partial u} (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta} &= -\frac{1}{\theta} (u^{-\theta} + v^{-\theta} - 1)^{-1/(\theta-1)} \frac{\partial}{\partial u} (u^{-\theta} + v^{-\theta} - 1) \\
&= -\frac{1}{\theta} (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta-1} \left[ \frac{\partial}{\partial u} u^{-\theta} + \frac{\partial}{\partial u} v^{-\theta} \right] \\
&= -\frac{1}{\theta} (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta-1} (-\theta u^{-\theta-1})
\end{aligned}$$

Da alle Variablen positiv sind, vereinfacht sich der Term durch Kürzen zu

$$\frac{\partial}{\partial u} (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta} = u^{-\theta-1} (u^{-\theta} + v^{-\theta} - 1)^{-\frac{\theta+1}{\theta}}.$$

Der zweite Schritt zur Dichtebestimmung erfolgt durch Ableiten der Funktion nach  $v$ , also

$$\begin{aligned}
c(u, v, \theta) &= \frac{\partial}{\partial v} u^{-\theta-1} (u^{-\theta} + v^{-\theta} - 1)^{-\frac{\theta+1}{\theta}} \\
&= \frac{\partial}{\partial v} u^{-\theta-1} (u^{-\theta} + v^{-\theta} - 1)^{-\frac{\theta+1}{\theta}} \\
&= -\frac{\theta+1}{\theta} u^{-\theta-1} (u^{-\theta} + v^{-\theta} - 1)^{-\frac{\theta+1}{\theta}-1} \frac{\partial}{\partial v} (u^{-\theta} + v^{-\theta} - 1) \\
&= -\frac{\theta+1}{\theta} u^{-\theta-1} (u^{-\theta} + v^{-\theta} - 1)^{-\frac{\theta+1}{\theta}-1} \frac{\partial}{\partial v} v^{-\theta}
\end{aligned}$$

$$\Leftrightarrow c(u, v, \theta) = (\theta + 1) (uv)^{-\theta-1} (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}-2} \quad (\text{A.3})$$

## A.2.2 Dichte der Gumbelcopula

Für die Gumbelcopula gilt nach der Formel (3.1)

$$c(u, v; \theta) = \frac{\partial C(u, v, \theta)}{\partial u \partial v} = \frac{\partial}{\partial v} \frac{\partial}{\partial u} \exp \left( - \left[ (-\ln(u))^\theta + (-\ln(v))^\theta \right]^{\frac{1}{\theta}} \right).$$

Zunächst wird die partielle Ableitung nach  $u$  gebildet.

$$\begin{aligned}
\frac{\partial C(u, v, \theta)}{\partial u} &= \frac{\partial}{\partial u} \exp \left( - \left[ (-\ln(u))^\theta + (-\ln(v))^\theta \right]^{\frac{1}{\theta}} \right) \\
&= \exp \left( - \left[ (-\ln(u))^\theta + (-\ln(v))^\theta \right]^{\frac{1}{\theta}} \right) \frac{\partial}{\partial u} \left( - \left[ (-\ln(u))^\theta + (-\ln(v))^\theta \right]^{\frac{1}{\theta}} \right) \\
&= -\frac{1}{\theta} \exp \left( - \left[ (-\ln(u))^\theta + (-\ln(v))^\theta \right]^{\frac{1}{\theta}} \right) \left[ (-\ln(u))^\theta + (-\ln(v))^\theta \right]^{\frac{1}{\theta}-1} \\
&\quad \cdot \frac{\partial}{\partial u} \left[ (-\ln(u))^\theta + (-\ln(v))^\theta \right]
\end{aligned}$$



$$\begin{aligned}
&= -\frac{1}{\theta} \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-1} \\
&\quad \cdot \left[\frac{\partial}{\partial u}(-\ln(u))^\theta + \frac{\partial}{\partial u}(-\ln(v))^\theta\right] \\
&= -\frac{1}{\theta} \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-1} \\
&\quad \cdot \theta(-\ln(u))^{\theta-1} \frac{\partial}{\partial u}(-\ln(u)) \\
&= \frac{1}{u}(-\ln(u))^{\theta-1} \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-1}
\end{aligned}$$

