seit 1558

Degenerierte U- und V-Statistiken

B A C H E L O R A R B E I T

zur Erlangung des akademischen Grades

Bachelor of Science (B. Sc.)

im Studiengang Mathematik

FRIEDRICH-SCHILLER-UNIVERSITÄT JENA
Fakultät für Mathematik und Informatik

eingereicht von Andreas Kühnapfel

geboren am 8.3.1989 in Oschatz

Betreuer: Prof. Dr. M. H. Neumann

Jena, 1.9.2011

Zusammenfassung

In dieser Arbeit werden wir uns mit zwei wichtigen Klassen von Statistiken - den sogenannten U- und V-Statistiken - befassen. Dabei beginnen wir mit den 1948 von Wassilij Hoeffding eingeführten erwartungstreuen U-Statistiken. Nachdem wir wichtige Eigenschaften hergeleitet haben, werden wir uns der Hoeffding-Zerlegung widmen. Anschließend gehen wir auf die Beziehung zwischen U- und den nach Richard von Mises benannten V-Statistiken ein, bevor wir zum Begriff der Degeneration übergehen und einige Beispiele betrachten werden.

Im darauffolgenden Kapitel untersuchen wir die Asymptotik von vollständig degenerierten U- und V-Statistiken vom Grad 2. Wir werden feststellen, dass sich die zentrierten und normierten Statistiken im Limes als unendliche Reihe von gewichteten chi-quadrat-verteilten Zufallsvariablen darstellen lassen.

Dass im degenerierten Fall die Bootstrap-Verteilung von U-Statistiken gegen dieselbe Grenzverteilung konvergiert wie die Verteilung der ursprünglichen U-Statistik, werden wir im letzten Kapitel überprüfen. Als Hilfsmittel wird dabei die Mallows-Metrik dienen.

Inhaltsverzeichnis

Bezeichnungen

$\mathbb{N}$	natürliche Zahlen ohne 0
$\mathbb{R}$	reelle Zahlen
Ω	Basismenge
$\mathcal{A}$	σ-Algebra
$\mathcal{B}$, $\mathcal{B}^d$	borelsche σ-Algebra von $\mathbb{R}$, $\mathbb{R}^d$
$\mathscr{F}, \mathscr{F}_2, \mathscr{F}_{2,t}, \mathscr{F}^*$	Familien von Verteilungsfunktionen auf $\mathbb{R}$
F	Verteilungsfunktion aus $\mathscr{F}$, $\mathscr{F}^*$
F_n	empirische Verteilungsunktion
D	Verteilungsfunktion des Dirac-Maßes
i.i.d.	unabhängig und identisch verteilt
$\mathscr{L}(X)$	Verteilung beziehungsweise Verteilungsfunktion von X
$\mathscr{L}_F(X)$	$\mathscr{L}(X)$ als Funktion in F
θ	reguläres statistisches Funktional
h	Kern von θ
$\sum_{\binom{n}{k}}$	summiert wird über alle Kombinationen $\{i_1, \ldots, i_k\}$ von $\{1, \ldots, n\}$
$h_c(x_1, \ldots, x_c)$	bedingter Erwartungswert von $h(X_1, \ldots, X_k)$ unter $X_1 = x_1, \ldots, X_c = x_c$
σ_c^2	Varianz von $h_c(X_1, \ldots, X_c)$
$\sigma_{c,d}^2$	Kovarianz von $h_c(X_1, \ldots, X_c)$ und $h_d(X_1, \ldots, X_d)$
$h^{(l)}$	Kern vom Grad l (Hoeffding-Zerlegung)
$U_n = U_n(h; X_1, \ldots, X_n)$	U-Statistik mit Kern h
$V_n = V_n(h; X_1, \ldots, X_n)$	V-Statistik mit Kern h
$S_{k\|l}$	Stirling-Zahl zweiter Art
$\xrightarrow{P_F\text{-f.s.}}$	Konvergenz P_F-fast sicher
$\xrightarrow{P_F}$	Konvergenz in Wahrscheinlichkeit (bezüglich P_F)
$\xrightarrow{d}$	Konvergenz in Verteilung
$\xrightarrow{w}$	schwache Konvergenz
λ_k	Eigenwert von Integralgleichung
f_k	Eigenfunktion zu λ_k
$\|M\|$	Kardinalität der Menge M
d_2	Mallows-Metrik
$d_{2,t}$	aus d_2 abgeleitete Halbmetrik

1 Grundlagen

In diesem ersten Kapitel werden wir die Grundbausteine für die weiteren Untersuchungen von U- und V-Statistiken legen. Beginnend mit einigen Vorüberlegungen führen wir anschließend die Klasse der U-Statistiken ein und betrachten deren Eigenschaften genauer. Nachdem wir die Hoeffding-Zerlegung für U-Statistiken kennengelernt haben, gehen wir zu der Klasse der V-Statistiken über und stellen eine Verbindung zwischen den beiden Statistiken her. Zum Abschluss dieses Kapitels werden wir uns mit dem Begriff der Degeneration und verschiedenen Beispielen für U- und V-Statistiken beschäftigen.

1.1 Vorbetrachtungen

Es sei $(\Omega, \mathcal{A}, \{P_F \colon F \in \mathscr{F}\})$ ein parametrisierter Wahrscheinlichkeitsraum, wobei

$$\mathscr{F} \subseteq \{G \colon G \text{ Verteilungsfunktion auf } \mathbb{R}\}.$$

Wir betrachten Zufallsvariablen $X_1, \ldots, X_n$ mit $X_i \colon (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$, $i = 1, \ldots, n$, wobei $\mathcal{B}$ die borelsche σ-Algebra von $\mathbb{R}$ bezeichnet. Im Folgenden sei zudem $\mathcal{B}^d$ die borelsche σ-Algebra von $\mathbb{R}^d$, für $d \in \mathbb{N}$. Weiterhin seien $X_1, \ldots, X_n$ unabhängig und identisch verteilt, jeweils mit Verteilungsfunktion $F \in \mathscr{F}$, was wir mit

$$\mathscr{L}(X_1) = F$$

kennzeichnen. Dabei sei zwar $\mathscr{F}$ bekannt, aber nicht das konkrete $F \in \mathscr{F}$.

Anzumerken ist, dass wir mit $\mathscr{L}$ nicht immer nur die Verteilungsfunktion, sondern auch die Verteilung einer Zufallsgröße bezeichnen werden. Natürlich ist beides nicht dasselbe. Aber aufgrund der gegebenen Beziehung zwischen Verteilung und Verteilungsfunktion und um unnötige Bezeichnungen zu vermeiden, werden wir eben diese Schreibweise verwenden. Die jeweilige Bedeutung wird sich leicht aus dem Kontext ergeben.

Schließlich ist festzuhalten, dass die folgenden Aussagen nicht nur für reellwertige Zufallsvariablen von Bedeutung sind. Analoge Resultate lassen sich auch für Zufallsgrößen beziehungsweise Zufallsvektoren mit Werten in anderen geeigneten Räumen herleiten. Der Einfachheit halber betrachten wir aber den oben genannten Fall.

Definition 1 (vgl. Lee (1990), S. 1)**.** *Es sei $\theta \colon \mathscr{F} \to \mathbb{R}$ ein Funktional. Falls für θ die Darstellung*

$$\theta(F) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, \ldots, x_k) dF(x_1) \cdots dF(x_k)$$

für $F \in \mathscr{F}$ und $k \in \mathbb{N}$ gilt, so heißt θ ein reguläres statistisches Funktional vom Grad k und $h \colon \mathbb{R}^k \to \mathbb{R}$ der Kern des Funktionals.

Es sei h $(\mathcal{B}^k\text{-}\mathcal{B})$-messbar, womit dann $h(X_1,\ldots,X_k)$ ein erwartungstreuer Schätzer für θ ist.

Ohne Einschränkung nehmen wir an, dass

$$E_F|h(X_1,\ldots,X_k)| < \infty.$$

Für $S_k := \{1,\ldots,k\}$ mit $k \in \mathbb{N}$ sei $\pi\colon S_k \to S_k$ eine Abbildung, die eine Permutation der Elemente von S_k beschreibt. Alle Permuationen π von S_k fassen wir in der Menge Π_k zusammen.

Unter Verwendung dieser Notation können wir weiter annehmen, dass h symmetrisch ist, das heißt es gilt:

$$h\left(x_{i_1},\ldots,x_{i_k}\right) = h\left(x_{i_{\pi(1)}},\ldots,x_{i_{\pi(k)}}\right)$$

für alle Kombinationen $\{i_1,\ldots,i_k\}$ von $\{1,\ldots,n\}$ mit $i_s \neq i_t$ für $s \neq t$ und für alle $\pi \in \Pi_k$. Sonst ersetzen wir $h(x_{i_1},\ldots,x_{i_k})$ durch

$$\widetilde{h}(x_{i_1},\ldots,x_{i_k}) := \frac{1}{k!}\sum_{\pi\in\Pi_k} h\left(x_{i_{\pi(1)}},\ldots,x_{i_{\pi(k)}}\right),$$

wobei über alle $k!$ Permutationen der Elemente von $\{i_1,\ldots,i_k\}$ summiert wird. Damit ist dann $\widetilde{h}$ symmetrisch, ebenfalls $(\mathcal{B}^k\text{-}\mathcal{B})$-messbar und $\widetilde{h}(X_1,\ldots,X_k)$ aufgrund der Konstruktion auch weiterhin erwartungstreu.

1.2 U-Statistiken

U-Statistiken bilden eine wichtige Klasse von Statistiken, da sie eine einfache Gestalt haben und sich viele bekannte Schätzer und Teststatistiken eben als eine solche Statistik darstellen lassen. U-Statistiken wurden aufgrund ihrer Erwartungstreue (englisch: „unbiasedness“) von Wassilij Hoeffding in seiner Arbeit „A class of statistics with asymptotically normal distribution“ von 1948 so benannt.

1.2.1 Definitionen und Eigenschaften

Definition 2 (vgl. Lee (1990), S. 7)**.** *Es seien $\theta\colon \mathscr{F} \to \mathbb{R}$ ein reguläres statistisches Funktional vom Grad k mit $k \leq n$ und h der Kern des Funktionals. Dann ist*

$$U_n = U_n(h; X_1,\ldots,X_n) := \binom{n}{k}^{-1} \sum_{\binom{n}{k}} h(X_{i_1},\ldots,X_{i_k})$$

die zugehörige U-Statistik. Dabei wird über alle $\binom{n}{k}$ Möglichkeiten, eine Teilmenge $\{i_1,\ldots,i_k\}$ von $\{1,\ldots,n\}$ mit $i_s \neq i_t$ für $s \neq t$ auszuwählen, summiert.

Aus der Definition kennen wir bereits den Erwartungswert von U-Statistiken: $E_F U_n = E_F h(X_1, \ldots, X_k) = \theta$. Wie sieht es aber nun beispielsweise mit der Varianz einer U-Statistik oder der Kovarianz von zwei U-Statistiken aus? Dies werden wir im Folgenden untersuchen, wobei wir zuerst wieder einige Vorbereitungen treffen werden.

Es gelte:

$$E_F h^2(X_1, \ldots, X_k) < \infty.$$

Für $c = 1, \ldots, k$ betrachten wir die bedingten Erwartungen $h_c(X_1, \ldots, X_c)$, wobei

$$h_c(x_1, \ldots, x_c) := E_F[h(X_1, \ldots, X_k) \mid X_1 = x_1, \ldots, X_c = x_c].$$

Entsprechend definieren wir

$$\sigma_c^2 := Var_F h_c(X_1, \ldots, X_c).$$

Mit diesen Größen wird es sich im Weiteren leichter rechnen und einige Resultate kompakter darstellen lassen.

In den folgenden zwei Lemmas geht es zunächst um Eigenschaften der eben definierten bedingten Erwartungen und deren Varianzen.

Lemma 1 (vgl. Lee (1990), S. 10)**.** *Die Zufallsvariablen $X_1, \ldots, X_n$ seien unabhängig und identisch verteilt mit $\mathscr{L}(X_1) = F \in \mathscr{F}$. Weiterhin seien $\theta\colon \mathscr{F} \to \mathbb{R}$ ein reguläres statistisches Funktional vom Grad k mit $1 \leq k \leq n$ und h der Kern des Funktionals. Dann gilt:*

(i) $h_c(x_1, \ldots, x_c) = E_F h_d(x_1, \ldots, x_c, X_{c+1}, \ldots X_d)$ *für* $1 \leq c < d \leq k$ *und*

(ii) $E_F h_c(X_1, \ldots, X_c) = E_F h(X_1, \ldots, X_k)$ *für* $1 \leq c \leq k$.

