

Kendalls tau für Gemeinsame Unabhängigkeit

Jan Krohn

INHALTSVERZEICHNIS

1. Motivation und Definition	1
2. Test auf gemeinsame Unabhängigkeit	3
3. Einige Simulationsversuche	16
Anhang A. Quellcode und Ausgabe	17
Literatur	19

1. Motivation und Definition

Gegeben sei ein stochastischer Prozess $(X_i)_{i=1,\dots,n}$. Weiterhin seien die X_i identisch verteilt mit der gleichen λ^1 -Dichte. Man möchte jetzt bei einer gegebenen Realisation (x_i) von (X_i) auf gemeinsame stochastische Unabhängigkeit der X_i testen. Hierzu kann man Kendalls τ verwenden. Die Basis für den Seminarvortrag bildet der Artikel [FGH].

DEFINITION 1.1 (Generalvoraussetzung). Im folgenden seien (bis auf wenige Ausnahmen, auf die explizit hingewiesen wird) alle stochastischen Prozesse (X_i) definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ und identisch verteilt mit der gleichen λ^1 -Dichte, n bezeichne jeweils die Länge des Prozesses.

DEFINITION 1.2. Sei (X_i) ein stochastischer Prozess wie oben.

- (1) Bezeichne mit $(X_{(i)})$ die Folge der zugehörigen Ordnungsstatistiken, für deren Realisationen gilt $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$.
- (2) Definiere den **Rang** der Zufallsvariablen X_i als $R_i := \sum_{j=1}^n \mathbf{1}_{(-\infty, X_i]}(X_j)$.

BEISPIEL 1.3. Sei $n = 5$ und

$$(x_i) = (6, 3, 10, 4, 2)$$

eine Realisation des Prozesses (X_i) . Dann ist

$$(x_{(i)}) = (2, 3, 4, 6, 10),$$

und somit

$$(R_i(\omega)) = (4, 2, 5, 3, 1)$$

die zugehörige Realisation der Ränge. Vergleiche auch [Har].

LEMMA 1.4. *Sei (X_i) ein Prozess wie in Definition 1.1, der zusätzlich unabhängig ist. Dann ist (R_1, \dots, R_n) P-f.s. eine Permutation von $(1, \dots, n)$ und besitzt eine Gleichverteilung auf allen Permutationen. Siehe [Háj].*

DEFINITION 1.5. Sei (X_i) ein Prozess wie in Definition 1.1. Setze $R_{n+1} := R_1$.

(1) Definiere

$$N^\circ := \sum_{i=1}^{n-1} \sum_{j=i+1}^n (\mathbf{1}_{\{R_i < R_j, R_{i+1} > R_{j+1}\}} + \mathbf{1}_{\{R_i > R_j, R_{i+1} < R_{j+1}\}}) \quad (1.1)$$

und

$$N := \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} (\mathbf{1}_{\{R_i < R_j, R_{i+1} > R_{j+1}\}} + \mathbf{1}_{\{R_i > R_j, R_{i+1} < R_{j+1}\}}). \quad (1.2)$$

(2) Zwei modifizierte Versionen von **Kendalls tau** werden definiert als

$$\tau_n^\circ := 1 - \frac{2N^\circ}{\binom{n}{2}} = 1 - \frac{4N^\circ}{n(n-1)}. \quad (1.3)$$

bzw.

$$\tau_n := 1 - \frac{2N}{\binom{n-1}{2}} = 1 - \frac{4N}{(n-1)(n-2)}. \quad (1.4)$$

Bei Verwendung von N° und τ_n° sprechen wir vom **zirkularen Fall**, bei Verwendung von N und τ_n vom **nichtzirkularen Fall**.

BEMERKUNG 1.6. Die Zufallsvariablen N und N° geben die Anzahl der Inversionen an, wobei ein Paar $((R_i, R_{i+1}), (R_j, R_{j+1}))$ genau dann eine **Inversion** heißt, wenn entweder $R_i < R_j, R_{i+1} > R_{j+1}$ oder $R_i > R_j, R_{i+1} < R_{j+1}$. Im nicht-zirkularen Fall werden die Paare mit $1 \leq i, j \leq n-1$ betrachtet, im zirkularen Fall die Paare mit $1 \leq i, j \leq n$.

LEMMA 1.7. *Es gilt*

$$N^\circ = \sum_{i=1}^n \sum_{j=1}^n \mathbf{1}_{\{R_i < R_j, R_{i+1} > R_{j+1}\}} \quad (1.5)$$

und

$$N = \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \mathbf{1}_{\{R_i < R_j, R_{i+1} > R_{j+1}\}}. \quad (1.6)$$

PROOF. Im zirkularen Fall ergibt sich

$$\begin{aligned}
N^\circ &= \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(\mathbf{1}_{\{R_i < R_j, R_{i+1} > R_{j+1}\}} + \mathbf{1}_{\{R_i > R_j, R_{i+1} < R_{j+1}\}} \right) \\
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbf{1}_{\{R_i < R_j, R_{i+1} > R_{j+1}\}} + \sum_{j=1}^{n-1} \sum_{i=j+1}^n \mathbf{1}_{\{R_i < R_j, R_{i+1} > R_{j+1}\}} \\
&= \sum_{i=1}^{n-1} \sum_{j=i}^n \mathbf{1}_{\{R_i < R_j, R_{i+1} > R_{j+1}\}} + \sum_{j=1}^{n-1} \sum_{i=j+1}^n \mathbf{1}_{\{R_i < R_j, R_{i+1} > R_{j+1}\}} \\
&= \sum_{i \leq j} \mathbf{1}_{\{R_i < R_j, R_{i+1} > R_{j+1}\}} + \sum_{i > j} \mathbf{1}_{\{R_i < R_j, R_{i+1} > R_{j+1}\}} \\
&= \sum_{i=1}^n \sum_{j=1}^n \mathbf{1}_{\{R_i < R_j, R_{i+1} > R_{j+1}\}}.
\end{aligned}$$

Den nichtzirkularen Fall beweist man analog. \square

BEISPIEL 1.8. Für $(R_i(\omega)) = (4, 2, 5, 3, 1)$ aus Beispiel 1.3 gibt es die Inversionen $((2, 5), (4, 2)), ((2, 5), (5, 3)), ((2, 5), (3, 1)), ((1, 4), (4, 2)), ((1, 4), (5, 3))$ und $((1, 4), (3, 1))$, also $N^\circ = 6$, $N = 3$, $\tau_5^\circ = 1 - 4 \cdot 6 / 5 \cdot 4 = -1/5$ und $\tau_5 = 1 - 4 \cdot 3 / 4 \cdot 3 = 0$.

Ziel: Benutze unter der Nullhypothese der gemeinsamen Unabhängigkeit von X_1, \dots, X_n das modifizierte Kendalls tau als Teststatistik gegen diese Hypothese.