Der zweite Schritt zur Bestimmung der Gumbeldichte erfolgt durch Ableiten der Funktion nach  $v$ , also

$$\begin{aligned}
c(u, v, \theta) &= \frac{\partial}{\partial v} \left[ \frac{1}{u}(-\ln(u))^{\theta-1} \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \right. \\
&\quad \left. \cdot \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-1} \right] \\
&= \frac{(-\ln(u))^{\theta-1}}{u} \frac{\partial}{\partial v} \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \\
&\quad \cdot \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-1} \\
&= \frac{(-\ln(u))^{\theta-1}}{u} \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-1} \\
&\quad \cdot \frac{\partial}{\partial v} \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \\
&\quad \cdot \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \frac{\partial}{\partial v} \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-1} \\
&= \frac{(-\ln(u))^{\theta-1}}{u} \left[ \left(\frac{1}{\theta} - 1\right) \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \right. \\
&\quad \left. \cdot \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-2} \frac{\partial}{\partial v} \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right] \right. \\
&\quad \left. + \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-1} \frac{\partial}{\partial v} \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \right] \\
&= \frac{(-\ln(u))^{\theta-1}}{u} \left[ \left(\frac{1}{\theta} - 1\right) \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \right. \\
&\quad \left. \cdot \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-2} \left[ \frac{\partial}{\partial v}(-\ln(u))^\theta + \frac{\partial}{\partial v}(-\ln(v))^\theta \right] \right. \\
&\quad \left. + \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-1} \frac{\partial}{\partial v} \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \right] \\
&= \frac{(-\ln(u))^{\theta-1}}{u} \left[ \left(\frac{1}{\theta} - 1\right) \theta(-\ln(v))^{\theta-1} \right. \\
&\quad \left. \cdot \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-2} \frac{\partial}{\partial v}(-\ln(v)) \right. \\
&\quad \left. + \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-1} \frac{\partial}{\partial v} \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \right]
\end{aligned}$$

$$\begin{aligned}
&= \frac{(-\ln(u))^{\theta-1}}{u} \left[ \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-1} \right. \\
&\quad \cdot \frac{\partial}{\partial v} \left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) - \left(\frac{1}{\theta} - 1\right) \theta (-\ln(v))^{\theta-1} \\
&\quad \cdot \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-2} \frac{\partial}{\partial v}(\ln(v)) \left. \right] \\
&= \frac{(-\ln(u))^{\theta-1}}{u} \left[ -\frac{1}{\theta} \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \right. \\
&\quad \cdot \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-2} \frac{\partial}{\partial v} \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right] - \left(\frac{1}{\theta} - 1\right) \theta (-\ln(v))^{\theta-1} \\
&\quad \cdot \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-2} \frac{\partial}{\partial v}(\ln(v)) \left. \right] \\
&= \frac{(-\ln(u))^{\theta-1}}{u} \left[ -\frac{1}{\theta} \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \right. \\
&\quad \cdot \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-2} \frac{\partial}{\partial v} (-\ln(v))^\theta - \left(\frac{1}{\theta} - 1\right) \theta (-\ln(v))^{\theta-1} \\
&\quad \cdot \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-2} \frac{\partial}{\partial v}(\ln(v)) \left. \right] \\
&= \frac{(-\ln(u))^{\theta-1}}{u} \left[ -(-\ln(v))^{\theta-1} \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \right. \\
&\quad \cdot \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-2} \frac{\partial}{\partial v} (-\ln(v)) - \left(\frac{1}{\theta} - 1\right) \theta (-\ln(v))^{\theta-1} \\
&\quad \cdot \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-2} \frac{\partial}{\partial v}(\ln(v)) \left. \right] \\
&= \frac{(-\ln(u))^{\theta-1}}{u} \left[ \frac{1}{v} (-\ln(v))^{\theta-1} \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \right. \\
&\quad \cdot \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-2} - \left(\frac{1}{\theta} - 1\right) \theta (-\ln(v))^{\theta-1} \\
&\quad \cdot \frac{1}{v} \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-2} \left. \right]
\end{aligned}$$

Damit gilt also schließlich:

$$\begin{aligned}
c(u, v, \theta) &= \frac{1}{uv} (-\ln(u))^{\theta-1} (-\ln(v))^{\theta-1} \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right) \\
&\quad \cdot \left( \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}} + \theta - 1 \right) \left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}-2} \quad (\text{A.4})
\end{aligned}$$