Beweis. Beides lässt sich einfach unter Zuhilfenahme des Satzes von Fubini (vgl. Dunford und Schwartz (1988a), S. 193) nachrechnen:

(i)

$$\begin{aligned}
&E_F h_d(x_1, \ldots, x_c, X_{c+1}, \ldots, X_d) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_d(x_1, \ldots, x_c, x_{c+1}, \ldots, x_d) dF(x_{c+1}) \cdots dF(x_d) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \\
&\quad \cdot \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, \ldots, x_d, x_{d+1}, \ldots, x_k) dF(x_{d+1}) \cdots dF(x_k) \right] \\
&\quad \cdot dF(x_{c+1}) \cdots dF(x_d) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, \ldots, x_k) dF(x_{c+1}) \cdots dF(x_k) \\
&= h_c(x_1, \ldots, x_c)
\end{aligned}$$

(ii)

$$
\begin{aligned}
& E_F h_c(X_1, \ldots, X_c) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_c(x_1, \ldots, x_c) dF(x_1) \cdots dF(x_c) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \\
&\qquad \cdot \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, \ldots, x_c, x_{c+1}, \ldots, x_k) dF(x_{c+1}) \cdots dF(x_k) \right] \\
&\qquad \cdot dF(x_1) \cdots dF(x_c) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, \ldots, x_k) dF(x_1) \cdots dF(x_k) \\
&= E_F h(X_1, \ldots, X_k)
\end{aligned}
$$

□

Lemma 2 (vgl. Lee (1990), S. 11)**.** *Die Zufallsgrößen $X_1, \ldots, X_n$ seien unabhängig und identisch verteilt mit $\mathscr{L}(X_1) = F \in \mathscr{F}$. Außerdem seien $\theta\colon \mathscr{F} \to \mathbb{R}$ ein reguläres statistisches Funktional vom Grad k mit $1 \leq k \leq n$ und h der Kern des Funktionals. Dann gilt:*

$$\sigma_c^2 = Cov_F(h(X_{i_1}, \ldots, X_{i_k}), h(X_{j_1}, \ldots, X_{j_k})),$$

wobei $\{i_1, \ldots, i_k\}, \{j_1, \ldots, j_k\} \subseteq \{1, \ldots, n\}$ mit $i_s \neq i_t$ und $j_u \neq j_v$ für $s \neq t$ beziehungsweise $u \neq v$ und $|\{i_1, \ldots, i_k\} \cap \{j_1, \ldots, j_k\}| = c$ ist.

Beweis. Da h symmetrisch und $X_1, \ldots, X_n$ unabhängig und identisch verteilt sind, betrachten wir ohne Beschränkung der Allgemeinheit als $\{i_1, \ldots, i_k\} := \{1, \ldots, k\}$ und als $\{j_1, \ldots, j_k\} := \{1, \ldots, c, k+1, \ldots, 2k-c\}$.

Es ist:

$$
\begin{aligned}
& Cov_F(h(X_1, \ldots, X_k), h(X_1, \ldots X_c, X_{k+1}, \ldots, X_{2k-c})) \\
&\quad = E_F[h(X_1, \ldots, X_k) h(X_1, \ldots X_c, X_{k+1}, \ldots, X_{2k-c})] \\
&\qquad - [E_F h(X_1, \ldots, X_k)][E_F h(X_1, \ldots X_c, X_{k+1}, \ldots, X_{2k-c})] \\
&\quad = E_F[h(X_1, \ldots, X_k) h(X_1, \ldots X_c, X_{k+1}, \ldots, X_{2k-c})] \\
&\qquad - [E_F h(X_1, \ldots, X_k)]^2.
\end{aligned}
$$

Wir wollen zeigen, dass

$$
\begin{aligned}
& Cov_F(h(X_1, \ldots, X_k), h(X_1, \ldots X_c, X_{k+1}, \ldots, X_{2k-c})) \\
&\quad = E_F h_c^2(X_1, \ldots, X_c) - [E_F h_c(X_1, \ldots, X_c)]^2 .
\end{aligned}
$$

Nach Lemma 1 (ii) genügt es also zu zeigen, dass

$$E_F[h(X_1,\ldots,X_k)h(X_1,\ldots,X_c,X_{k+1},\ldots,X_{2k-c})] = E_F h_c^2(X_1,\ldots,X_c):$$

$$\begin{aligned}
&E_F[h(X_1,\ldots,X_k)h(X_1,\ldots,X_c,X_{k+1},\ldots,X_{2k-c})] \\
= \;& \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h(x_1,\ldots,x_k)h(x_1,\ldots,x_c,x_{k+1},\ldots,x_{2k-c})dF(x_1)\cdots dF(x_{2k-c}) \\
= \;& \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \\
&\quad \cdot\left[\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h(x_1,\ldots,x_k)dF(x_{c+1})\cdots dF(x_k)\right] \\
&\quad \cdot\left[\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h(x_1,\ldots,x_c,x_{k+1},\ldots,x_{2k-c})dF(x_{k+1})\cdots dF(x_{2k-c})\right] \\
&\quad \cdot dF(x_1)\cdots dF(x_c) \\
= \;& \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h_c^2(x_1,\ldots,x_c)dF(x_1)\cdots dF(x_c) \\
= \;& E_F h_c^2(X_1,\ldots,X_c)
\end{aligned}$$

unter Verwendung des Satzes von Fubini (vgl. Dunford und Schwartz (1988a), S. 193). □

Mit der Hilfe unserer Vorbereitungen können wir nun die Varianz einer U-Statistik folgendermaßen darstellen:

Satz 1 (vgl. Lee (1990), S. 12)**.** *Die Zufallsvariablen $X_1,\ldots,X_n$ seien unabhängig und identisch verteilt mit $\mathscr{L}(X_1) = F \in \mathscr{F}$. Weiterhin seien $\theta\colon \mathscr{F} \to \mathbb{R}$ ein reguläres statistisches Funktional vom Grad k mit $1 \leq k \leq n$ und h der Kern des Funktionals. Dann gilt für die entsprechende U-Statistik U_n:*

$$Var_F U_n = \binom{n}{k}^{-1}\sum_{c=1}^{k}\binom{k}{c}\binom{n-k}{k-c}\sigma_c^2.$$

Beweis. Für die Varianz gilt:

$$\begin{aligned}
Var_F U_n &= Var_F\left[\binom{n}{k}^{-1}\sum_{\binom{n}{k}} h(X_{i_1},\dots,X_{i_k})\right]\\
&= Cov_F\left(\binom{n}{k}^{-1}\sum_{\binom{n}{k}} h(X_{i_1},\dots,X_{i_k}),\binom{n}{k}^{-1}\sum_{\binom{n}{k}} h(X_{j_1},\dots,X_{j_k})\right)\\
&= \binom{n}{k}^{-2}\sum_{\binom{n}{k}}\sum_{\binom{n}{k}} Cov_F(h(X_{i_1},\dots,X_{i_k}),h(X_{j_1},\dots,X_{j_k})).
\end{aligned}$$

Im nächsten Schritt werden wir zählen, bei wie viel der $\binom{n}{k}^2$ Kovarianzen die Funktionen h $c = 1,\dots,k$ Elemente gemeinsam haben, damit wir Lemma 2 anwenden können.

Für h im ersten Argument der Kovarianz gibt es $\binom{n}{k}$ Möglichkeiten k Elemente von $\{1,\dots,n\}$ auszuwählen. Für h im zweiten Argument sind es genau $\binom{k}{c}$ Möglichkeiten die c gemeinsamen Elemente und $\binom{n-k}{k-c}$ Möglichkeiten die $k-c$ verschiedenen Elemente zu h im ersten Argument auszuwählen. Falls h im ersten und h im zweiten Argument keine gemeinsamen Elemente besitzen, ist die Kovarianz aufgrund der Unabhängigkeit von $X_1,\dots,X_n$ gleich 0.

Mit Lemma 2 folgt nun:

$$\begin{aligned}
Var_F U_n &= \binom{n}{k}^{-2}\sum_{c=1}^{k}\binom{n}{k}\binom{k}{c}\binom{n-k}{k-c}\sigma_c^2\\
&= \binom{n}{k}^{-1}\sum_{c=1}^{k}\binom{k}{c}\binom{n-k}{k-c}\sigma_c^2.
\end{aligned}$$

□

Zum Abschluss dieses Abschnitts werden wir uns jetzt noch mit der Kovarianz von zwei U-Statistiken befassen. Dazu betrachten wir die U-Statistiken $U_n^{(1)}$ und $U_n^{(2)}$ mit Kernen $h^{(1)}$ und $h^{(2)}$ vom Grad k_1 beziehungsweise k_2 mit $1 \le k_1 \le k_2 \le n$.

Dafür definieren wir für $1 \le c \le k_1$ und $1 \le d \le k_2$ die Größe

$$\sigma_{c,d}^2 := Cov_F\left(h_c^{(1)}(X_1,\dots,X_c),h_d^{(2)}(X_1,\dots,X_d)\right).$$

Nun halten wir fest:

Lemma 3 (vgl. Lee (1990), S. 17)**.** *Die Zufallsgrößen* $X_1, \ldots, X_n$ *seien unabhängig und identisch verteilt mit* $\mathscr{L}(X_1) = F \in \mathscr{F}$*.* $\theta\colon \mathscr{F} \to \mathbb{R}$ *sei ein reguläres statistisches Funktional mit Kernen* $h^{(1)}$ *und* $h^{(2)}$ *vom Grad* k_1 *beziehungsweise* k_2 *mit* $1 \leq k_1 \leq k_2 \leq n$*. Weiterhin seien* $\{i_1, \ldots, i_{k_1}\}, \{j_1, \ldots, j_{k_2}\} \subseteq \{1, \ldots, n\}$ *mit* $i_s \neq i_t$ *und* $j_u \neq j_v$ *für* $s \neq t$ *und* $u \neq v$*, wobei* $|\{i_1, \ldots, i_{k_1}\} \cap \{j_1, \ldots, j_{k_2}\}| = c$ *sei. Schließlich gelte* $c \leq d$*, mit* $1 \leq c \leq k_1$ *und* $1 \leq d \leq k_2$*. Dann gilt:*

$$\sigma^2_{c,d} = Cov_F\left(h^{(1)}\left(X_{i_1}, \ldots, X_{i_{k_1}}\right), h^{(2)}\left(X_{j_1}, \ldots, X_{j_{k_2}}\right)\right).$$

Beweis. Der Beweis dieses Lemmas erfordert lediglich einige Modifikationen des Beweises von Lemma 2.

Aufgrund der Symmetrie von $h^{(1)}$ und $h^{(2)}$ und der unabhängigen, identischen Verteilung von $X_1, \ldots, X_n$ betrachten wir ohne Einschränkung als $\{i_1, \ldots, i_{k_1}\} := \{1, \ldots, k_1\}$ und als $\{j_1, \ldots, j_{k_2}\} := \{1, \ldots, c, k_1 + 1, \ldots, k_1 + k_2 - c\}$.

Mit den Überlegungen aus dem Beweis von Lemma 2 und unter Zuhilfenahme von Lemma 1 (ii) bleibt zu zeigen, dass

$$\begin{aligned} & E_F\left[h^{(1)}(X_1, \ldots, X_{k_1}) h^{(2)}(X_1, \ldots, X_c, X_{k_1+1}, \ldots, X_{k_1+k_2-c})\right] \\ = \ & E_F\left[h^{(1)}_c(X_1, \ldots, X_c) h^{(2)}_d(X_1, \ldots, X_d)\right]. \end{aligned}$$

Mit dem Satz von Fubini (vgl. Dunford und Schwartz (1988a), S. 193) gilt:

$$
\begin{aligned}
& E_F\left[h^{(1)}(X_1,\ldots,X_{k_1})h^{(2)}(X_1,\ldots,X_c,X_{k_1+1},\ldots,X_{k_1+k_2-c})\right] \\
= \; & \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h^{(1)}(x_1,\ldots,x_{k_1})h^{(2)}(x_1,\ldots,x_c,x_{k_1+1},\ldots,x_{k_1+k_2-c}) \\
& \quad \cdot dF(x_1)\cdots dF(x_{k_1+k_2-c}) \\
= \; & \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \\
& \quad \cdot \left[\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h^{(1)}(x_1,\ldots,x_{k_1})dF(x_{c+1})\cdots dF(x_{k_1})\right] \\
& \quad \cdot \left[\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h^{(2)}(x_1,\ldots,x_c,x_{k_1+1},\ldots,x_{k_1+k_2-c})dF(x_{k_1+1})\cdots dF(x_{k_1+k_2-c})\right] \\
& \quad \cdot dF(x_1)\cdots dF(x_c) \\
= \; & \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h_c^{(1)}(x_1,\ldots,x_c)h_c^{(2)}(x_1,\ldots,x_c)dF(x_1)\cdots dF(x_c) \\
= \; & \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h_c^{(1)}(x_1,\ldots,x_c) \\
& \quad \cdot \left[\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h_d^{(2)}(x_1,\ldots,x_c,x_{c+1},\ldots,x_d)dF(x_{c+1})\cdots dF(x_d)\right] \\
& \quad \cdot dF(x_1)\cdots dF(x_c) \\
= \; & \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h_c^{(1)}(x_1,\ldots,x_c)h_d^{(2)}(x_1,\ldots,x_d)dF(x_1)\cdots dF(x_d) \\
= \; & E_F\left[h_c^{(1)}(X_1,\ldots,X_c)h_d^{(2)}(X_1,\ldots,X_d)\right],
\end{aligned}
$$

wobei bei den letzten Schritten der Rechnung Lemma 1 (i) Anwendung findet. □

Aus Lemma 3 können wir folgern, dass für die Kovarianzen der bedingten Erwartungen für $c = 1,\ldots,k_1$ gilt:

$$\sigma^2_{c,c} = \cdots = \sigma^2_{c,k_2}.$$

Dies wird aus der Rechnung im Beweis des Lemmas und unter Verwendung von Lemma 1 (i) ersichtlich.