2. Test auf gemeinsame Unabhängigkeit

2.1. Erwartungswert und Varianz von Kendalls tau.

DEFINITION 2.1. Definiere die Teststatistiken

$$T_n := \frac{\tau_n - E\tau_n}{\sqrt{\frac{\text{Var}\tau_n}{n}}}$$

und

$$T_n^\circ := \frac{\tau_n^\circ - E\tau_n^\circ}{\sqrt{\frac{\text{Var}\tau_n^\circ}{n}}}$$

Zum Test der Nullhypothese H_0 der gemeinsamen Unabhängigkeit der X_i gegen die Alternative H_1 der positiven Abhängigkeit weise H_0 zurück, falls $T_n > t_{n,\alpha}$ bzw. $T_n^\circ > t_{n,\alpha}^\circ$ für $t_{n,\alpha}$ bzw. $t_{n,\alpha}^\circ$ so gewählt, dass unter H_0 , $P(T_n > t_{n,\alpha}) = \alpha$ bzw. $P(T_n^\circ > t_{n,\alpha}^\circ) = \alpha$ gilt.

Der Artikel [FGH] lässt es unklar, was genau die positive Abhängigkeit ist. Anders als in den üblichen Fällen ist hier H_1 nicht das Komplement zu H_0 . Die Ähnlichkeit zum Einstichproben- t -Test ist erkennbar. Unter der Nullhypothese

H_0 sei also $\mu_0 := E\tau_n$ und $\mu_0^\circ := E\tau_n^\circ$. Es ist vernünftig anzunehmen, dass unter der Alternativhypothese die Wahrscheinlichkeit, dass wenig Inversionen vorkommen, größer ist als bei einem unabhängigen Prozess. Damit ist analog zum t -Test eine vernünftige Formulierung der Alternative etwa gegeben durch $E\tau_n > \mu_0$ bzw. $E\tau_n^\circ > \mu_0^\circ$.

Um diese Prozedur durchzuführen, benötigen wir Erwartungswert und Varianz von τ_n° und τ_n .

LEMMA 2.2. *Sei (X_i) ein Prozess wie in Definition 1.1 und $n \geq 8$. Unter der Nullhypothese der gemeinsamen stochastischen Unabhängigkeit von X_1, \dots, X_n gilt:*

(1)

$$EN^\circ = \frac{n(3n-1)}{12},$$

(2)

$$\text{Var } N^\circ = \frac{10n^3 - 7n^2 - 4n - 9}{360},$$

(3)

$$EN = \frac{(3n-1)(n-2)}{12},$$

(4)

$$\text{Var } N = \frac{10n^3 - 37n^2 + 27n + 74}{360}.$$

PROOF. Da bei unabhängigen X_1, \dots, X_n nach Lemma 1.4 jede Reihenfolge der R_i gleichwahrscheinlich ist, gilt für alle $1 \leq i, j \leq n$:

$$\begin{aligned} \mathbb{P}(R_i < R_j, R_{i+1} > R_{j+1}) &= \begin{cases} \mathbb{P}(R_1 < R_1, R_2 > R_2), & i = j \\ \mathbb{P}(R_1 < R_2, R_2 > R_3), & |i - j| = 1 \\ \mathbb{P}(R_1 < R_2, R_3 > R_4), & \text{sonst} \end{cases} \\ &= \begin{cases} 0 \\ \frac{1}{3} \\ \mathbb{P}(R_1 < R_2) \mathbb{P}(R_3 > R_4) = \frac{1}{4} \end{cases}. \end{aligned}$$

Hierbei können die Wahrscheinlichkeit im ersten Fall aus Tabelle 1 abgelesen werden. Der zweite Fall ist trivial.

Im zirkularen Fall errechnet sich der Erwartungswert also folgendermaßen:

$$\begin{aligned} EN^\circ &= \sum_{i=1}^n \sum_{j=1}^n \mathbb{P}(R_i < R_j, R_{i+1} > R_{j+1}) \\ &= \sum_{i=1}^n \left(2 \cdot \frac{1}{3} + (n-3) \frac{1}{4} \right) = n \left(\frac{2}{3} + \frac{n-3}{4} \right) = \frac{n(3n-1)}{12}. \end{aligned} \quad (2.1)$$

$(R_i(\omega))$	Inversionen	N	N°
$(1, 2, 3)$	$((1, 2), (3, 1)), ((2, 3), (3, 1))$	0	2
$(1, 3, 2)$	$((1, 3), (3, 2)), ((1, 3), (2, 1))$	1	2
$(2, 1, 3)$	$((2, 1), (1, 3)), ((1, 3), (3, 2))$	1	2
$(2, 3, 1)$	$((2, 3), (3, 1)), ((3, 1), (1, 2))$	1	2
$(3, 1, 2)$	$((3, 1), (1, 2)), ((3, 1), (2, 3))$	1	2
$(3, 2, 1)$	$((3, 2), (1, 3)), ((2, 1), (1, 3))$	0	2

TABELLE 1. Permutationen der Menge $\{1, 2, 3\}$

Im nichtzirkulären Fall ergibt sich:

$$\begin{aligned}
EN &= \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} P(R_i < R_j, R_{i+1} > R_{j+1}) \\
&= \sum_{i=2}^{n-2} \left(2 \cdot \frac{1}{3} + (n-4) \frac{1}{4} \right) + 2 \left(\frac{1}{3} + (n-3) \frac{1}{4} \right) \\
&= (n-3) \left(\frac{2}{3} + \frac{n-3}{4} \right) + \frac{2}{3} + \frac{n-4}{2} = \frac{(3n-1)(n-2)}{12}. \quad (2.2)
\end{aligned}$$

Berechnung der zweiten Momente:

$$\begin{aligned}
EN^{\circ 2} &= E \left(\sum_{i=1}^n \sum_{j=1}^n \mathbf{1}_{\{R_i < R_j, R_{i+1} > R_{j+1}\}} \right)^2 \\
&= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n E \mathbf{1}_{\{R_i < R_j, R_{i+1} > R_{j+1}, R_k < R_l, R_{k+1} > R_{l+1}\}} \\
&= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n P(R_i < R_j, R_{i+1} > R_{j+1}, R_k < R_l, R_{k+1} > R_{l+1}).
\end{aligned}$$

Eine analoge Formel ergibt sich für EN mit dem Unterschied, dass dort alle Summen bis $n-1$ laufen und nicht bis n . Im folgenden werden beide Fälle jeweils hintereinander betrachtet. Die Summanden können in vier Klassen unterteilt werden, die erste mit $i = k$ und $j = l$, die zweite mit entweder $i = k$ oder $j = l$, die dritte mit $i \neq k, j \neq l$ und entweder $i = l$ oder $k = j$ und die vierte mit i, j, k, l paarweise verschieden. Beachte, dass alle anderen Summanden Wahrscheinlichkeiten der Form $P(R_i < R_j, R_i > R_j) = 0$ oder $P(R_i < R_i, R_{i+1} > R_{i+1}) = 0$ sind. Damit ergibt sich:

(1, 2, 3, 4)	(1, 2, 4, 3)	(1, 4, 2, 3)	(4, 1, 2, 3)
(1, 3, 2, 4)	(1, 3, 4, 2)	(1, 4, 3, 2)	(4, 1, 3, 2)
(2, 1, 3, 4)	(2, 1, 4, 3)	(2, 4, 1, 3)	(4, 2, 1, 3)
(2, 3, 1, 4)	(2, 3, 4, 1)	(2, 4, 3, 1)	(4, 2, 3, 1)
(3, 1, 2, 4)	(3, 1, 4, 2)	(3, 4, 1, 2)	(4, 3, 1, 2)
(3, 2, 1, 4)	(3, 2, 4, 1)	(3, 4, 2, 1)	(4, 3, 2, 1)

TABELLE 2. Permutationen der Menge $\{1, 2, 3, 4\}$

$$EN^{\circ 2} = \sum_i \sum_j P(R_i < R_j, R_{i+1} > R_{j+1}) \quad (2.3)$$

$$+ 2 \sum_i \sum_j \sum_{l \neq j} P(R_i < R_j, R_{i+1} > R_{j+1}, R_i < R_l, R_{i+1} > R_{l+1}) \quad (2.4)$$

$$+ 2 \sum_i \sum_{k \neq i} \sum_{j \neq k} P(R_i < R_j, R_{i+1} > R_{j+1}, R_k < R_i, R_{k+1} > R_{i+1}) \quad (2.5)$$

$$+ \sum_i \sum_{j \neq i} \sum_{k \neq i, j} \sum_{l \neq i, j, k} P(R_i < R_j, R_{i+1} > R_{j+1}, R_k < R_l, R_{k+1} > R_{l+1}). \quad (2.6)$$

Der Term (2.3) ist EN° , die anderen werden analog ausgerechnet.

Berechnung von (2.4): Für die einzelnen Summanden können sich folgende Fälle ergeben, wobei in Klammern zusammengefasste Indizes die Fälle bezeichnen, dass die entsprechenden Werte direkt aufeinander folgen, und alle einzelnen Werte und Klammern paarweise mindestens den Abstand 1 voneinander haben. Das kann auch modulo n sein. So werden beispielsweise mit (i, j, l) die Fälle $(i, j, j) = (1, 2, 3), \dots, (i, j, j) = (n-2, n-1, n)$ erfasst, und im zirkularen Fall zusätzlich $(i, j, j) = (n-1, n, 1)$ und $(i, j, j) = (n, 1, 2)$.

Die Wahrscheinlichkeiten können nach Lemma 1.4 in den einfachen Fällen noch durch Ablesen aus den Tabellen berechnet werden, für schwierige Fälle gibt es im Anhang A.1 ein MAPLE-Programm, das alle Permutationen untersucht und die passenden zählt.

- (1) (i, j, l) : $P(R_1 < R_2, R_2 > R_3, R_1 < R_3, R_2 > R_4) = 1/8$.
- (2) (j, i, l) : $P(R_2 < R_1, R_3 > R_2, R_2 < R_3, R_3 > R_4) = 5/24$.
- (3) (i, l, j) : $P(R_1 < R_3, R_2 > R_4, R_1 < R_2, R_2 > R_3) = 1/8$.
- (4) (j, l, i) : $P(R_3 < R_1, R_4 > R_2, R_3 < R_2, R_4 > R_3) = 1/8$.
- (5) (l, i, j) : $P(R_2 < R_3, R_3 > R_4, R_2 < R_1, R_3 > R_2) = 5/24$.
- (6) (l, j, i) : $P(R_3 < R_2, R_4 > R_3, R_3 < R_1, R_4 > R_2) = 1/8$.

Im zirkularen Fall kommen diese Summanden jeweils n mal vor, im nichtzirkularen Fall jeweils $n-3$ mal.

- (7) $(i, j), l$: $P(R_1 < R_2, R_2 > R_3, R_1 < R_4, R_2 > R_5) = 3/20$.

- (8) $(j, i), l: P(R_2 < R_1, R_3 > R_2, R_2 < R_4, R_3 > R_5) = 3/20.$
(9) $(i, l), j: P(R_1 < R_4, R_2 > R_5, R_1 < R_2, R_2 > R_3) = 3/20$
(10) $(l, i), j: P(R_2 < R_4, R_3 > R_5, R_2 < R_1, R_3 > R_2) = 3/20.$
(11) $(j, l), i: P(R_4 < R_1, R_5 > R_2, R_4 < R_2, R_5 > R_3) = 11/120.$
(12) $(l, j), i: P(R_4 < R_2, R_5 > R_3, R_4 < R_1, R_5 > R_2) = 11/120.$

Im zirkularen Fall kommen diese Summanden jeweils $n(n-4)$ mal vor, im nichtzirkularen Fall jeweils $(n-3)(n-4)$ mal.

- (13) $i, j, l: P(R_1 < R_3, R_2 > R_4, R_1 < R_5, R_2 > R_6) = 1/9.$

Es gibt hier $n(2(n-5) + (n-5)(n-6)) = n(n-4)(n-5)$ Summanden im zirkularen Fall und

$$2(2(n-5) + (n-5)(n-6)) + 2(2(n-6) + (n-6)(n-7)) + \\ 2(n-5 + 2(n-6) + (n-7)^2) + (n-7)(4(n-6) + (n-8)(n-7)) = \\ (n-3)(n-4)(n-5)$$

im nichtzirkularen Fall.

Insgesamt ergibt sich im zirkularen Fall für den Term (2.4):

$$2n \left(\frac{11}{12} + (n-4) \frac{47}{60} + (n-4)(n-5) \frac{1}{9} \right) = \frac{n(20n^2 - 39n + 1)}{90} \quad (2.7)$$

und im nichtzirkularen Fall

$$2(n-3) \left(\frac{11}{12} + (n-4) \frac{47}{60} + (n-4)(n-5) \frac{1}{9} \right) = \frac{(n-3)(20n^2 - 39n + 1)}{90}. \quad (2.8)$$

Berechnung von (2.5) erfolgt nach der gleichen Fallunterscheidung, substituieren nur l durch k . Die Anzahl der Summanden ist jeweils ebenfalls die gleiche.