### A.2.3 Vergleich der Formeln 6.4 und 6.12

Aus Formel 6.4 folgt:

$$\mathbb{P}\left(\sum_{i=1}^n R_i = x\right) = \binom{2n}{x} 0.5^x (1 - 0.5)^{2n-x} = \frac{(2n)!}{(2n-x)!x!} 0.5^{2n},$$

Formel 6.12 besagt:

$$P^*\left(\sum_{i=1}^n R_i = x\right) = \sum_{j=0}^{\lfloor x/2 \rfloor} n! \frac{1}{(n-x+j)!(x-2j)!j!} p_1^{(n-x+j)} (1-p_1-p_2)^{(x-2j)} p_2^j$$

Wobei für  $p_1 = p_2 = 0.25$  beide Formeln gleiche Resultate liefern. Diese Aussage wird hier exemplarisch für  $x = 0$ ,  $x = 1$  und  $x = 2$  gezeigt. Der Beweis durch vollständige Induktion wäre - durch die verwendete Summe - sehr aufwändig und wird daher hier nicht erbracht.

Zunächst ist für  $p_1 = p_2 = 0.25$ :

$$P^*\left(\sum_{i=1}^n R_i = x\right) = \sum_{j=0}^{\lfloor x/2 \rfloor} n! \frac{1}{(n-x+j)!(x-2j)!j!} 0.5^{2n-x+2j}$$

Damit folgt

$$P^*\left(\sum_{i=1}^n R_i = 0\right) = \frac{n!}{n!0!0!} 0.5^{2n} = \frac{(2n)!}{(2n)!0!} 0.5^{2n} = P\left(\sum_{i=1}^n R_i = 0\right)$$

$$\begin{aligned} P^*\left(\sum_{i=1}^n R_i = 1\right) &= \frac{n!}{(n-1)!1!0!} 0.5^{2n-1} = n 0.5^{2n-1} = n \frac{0.5^{2n}}{0.5} = 2n 0.5^{2n} \\ &= \frac{(2n)!}{(2n-1)!1!} 0.5^{2n} = P\left(\sum_{i=1}^n R_i = 1\right) \end{aligned}$$

$$\begin{aligned} P^*\left(\sum_{i=1}^n R_i = 2\right) &= \frac{n!}{(n-2)!2!0!} 0.5^{2n-2} + \frac{n!}{(n-1)!0!1!} 0.5^{2n-2+2} \\ &= \frac{n(n-1)}{2} 0.5^{2n-2} + n 0.5^{2n} \\ &= 0.5n(n-1) \frac{0.5^{2n}}{0.5^2} + n 0.5^{2n} \\ &= 2n(n-1) 0.5^{2n} + n 0.5^{2n} \\ &= (2n^2 - 2n + n) 0.5^{2n} = (2n^2 - n) 0.5^{2n} \\ &= n(2n-1) 0.5^{2n} = \frac{2n(2n-1)}{2} 0.5^{2n} \\ &= \frac{2n!}{(2n-2)!2!} 0.5^{2n} = P\left(\sum_{i=1}^n R_i = 2\right) \end{aligned}$$