Damit können wir nun folgende Aussage für die Kovarianz zwischen zwei U-Statistiken treffen:

Satz 2 (vgl. Lee (1990), S. 17)**.** *Die Zufallsvariablen $X_1,\ldots,X_n$ seien unabhängig und identisch verteilt mit $\mathscr{L}(X_1) = F \in \mathscr{F}$. $\theta\colon \mathscr{F} \to \mathbb{R}$ sei ein reguläres statistisches Funktional mit Kernen $h^{(1)}$ und $h^{(2)}$ vom Grad k_1 beziehungsweise k_2 mit $1 \le k_1 \le k_2 \le n$.*

Dann gilt für die entsprechenden U-Statistiken $U_n^{(1)}$ und $U_n^{(2)}$:

$$Cov_F\left(U_n^{(1)}, U_n^{(2)}\right) = \binom{n}{k_1}^{-1} \sum_{c=1}^{k_1} \binom{k_2}{c}\binom{n-k_2}{k_1-c}\sigma_{c,c}^2.$$

Beweis. Es gilt:

$$\begin{aligned} &Cov_F\left(U_n^{(1)}, U_n^{(2)}\right) \\ &= \binom{n}{k_1}^{-1}\binom{n}{k_2}^{-1} \sum_{\binom{n}{k_1}}\sum_{\binom{n}{k_2}} Cov_F\left(h^{(1)}\left(X_{i_1},\ldots,X_{i_{k_1}}\right), h^{(2)}\left(X_{j_1},\ldots,X_{j_{k_2}}\right)\right). \end{aligned}$$

Bei der Summierung gehen aufgrund der Unabhängigkeit von $X_1, \ldots, X_n$ nur all jene Kovarianzen ein, bei denen $h^{(1)}$ und $h^{(2)}$ $c = 1, \ldots, k_1$ gemeinsame Elemente besitzen. Also haben wir:

$$\begin{aligned} &Cov_F\left(U_n^{(1)}, U_n^{(2)}\right) \\ &= \binom{n}{k_1}^{-1}\binom{n}{k_2}^{-1} \\ &\quad \cdot \sum_{c=1}^{k_1} \sum_{\left|\binom{n}{k_1}\cap\binom{n}{k_2}\right|=c} Cov_F\left(h^{(1)}\left(X_{i_1},\ldots,X_{i_{k_1}}\right), h^{(2)}\left(X_{j_1},\ldots,X_{j_{k_2}}\right)\right), \end{aligned}$$

wobei bei $\sum_{\left|\binom{n}{k_1}\cap\binom{n}{k_2}\right|=c}$ über alle $\{i_1, \ldots, i_{k_1}\}, \{j_1, \ldots, j_{k_2}\} \subseteq \{1, \ldots, n\}$ mit c gemeinsamen Elementen summiert wird und $i_s \neq i_t$ und $j_u \neq j_v$ für $s \neq t$ beziehungsweise $u \neq v$ ist.

Nun überlegen wir uns wie viele Kombinationen es gibt, sodass $h^{(1)}$ und $h^{(2)}$ c gemeinsame Elemente besitzen: Für $h^{(2)}$ gibt es $\binom{n}{k_2}$ Möglichkeiten k_2 Elemente auszuwählen. Bei $h^{(1)}$ sind es damit $\binom{k_2}{c}$ Möglichkeiten für die gemeinsamen und $\binom{n-k_2}{k_1-c}$ Kombinationen für die verschiedenen Elemente. Mit der Folgerung nach Lemma 3 haben wir somit insgesamt:

$$\begin{aligned} Cov_F\left(U_n^{(1)}, U_n^{(2)}\right) &= \binom{n}{k_1}^{-1}\binom{n}{k_2}^{-1} \sum_{c=1}^{k_1} \binom{n}{k_2}\binom{k_2}{c}\binom{n-k_2}{k_1-c}\sigma_{c,c}^2 \\ &= \binom{n}{k_1}^{-1} \sum_{c=1}^{k_1} \binom{k_2}{c}\binom{n-k_2}{k_1-c}\sigma_{c,c}^2. \end{aligned}$$

□

1.2.2 Hoeffding-Zerlegung

Nachdem wir die grundlegenden stochastischen Kenngrößen von U-Statistiken untersucht haben, können wir nun zum nächsten Schritt übergehen - der sogenannten H-Zerlegung, benannt nach Wassilij Hoeffding. Dabei geht es darum, eine U-Statistik vom Grad k als Summe von k unkorrelierten U-Statistiken vom Grad $c = 1, \dots, k$ darzustellen. Dies kann beispielsweise dann nützlich sein, wenn man mit Projektionen arbeiten möchte.

Gegeben sei ein reguläres statistisches Funktional $\theta \colon \mathscr{F} \to \mathbb{R}$ vom Grad k mit $1 \leq k \leq n$ mit Kern h. U_n sei die entsprechende U-Statistik.

Zunächst definieren wir für die U-Statistiken $U_n^{(c)}$ vom Grad $c = 1, \dots, k$ die entsprechenden Kerne $h^{(c)}$ rekursiv:

$$h^{(1)} := h_1(x_1) - \theta$$

und

$$h^{(c)}(x_1, \dots, x_c) := h_c(x_1, \dots, x_c) - \left[\sum_{l=1}^{c-1} \sum_{\binom{c}{l}} h^{(l)}(x_{j_1}, \dots, x_{j_l}) \right] - \theta \tag{1}$$

für $c = 2, \dots, k$.

Weiterhin gilt:

$$\sum_{\binom{n}{k}} \sum_{\binom{k}{l}} h^{(l)}(x_{j_1}, \dots, x_{j_l}) = \binom{n-l}{k-l} \sum_{\binom{n}{l}} h^{(l)}(x_{j_1}, \dots, x_{j_l}). \tag{2}$$

Begründung: Wegen der Symmetrie von h sind auch die bedingten Erwartungswerte h_c und damit die oben konstruierten Kerne $h^{(c)}$ symmetrisch, für $c = 1, \dots, k$. Somit brauchen wir nur noch die Anzahl der Summanden vergleichen:

$$\begin{aligned}
\sum_{\binom{n}{k}} \sum_{\binom{k}{l}} 1 &= \binom{n}{k}\binom{k}{l} \\
&= \frac{n!}{k!(n-k)!}\frac{k!}{l!(k-l)!} \\
&= \frac{n!}{(n-k)!l!(k-l)!} \\
&= \frac{(n-l)!}{(k-l)!(n-k)!}\frac{n!}{l!(n-l)!} \\
&= \binom{n-l}{k-l}\binom{n}{l} \\
&= \binom{n-l}{k-l}\sum_{\binom{n}{l}} 1.
\end{aligned}$$

Aus dieser Rechnung ergibt sich die Identität:

$$\binom{n}{k}^{-1}\binom{n-l}{k-l} = \binom{k}{l}\binom{n}{l}^{-1}. \tag{3}$$

Mit (1) für $c = k$ und damit $h_c = h$ und (2) folgt:

$$\begin{aligned}
U_n = U_n(h; x_1, \ldots, x_n) &= \binom{n}{k}^{-1} \sum_{\binom{n}{k}} h(x_{i_1}, \ldots, x_{i_k}) \\
&= \binom{n}{k}^{-1} \sum_{\binom{n}{k}} \left[\left(\sum_{l=1}^{k} \sum_{\binom{k}{l}} h^{(l)}(x_{j_1}, \ldots, x_{j_l})\right) + \theta\right] \\
&= \theta + \binom{n}{k}^{-1} \sum_{l=1}^{k} \sum_{\binom{n}{k}} \sum_{\binom{k}{l}} h^{(l)}(x_{j_1}, \ldots, x_{j_l}) \\
&= \theta + \binom{n}{k}^{-1} \sum_{l=1}^{k} \binom{n-l}{k-l} \sum_{\binom{n}{l}} h^{(l)}(x_{j_1}, \ldots, x_{j_l}).
\end{aligned}$$

Für $l = 1, \ldots, k$ definieren wir

$$U_n^{(l)} = U_n^{(l)}\left(h^{(l)}; x_1, \ldots, x_n\right) := \binom{n}{l}^{-1} \sum_{\binom{n}{l}} h^{(l)}(x_{j_1}, \ldots, x_{j_l})$$

und erhalten unter Verwendung von (3):

Satz 3 (vgl. Lee (1990), S. 26)**.** *Es seien $\theta\colon \mathscr{F} \to \mathbb{R}$ ein reguläres statistisches Funktional vom Grad k mit $1 \leq k \leq n$ und h der Kern des Funktionals. Dann gilt für die zugehörige U-Statistik U_n:*

$$U_n - \theta = \sum_{l=1}^{k} \binom{k}{l} U_n^{(l)}.$$

Beweis. siehe Herleitung H-Zerlegung □

Nach Definition sind die $U_n^{(l)}$, $l = 1, \ldots, k$ selbst wieder U-Statistiken vom Grad l und mit Kern $h^{(l)}$. Somit können wir bereits eine U-Statistik vom Grad k als Summe von U-Statistiken vom Grad $l = 1, \ldots, k$ darstellen. Wie zu Beginn des Abschnitts genannt, sollen die einzelnen Summanden allerdings unkorreliert sein, was es noch zu zeigen gilt.

Dafür überlegen wir uns zunächst noch eine alternative Darstellung der Kerne $h^{(l)}$ für $l = 1, \ldots k$.

Es gilt:

$$\begin{aligned}
h^{(1)}(x_1) &= h_1(x_1) - \theta \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, y_2, \ldots, y_k) dF(y_2) \cdots dF(y_k) \\
&\qquad - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_1, \ldots, y_k) dF(y_1) \cdots dF(y_k) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_1, \ldots, y_k) \left[dD_{x_1}(y_1) - dF(y_1)\right] dF(y_2) \cdots dF(y_k),
\end{aligned}$$

wobei D die Verteilungsfunktion des Dirac-Maßes bezeichnet.

Allgemein gilt für $l = 1, \ldots, k$:

$$\begin{aligned}
&h^{(l)}(x_1, \ldots, x_l) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_1, \ldots, y_k) \\
&\qquad \cdot \left[dD_{x_1}(y_1) - dF(y_1)\right] \cdots \left[dD_{x_l}(y_l) - dF(y_l)\right] dF(y_{l+1}) \cdots dF(y_k).
\end{aligned}$$

Das sieht man folgendermaßen:

$$\begin{aligned}
&dD_{x_1}(y_1) \cdots dD_{x_k}(y_k) \\
&= \left[\left[dD_{x_1}(y_1) - dF(y_1)\right] + dF(y_1)\right] \cdots \left[\left[dD_{x_k}(y_k) - dF(y_k)\right] + dF(y_k)\right]
\end{aligned}$$

und damit ist

$$
\begin{aligned}
& dD_{x_1}(y_1)\cdots dD_{x_k}(y_k) \\
& \quad = \sum_{c=0}^{k}\sum_{\binom{k}{c}} \left[dD_{x_{i_1}}(y_{i_1}) - dF(y_{i_1})\right]\cdots\left[dD_{x_{i_c}}(y_{i_c}) - dF(y_{i_c})\right] \\
& \qquad\qquad \cdot dF(y_{i_{c+1}})\cdots dF(y_{i_k}).
\end{aligned}
$$

Dies wird klar, indem wir den Ausdruck vereinfachen:

Es seien $a_i, b_i \in \mathbb{R}$ für $i = 1, \ldots, k$. Dann ist

$$
\begin{aligned}
\prod_{i=1}^{k}(a_i + b_i) &= (a_1 + b_1)\cdots(a_k + b_k) \\
&= \sum_{c=0}^{k}\sum_{\binom{k}{c}} \left[\prod_{j=1}^{c} a_{i_j}\right]\left[\prod_{j=c+1}^{k} b_{i_j}\right].
\end{aligned}
$$