- (1) $P(R_1 < R_2, R_2 > R_3, R_3 < R_1, R_4 > R_2) = 1/24.$
(2) $P(R_1 < R_3, R_2 > R_4, R_2 < R_1, R_3 > R_2) = 1/24.$
(3) $P(R_2 < R_1, R_3 > R_2, R_3 < R_2, R_4 > R_3) = 0.$
(4) $P(R_3 < R_1, R_4 > R_2, R_2 < R_3, R_3 > R_4) = 1/24.$
(5) $P(R_2 < R_3, R_3 > R_4, R_1 < R_2, R_2 > R_3) = 0.$
(6) $P(R_3 < R_2, R_4 > R_3, R_1 < R_3, R_2 > R_4) = 1/24.$
(7) $P(R_1 < R_2, R_2 > R_3, R_4 < R_1, R_5 > R_2) = 1/40.$
(8) $P(R_2 < R_1, R_3 > R_2, R_4 < R_2, R_5 > R_3) = 1/40.$
(9) $P(R_1 < R_4, R_2 > R_5, R_2 < R_1, R_3 > R_2) = 1/40.$
(10) $P(R_2 < R_4, R_3 > R_5, R_1 < R_2, R_2 > R_3) = 1/40.$
(11) $P(R_4 < R_1, R_5 > R_2, R_2 < R_4, R_3 > R_5) = 1/20.$
(12) $P(R_4 < R_2, R_5 > R_3, R_1 < R_4, R_2 > R_5) = 1/20.$
(13) $P(R_1 < R_3, R_2 > R_4, R_5 < R_1, R_6 > R_2) = 1/36.$

Für den Term (2.5) ergibt sich also im zirkularen Fall

$$2n \left(\frac{1}{6} + (n-4) \frac{1}{5} + (n-4)(n-5) \frac{1}{36} \right) = \frac{n(n+1)(5n-14)}{90} \quad (2.9)$$

und im nichtzirkularen Fall

$$2(n-3) \left(\frac{1}{6} + (n-4) \frac{1}{5} + (n-4)(n-5) \frac{1}{36} \right) = \frac{(n-3)(n+1)(5n-14)}{90}. \quad (2.10)$$

Berechnung von (2.6) nach den gewohnten Fallunterscheidungen:

(1) (i, j, k, l) oder (k, l, i, j) :

$$P(R_1 < R_2, R_2 > R_3, R_3 < R_4, R_4 > R_5) = 2/15.$$

(2) (j, i, k, l) oder (l, k, j, i) :

$$P(R_2 < R_1, R_3 > R_2, R_3 < R_4, R_4 > R_5) = 11/120.$$

(3) (i, j, l, k) oder (k, l, j, i) :

$$P(R_2 < R_1, R_3 > R_2, R_4 < R_3, R_5 > R_4) = 11/120.$$

(4) (j, i, l, k) oder (l, k, j, i) :

$$P(R_2 < R_1, R_3 > R_2, R_4 < R_3, R_5 > R_4) = 2/15.$$

(5) (i, k, l, j) oder (k, i, j, l) :

$$P(R_1 < R_4, R_2 > R_5, R_2 < R_3, R_3 > R_4) = 1/20.$$

(6) (j, k, l, i) oder (l, i, j, k) :

$$P(R_4 < R_1, R_5 > R_2, R_2 < R_3, R_3 > R_4) = 2/15.$$

(7) (i, l, k, j) oder (k, j, i, l) :

$$P(R_1 < R_4, R_2 > R_5, R_3 < R_2, R_4 > R_3) = 2/15.$$

(8) (j, l, k, i) oder (l, j, i, k) :

$$P(R_4 < R_1, R_5 > R_2, R_3 < R_2, R_4 > R_3) = 1/20.$$

(9) (i, k, j, l) oder (k, i, l, j) :

$$P(R_1 < R_3, R_2 > R_4, R_2 < R_4, R_3 > R_5) = 0.$$

(10) (j, k, i, l) oder (l, i, k, j) :

$$P(R_3 < R_1, R_4 > R_2, R_2 < R_4, R_3 > R_5) = 1/12.$$

(11) (i, l, j, k) oder (k, j, l, i) :

$$P(R_1 < R_3, R_2 > R_4, R_4 < R_2, R_5 > R_3) = 1/12.$$

(12) (j, l, i, k) oder (l, j, k, i) :

$$P(R_3 < R_1, R_4 > R_2, R_4 < R_2, R_5 > R_3) = 0.$$

In diesen Fällen ist die Anzahl der Summanden im zirkularen Fall jeweils $2n$ und $2(n-4)$ im nichtzirkularen Fall.

(13) $(i, j, k), l$ oder $(k, l, i), j$:

$$P(R_1 < R_2, R_2 > R_3, R_3 < R_5, R_4 > R_6) = 5/48.$$

(14) $(i, k, j), l$ oder $(k, i, l), j$:

$$P(R_1 < R_3, R_2 > R_4, R_2 < R_5, R_3 > R_6) = 1/18.$$

(15) $(j, i, k), l$ oder $(l, k, i), j$:

$$P(R_2 < R_1, R_3 > R_4, R_3 < R_5, R_4 > R_6) = 1/16.$$

(16) $(j, k, i), l$ oder $(l, i, k), j$:

$$P(R_3 < R_1, R_4 > R_2, R_2 < R_5, R_3 > R_6) = 1/18.$$

(17) $(k, i, j), l$ oder $(i, k, l), j$:

$$P(R_2 < R_3, R_3 > R_4, R_1 < R_5, R_2 > R_6) = 1/16.$$

(18) $(k, j, i), l$ oder $(i, l, k), j$:

$$P(R_3 < R_2, R_4 > R_3, R_1 < R_5, R_2 > R_6) = 5/48.$$

(19) $(i, j, l), k$ oder $(k, l, j), i$:

$$P(R_1 < R_2, R_2 > R_3, R_3 < R_5, R_4 > R_6) = 1/16.$$

(20) $(i, l, j), k$ oder $(k, j, l), i$:

$$P(R_1 < R_3, R_2 > R_4, R_2 < R_5, R_3 > R_6) = 1/18.$$

(21) $(j, i, l), k$ oder $(l, k, j), i$:

$$P(R_2 < R_1, R_3 > R_4, R_3 < R_5, R_4 > R_6) = 5/48.$$

(22) $(j, l, i), k$ oder $(l, j, k), i$:

$$P(R_3 < R_1, R_4 > R_2, R_2 < R_5, R_3 > R_6) = 1/18.$$

(23) $(l, i, j), k$ oder $(j, k, l), i$:

$$P(R_2 < R_3, R_3 > R_4, R_1 < R_5, R_2 > R_6) = 5/48.$$

(24) $(l, j, i), k$ oder $(j, l, k), i$:

$$P(R_3 < R_2, R_4 > R_3, R_1 < R_5, R_2 > R_6) = 1/16.$$

In den Fällen (13) bis (24) gibt es jeweils $2n(n-5)$ Summanden im zirkularen Fall und $2(2(n-5) + (n-5)(n-6)) = 2(n-4)(n-5)$ im nichtzirkularen Fall.