## B Zusätzliche Tabellen und Abbildungen

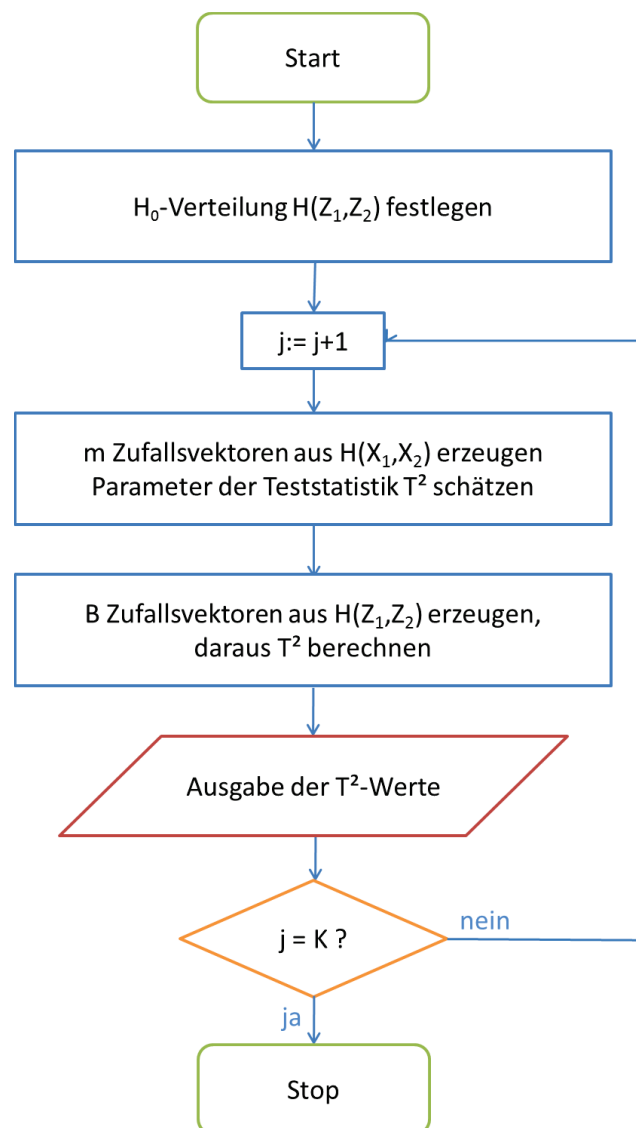


Abbildung B.1: Flussdiagramm zur Simulation der Grenzwerte

Schritt	Beschreibung
1	$X_i = (X_{i1}, X_{i2})$ umformen zu $(U_{i1}, U_{i2})$ $r(U_{i1}, U_{i2})$ und $e(U_{i1}, U_{i2})$ berechnen
2	$R_i$ aus $[r(U_{i1}, U_{i2}), e(U_{i1}, U_{i2})]$ ableiten Summieren der $R_i$ zur Teststatistik $R_n$

Tabelle B.1: Zweistufiges Vorgehen zur Berechnung der Teststatistik

	ARL			SDRL			OC		
$\tau_0$	-0.1	$\tau_0$	+0.1	-0.1	$\tau_0$	+0.1	-0.1	$\tau_0$	+0.1
0.1	22.03	36.71	21.27	22.01	36.45	20.71	0.955	0.9728	0.953
0.2	17.83	36.39	17.13	17.39	36.24	16.78	0.944	0.9725	0.942
0.3	14.35	36.63	13.44	13.81	36.04	13.10	0.930	0.9727	0.926
0.4	11.27	36.54	9.97	10.95	36.17	9.52	0.911	0.9726	0.900
0.5	8.51	36.55	7.04	8.00	36.58	6.48	0.882	0.9726	0.858
0.6	5.98	36.60	4.33	5.49	36.26	3.85	0.833	0.9727	0.769
0.7	3.84	36.47	2.22	3.30	36.45	1.64	0.740	0.9726	0.549
0.8	2.17	36.64	1.09	1.57	36.25	0.31	0.538	0.9727	0.080
0.9	1.15	36.49	1.00	0.43	35.69	0.00	0.133	0.9726	0.000
E[.]	-	36.58	-	-	-	-	-	0.973	-

Tabelle B.2: Kennwerte der entwickelten Kontrollkarte für 20.000 simulierte Lauflängen unter Voraussetzung einer Claytoncopula

	ARL			SDRL			OC		
$\tau_0$	-0.1	$\tau_0$	+0.1	-0.1	$\tau_0$	+0.1	-0.1	$\tau_0$	+0.1
0.1	22.00	36.47	22.76	22.00	35.64	22.14	0.955	0.973	0.956
0.2	18.00	36.54	19.69	17.32	35.33	19.23	0.944	0.973	0.949
0.3	14.70	36.59	16.36	14.20	35.92	15.88	0.932	0.973	0.939
0.4	11.73	36.47	12.46	11.34	35.37	11.88	0.915	0.973	0.920
0.5	8.77	36.57	8.82	8.27	35.12	8.17	0.886	0.973	0.887
0.6	6.37	36.50	5.36	5.85	35.00	4.84	0.843	0.973	0.814
0.7	3.99	36.54	2.59	3.43	35.04	2.03	0.750	0.973	0.614
0.8	2.21	36.41	1.13	1.65	34.63	0.38	0.547	0.973	0.114
0.9	1.16	36.46	1.00	0.43	34.82	0.00	0.139	0.973	0.000
E[.]	-	36.58	-	-	-	-	-	0.973	-