Denn: Beim Ausmultiplizieren der ersten Zeile entsteht eine Summe. Betrachten wir nun einen konkreten Summanden von dieser Summe. Mit c bezeichnen wir die Anzahl der a_i, die in diesem Summanden vorkommen, mit $c \in \{0, \ldots, k\}$. Dabei gibt es $\binom{k}{c}$ Möglichkeiten, c a_i von $a_1, \ldots, a_k$ auszuwählen. Die übrigen $(k - c)$ Faktoren b_j ergeben sich damit sofort, da es nicht möglich ist, das ein Faktorpaar $a_i b_j$ mit $i = j$ in einem Summanden vorkommt. Somit gilt für die Indizes der Faktoren eines Summanden:

$$
\{i_1, \ldots, i_c\} \cup \{i_{c+1}, \ldots, i_k\} = \{1, \ldots, k\}.
$$

Also ist nun klar:

$$
\begin{aligned}
& dD_{x_1}(y_1)\cdots dD_{x_k}(y_k) \\
& \quad = \sum_{c=0}^{k}\sum_{\binom{k}{c}} \left[dD_{x_{i_1}}(y_{i_1}) - dF(y_{i_1})\right]\cdots\left[dD_{x_{i_c}}(y_{i_c}) - dF(y_{i_c})\right] \\
& \qquad\qquad \cdot dF(y_{i_{c+1}})\cdots dF(y_{i_k}).
\end{aligned}
$$

Nun multiplizieren wir beide Seiten mit $h(y_1, \ldots, y_k)$, integrieren und erhalten:

$$
\begin{aligned}
& h(x_1, \ldots, x_k) \\
= \; & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_1, \ldots, y_k) dD_{x_1}(y_1) \cdots dD_{x_k}(y_k) \\
= \; & \sum_{c=0}^{k} \sum_{\binom{k}{c}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_{i_1}, \ldots, y_{i_k}) \\
& \qquad \cdot \left[dD_{x_{i_1}}(y_{i_1}) - dF(y_{i_1})\right] \cdots \left[dD_{x_{i_c}}(y_{i_c}) - dF(y_{i_c})\right] dF(y_{i_{c+1}}) \cdots dF(y_{i_k}) \\
= \; & \sum_{c=1}^{k} \sum_{\binom{k}{c}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_{i_1}, \ldots, y_{i_k}) \\
& \qquad \cdot \left[dD_{x_{i_1}}(y_{i_1}) - dF(y_{i_1})\right] \cdots \left[dD_{x_{i_c}}(y_{i_c}) - dF(y_{i_c})\right] dF(y_{i_{c+1}}) \cdots dF(y_{i_k}) \\
& + \theta.
\end{aligned}
$$

Mit $c = k$ für unsere rekursiv definierten Kerne $h^{(c)}$ gilt nach Umstellung:

$$
\begin{aligned}
h(x_1, \ldots, x_k) &= h_k(x_1, \ldots, x_k) \\
&= \sum_{l=1}^{k} \sum_{\binom{k}{l}} h^{(l)}\left(x_{i_1}, \ldots, x_{i_l}\right) + \theta
\end{aligned}
$$

und somit, wenn wir beide Gleichungen miteinander vergleichen:

$$
\begin{aligned}
& h^{(l)}(x_1, \ldots, x_l) \\
= \; & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_1, \ldots, y_k) \\
& \cdot \left[dD_{x_1}(y_1) - dF(y_1)\right] \cdots \left[dD_{x_l}(y_l) - dF(y_l)\right] dF(y_{l+1}) \cdots dF(y_k)
\end{aligned}
$$

für $l = 1, \ldots, k$.

Wie wir gleich sehen werden, wird es sich mit dieser alternativen Darstellung der Kerne $h^{(l)}$, $l = 1, \ldots k$ einfacher rechnen lassen.

Lemma 4 (vgl. Lee (1990), S. 28)**.** *Die Zufallsgrößen $X_1, \ldots, X_n$ seien unabhängig und identisch verteilt mit $\mathscr{L}(X_1) = F \in \mathscr{F}$. Für die H-Zerlegung einer U-Statistik U_n vom Grad k mit $1 \leq k \leq n$ mit Kern h gilt:*

(i) $h_c^{(l)}(x_1, \ldots, x_c) = 0$ *für* $l = 2, \ldots, k$ *und* $c = 1, \ldots, l-1$ *und*

(ii) $E_F h^{(l)}(X_1, \ldots, X_l) = 0$ *für* $l = 1, \ldots, k$.

Beweis.
(i) Es gilt für $l = 2, \dots, k$:

$$\begin{aligned}
& h^{(l)}(x_1, \dots, x_l) \\
= & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_1, \dots, y_k) \\
& \quad \cdot [dD_{x_1}(y_1) - dF(y_1)] \cdots [dD_{x_l}(y_l) - dF(y_l)] \, dF(y_{l+1}) \cdots dF(y_k) \\
= & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_1, \dots, y_{l-1}, x_l, y_{l+1}, \dots, y_k) \\
& \qquad \cdot [dD_{x_1}(y_1) - dF(y_1)] \cdots \left[dD_{x_{l-1}}(y_{l-1}) - dF(y_{l-1})\right] \\
& \qquad \cdot dF(y_{l+1}) \cdots dF(y_k) \\
& \quad - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_1, \dots, y_k) \\
& \qquad \cdot [dD_{x_1}(y_1) - dF(y_1)] \cdots \left[dD_{x_{l-1}}(y_{l-1}) - dF(y_{l-1})\right] \\
& \qquad \cdot dF(y_l) \cdots dF(y_k) \\
= & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_1, \dots, y_{l-1}, x_l, y_{l+1}, \dots, y_k) \\
& \qquad \cdot [dD_{x_1}(y_1) - dF(y_1)] \cdots \left[dD_{x_{l-1}}(y_{l-1}) - dF(y_{l-1})\right] \\
& \qquad \cdot dF(y_{l+1}) \cdots dF(y_k) \\
& \quad - h^{(l-1)}(x_1, \dots, x_{l-1}),
\end{aligned}$$

wobei bei dem letzten Rechenschritt die alternative Darstellung des Kerns $h^{(l-1)}$ benutzt wird.

Wenn wir nun auf beiden Seiten bezüglich F in der l-ten Komponente integrieren, erhalten wir:

$$h_{l-1}^{(l)}(x_1, \dots, x_{l-1}) = h^{(l-1)}(x_1, \dots, x_{l-1}) - h^{(l-1)}(x_1, \dots, x_{l-1}) = 0.$$

Mit Lemma 1 (i) folgt nun, dass

$$h_c^{(l)}(x_1, \dots, x_c) = E_F h_{l-1}^{(l)}(x_1, \dots, x_c, X_{c+1}, \dots, X_{l-1}) = 0$$

für $c = 1, \dots, l-2$, womit (i) gezeigt ist.

(ii) Mit (i) und Anwendung von Lemma 1 (ii) haben wir für $l = 1, \dots, k$:

$$\begin{aligned}
E_F h^{(l)}(X_1, \dots, X_l) &= E_F h_c^{(l)}(X_1, \dots, X_c) \\
&= 0
\end{aligned}$$

für $c \in \{1, \dots, l-1\}$. □

Mit diesem Lemma sind wir nun am Ziel:

Satz 4 (vgl. Lee (1990), S. 30). *Die Zufallsvariablen $X_1, \dots, X_n$ seien unabhängig und identisch verteilt mit $\mathscr{L}(X_1) = F \in \mathscr{F}$. Außerdem sei U_n eine U-Statistik vom Grad k mit $1 \leq k \leq n$ mit Kern h. Dann gilt für deren H-Zerlegung Folgendes:*

(i) Es sei $1 \leq l < m \leq k$. Dann gilt:

$$Cov_F\left(U_n^{(l)}, U_n^{(m)}\right) = Cov_F\left(h^{(l)}(X_{i_1}, \dots, X_{i_l}), h^{(m)}(X_{j_1}, \dots, X_{j_m})\right) = 0,$$

wobei $\{i_1, \dots, i_l\}, \{j_1, \dots, j_m\} \subseteq \{1, \dots, n\}$ und $i_s \neq i_t$ für $s \neq t$ beziehungsweise $j_u \neq j_v$ für $u \neq v$.

(ii) Für $\{i_1, \dots, i_l\}, \{i'_1, \dots, i'_l\} \subseteq \{1, \dots, n\}$ mit $i_s \neq i_t$ für $s \neq t$ beziehungsweise $i'_{s'} \neq i'_{t'}$ für $s' \neq t'$ und $\{i_1, \dots, i_l\} \neq \{i'_1, \dots, i'_l\}$ ist

$$Cov_F\left(h^{(l)}(X_{i_1}, \dots, X_{i_l}), h^{(l)}\left(X_{i'_1}, \dots, X_{i'_l}\right)\right) = 0,$$

für $l = 1, \dots, k$.

Beweis.
(i)

$$\begin{aligned} & Cov_F\left(h^{(l)}(X_{i_1}, \dots, X_{i_l}), h^{(m)}(X_{j_1}, \dots, X_{j_m})\right) \\ = \;& E_F\left[h^{(l)}(X_{i_1}, \dots, X_{i_l}) h^{(m)}(X_{j_1}, \dots, X_{j_m})\right] \\ & - \left[E_F h^{(l)}(X_{i_1}, \dots, X_{i_l})\right]\left[E_F h^{(m)}(X_{j_1}, \dots, X_{j_m})\right]. \end{aligned}$$

Wegen Lemma 4 (ii) genügt es zu zeigen, dass

$$E_F\left[h^{(l)}(X_{i_1}, \dots, X_{i_l}) h^{(m)}(X_{j_1}, \dots, X_{j_m})\right] = 0.$$

Da $l < m$ gilt, existiert ein $j_u \in \{j_1, \dots, j_m\}$ mit $j_u \notin \{i_1, \dots, i_l\}$. Somit existiert eine Zufallsvariable X_{j_u}, welche in $h^{(m)}$ aber nicht in $h^{(l)}$ vorkommt. Wegen der Unabhängigkeit von X_{j_u} und $h^{(l)}$ und mit Lemma 4 (i) folgt

$$\begin{aligned} & E_F\left[h^{(l)}(X_{i_1}, \dots, X_{i_l}) h^{(m)}(X_{j_1}, \dots, X_{j_m})\right] \\ = \;& E_F\left[E_F\left[h^{(l)}(X_{i_1}, \dots, X_{i_l}) h^{(m)}(X_{j_1}, \dots, X_{j_m}) \mid X_{j_u}\right]\right] \\ = \;& E_F\left[h^{(l)}(X_{i_1}, \dots, X_{i_l})\left[E_F\left[h^{(m)}(X_{j_1}, \dots, X_{j_m}) \mid X_{j_u}\right]\right]\right] \\ = \;& E_F\left[h^{(l)}(X_{i_1}, \dots, X_{i_l}) h_1^{(m)}(X_{j_u})\right] \\ = \;& 0. \end{aligned}$$

Außerdem ist

$$
\begin{aligned}
&Cov_F\left(U_n^{(l)}, U_n^{(m)}\right) \\
&= Cov_F\left(\binom{n}{l}^{-1}\sum_{\binom{n}{l}} h^{(l)}(X_{i_1},\ldots,X_{i_l}), \binom{n}{m}^{-1}\sum_{\binom{n}{m}} h^{(m)}(X_{j_1},\ldots,X_{j_m})\right) \\
&= \binom{n}{l}^{-1}\binom{n}{m}^{-1}\sum_{\binom{n}{l}}\sum_{\binom{n}{m}} Cov_F\left(h^{(l)}(X_{i_1},\ldots,X_{i_l}), h^{(m)}(X_{j_1},\ldots,X_{j_m})\right) \\
&= 0.
\end{aligned}
$$

(ii)

Falls $\{i_1,\ldots,i_l\}\cap\{i'_1,\ldots,i'_l\}=\emptyset$, dann ist aufgrund der Unabhängigkeit klar, dass

$$Cov_F\left(h^{(l)}(X_{i_1},\ldots,X_{i_l}), h^{(l)}\left(X_{i'_1},\ldots,X_{i'_l}\right)\right)=0.$$

Falls $\{i_1,\ldots,i_l\}\cap\{i'_1,\ldots,i'_l\}\neq\emptyset$, betrachten wir

$$\{i_1,\ldots,i_c\}:=\{i_1,\ldots,i_l\}\setminus(\{i_1,\ldots,i_l\}\cap\{i'_1,\ldots,i'_l\})$$

mit $1\le c\le l-1$. Wie schon in (i) genügt es zu zeigen, dass

$$E_F\left[h^{(l)}(X_{i_1},\ldots,X_{i_l})h^{(l)}\left(X_{i'_1},\ldots,X_{i'_l}\right)\right]=0.$$

Dann ist aber wegen der Unabhängigkeit von $X_{i_1},\ldots,X_{i_c}$ und $h^{(l)}\left(X_{i'_1},\ldots,X_{i'_l}\right)$ und wieder wegen Lemma 4 (i)

$$
\begin{aligned}
&E_F\left[h^{(l)}(X_{i_1},\ldots,X_{i_l})h^{(l)}\left(X_{i'_1},\ldots,X_{i'_l}\right)\right] \\
&= E_F\left[E_F\left[h^{(l)}(X_{i_1},\ldots,X_{i_l})h^{(l)}\left(X_{i'_1},\ldots,X_{i'_l}\right)\mid X_{i_1},\ldots,X_{i_c}\right]\right] \\
&= E_F\left[h^{(l)}\left(X_{i'_1},\ldots,X_{i'_l}\right)\left[E_F\left[h^{(l)}(X_{i_1},\ldots,X_{i_l})\mid X_{i_1},\ldots,X_{i_c}\right]\right]\right] \\
&= E_F\left[h^{(l)}\left(X_{i'_1},\ldots,X_{i'_l}\right)h_c^{(l)}(X_{i_1},\ldots,X_{i_c})\right] \\
&= 0.
\end{aligned}
$$

Eine analoge Rechnung ergibt sich für

$$\{i'_1,\ldots,i'_{c'}\}:=\{i'_1,\ldots,i'_l\}\setminus(\{i_1,\ldots,i_l\}\cap\{i'_1,\ldots,i'_l\})$$

mit $1\le c'\le l-1$. □

1.3 V-Statistiken

Um eine weitere Klasse von Statistiken, den sogenannten V-Statistiken, die von Richard von Mises 1947 in „On the asymptotic distribution of differentiable statistical functions“ eingeführt wurden, wird es im nun folgenden Abschnitt gehen. Auch für V-Statistiken gilt, dass sich sehr viele Statistiken exakt oder zumindest approximativ als V-Statistik darstellen lassen.