(25) $(i, j), (k, l)$ oder $(j, i), (k, l)$:

$$P(R_1 < R_2, R_2 > R_3, R_4 < R_5, R_5 > R_6) = 1/9.$$

(26) $(i, j), (l, k)$ oder $(j, i), (k, l)$:

$$P(R_1 < R_2, R_2 > R_3, R_4 < R_5, R_5 > R_6) = 1/9.$$

(27) $(i, k), (j, l)$ oder $(k, i), (l, j)$:

$$P(R_1 < R_4, R_2 > R_5, R_2 < R_5, R_3 > R_6) = 0.$$

(28) $(i, k), (l, j)$ oder $(k, i), (j, l)$:

$$P(R_1 < R_5, R_2 > R_6, R_2 < R_4, R_3 > R_5) = 1/36.$$

(29) $(i, l), (j, k)$ oder $(k, j), (l, i)$:

$$P(R_1 < R_4, R_2 > R_5, R_5 < R_2, R_6 > R_3) = 1/8.$$

(30) $(i, l), (k, j)$ oder $(k, j), (i, l)$:

$$P(R_1 < R_5, R_2 > R_6, R_4 < R_2, R_5 > R_3) = 1/9.$$

Damit erhält man in den Fällen (25) bis (30) $2n(n-5)$ Summanden im zirkularen Fall und im nichtzirkularen Fall

$$2(2(n-5) + 2(n-6) + 2(n-7) + (n-8)(n-7)) = 2(n-4)(n-5)$$

Summanden.

(31) $(i, j), k, l$ oder $(k, l), i, j$:

$$P(R_1 < R_2, R_2 > R_3, R_4 < R_6, R_5 > R_7) = 1/12.$$

(32) $(i, k), j, l$ oder $(k, i), j, l$:

$$P(R_1 < R_4, R_2 > R_5, R_2 < R_6, R_3 > R_7) = 1/24.$$

(33) $(j, i), k, l$ oder $(l, k), i, j$:

$$P(R_2 < R_1, R_3 > R_2, R_4 < R_6, R_5 > R_7) = 1/12.$$

(34) $(i, l), j, k$ oder $(k, j), i, l$:

$$P(R_1 < R_4, R_2 > R_5, R_6 < R_2, R_7 > R_3) = 1/12.$$

(35) $(l, i), j, k$ oder $(j, k), i, l$:

$$P(R_2 < R_4, R_3 > R_5, R_6 < R_1, R_7 > R_2) = 1/12.$$

(36) $(j, l), i, k$ oder $(l, j), i, k$:

$$P(R_4 < R_1, R_5 > R_2, R_6 < R_2, R_7 > R_3) = 1/24.$$

Zählen der Summanden in den Fällen (31) bis (36) ergibt im zirkularen Fall $2n(2(n-6) + (n-6)(n-7)) = 2n(n-5)(n-6)$ und

$$\begin{aligned} &2(2(2(n-6) + (n-6)(n-7)) + 2(2(n-7) + (n-7)(n-8)) + \\ &2((n-6) + 2(n-7) + (n-8)^2) + (n-8)(4(n-7) + (n-9)(n-8))) = \\ &2(n-4)(n-5)(n-6) \end{aligned}$$

im nichtzirkularen Fall.

(37) i, j, k, l :

$$P(R_1 < R_3, R_2 > R_4, R_5 < R_7, R_6 > R_8) = 1/16.$$

Es ergibt sich $n(n-5)(n-6)(n-7)$ für die Anzahl der Summanden im zirkularen Fall und

$$(n-1)(n-2)(n-3)(n-4) - 24(n-4) - 24(n-4)(n-5) - 12(n-4)(n-5) - 12(n-4)(n-5)(n-6) = (n-4)(n-5)(n-6)(n-7)$$

im nichtzirkularen Fall.

Insgesamt ergibt die Summe (2.5) also

$$n \left(\frac{59}{30} + \frac{11}{4}(n-5) + \frac{5}{6}(n-5)(n-6) + \frac{1}{16}(n-5)(n-6)(n-7) \right) = \frac{n(15n^3 - 70n^2 + 65n + 22)}{240} \quad (2.11)$$

im zirkularen Fall oder

$$(n-4) \left(\frac{59}{30} + \frac{11}{4}(n-5) + \frac{5}{6}(n-5)(n-6) + \frac{1}{16}(n-5)(n-6)(n-7) \right) = \frac{(n-4)(15n^3 - 70n^2 + 65n + 22)}{240} \quad (2.12)$$

im nichtzirkularen Fall.

Jetzt können wir $EN^{\circ 2}$ durch Aufsummieren von (2.1), (2.7), (2.9) und (2.11) berechnen, analog EN^2 durch Addieren von (2.2), (2.8), (2.10) und (2.12). Es ergibt sich:

$$\begin{aligned} EN^{\circ 2} &= \\ &= \frac{n(3n-1)}{12} + \frac{n(20n^2-39n+1)}{90} \\ &+ \frac{n(n+1)(5n-14)}{90} + \frac{n(15n^3-70n^2+65n+22)}{240} \\ &= \frac{n(45n^3-10n^2-9n-98)}{720} \end{aligned}$$

und

$$\begin{aligned} EN^2 &= \\ &= \frac{(n-2)(3n-1)}{12} + \frac{(n-3)(20n^2-39n+1)}{90} \\ &+ \frac{(n-3)(n+1)(5n-14)}{90} + \frac{(n-4)(15n^3-70n^2+65n+22)}{240} \\ &= \frac{45n^4-190n^3+231n^2-86n+168}{720}, \end{aligned}$$

also

$$\text{Var } N^\circ = EN^{\circ 2} - (EN^\circ)^2 = \frac{n(10n^2 - 7n - 49)}{360}$$

und

$$\text{Var } N = EN^2 - (EN)^2 = \frac{(n+1)(10n^2 - 47n + 74)}{360}.$$

□

SATZ 2.3. *Unter H_0 gilt für den Erwartungswert und die Varianz*

$$E\tau_n^\circ = E\tau_n = -\frac{2}{3(n-1)} = O\left(\frac{1}{n}\right),$$

$$\text{Var } \tau_n^\circ = \frac{20n^3 - 14n^2 - 98}{45n^2(n-1)^2} = \frac{4}{9n} + o\left(\frac{1}{n}\right)$$

und

$$\text{Var } \tau_n = \frac{(n+1)(20n^2 - 94n + 148)}{45(n-1)^2(n-2)^2} = \frac{4}{9n} + o\left(\frac{1}{n}\right),$$

jeweils für $n \geq 8$.