Tabelle B.3: Kennwerte der entwickelten Kontrollkarte für 20.000 simulierte Lauflängen unter Voraussetzung einer Gumbelcopula

$n$	$p_1$	$p_2$	LCL	UCL	Niveau	$n$	$p_1$	$p_2$	LCL	UCL	Niveau
20	0.37	0.37	13	27	.04990	46	0.32	0.32	35	55	.04981
21	0.54	0.39	10	26	.04952	47	0.16	0.68	60	79	.04978
22	0.26	0.33	17	30	.04980	48	0.15	0.43	51	69	.04975
23	0.12	0.66	29	41	.04983	49	0.36	0.36	37	59	.04989
24	0.11	0.67	31	43	.04931	50	0.49	0.49	36	62	.04964
25	0.48	0.48	16	34	.04973	51	0.37	0.38	39	62	.04950
26	0.44	0.38	16	33	.04975	52	0.51	0.43	34	60	.04989
27	0.52	0.31	13	30	.04966	53	0.21	0.64	63	85	.04981
28	0.14	0.58	33	47	.04930	54	0.23	0.72	67	90	.04965
29	0.13	0.59	35	49	.04934	55	0.32	0.32	43	65	.04995
30	0.42	0.30	18	35	.04954	56	0.28	0.52	56	80	.04992
31	0.38	0.38	22	40	.04947	57	0.19	0.57	66	88	.04953
32	0.24	0.48	31	48	.04968	58	0.35	0.43	49	74	.04980
33	0.13	0.61	41	56	.04978	59	0.43	0.35	40	65	.04980
34	0.49	0.37	20	40	.04940	60	0.35	0.49	53	79	.04987
35	0.19	0.80	47	64	.04986	61	0.49	0.35	37	63	.04987
36	0.40	0.40	26	46	.04967	62	0.29	0.35	51	74	.04973
37	0.39	0.39	27	47	.04993	63	0.34	0.34	48	72	.04991
38	0.31	0.62	39	60	.04984	64	0.45	0.45	47	75	.04993
39	0.37	0.37	29	49	.04996	65	0.23	0.76	82	107	.04967
40	0.41	0.55	34	57	.04978	66	0.20	0.78	87	111	.04967
41	0.49	0.49	29	53	.04976	67	0.15	0.41	70	91	.04959
42	0.34	0.51	38	60	.04953	68	0.36	0.43	56	83	.04961
43	0.51	0.34	24	46	.04953	69	0.36	0.36	53	79	.04979
44	0.44	0.31	27	48	.04945	70	0.50	0.50	52	82	.04980
45	0.43	0.50	35	59	.04951						

Tabelle B.4: Auf das Niveau  $\alpha = 0.05$  optimierte  $p_1$  und  $p_2$  (jeweils auf zwei Nachkommastellen genau) für verschiedene Stichprobengrößen  $n$  mit Angabe des erreichten Niveaus  $\alpha$