1.3.1 Definition und Beziehung zu U-Statistiken

Definition 3 (vgl. Lee (1990), S. 183)**.** *Es seien $\theta\colon \mathscr{F} \to \mathbb{R}$ ein reguläres statistisches Funktional vom Grad k und h der Kern des Funktionals. Dann ist*

$$V_n = V_n(h; X_1, \ldots, X_n) := \frac{1}{n^k} \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} h(X_{i_1}, \ldots, X_{i_k})$$

die zugehörige V-Statistik.

Zu erkennen ist, dass V-Statistiken eine gewisse Ähnlichkeit zu U-Statistiken aufweisen. Bei Ersteren wird allerdings zusätzlich über Schätzer mit $c = 1, \ldots, k$ gleichen Argumenten summiert und entsprechend normiert.

Wie nun der genaue Zusammenhang zwischen U- und V-Statistiken aussieht, wird uns der nächste Satz zeigen.

Dazu noch eine kurze Bemerkung zu den sogenannten Stirling-Zahlen, konkret zu den Stirling-Zahlen zweiter Art:

Die Stirling-Zahl zweiter Art $S_{k|l}$ mit $1 \leq l \leq k$ gibt die Anzahl der Möglichkeiten an, eine k-elementige Menge in l nichtleere, disjunkte Teilmengen aufzuteilen.

Eine Formel zur Berechnung von $S_{k|l}$ werden wir hier nur angeben und nicht beweisen:

$$S_{k|l} = \frac{1}{l!} \sum_{i=0}^{l} (-1)^{l-i} \binom{l}{i} i^k.$$

Eine genaue Herleitung findet man beispielsweise in Comtet (1974) (S. 204).

Satz 5 (vgl. Lee (1990), S. 183)**.** *Es sei V_n eine V-Statistik vom Grad k mit $1 \leq k \leq n$ mit Kern h. Dann gilt:*

$$V_n = \frac{1}{n^k} \sum_{l=1}^{k} l! S_{k|l} \binom{n}{l} U_n^{(l)}.$$

Dabei ist $U_n^{(l)}$ eine U-Statistik vom Grad l mit

$$U_n^{(l)} = \binom{n}{l}^{-1} \sum_{\binom{n}{l}} h^{(l)}(X_{j_1}, \ldots, X_{j_l}).$$

Der Kern $h^{(l)}$ von $U_n^{(l)}$ ist gegeben durch

$$h^{(l)}(x_{j_1}, \ldots, x_{j_l}) = \frac{1}{l! S_{k|l}} \sum_{\substack{(i_1,\ldots,i_k)\in\{j_1,\ldots,j_l\}^k: \\ |\{i_1,\ldots,i_k\}|=l}} h(x_{i_1}, \ldots, x_{i_k}),$$

wobei über alle k-Tupel $(i_1, \ldots, i_k)$ mit l verschiedenen Elementen summiert wird, deren Elemente aus der Menge $\{j_1, \ldots, j_l\} \subseteq \{1, \ldots, n\}$ mit $j_u \neq j_v$ für $u \neq v$ sind.

Beweis. Zuerst zeigen wir, dass gilt:

$$n^k = \sum_{l=1}^{k} l! S_{k|l} \binom{n}{l}.$$

Die Menge aller möglichen k-Tupel von $\{1, \ldots, n\}$ kann man in k disjunkte Teilmengen zerlegen, wobei eine solche Teilmenge dann nur k-Tupel mit l verschiedenen Elementen enthält, mit $l = 1, \ldots, k$. Für die Menge aller k-Tupel mit l verschiedenen Elementen gibt es $S_{k|l}$ Möglichkeiten. Dabei ist zu beachten, dass ein k-Tupel $(i_1, \ldots, i_k)$ als Menge $\{i_1, \ldots, i_k\}$ betrachtet wird, auch wenn gilt: $i_s = i_t$ für ein $s \neq t$. Für die Auswahl dieser l verschiedenen Elemente gibt es wiederum $n(n-1)\cdots(n-l+1)$ Möglichkeiten, also insgesamt:

$$\begin{aligned}
n^k &= \sum_{l=1}^{k} S_{k|l} n(n-1)\cdots(n-l+1) \\
&= \sum_{l=1}^{k} l! S_{k|l} n(n-1)\cdots(n-l+1) \frac{(n-l)!}{l!(n-l)!} \\
&= \sum_{l=1}^{k} l! S_{k|l} \frac{n!}{l!(n-l)!} \\
&= \sum_{l=1}^{k} l! S_{k|l} \binom{n}{l}.
\end{aligned}$$

Im zweiten Schritt überprüfen wir, dass

$$\sum_{\substack{(i_1,\ldots,i_k)\in\{j_1,\ldots,j_l\}^k: \\ |\{i_1,\ldots,i_k\}|=l}} 1 = l! S_{k|l}.$$

Dies folgt mit einer ähnlichen Überlegung wie im ersten Schritt des Beweises. Der Unterschied hier ist, dass die zugrundeliegende Menge nicht $\{1, \ldots, n\}$ sondern $\{j_1, \ldots, j_l\}$

ist. Für $(i_1, \ldots, i_k)$ gibt es nun $S_{k|l}$ Möglichkeiten für die Wahl von $i_1, \ldots, i_k$, sodass gilt: $|\{i_1, \ldots, i_k\}| = l$. Gleichzeitig gibt es $l!$ Möglichkeiten für die Anordnung der l Elemente aus $\{j_1, \ldots, j_l\}$ in $(i_1, \ldots, i_k)$.

Somit können wir den Beweis abschließen:

$$\begin{aligned}
n^k V_n &= \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} h(X_{i_1}, \ldots, X_{i_k}) \\
&= \sum_{l=1}^{k} \sum_{\substack{(i_1,\ldots,i_k)\in\{1,\ldots,n\}^k: \\ |\{i_1,\ldots,i_k\}|=l}} h(X_{i_1}, \ldots, X_{i_k}) \\
&= \sum_{l=1}^{k} \sum_{\binom{n}{l}} l! S_{k|l} h^{(l)}(X_{j_1}, \ldots, X_{j_l}) \\
&= \sum_{l=1}^{k} l! S_{k|l} \sum_{\binom{n}{l}} h^{(l)}(X_{j_1}, \ldots, X_{j_l}) \\
&= \sum_{l=1}^{k} l! S_{k|l} \binom{n}{l} U_n^{(l)}.
\end{aligned}$$

□

Auf die Eigenschaften von V-Statistiken werden wir hier nicht weiter eingehen. Zu sagen, dass die Herleitung der Charakteristika aufgrund der Darstellung einer V-Statistik als Summe von U-Statistiken trivial ist, wäre sicherlich nicht korrekt. Aber über längere Rechnungen mit einigen Modifikationen, vor allem bezüglich der Korrelation zwischen den einzelnen U-Statistiken, beziehungsweise über neue Überlegungen von Grund auf, kommt man zu analogen Resultaten.

Vielmehr werden wir uns jetzt noch in diesem Kapitel mit dem Begriff der Degeneration und konkreten Beispielen für U- und V-Statistiken befassen.

1.4 Degeneration

Nachdem wir uns bis jetzt ausschließlich mit gewöhnlichen U- und V-Statistiken beschäftigt haben, kommen wir nun zum degenerierten Fall. Dazu zunächst eine Definition zum Begriff:

Definition 4 (vgl. Lee (1990), S. 78/83)**.** *Es sei U_n eine U-Statistik vom Grad k mit $1 \leq k \leq n$ mit Kern h. Dann besitzt U_n Degeneration d-ter Ordnung, falls gilt:*

$$\sigma_1^2 = \cdots = \sigma_d^2 = 0$$

und

$$\sigma^2_{d+1} > 0$$

für $d \in \{0, \ldots, k-1\}$.

Analoges gilt für eine V*-Statistik* V_n *vom Grad* k *mit Kern* h.

Für den Sonderfall $d = 0$ liegt keine Degeneration vor. Für $d = k-1$ hingegen nennt man die U- beziehungsweise V-Statistik auch vollständig degeneriert.

1.5 Beispiele

Ausgangspunkt sind auch hier die Zufallsgrößen $X_1, \ldots, X_n$, welche unabhängig und identisch verteilt seien mit $\mathscr{L}(X_1) = F \in \mathscr{F}$.

(U1) Es sei $\theta = E_F X_1$. Dann ist für $h(x) := x$

$$\begin{aligned} U_n &= \binom{n}{1}^{-1} \sum_{\binom{n}{1}} h(X_i) \\ &= \frac{1}{n} \sum_{i=1}^{n} X_i \\ &= \overline{X}_n, \end{aligned}$$

wobei $\overline{X}_n$ das Stichprobenmittel bezeichnet.

(U2) Für $\theta = Var_F X_1$ und $h(x_1, x_2) := \frac{(x_1-x_2)^2}{2}$ ist

$$\begin{aligned}
U_n &= \binom{n}{2}^{-1} \sum_{\binom{n}{2}} h(X_i, X_j) \\
&= \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \frac{(X_i - X_j)^2}{2} \\
&= \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} (X_i - X_j)^2 \\
&= \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} (X_i - X_j)^2 \\
&= \frac{1}{2n(n-1)} \left[n \sum_{i=1}^{n} X_i^2 - 2 \sum_{i=1}^{n} \sum_{j=1}^{n} X_i X_j + n \sum_{j=1}^{n} X_j^2 \right] \\
&= \frac{1}{n-1} \sum_{i=1}^{n} X_i^2 - \frac{1}{n(n-1)} \left[\sum_{i=1}^{n} X_i \right] \left[\sum_{j=1}^{n} X_j \right] \\
&= \frac{1}{n-1} \sum_{i=1}^{n} X_i^2 - \frac{n}{n-1} \overline{X}_n^2 \\
&= \frac{1}{n-1} \left[\sum_{i=1}^{n} X_i^2 - n\overline{X}_n^2 \right] \\
&= \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n \right)^2 \\
&= S_n^2,
\end{aligned}$$

mit S_n^2 als (korrigierte) Stichprobenvarianz.

(U3) Für $\theta = F(q_{p_0})$, wobei q_{p_0} das p_0-Quantil bezüglich F und $p_0 \in [0,1]$ ist, und $h(x) := 1_{(-\infty, q_{p_0}]}(x)$ ist

$$\begin{aligned}
U_n &= \binom{n}{1}^{-1} \sum_{\binom{n}{1}} h(X_i) \\
&= \frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty, q_{p_0}]}(X_i) \\
&= F_n(q_{p_0}),
\end{aligned}$$

wobei F_n die empirische Verteilungsfunktion bezeichnet.