PROOF. Nach bekannten Formeln aus der Einführung in die Stochastik und Lemma 2.2 erhält man

$$E\tau_n^\circ = E\left(1 - \frac{4N^\circ}{n(n-1)}\right) = 1 - \frac{4EN^\circ}{n(n-1)} = -\frac{2}{3(n-1)},$$

$$E\tau_n = E\left(1 - \frac{4N}{(n-1)(n-2)}\right) = 1 - \frac{4EN}{(n-1)(n-2)} = -\frac{2}{3(n-1)},$$

$$\text{Var } \tau_n^\circ = \text{Var}\left(1 - \frac{4N^\circ}{n(n-1)}\right) = \frac{16\text{Var } N^\circ}{n^2(n-1)^2} = \frac{20n^3 - 14n^2 - 98}{45n^2(n-1)^2}$$

und

$$\begin{aligned} \text{Var } \tau_n &= \text{Var}\left(1 - \frac{4N}{(n-1)(n-2)}\right) \\ &= \frac{16\text{Var } N}{(n-1)^2(n-2)^2} = \frac{(n+1)(20n^3 - 94n + 148)}{45(n-1)^2(n-2)^2}. \end{aligned}$$

□

2.2. Asymptotisches Verhalten für große n .

DEFINITION 2.4. Sei (X_i) ein möglicherweise vektorwertiger Prozess mit Parametermenge \mathbb{N} oder $\{1, \dots, n\}$.

- (1) Existiert eine Zahl $m \in \mathbb{N}$, so dass X_i und X_j stochastisch unabhängig sind für alle i und j mit $|i - j| > m$, so heißt (X_i) **m -abhängig**. Siehe [Ser].
- (2) Der stochastische Prozess (X_i) heißt ***-mixing**, falls ein $M \in \mathbb{N}$ existiert und eine monoton fallende Funktion $f : \{M, M + 1, \dots\} \rightarrow \mathbb{R}$, so dass $\lim_{n \rightarrow \infty} f(n) = 0$ und für alle $m \in \mathbb{N}$ gilt

$$|\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)| \leq f(n)\mathbb{P}(A)\mathbb{P}(B)$$

für alle $n \geq M$, $A \in \sigma(X_1, \dots, X_m)$ und $B \in \sigma(X_{m+1}, \dots, X_n)$. Siehe auch [BHK].

BEMERKUNG 2.5. Die obigen Definitionen der m -Abhängigkeit und der *-mixing-Eigenschaft sind vernünftige Forderungen für stochastische Prozesse, da man häufig davon ausgehen kann, dass zwei Werte im Prozess nur noch sehr wenig, oder gar nicht mehr, abhängig voneinander sind, wenn genug Zeit zwischendurch vergangen ist.

Es ist leicht zu sehen, dass aus der m -Abhängigkeit die *-mixing-Eigenschaft folgt: Sei (X_i) m -abhängig, dann wähle $M = m + 1$ und $f \equiv 0$. Dann sind alle A und B wie oben stochastisch unabhängig und daher $|\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)| = 0$.

DEFINITION 2.6. Sei (X_i) ein (möglicherweise vektorwertiger, etwa $X_i \in \mathbb{R}^d$ für alle i) identisch verteilter *-mixing Prozess der Länge n . Sei $m \leq n$ eine natürliche Zahl.

Sei $h : (\mathbb{R}^d)^m \rightarrow \mathbb{R}$ eine Funktion mit folgenden Eigenschaften:

- h ist symmetrisch, d.h. für jedes $(x_1, \dots, x_m) \in (\mathbb{R}^d)^m$ und jede Permutation π von $\{1, \dots, m\}$ gilt $h(x_1, \dots, x_m) = h(x_{\pi(1)}, \dots, x_{\pi(m)})$.
- $Eh(X_{i_1}, \dots, X_{i_m}) < \infty$ für alle $\{i_1, \dots, i_m\} \subseteq \{1, \dots, n\}$.

Dann ist die mit h korrespondierende **U -Statistik** auf der Basis der Stichproben X_1, \dots, X_n definiert als

$$U_n := U(X_1, \dots, X_n) := \frac{1}{\binom{n}{m}} \sum_{\{i_1, \dots, i_m\} \subseteq \{1, \dots, n\}} h(X_{i_1}, \dots, X_{i_m}).$$

Die Funktion h heißt **Kern** von U_n . Siehe auch [Ser].

SATZ 2.7 (Sen). *Wir geben hier nur den Spezialfall für $m = 2$ an. Gegeben seien die Voraussetzungen von Definition 2.6. Sei zusätzlich (X_i) stationär. Sei $h_1 : \mathbb{R}^d \rightarrow \mathbb{R}$ definiert durch $h_1(x) := Eh(x, X_1)$, $\mu := Eh_1(X_1)$ und $\sigma^2 := \text{Var} h_1(X_1) + 2\text{Cov}(h_1(X_1), h_1(X_2))$. Falls $0 < \sigma^2 < \infty$, dann gilt*

$$\sqrt{n}(U_n - \mu) \xrightarrow{d} N(0, 4\sigma^2).$$

Siehe [Sen].

SATZ 2.8 (Lemma von Slutsky). *Gegeben seien zwei Prozesse (X_n) und (Y_n) , eine Zufallsvariable X und eine Konstante c , so dass $X_n \xrightarrow{d} X$ und $Y_n \xrightarrow{stoch.} c$. Dann gilt auch*

$$X_n + Y_n \xrightarrow{d} X + c.$$

Siehe [Ser].

SATZ 2.9. *Unter H_0 und für große n sind $\sqrt{n}\tau_n^\circ$ und $\sqrt{n}\tau_n$ approximativ normalverteilt mit Erwartungswert 0 und Varianz $4/9$.*

PROOF. Betrachten wir zunächst den nichtzirkularen Fall. Da N unabhängig von der zugrundeliegenden Verteilung, wähle o.B.d.A. $X_i \sim U[-1/2, 1/2]$ für alle $1 \leq i \leq n$. Sei $Y_i := (X_i, X_{i+1})$ für $1 \leq i \leq n-1$, und definiere

$$U_n := \frac{N}{\binom{n-1}{2}} \stackrel{(1.2)}{=} \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \frac{g(Y_i, Y_j)}{\binom{n}{2}},$$

wobei

$$g(Y_i, Y_j) := \mathbf{1}_{\{X_i < X_j, X_{i+1} > X_{j+1}\}} + \mathbf{1}_{\{X_i > X_j, X_{i+1} < X_{j+1}\}}.$$