# Literaturverzeichnis

- DIN 1319-1. Grundlagen der Meßtechnik - Teil 1: Grundbegriffe. 1995.
- F. Aparasi, J. Jabaloyes, und A. Carrion. Statistical Properties of the |S| Multivariate Control Chart. *Communications in Statistics - Theory and Methods* 28 (11), S.2671-2686, 1999.
- D. Brigo, A. Pallavicini, und R. Torresetti. *Credit Models and the Crisis: A Journey into CDOs, Copulas, Correlations and Dynamic Models*. Wiley & Sons, 2010.
- B.E. Brodski, H.I. Penikas, und I.A. Safaryan. Copula structural shift identification. *BASIC RESEARCH PROGRAM WORKING PAPERS, Series: Financial Economics. 05FE2012*, 2012.
- R.A.J. Campbell, C.S. Forbes, K.G. Koedijk, und P. Kofman. Increasing correlations or just fat tails? *Journal of empirical finance*, Vol. 15, S.287-309, 2008.
- C.C. Caprani, A. González, P.H. Rattigan, und E.J. OBrien. Assessment dynamic ratio for traffic loading on bridges. *Structure and Infrastructure Engineering*, 8(3), S.295-304, 2012.
- P. Cizek, W. Härdle, und R. Weron. *Statistical tools for finance and insurance*. Springer, 2005.
- P. H. A. J. M. de Waal, D. J. und van Gelder. Modelling of extreme wave heights and periods through copulas. *Extremes*, Vol. 8, No. 4, S.345-356, 2005.
- W. E. Deming. On probability as a basis for action. *The American Statistician*, Vol. 29, No. 4, S.146-152, 1975.
- C. Donnelly und P. Embrechts. The devil is in the tails: actuarial mathematics and the subprime mortgage crisis. *ASTIN Bulletin*, Vol. 40, No. 1, S.1-33, 2010.
- H. Dresig und F. Holzweißig. *Maschinendynamik, 10. Auflage*. Springer, 2011.
- P. Embrechts, A. McNeil, und D. Straumann. Correlation and Dependence in Risk Management: Properties and Pitfalls. *Risk Management . Value at Risk and Beyond*, S.146-152, 2010.
- EMPA. Geregelte Dämpfung von Schrägseilbrücken (Stand: 10.07.2012). [http://www.empa.ch/plugin/template/empa/\\*/2956](http://www.empa.ch/plugin/template/empa/*/2956), 2012.
- L. Fahrmeir, A. Hamerle, und G. Tutz. *Multivariate statistische Verfahren, 2. überarbeitete Auflage*. Walter de Gruyter, 1996.

- C. Genest und A.-C. Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, No.12, S.347-368, 2007.
- C. Genest und J. MacKay. The Joy of Copulas : Bivariate Distributions With Uniform Marginals. *The American Statistician*, Vol. 40, No. 4, S.280-283, 1986.
- W. H. Greene. *Econometric Analysis*, 7 ed. Pearson, 2012.
- J. Hartung, B. Epelt, und K.-H. Klösener. *Multivariate Statistik . Lehr- und Handbuch der angewandten Statistik*, 14. unwesentlich veränderte Auflage. Oldenbourg, 2005.
- F. Hofmann. Grundlagen der Ultraschall-Durchflussmessung für den industriellen Einsatz. *KROHNE Messtechnik GmbH & Co. KG, Duisburg*, 2000.
- H. Hotelling. *Multivariate Quality Control. Techniques of Statistical Analysis*. McGraw-Hill, 1947.
- L. Jaupi. *Multivariate Control Charts for Complex Processes*. S.125-136 in: C. Lauro, J. Antoch, V. E. Vinzi, G. Saporta (2002): *Multivariate Total Quality Control, Foundation and Recent Advances*. Physica-Verlag, 2002.
- H. Joe. *Multivariate Models and Dependence Concepts (Monographs on Statistics and Applied Probability)*. Chapman & Hall/CRC Mathematics Series. Chapman & Hall/CRC, 2001.
- W. Krämer und M. van Kampen. A simple nonparametric test for structural change in joint tail probabilities. *Economics Letters* 110, S.245-247, 2011.
- F. Lindskog. Modelling Dependence with Copulas and Applications to Risk Management. *Department of Mathematics · ETH, Zürich*, 2000.
- F. Lindskog, A. McNeil, und U. Schmock. Kendall's Tau for Elliptical Distributions. *Department of Mathematics · ETH, Zürich*, 2001.
- K. V. Mardia, J. T. Kent, und J. M. Bibby. *Multivariate Analysis*. Academic Press Limited, 1995.
- G. Mehlhorn. *Handbuch Brücken*. Springer, 2007.
- O. Melsheimer, G. Sebastiani, S. Wenzel, A. E. Tekkaya, J. Kunert, A. Brosis, und L. Kwiatkowski. Aufbereitung von optischen Messdaten zur Analyse der asymmetrischen inkrementellen Blechumformung (AIBU). *Discussion papers SFB 823*, 52/11, 2011.
- D. Meneguzzo und W. Vecchiato. Copula sensitivity in collateralized debt obligations and basket default swaps. *Journal of Futures Markets* 24, Vol. 1, S.37-70, 2003.
- A. Meucci. A New Breed of Copulas for Risk and Portfolio Management. *Risk* 24, Vol. 9, S.122-126, 2011.