(U4) Es seien $E_F X_1 = \mu$ und $Var_F X_1 = \sigma^2 > 0$. Für $\theta = \mu^2$ und $h(x_1, x_2) := x_1 x_2$ ist

$$\begin{aligned} U_n &= \binom{n}{2}^{-1} \sum_{\binom{n}{2}} h(X_i, X_j) \\ &= \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} X_i X_j \\ &= \frac{1}{n(n-1)} \left[\sum_{i=1}^{n} \sum_{j=1}^{n} X_i X_j - \sum_{i=1}^{n} X_i^2 \right] \\ &= \frac{1}{n-1} \left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \right)^2 - \frac{1}{n} \sum_{i=1}^{n} X_i^2 \right]. \end{aligned}$$

Betrachten wir nun

$$\begin{aligned} \sigma_1^2 &= Cov_F(h(X_1, X_2), h(X_2, X_3)) \\ &= Cov_F(X_1 X_2, X_2 X_3) \\ &= E_F \left(X_1 X_2^2 X_3 \right) - [E_F(X_1 X_2)][E_F(X_2 X_3)] \\ &= [E_F X_1] \left[E_F X_2^2 \right] [E_F X_3] - [E_F X_1][E_F X_2][E_F X_2][E_F X_3] \\ &= [E_F X_1][E_F X_3] \left[E_F X_2^2 - [E_F X_2]^2 \right] \\ &= [E_F X_1][E_F X_3][Var_F X_2] \\ &= [E_F X_1][E_F X_1][Var_F X_1] \\ &= \mu^2 \sigma^2 \end{aligned}$$

und

$$\begin{aligned} \sigma_2^2 &= Cov_F(h(X_1, X_2), h(X_1, X_2)) \\ &= Var_F h(X_1, X_2) \\ &= Var_F(X_1 X_2) \\ &= E_F \left(X_1^2 X_2^2 \right) - [E_F(X_1 X_2)]^2 \\ &= \left[E_F X_1^2 \right] \left[E_F X_2^2 \right] - [E_F X_1][E_F X_2][E_F X_1][E_F X_2] \\ &= \left[E_F X_1^2 \right]^2 - [E_F X_1]^4 \\ &= \left(\sigma^2 + \mu^2 \right)^2 - \mu^4, \end{aligned}$$

wobei bei der Berechnung von σ_i^2, $i = 1, 2$, Lemma 2 verwendet wird.

Falls nun $\mu = 0$ gilt, so besitzt U_n Degeneration erster Ordnung beziehungsweise ist U_n vollständig degeneriert, da $\sigma_1^2 = 0$ und $\sigma_2^2 = \sigma^4 > 0$ sind.

Nach dem zentralen Grenzwertsatz (vgl. Billingsley (1995), S. 357) gilt, dass

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \xrightarrow{d} \widetilde{Z} \sim N\left(0, \sigma^2\right),$$

wobei „$\xrightarrow{d}$“ die Konvergenz in Verteilung bezeichnet.

Weiterhin gilt nach dem schwachen Gesetz der großen Zahlen von Chintschin (vgl. Georgii (2009), S. 123):

$$\frac{1}{n} \sum_{i=1}^{n} X_i^2 \xrightarrow{P_F} E_F X_1^2 = Var_F X_1 = \sigma^2,$$

wobei „$\xrightarrow{P_F}$“ die Konvergenz in Wahrscheinlichkeit (bezüglich P_F) bezeichnet.

Insgesamt erhalten wir mit dem Continuous Mapping Theorem (vgl. Billingsley (1999), S. 21) und dem Lemma von Sluzki (vgl. Georgii (2009), S. 292):

$$nU_n \xrightarrow{d} \widetilde{Z}^2 - \sigma^2 = \sigma^2\left(Z^2 - 1\right),$$

mit $Z \sim N(0,1)$.

(V1) Es sei $\theta = Var_F X_1$. Dann ist für $h(x_1, x_2) := \frac{(x_1 - x_2)^2}{2}$

$$\begin{aligned}
V_n &= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} h(X_i, X_j) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(X_i - X_j)^2}{2} \\
&= \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(X_i^2 - 2X_iX_j + X_j^2\right) \\
&= \frac{1}{2n^2} \left[n \sum_{i=1}^{n} X_i^2 - 2 \sum_{i=1}^{n} \sum_{j=1}^{n} X_iX_j + n \sum_{j=1}^{n} X_j^2 \right] \\
&= \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \frac{1}{n^2} \left[\sum_{i=1}^{n} X_i \right] \left[\sum_{j=1}^{n} X_j \right] \\
&= \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \overline{X}_n^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2 \\
&= \widetilde{S}_n^2
\end{aligned}$$

mit $\widetilde{S}_n^2$ als (unkorrigierte) Stichprobenvarianz.

(V2) Wenn wir überprüfen wollen, ob die Zufallsvariablen $X_1, \ldots, X_n$ auf eine bestimmte Weise verteilt sind, können wir den „Chi-Quadrat-Anpassungstest“ verwenden. Dazu betrachten wir eine Partition $A_1, \ldots, A_L$ von $\mathbb{R}$ und bezeichnen mit n_l die Anzahl der X_i, die in A_l vorkommen, $l = 1, \ldots, L$. Unter der Nullhypothese $H_0\colon F = F_0$ mit $F_0 \in \mathscr{F}$ gilt: $E_F n_l = n_l^{(0)}$, wobei sich $n_l^{(0)}$ als $n_l^{(0)} = np_l^{(0)}$ ergibt, mit $p_l^{(0)} = \int_{A_l} dF_0(x_1)$, für $l = 1, \ldots, L$. Weiterhin sei sinnvollerweise $\sum_{l=1}^{L} p_l^{(0)} = 1$ und $\sum_{l=1}^{L} n_l^{(0)} = \sum_{l=1}^{L} n_l = n$. Dann ist

$$\chi_n^2 := \sum_{l=1}^{L} \frac{\left[n_l - n_l^{(0)}\right]^2}{n_l^{(0)}}$$

die entsprechende Teststatistik für den Chi-Quadrat-Anpassungstest.

Damit ist

$$\begin{aligned}
\frac{1}{n}\chi_n^2 &= \frac{1}{n}\sum_{l=1}^{L} \frac{\left[n_l - n_l^{(0)}\right]^2}{n_l^{(0)}} \\
&= \frac{1}{n}\sum_{l=1}^{L} \frac{n_l^2 - 2n_l n_l^{(0)} + \left[n_l^{(0)}\right]^2}{n_l^{(0)}} \\
&= \frac{1}{n}\left[\sum_{l=1}^{L} \frac{n_l^2}{n_l^{(0)}} - \left(2\sum_{l=1}^{L} n_l - \sum_{l=1}^{L} n_l^{(0)}\right)\right] \\
&= \frac{1}{n}\left[\left(\frac{1}{n}\sum_{l=1}^{L} \frac{n_l^2}{p_l^{(0)}}\right) - n\right] \\
&= \frac{1}{n^2}\left[\sum_{l=1}^{L} \frac{n_l^2}{p_l^{(0)}}\right] - 1 \\
&= \frac{1}{n^2}\left[\sum_{l=1}^{L} \frac{1}{p_l^{(0)}}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} 1_{A_l}(X_i)1_{A_l}(X_j)\right)\right] - 1 \\
&= \frac{1}{n^2}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{L} \frac{1}{p_l^{(0)}} 1_{A_l}(X_i)1_{A_l}(X_j)\right] - 1 \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\left(\sum_{l=1}^{L} \frac{1}{p_l^{(0)}} 1_{A_l}(X_i)1_{A_l}(X_j)\right) - 1\right].
\end{aligned}$$

Somit lässt sich die (normierte) Teststatistik mit

$$h(x_1, x_2) := \left[\sum_{l=1}^{L} \frac{1}{p_l^{(0)}} 1_{A_l}(x_1)1_{A_l}(x_2)\right] - 1$$

als V-Statistik darstellen:

$$\frac{1}{n}\chi_n^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} h(X_i, X_j) = V_n.$$

Nun ist unter $H_0\colon F = F_0$

$$\begin{aligned} h_1(x_1) &= \int_{-\infty}^{\infty} h(x_1, x_2) dF(x_2) \\ &= \int_{-\infty}^{\infty} \left[\left(\sum_{l=1}^{L} \frac{1}{p_l^{(0)}} 1_{A_l}(x_1) 1_{A_l}(x_2) \right) - 1 \right] dF(x_2) \\ &= \left[\sum_{l=1}^{L} \frac{1}{p_l^{(0)}} 1_{A_l}(x_1) \int_{-\infty}^{\infty} 1_{A_l}(x_2) dF(x_2) \right] - 1 \\ &= \left[\sum_{l=1}^{L} 1_{A_l}(x_1) \right] - 1 \\ &= 0. \end{aligned}$$

Somit ist $\sigma_1^2 = 0$, das heißt V_n besitzt Degeneration erster Ordnung, was in diesem Fall der vollständigen Degeneration von V_n entspricht.

(V3) In diesem letzten Beispiel soll es um die sogenannte „Cramér-von Mises-Statistik“ gehen. Diese kann verwendet werden, wenn wir testen wollen, ob die Verteilungsfunktion $F \in \mathscr{F}$ von den unabhängig und identisch verteilten Zufallsgrößen $X_1, \ldots, X_n$ mit einer bekannten Verteilungsfunktion $F_0 \in \mathscr{F}$ übereinstimmt. Dabei sei $H_0\colon F = F_0$ die Nullhypothese. Dann ist

$$\omega_n^2 := n \int_{-\infty}^{\infty} [F_0(t) - F_n(t)]^2 dF_0(t)$$

die zugehörige Cramér-von Mises-Statistik, wobei F_n wieder die empirische Verteilungsfunktion bezeichnet.

Eine alternative Darstellung ist Folgende:

$$\begin{aligned} \frac{1}{n}\omega_n^2 &= \int_{-\infty}^{\infty} [F_n(t) - F_0(t)]^2 dF_0(t) \\ &= \int_{-\infty}^{\infty} \left[\frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,t]}(X_i) - F_0(t) \right] \left[\frac{1}{n}\sum_{j=1}^{n} 1_{(-\infty,t]}(X_j) - F_0(t) \right] dF_0(t) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \int_{-\infty}^{\infty} \left[1_{(-\infty,t]}(X_i) - F_0(t)\right] \left[1_{(-\infty,t]}(X_j) - F_0(t)\right] dF_0(t). \end{aligned}$$

Mit

$$h(x_1, x_2) := \int_{-\infty}^{\infty} \left[1_{(-\infty,t]}(x_1) - F_0(t)\right] \left[1_{(-\infty,t]}(x_2) - F_0(t)\right] dF_0(t)$$

lässt sich auch die (normierte) Cramér-von Mises-Statistik als V-Statistik darstellen:

$$\frac{1}{n}\omega_n^2 = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} h(X_i, X_j) = V_n.$$

Dass V_n unter $H_0\colon F = F_0$ Degeneration erster Ordnung besitzt beziehungsweise vollständig degeneriert ist, rechnen wir leicht unter Verwendung des Satzes von Fubini (vgl. Dunford und Schwartz (1988a), S. 193) nach:

$$\begin{aligned}
& h_1(x_1) \\
= & \int_{-\infty}^{\infty} h(x_1, x_2) dF(x_2) \\
= & \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \left[1_{(-\infty,t]}(x_1) - F(t)\right] \left[1_{(-\infty,t]}(x_2) - F(t)\right] dF(t)\right] dF(x_2) \\
= & \int_{-\infty}^{\infty} 1_{(-\infty,t]}(x_1) \left[\int_{-\infty}^{\infty} 1_{(-\infty,t]}(x_2) dF(x_2)\right] dF(t) \\
& \quad - \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} dF(x_2)\right] 1_{(-\infty,t]}(x_1) F(t) dF(t) \\
& \quad - \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} 1_{(-\infty,t]}(x_2) dF(x_2)\right] F(t) dF(t) \\
& \quad + \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} dF(x_2)\right] F^2(t) dF(t) \\
= & \int_{-\infty}^{\infty} 1_{(-\infty,t]}(x_1) F(t) dF(t) \\
& \quad - \int_{-\infty}^{\infty} 1_{(-\infty,t]}(x_1) F(t) dF(t) \\
& \quad - \int_{-\infty}^{\infty} F^2(t) dF(t) \\
& \quad + \int_{-\infty}^{\infty} F^2(t) dF(t) \\
= & \ 0,
\end{aligned}$$

und damit gilt: $\sigma_1^2 = 0$.

2 Asymptotik - Konvergenz in Verteilung

Man kann für die Verteilung nicht-degenerierter U- und V-Statistiken zeigen, dass diese schwach gegen eine Normalverteilung konvergiert.

Da der Schwerpunkt dieser Arbeit allerdings bei degenerierten U- und V-Statistiken liegt, werden wir diese Fälle genauer untersuchen. Wir werden uns dabei auf U- und V-Statistiken vom Grad 2 beschränken, da diese in der Praxis am meisten Anwendung finden, wie es auch teilweise schon aus den Beispielen vom Ende von Kapitel 1 hervorging. Dementsprechend betrachten wir vollständig degenerierte U- und V-Statistiken.

2.1 Degenerierte U-Statistiken

Satz 6 (vgl. Lee (1990), S. 79)**.** *Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängig und identisch verteilter Zufallsvariablen mit $X_n\colon (\Omega,\mathcal{A}) \to (\mathbb{R},\mathcal{B})$ für alle $n \in \mathbb{N}$ und $\mathscr{L}(X_1) = F \in \mathscr{F}$. Weiterhin seien $\theta\colon \mathscr{F} \to \mathbb{R}$ ein reguläres statistisches Funktional vom Grad 2 mit Kern h und U_n die entsprechende U-Statistik. Für U_n gelte:*

(i) $E_F h^2(X_1, X_2) < \infty$ *und*

(ii) $\sigma_1^2 = Var_F h_1(X_1) = 0.$

Dann gilt:

$$n(U_n - \theta) \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k \left(Z_k^2 - 1\right),$$

wobei $Z_1, Z_2, \ldots$ unabhängig standardnormalverteilt und $\lambda_1, \lambda_2, \ldots$ Eigenwerte der Integralgleichung

$$\int_{-\infty}^{\infty} h(x_1, x_2) f(x_2) dF(x_2) = \lambda f(x_1)$$

sind.