Hierbei geht ein, dass $R_i < R_j$ genau dann, wenn $X_i < X_j$. Aus der Definition ist sofort klar, dass g symmetrisch ist. Also ist U_n eine U -Statistik mit Kern g für den Prozess (Y_i) , welcher unter der Nullhypothese stationär und 1-abhängig, daher auch *-mixing ist. Außerdem ist $Eg^2(Y_1, Y_2) < \infty$, weil g die Summe zweier Indikatorfunktionen ist. Daher ist Satz 2.7 anwendbar, und es gilt unter H_0

$$\sqrt{n}(U_n - \mu) \xrightarrow{d} N(0, 4\sigma^2) \quad (2.13)$$

mit

$$\begin{aligned} g_1(y_1) &= Eg(y_1, Y_1) = P(x_1 < X_1, x_2 > X_2) + P(x_1 > X_1, x_2 < X_2) \\ &= \left(\frac{1}{2} - x_1\right) \left(x_2 - \frac{1}{2}\right) + \left(x_1 - \frac{1}{2}\right) \left(\frac{1}{2} - x_2\right) = \frac{1}{2} - 2x_1x_2, \end{aligned}$$

$$\begin{aligned} \mu &= Eg_1(Y_1) = E\left(\frac{1}{2} - 2X_1X_2\right) = \frac{1}{2} - 2EX_1X_2 \\ &= \frac{1}{2} - 2 \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} x_1x_2 dx_1 dx_2 = \frac{1}{2}, \end{aligned}$$

$$\begin{aligned} \text{Var } g(Y_1) &= \text{Var}\left(\frac{1}{2} - 2X_1X_2\right) = 4\text{Var } X_1X_2 \\ &= 4EX_1^2X_2^2 = 4 \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} x_1^2x_2^2 dx_1 dx_2 = \frac{1}{36}, \end{aligned}$$

$$\begin{aligned} \text{Cov}(g_1(Y_1), g_1(Y_2)) &= \text{Cov}\left(\frac{1}{2} - 2X_1X_2, \frac{1}{2} - 2X_2X_3\right) = 4\text{Cov}(X_1X_2, X_2X_3) \\ &= EX_1X_2^2X_3 - EX_1X_2EX_2X_3 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} x_1x_2^2x_3dx_1dx_2dx_3 = 0 \end{aligned}$$

und

$$\sigma^2 = \text{Var} g(Y_1) + \text{Cov}(g_1(Y_1), g_1(Y_2)) = \frac{1}{36}.$$

Damit folgt mit (2.13):

$$\sqrt{n} \left(U_n - \frac{1}{2} \right) = \sqrt{n} \left(\frac{N}{\binom{n-1}{2}} - \frac{1}{2} \right) = \sqrt{n} \left(\frac{2N}{(n-1)(n-2)} - \frac{1}{2} \right) \xrightarrow{d} N\left(0, \frac{1}{9}\right).$$

Durch Ausnutzen der Symmetrie der Normalverteilung und Multiplizieren mit 2 erhält man das Resultat für den nichtzirkularen Fall

$$\sqrt{n}\tau_n = \sqrt{n} \left(1 - \frac{4N}{(n-1)(n-2)} \right) \xrightarrow{d} N\left(0, \frac{4}{9}\right).$$

Analog zu U_n definiere jetzt $U_n^\circ = N^\circ / \binom{n}{2}$. Es konvergieren die Verteilungen von U_n und U_n° gegen die gleiche Verteilung, denn

$$\begin{aligned} |U_n - U_n^\circ| &= \left| \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \frac{g(Y_i, Y_j)}{\binom{n-1}{2}} - \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{g(Y_i, Y_j)}{\binom{n}{2}} \right| \\ &\leq \left| \frac{1}{\binom{n-1}{2}} - \frac{1}{\binom{n}{2}} \right| \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} g(Y_i, Y_j) + \frac{1}{\binom{n}{2}} \sum_{i=1}^n g(Y_i, Y_n) \\ &\leq \left| \frac{2}{(n-1)(n-1)} - \frac{2}{n(n-1)} \right| 2 \sum_{i=1}^{n-2} (n-1-i) + \frac{4(n-1)}{n(n-1)} \\ &= \frac{2n-2(n-2)}{n(n-1)(n-2)} (n-1)(n-2) + \frac{4}{n} = \frac{2}{n}. \end{aligned}$$

Mit Satz 2.8 folgt

$$U_n^\circ = \underbrace{U_n}_{\xrightarrow{d} N(0, \frac{4}{9})} + \underbrace{U_n^\circ - U_n}_{\xrightarrow{\text{stoch.}} 0} \xrightarrow{d} N\left(0, \frac{4}{9}\right).$$

□

Damit wissen wir, dass approximativ $T_n, T_n^\circ \sim N(0, 1)$ gilt. [FGH] empfehlen diese Approximation für $n > 10$. Es ergibt sich also für die Quantile $t_{n,\alpha} = t_{n,\alpha}^\circ = \Phi^{-1}(1 - \alpha)$ das $(1 - \alpha)$ -Quantil der Standardnormalverteilung mit Ablehnungsbereich $T_n > \Phi^{-1}(1 - \alpha)$ bzw. $T_n^\circ > \Phi^{-1}(1 - \alpha)$, denn große T_n und T_n° bedeuten wenige Inversionen, was auf die Alternativhypothese hindeutet.

3. Einige Simulationsversuche

[FGH] haben umfangreiche Simulationsstudien durchgeführt, und τ_n° mit anderen Tests verglichen. Dazu sei (X_i) ein $AR(1)$ -Prozess erster Ordnung mit $X_i - \theta X_{i-1} = \varepsilon_i$ für $2 \leq i \leq n$ und Startwert $X_1 = (1 - \theta^2)^{-1/2} \varepsilon_1$, wobei $\theta \in \{1/4, 1/8, 1/16, 1/32, 0\}$ und ε_i stochastisch unabhängige Zufallsvariablen sind, die einer der folgenden Wahrscheinlichkeitsverteilungen entstammen:

- Standardnormalverteilung: $\varepsilon_i \sim N(0, 1)$, $1 \leq i \leq n$, mit Dichte

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in \mathbb{R},$$

- Logistische Verteilung: $\varepsilon_i \sim \text{Log}(0, 1)$, $1 \leq i \leq n$, mit Dichte

$$f(x) = \frac{e^x}{(1 + e^x)^2}, x \in \mathbb{R},$$

- Laplace-Verteilung: ε_i , $1 \leq i \leq n$, hat Dichte

$$f(x) = \frac{1}{2} e^{-|x|}, x \in \mathbb{R},$$

- Cauchy-Verteilung: ε_i , $1 \leq i \leq n$, hat Dichte

$$f(x) = \frac{1}{\pi(1 + x^2)}, x \in \mathbb{R}.$$

Der Fall $\theta = 0$ hat die Unabhängigkeit der X_i zur Folge, die Nullhypothese ist dann also erfüllt. Kleine θ haben nur eine geringe Abhängigkeit zur Folge, weshalb in diesem Fall die Fehlerwahrscheinlichkeit 2. Art sehr hoch zu erwarten ist. Die Tests, mit denen verglichen wird, sind sowohl andere nichtparametrische Tests, die ebenfalls auf Rängen basieren (van der Waerden, Wilcoxon, Laplace und ρ_n), als auch parametrische, die auf den Autokorrelationskoeffizienten basieren (Moran, Ljung-Box und Dufour-Roy). Von diesen vier Verteilungen wurden jeweils 5000 Prozesse der Längen $n = 20, 50, 100$ generiert, aus welchen die fünf autoregressiven Prozesse errechnet wurden, auf die schließlich alle Tests zum Niveau $\alpha = 5\%$ angewendet wurden. Die Tabellen in [FGH] Seite 592ff zeigen, dass τ_n° zu den besseren Teststatistiken gehört, da es bei abhängigen X_i immer mit an der Spitze liegt, und bei unabhängigen X_i das 5%-Niveau ziemlich genau ausschöpft.