- H. J. Mittag. *Qualitätsregelkarten*. Hanser, 1993.
- D. C. Montgomery. *Statistical Process Control. A modern introduction, 6 ed.* International Student Version, John Wiley & Sons, 2009.
- R. B. Nelson. *An Introduction to Copulas (Lecture Notes in Statistics)*. Springer, 2006.
- A. Onken, S. Grünewälder, M.H. Munk, und K. Obermayer. Analyzing Short-Term Noise Dependencies of Spike-Counts in Macaque Prefrontal Cortex Using Copulas and the Flashlight Transformation. *PLoS Computational Biology* 5 (11), e1000577, 2009.
- G. M. Oyeyemi. Principal Component Chart for Multivariate Statistical Process Control. *The Online Journal of Science and Technology, Vol. 1, Issue 2*, 2011.
- E. S. Page. Continuous inspection schemes. *Biometrika* 41, S.100-115, 1954.
- D. Qu. Pricing Basket Options With Skew. *Wilmott Magazine (July)*, 2005.
- S. W. Roberts. Control-charts-tests based on geometric moving averages. *Technometrics* 1, S.239-250, 1959.
- T. P. Ryan. *Statistical Methods for Quality Improvement (Wiley Series in Probability and Statistics)*. John Wiley & Sons, 2011.
- F. Salmon. Recipe for Disaster: The Formula That Killed Wall Street. *Wired*, 2/23/2009, 2009.
- G. Salvadori und C. De Michele. On the Use of Copulas in Hydrology: Theory and Practice. *Journal of Hydrological Engineering, Vol. 12, S.369-381*, 2007.
- C. Savu. *Modellierung multivariater Abhängigkeitsstrukturen auf Finanzmärkten mit archimedischen und hierarchischen archimedischen Copulas*. Logos · Berlin, 2007.
- C. Schölzel und P. Friederichs. Multivariate nonnormally distributed random variables in climate research . Introduction to the copula approach. *Nonlinear Processes in Geophysics, 15, S.761-772*, 2008.
- W. A. Shewhart. *Economic Control of Quality of Manufactured Product*. ASQ Quality Press, 1931.
- S. Srinivas. *Copula based dependence and probabilistic load modeling of highway bridges*. Ph. D. Thesis, Indian Institute of Technology Madras, Chennai, 2006.
- S. Srinivas, D. Menon, und A. Meher Prasad. Multivariate simulation and multimodal dependence modeling of vehicle axle weights with copulas. *Journal of Transportation Engineering, ASCE, Vol. 132, Issue 12, S.945-955*, 2006.
- S. Sriramula, D. Menon, und A. M. Prasad. Axle load variations and vehicle growth projection models for safety assessment of transportation structures. *Transport, 22:1, S.31-37*, 2005.

- J.H. Steiger. Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, Vol. 87, S.245-251, 1980.
- R. Taylor und R. Phillips. *The Philosophical Magazine: or Annals of chemistry, mathematics, astronomie, natural history, and general sciende*, Vol. IX. Richard Taylor, London, 1881.
- D. Thompson und R. Kilgore. Estimating Joint Flow Probabilities at Stream Confluences using Copulas. *Transportation Research Record*, Vol. 2262, S.200-206, 2011.
- Q. H. Tran. *Copulas im Risikomanagement von Versicherungsunternehmen*. Grin. Verlag für akademische Texte, 2009.
- C. Weihs und J. Jessenberger. *Statistische Methoden zur Qualitätssicherung und -optimierung in der Industrie*. Wiley-VCH, 1999.
- D. Wied und P. Galeano. Monitoring correlation change in a sequence of random variables. *Discussion papers SFB 823*, 12/12, 2012.
- D. Wied, W. Krämer, und H. Dehling. Testing for a change in correlation at an unknown point in time using an extended functional delta method. *Econometric Theory* 28, 2012.
- J. Yan. Multivariate Modeling with Copulas and Engineering Applications. *Springer Handbook of Engineering Statistics*, S.973-989, 2007.
- K. Yang und B. El-Haik. *Design for Six Sigma: A Roadmap for Product Development*. Mcgraw-Hill Professional, 2003.

### **Eigenständigkeitserklärung**

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken (dazu zählen auch Internetquellen) entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.

Dortmund, den 15. Februar 2013

-----