Beweis. Ohne Beschränkung der Allgemeinheit sei unter Verwendung von Lemma 1 (ii)

$$\theta = E_F U_n = E_F h(X_1, X_2) = E_F h_1(X_1) = 0$$

und somit, wegen $\sigma_1^2 = 0$:

$$h_1 = 0 \ P_F\text{-fast überall.}$$

Aus der Fredholm-Theorie für Integralgleichungen (vgl. Dunford und Schwartz (1988b), S. 1009), auf die hier nur verwiesen wird, folgt, dass λ_k und f_k für $k \in \mathbb{N}$ existieren, mit

$$\int_{-\infty}^{\infty} h(x_1, x_2) f_k(x_2) dF(x_2) = \lambda_k f_k(x_1).$$

Dabei ist $f_k\colon \mathbb{R} \to \mathbb{R}$ für alle $k \in \mathbb{N}$ ($\mathcal{B}$-$\mathcal{B}$)-messbar.

Dann lässt sich h darstellen als

$$h(x_1, x_2) = \sum_{k=1}^{\infty} \lambda_k f_k(x_1) f_k(x_2),$$

wobei für $K \in \mathbb{N}$ gilt:

$$\int_{-\infty}^{\infty} \left| h(x_1, x_2) - \sum_{k=1}^{K} \lambda_k f_k(x_1) f_k(x_2) \right|^2 dF(x_1) dF(x_2) \xrightarrow[K \to \infty]{} 0.$$

Schließlich bilden die Eigenfunktionen eine Orthonormalbasis, das heißt

$$\int_{-\infty}^{\infty} f_k(x_1) f_l(x_1) dF(x_1) = \begin{cases} 1, & \text{falls } k = l \\ 0, & \text{falls } k \neq l \end{cases} .$$

Für $K \in \mathbb{N}$ definieren wir nun die Kerne

$$h_K(x_1, x_2) := \sum_{k=1}^{K} \lambda_k f_k(x_1) f_k(x_2)$$

und die entsprechenden U-Statistiken

$$\begin{aligned} U_{n,K} &:= \binom{n}{2}^{-1} \sum_{\binom{n}{2}} h_K(X_i, X_j) \\ &= \binom{n}{2}^{-1} \sum_{\binom{n}{2}} \sum_{k=1}^{K} \lambda_k f_k(X_i) f_k(X_j) \\ &= \sum_{k=1}^{K} \lambda_k \binom{n}{2}^{-1} \sum_{\binom{n}{2}} f_k(X_i) f_k(X_j). \end{aligned}$$

Mit den U-Statistiken

$$T_{n,k} := \binom{n}{2}^{-1} \sum_{\binom{n}{2}} f_k(X_i) f_k(X_j)$$

mit Kern $f_k(x_1) f_k(x_2)$ für $k = 1, \ldots, K$ erhalten wir:

$$U_{n,K} = \sum_{k=1}^{K} \lambda_k T_{n,k}.$$

Weiterhin können wir für $nT_{n,k}$ schreiben:

$$\begin{aligned} nT_{n,k} &= n\frac{2}{n(n-1)}\sum_{1\leq i<j\leq n} f_k(X_i)f_k(X_j) \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} f_k(X_i)f_k(X_j) - \sum_{i=1}^{n} f_k^2(X_i)\right] \\ &= \frac{n}{n-1}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_k(X_i)\right)^2 - \frac{1}{n}\sum_{i=1}^{n} f_k^2(X_i)\right]. \end{aligned}$$

Für $k = 1, \ldots, K$ und $i = 1, \ldots, n$ setzen wir

$$\widetilde{Z}_{k,i} := f_k(X_i)$$

und fassen die $\widetilde{Z}_{k,i}$ in den Vektoren

$$\widetilde{Z}_i := \left(\widetilde{Z}_{1,i}, \ldots, \widetilde{Z}_{K,i}\right)^T$$

zusammen.

Die Zufallsvektoren $\widetilde{Z}_1, \ldots, \widetilde{Z}_n$ sind wegen der Unabhängigkeit und der identischen Verteilung von $X_1, \ldots, X_n$ ebenfalls unabhängig und identisch verteilt, wobei

$$E_F\widetilde{Z}_i = 0_K$$

und

$$Cov_F\widetilde{Z}_i = I_{K\times K}$$

ist, für $i = 1, \ldots, n$, mit 0_K als Nullvektor der Länge K und $I_{K\times K}$ als Einheitsmatrix der Dimension $K \times K$.

Dies wird folgendermaßen klar:

Annahmegemäß gilt:

$$0 = h_1(x_1) = E_F[h(X_1, X_2) \mid X_1 = x_1] = \int_{-\infty}^{\infty} h(x_1, x_2)dF(x_2).$$

Dies erfüllt die Eigenfunktion $f_{k'} \equiv 1$ zu dem Eigenwert $\lambda_{k'} = 0$ für ein $k' \in \mathbb{N}$, sodass gilt:

$$\int_{-\infty}^{\infty} h(x_1, x_2)f_{k'}(x_2)dF(x_2) = \lambda_{k'}f_{k'}(x_1).$$

Da die Eigenfunktionen f_k für $k \in \mathbb{N}$ ein Orthonormalsystem bilden, gilt nun:

$$\begin{aligned} E_F \widetilde{Z}_{k,i} &= E_F f_k(X_i) \\ &= \int_{-\infty}^{\infty} f_k(x_1) dF(x_1) \\ &= \int_{-\infty}^{\infty} 1 \cdot f_k(x_1) dF(x_1) \\ &= \int_{-\infty}^{\infty} f_{k'}(x_1) f_k(x_1) dF(x_1) \\ &= \begin{cases} 1, & \text{falls } k = k' \\ 0, & \text{falls } k \neq k' \end{cases}, \end{aligned}$$

für $i = 1, \ldots, n$.

In der Darstellung

$$h(x_1, x_2) = \sum_{k=1}^{\infty} \lambda_k f_k(x_1) f_k(x_2)$$

spielt der Summand $\lambda_{k'} f_{k'}(x_1) f_{k'}(x_2) = 0$ aber ersichtlicherweise keine Rolle, weshalb wir $f_{k'}$ als mögliche Komponente von $\widetilde{Z}_i$, $i = 1, \ldots, n$, vernachlässigen können.

Mit Verwendung dessen gilt dann auch:

$$\begin{aligned} Cov_F\left(\widetilde{Z}_{k,i}, \widetilde{Z}_{l,i}\right) &= E_F\left[\widetilde{Z}_{k,i}\widetilde{Z}_{l,i}\right] - \left[E_F \widetilde{Z}_{k,i}\right]\left[E_F \widetilde{Z}_{l,i}\right] \\ &= E_F\left[\widetilde{Z}_{k,i}\widetilde{Z}_{l,i}\right] \\ &= \int_{-\infty}^{\infty} f_k(x_1) f_l(x_1) dF(x_1) \\ &= \begin{cases} 1, & \text{falls } k = l \\ 0, & \text{falls } k \neq l \end{cases} \end{aligned}$$

für $i = 1, \ldots, n$.

Für den Zufallsvektor

$$Z_n = (Z_{n,1}, \ldots, Z_{n,K})^T := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{Z}_i$$

gilt somit nach dem multivariaten zentralen Grenzwertsatz (vgl. Billingsley (1995), S. 385), dass

$$Z_n \xrightarrow{d} Z = (Z_1, \ldots, Z_K)^T \sim N(0_K, I_{K \times K}).$$

Als Nächstes definieren wir für $k = 1, \dots, K$

$$S_{n,k} := \frac{1}{n} \sum_{i=1}^{n} f_k^2(X_i).$$

Es gilt nach dem schwachen Gesetz der großen Zahlen von Chintschin (vgl. Georgii (2009), S. 123) und der Tatsache, dass die Eigenfunktionen f_k eine Orthonormalbasis bilden:

$$S_{n,k} \xrightarrow{P_F} E_F f_k^2(X_1) = 1.$$

Somit haben wir:

$$nT_{n,k} = \frac{n}{n-1} \left(Z_{n,k}^2 - S_{n,k} \right)$$

und damit

$$\begin{aligned} nU_{n,K} &= \sum_{k=1}^{K} \lambda_k n T_{n,k} \\ &= \frac{n}{n-1} \sum_{k=1}^{K} \lambda_k Z_{n,k}^2 - \frac{n}{n-1} \sum_{k=1}^{K} \lambda_k S_{n,k}. \end{aligned}$$

Dabei gilt:

$$\frac{n}{n-1} \sum_{k=1}^{K} \lambda_k Z_{n,k}^2 \xrightarrow{d} \sum_{k=1}^{K} \lambda_k Z_k^2$$

unter Zuhilfenahme des Continuous Mapping Theorems (vgl. Billingsley (1999), S. 21) und

$$\frac{n}{n-1} \sum_{k=1}^{K} \lambda_k S_{n,k} \xrightarrow{P_F} \sum_{k=1}^{K} \lambda_k$$

und somit nach dem Lemma von Sluzki (vgl. Georgii (2009), S. 292):

$$nU_{n,K} \xrightarrow{d} \sum_{k=1}^{K} \lambda_k \left(Z_k^2 - 1 \right).$$

Im letzten Schritt des Beweises werden wir nun noch das asymptotische Verhalten von $nU_n - nU_{n,K}$ untersuchen.

Wenn wir die Definition der $T_{n,k}$ auf alle $k > K$ erweitern, haben wir:

$$\begin{aligned}
E_F|nU_n - nU_{n,K}|^2 &= E_F\left|n\sum_{k=K+1}^{\infty}\lambda_k T_{n,k}\right|^2 \\
&= E_F\left[n\sum_{k=K+1}^{\infty}\lambda_k T_{n,k}\right]^2 \\
&= n^2 E_F\left[\sum_{k=K+1}^{\infty}\sum_{l=K+1}^{\infty}\lambda_k\lambda_l T_{n,k}T_{n,l}\right] \\
&= n^2\sum_{k=K+1}^{\infty}\sum_{l=K+1}^{\infty}\lambda_k\lambda_l E_F(T_{n,k}T_{n,l}) \\
&= n^2\sum_{k=K+1}^{\infty}\lambda_k^2\binom{n}{2}^{-1} \\
&= \frac{2n}{n-1}\sum_{k=K+1}^{\infty}\lambda_k^2 \xrightarrow[K\to\infty]{} 0,
\end{aligned}$$

denn:

$$\begin{aligned}
E_F[T_{n,k}T_{n,l}] &= E_F\left[\left(\binom{n}{2}^{-1}\sum_{\binom{n}{2}} f_k(X_i)f_k(X_j)\right)\left(\binom{n}{2}^{-1}\sum_{\binom{n}{2}} f_l(X_{i'})f_l\left(X_{j'}\right)\right)\right] \\
&= \binom{n}{2}^{-2}\sum_{\binom{n}{2}}\sum_{\binom{n}{2}} E_F\left[f_k(X_i)f_k(X_j)f_l(X_{i'})f_l\left(X_{j'}\right)\right]
\end{aligned}$$

und

$$E_F\left[f_k(X_i)f_k(X_j)f_l(X_{i'})f_l\left(X_{j'}\right)\right] = \begin{cases} 1, & \text{falls } i = i' \text{ und } j = j' \text{ und } k = l \\ 1, & \text{falls } i = j' \text{ und } j = i' \text{ und } k = l \\ 0, & \text{sonst} \end{cases},$$

was wegen der Unabhängigkeit von $X_1, \ldots, X_n$, der Orthonormalität von $f_1, f_2, \ldots$ und wegen $E_F f_k(X_1) = 0$, für $k \in \mathbb{N} \setminus \{k'\}$, folgt.

Damit ist:

$$\begin{aligned}
E_F[T_{n,k}T_{n,l}] &= \binom{n}{2}^{-2}\sum_{\binom{n}{2}} 1 \\
&= \binom{n}{2}^{-1}.
\end{aligned}$$

Insgesamt haben wir nun, dass $nU_{n,K}$ im quadratischen Mittel für $K \to \infty$ gegen nU_n konvergiert, womit auch die Konvergenz in Verteilung folgt.

Somit bleibt noch der Fall $n \to \infty$ zu betrachten. Dazu bezeichnen wir mit φ_n und $\varphi_{n,K}$ die charakteristischen Funktionen von nU_n beziehungsweise $nU_{n,K}$. Dann ist für $t \in \mathbb{R}$:

$$\begin{aligned}
|\varphi_n(t) - \varphi_{n,K}(t)| &= \left|E_F e^{itnU_n} - E_F e^{itnU_{n,K}}\right| \\
&= \left|E_F \left[e^{itnU_n} - e^{itnU_{n,K}}\right]\right| \\
&= \left|E_F \left[e^{itnU_{n,K}} \left(e^{it(nU_n - nU_{n,K})} - 1\right)\right]\right| \\
&\leq E_F \left|e^{it(nU_n - nU_{n,K})} - 1\right| \\
&\leq |t| E_F |nU_n - nU_{n,K}| \\
&\leq |t| \sqrt{E_F |nU_n - nU_{n,K}|^2} \sqrt{E_F 1^2} \\
&= |t| \sqrt{\frac{2n}{n-1} \sum_{k=K+1}^{\infty} \lambda_k^2},
\end{aligned}$$

wobei dieser Ausdruck, wie eben gezeigt, für $K \to \infty$ gegen 0 konvergiert.