BEISPIEL 3.1. Einige Beispiele zur Anwendung des modifizierten Kendalls tau:

- (1) Sei (X_i) der Prozess des Ausgabepreises des UniDynamicFonds Europa A im Zeitraum vom 19.11. bis 19.12.2001. (Quelle: Union-Investment GmbH.) Sowohl τ_n als auch τ_n° , angewendet zum Niveau $\alpha = 5\%$ verwerfen die Nullhypothese, denn das Programm `tau.r`, vergleiche Anhang A.2, errechnet $T_n \doteq 3.456395$ und $T_n^\circ \doteq 2.715739$ bei $\Phi^{-1}(0.95) \doteq 1.644854$.

Jahr	Anzahl	Jahr	Anzahl
1989	731	1995	643
1990	710	1996	655
1991	663	1997	635
1992	654	1998	665
1993	624	1999	681
1994	667	2000	623

TABELLE 3. Verkehrsverunglückte im Saarland pro 100000 Einwohner. Quelle: [Vu1999], [Vu2000]

- (2) Sei (X_i) die Anzahl der Verkehrsverunglückten pro 100000 Einwohner im Saarland in den Jahren 1989 bis 2000, siehe Tabelle 3. Mit einem modifiziertem Programm ergibt sich $T_n \doteq -0.0944755$ und $T_n^\circ \doteq -0.944755$. Da $\Phi^{-1}(0.95) \doteq 1.644854$, wird die Nullhypothese in beiden Fällen nicht verworfen.

Anhang A. Quellcode und Ausgabe

Auf meiner Seite <http://maths.vic-fontaine.de/> stehen die Programme zum Download zur Verfügung.

A.1. Permutationen abzählen.

```
> #
> # perm.mws (Permutationen abzaehlen)
> #
> with(combinat): # Zusatzpaket Kombinatorik laden
Warning, the protected name Chi has been redefined and unprotected

> n:=6; # n setzen

                                n := 6

> G:=permute(n): # Alle moeglichen Permutationen generieren
> zaehler:=0: # Zaehler zuruecksetzen
> i:=1: j:=2: k:=4: l:=5:
> for a from 1 to numbperm(n) do # Alle Permutationen durchlaufen
>   R:=G[a]; # a-te Permutation betrachten
>   if (R[i]<R[j] and R[i+1]>R[j+1] and
>       R[k]<R[l] and R[k+1]>R[l+1])
>     then zaehler:=zaehler+1 end if; # ggf Zaehler erhoehen
> end do:
> zaehler/numbperm(n); # Wahrscheinlichkeit berechnen und ausgeben
```

1/9

A.2. Beispiel für Kendalls tau.

```

#
# tau.r
#
cat("Beispiel fuer Kendalls tau\n")
cat("Daten: Ausgabepreis UniDynamicFonds Europa A ")
cat("19.11. bis 20.12.2001\n\n")
# Daten einlesen
n <- 24
X <- c(62.37, 61.68, 60.14, 60.62, 60.08, 61.34, 61.01, 59.13,
      58.70, 59.32, 58.20, 58.88, 60.58, 62.64, 61.98, 60.75,
      59.91, 60.74, 58.57, 57.43, 57.80, 59.14, 58.22, 57.83)
# Raenge ausrechnen
R <- rank(X)
R[n+1] <- R[1]
# Zaehler zuruecksetzen
N <- 0
Nz <- 0
# Inversionen zaehlen
for (i in 1:n)
{
  for (j in 1:n)
  {
    if (R[i]<R[j] & R[i+1]>R[j+1])
    {
      if (i<n & j<n) N <- N + 1
      Nz <- Nz + 1
    }
  }
}
# tau ausrechnen
tau <- 1 - 4*N / (n-1)/(n-2)
tauz <- 1 - 4*Nz / n/(n-1)
T <- tau * 3 / 2 * sqrt(n)
Tz <- tauz * 3 / 2 * sqrt(n)
cat("Zirkularer Fall, alpha = 5%:\nN =", N, "\ntau =")
cat(tauz, "\nT =", T, "\nNullhypothese wird ")
# Zum 5%-Niveau testen
if (Tz > qnorm(0.95)) cat("verworfen.\n\n") else
  cat("nicht verworfen.\n\n")
cat("Nichtzirkularer Fall, alpha = 5%:\nN =", N, "\ntau =")

```

```

cat(tau, "\nT =", T, "\nNullhypothese wird ")
if (T > qnorm(0.95)) cat("verworfen.\n") else
    cat("nicht verworfen.\n")

```

A.2.1. Ausgabe.

Beispiel fuer Kendalls tau

Daten: Ausgabepreis UniDynamicFonds Europa A 19.11. bis 20.12.2001

Zirkularer Fall, alpha = 5%:

N = 87

tau =0.3695652

T = 2.715739

Nullhypothese wird verworfen.

Nichtzirkularer Fall, alpha = 5%:

N = 67

tau =0.4703557

T = 3.456395

Nullhypothese wird verworfen.

Literatur

- [BHK] J. R. Blum, D. L. Hanson, L. H. Koopmans, *On the strong law of large numbers for a class of stochastic processes* (Z. Wahrscheinlichkeitstheorie 2, S. 1-11, 1963)
 - [FGH] T. S. Ferguson, C. Genest, M. Hallin, *Kendall's tau for serial dependence* (The Canadian Journal of Statistics 28, S. 587-604, 2000)
 - [Háj] J. Hájek, *A course in nonparametric statistics* (Holden-Day 1969)
 - [Har] J. Hartung, *Statistik, 11. Auflage* (Oldenbourg Verlag 1998)
 - [Sen] P. K. Sen, *Limiting behaviour of regular functionals of empirical distributions for stationary *-mixing Processes* (Z. Wahrscheinlichkeitstheorie 25, S. 71-82, 1972)
 - [Ser] R. J. Serfling, *Approximation theorems of mathematical statistics* (Wiley 1980)
 - [Vu1999] *Polizeiliche Verkehrsunfallstatistik des Saarlandes für 1999* (Saarländisches Ministerium für Inneres und Sport 2000)
 - [Vu2000] *Polizeiliche Verkehrsunfallstatistik des Saarlandes für 2000* (Saarländisches Ministerium für Inneres und Sport 2001)
- E-mail address: Jan.Krohn@post.rwth-aachen.de*