Bei der Rechnung wurde die Tatsache, dass $|e^{ix}| = 1$ für alle $x \in \mathbb{R}$ ist, die Restgliedabschätzung der Exponentialreihe im Komplexen (vgl. Meintrup und Schäffler (2005), S. 187) und die Cauchy-Schwarz-Ungleichung (vgl. Dunford und Schwartz (1988a), S. 119) verwendet.

Damit existiert für ein beliebiges $\varepsilon > 0$ ein $K_1 \in \mathbb{N}$, sodass

$$|\varphi_n(t) - \varphi_{n,K}(t)| < \frac{\varepsilon}{3}$$

für alle $K > K_1$ und für alle $n \geq 2$ ist.

Weiterhin seien $\varphi(t)$ und $\varphi_K(t)$ die charakteristischen Funktionen von $\sum_{k=1}^{\infty} \lambda_k \left(Z_k^2 - 1\right)$ beziehungsweise $\sum_{k=1}^{K} \lambda_k \left(Z_k^2 - 1\right)$. Dann existiert ein $K_2 \in \mathbb{N}$, sodass für alle $K > K_2$ gilt:

$$|\varphi(t) - \varphi_K(t)| < \frac{\varepsilon}{3}.$$

Wir wissen, dass für alle $K \in \mathbb{N}$ gilt:

$$nU_{n,K} \xrightarrow{d} \sum_{k=1}^{K} \lambda_k \left(Z_k^2 - 1\right).$$

Für $K_0 := \max\{K_1, K_2\}$ existiert ein $N \in \mathbb{N}$, so dass für alle $n > N$ gilt:

$$|\varphi_{K_0}(t) - \varphi_{n,K_0}(t)| < \frac{\varepsilon}{3}.$$

Somit ist für alle $n > N$ unter Verwendung der Dreiecksungleichung:

$$\begin{aligned}|\varphi(t) - \varphi_n(t)| &\leq |\varphi(t) - \varphi_{K_0}(t)| \\ &\quad + |\varphi_{K_0}(t) - \varphi_{n,K_0}(t)| \\ &\quad + |\varphi_{n,K_0}(t) - \varphi_n(t)| \\ &< \varepsilon,\end{aligned}$$

mit $\varepsilon > 0$ beliebig, womit alles bewiesen ist. □

2.2 Degenerierte V-Statistiken

Satz 7 (vgl. Serfling (1980), S. 226)**.** *Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängig und identisch verteilter Zufallsvariablen mit $X_n\colon (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ für alle $n \in \mathbb{N}$ und $\mathscr{L}(X_1) = F \in \mathscr{F}$. Weiterhin seien $\theta\colon \mathscr{F} \to \mathbb{R}$ ein reguläres statistisches Funktional vom Grad 2 mit Kern h und V_n die entsprechende V-Statistik. Für V_n gelte:*

(i) $E_F h^2(X_1, X_2) < \infty$ *(falls $n \geq 2$),*

(ii) $E_F|h(X_1, X_1)| < \infty$ *und*

(iii) $\sigma_1^2 = Var_F h_1(X_1) = 0$.

Dann gilt:

$$n(V_n - \theta) \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k Z_k^2,$$

wobei $Z_1, Z_2, \ldots$ unabhängig standardnormalverteilt und $\lambda_1, \lambda_2, \ldots$ Eigenwerte der Integralgleichung

$$\int_{-\infty}^{\infty} h(x_1, x_2) f(x_2) dF(x_2) = \lambda f(x_1)$$

sind.

Beweis. Ohne Einschränkung sei unter Zuhilfenahme von Lemma 1 (ii)

$$\theta = E_F h(X_1, X_2) = E_F h_1(X_1) = 0$$

und somit, wegen $\sigma_1^2 = 0$:

$$h_1 = 0 \ P_F\text{-fast überall.}$$

Es ist

$$E_F h(X_1, X_1) = \sum_{k=1}^{\infty} \lambda_k$$

(vgl. Dunford und Schwartz (1988b), S. 1087), wobei dieser Erwartungswert nach Voraussetzung endlich ist.

Die zu dem Kern h entsprechende U-Statistik U_n ist

$$U_n = \binom{n}{2}^{-1} \sum_{\binom{n}{2}} h(X_i, X_j).$$

Nun gilt:

$$\begin{aligned}
n(V_n - U_n) &= n\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} h(X_i, X_j) - \binom{n}{2}^{-1}\sum_{\binom{n}{2}} h(X_i, X_j)\right] \\
&= n\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} h(X_i, X_j) - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} h(X_i, X_j)\right] \\
&= n\left[\frac{1}{n^2}\sum_{i=1}^{n} h(X_i, X_i) + \left(\frac{1}{n^2} - \frac{1}{n(n-1)}\right)\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} h(X_i, X_j)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} h(X_i, X_i) - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} h(X_i, X_j).
\end{aligned}$$

Nach dem starken Gesetz der großen Zahlen von Etemadi (vgl. Billingsley (1995), S. 282) gilt dabei:

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i, X_i) \xrightarrow{P_F\text{-f.s.}} E_F h(X_1, X_1) = \sum_{k=1}^{\infty} \lambda_k,$$

wobei „$\xrightarrow{P_F\text{-f.s.}}$“ die Konvergenz P_F-fast sicher bezeichnet.

Nach dem starken Gesetz der großen Zahlen für U-Statistiken (vgl. Serfling (1980), S. 190) gilt:

$$\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} h(X_i, X_j) \xrightarrow{P_F\text{-f.s.}} E_F h(X_1, X_2) = 0,$$

Also haben wir insgesamt:

$$n(V_n - U_n) \xrightarrow{P_F\text{-f.s.}} \sum_{k=1}^{\infty} \lambda_k,$$

womit auch die Konvergenz in Wahrscheinlichkeit (bezüglich P_F) folgt.

Die Voraussetzungen von Satz 6 sind erfüllt und mit den Größen aus Satz 6 und wegen $\theta = E_F U_n = E_F h(X_1, X_2) = 0$ gilt:

$$nU_n \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k \left(Z_k^2 - 1\right).$$

Unter Verwendung des Lemmas von Sluzki (vgl. Georgii (2009), S. 292) folgt schließlich:

$$nV_n \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k Z_k^2.$$

$\square$

3 Bootstrap-Verteilung

In diesem letzten Kapitel soll es nun noch um die asymptotische Bootstrap-Verteilung von vollständig degenerierten U-Statistiken vom Grad 2 gehen. Als erstes befassen wir uns dazu kurz mit der Idee des Bootstrapping. Danach werden wir einige Vorüberlegungen tätigen, bevor wir schließlich zu den Konvergenzaussagen kommen. Auf ein entsprechendes Resultat für V-Statistiken werden wir am Ende nur verweisen.

3.1 Zum Begriff des Bootstrapping

Gegeben seien weiterhin die Zufallsvariablen $X_1, \ldots, X_n$ mit $X_i \colon (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ für $i = 1, \ldots, n$ und $n \geq 2$, welche unabhängig und identisch verteilt seien mit $\mathscr{L}(X_1) = F \in \mathscr{F}$. Mit F_n bezeichnen wir wieder die empirische Verteilungsfunktion von $X_1, \ldots, X_n$. Außerdem seien $\theta \colon \mathscr{F} \to \mathbb{R}$ ein reguläres statistisches Funktional vom Grad 2 mit Kern h und U_n die entsprechende U-Statistik, welche vollständig degeneriert sei.

Die Bootstrap-Methode beschrieb erstmals Bradley Efron in seiner gleichnamigen Arbeit „Bootstrap Methods: Another look at the jackknife“ von 1979.

Als Bootstrap-Stichprobe bezeichnen wir nun eine Stichprobe $X_1^*, \ldots, X_n^*$, wobei die Zufallsgrößen unabhängig und identisch verteilt sind, mit $\mathscr{L}(X_1^*) = F_n$. Die Bootstrap-Verteilung von

$$U_n - \theta = U_n - E_F U_n = \left[\binom{n}{2}^{-1} \sum_{\binom{n}{2}} h(X_i, X_j) \right] - E_F h(X_1, X_2)$$

ist dann die Verteilung von

$$\left[\binom{n}{2}^{-1} \sum_{\binom{n}{2}} h\left(X_i^*, X_j^*\right) \right] - E_{F_n} h(X_1^*, X_2^*).$$

Im Folgenden betrachten wir die Verteilung von $U_n - \theta$ als Funktion der zugrundeliegenden Verteilungsfunktion F und bezeichnen dies mit

$$\mathscr{L}_F(U_n - \theta_F),$$

wobei θ_F verdeutlichen soll, dass die Erwartungswertbildung bezüglich F geschieht. Für $\mathscr{L}_G(U_n - \theta_G)$ mit $G \in \mathscr{F}$ sind dann $X_1, \ldots, X_n$ unabhängig und identisch verteilt mit $\mathscr{L}(X_1) = G$, und es gilt: $\theta_G = E_G U_n = E_G h(X_1, X_2)$.

Warum macht man nun das Ganze? Nach Voraussetzung ist $F \in \mathscr{F}$ unbekannt. Beim Bootstrapping ersetzen wir F durch F_n, welches für eine Realisierung von $X_1, \ldots, X_n$

bekannt ist, mit der Hoffnung, dass $\mathscr{L}_{F_n}(U_n - \theta_{F_n})$ eine „gute“ Approximation für $\mathscr{L}_F(U_n - \theta_F)$ ist.

3.2 Vorbereitungen

Die Grundlage, auf der man hier aufbauen würde, ist der Satz von Gliwenko-Cantelli (vgl. Billingsley (1995), S. 269), der besagt, dass gilt:

$$\sup_{x\in\mathbb{R}} |F(x) - F_n(x)| \xrightarrow{P_F\text{-f.s.}} 0.$$

Dies setzt jedoch voraus, dass $\mathscr{L}_F(n(U_n - \theta_F))$ gleichmäßig stetig bezüglich F ist. Und genau das ist hier nicht unbedingt der Fall: schon kleine Änderungen bezüglich F können zum Verlust der Degeneration führen. Dann wiederum wäre nicht mehr n sondern $\sqrt{n}$ die richtige Normierung, um zu einer asymptotischen Verteilung zu gelangen. Dementsprechend werden wir zunächst die U-Statistiken etwas modifizieren.

Um die Degeneration beim Bootstrapping nicht zu verlieren, definieren wir die Kerne h_F, für $F \in \mathscr{F}$, folgendermaßen:

$$\begin{aligned} h_F(x_1, x_2) \quad := \quad & h(x_1, x_2) \\ & - \int_{-\infty}^{\infty} h(x_1, x_2) dF(x_1) - \int_{-\infty}^{\infty} h(x_1, x_2) dF(x_2) \\ & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x_1, x_2) dF(x_1) dF(x_2). \end{aligned}$$

Dann sei:

$$U_n^* = U_n^*(h_{F_n}; X_1^*, \ldots, X_n^*) := \binom{n}{2}^{-1} \sum_{\binom{n}{2}} h_{F_n}\left(X_i^*, X_j^*\right),$$

womit aufgrund der Definition von h_{F_n} gilt:

$$E_{F_n} U_n^* = 0.$$

Die im Folgenden verwendenten Zufallsgrößen $Y, Y_1, \ldots, Y_n$ und $Z, Z_1, \ldots, Z_n$ seien messbare Abbildungen von $(\Omega, \mathcal{A})$ nach $(\mathbb{R}, \mathcal{B})$.

Um zu beweisen, dass nU_n^* für P_F-fast alle Realisierungen von $X_1, \ldots, X_n$ die gleiche Grenzverteilung besitzt wie $n(U_n - \theta)$, werden wir die sogenannte „Mallows-Metrik“ d_2 (vgl. Mallows (1972)) verwenden:

$$d_2(G, H) := \inf \left\{ E_{G,H} |Y - Z|^2 \colon \mathscr{L}(Y) = G \text{ und } \mathscr{L}(Z) = H \right\}$$

für $G, H \in \mathscr{F}_2$ mit

$$\mathscr{F}_2 := \left\{F\colon F \text{ Verteilungsfunktion auf } \mathbb{R} \text{ mit } E_F X^2 < \infty, \text{ falls } \mathscr{L}(X) = F\right\}.$$

Auf den Nachweis, dass es sich bei d_2 wirklich um eine Metrik handelt, werden wir hier verzichten. Auch verweisen wir nur darauf, dass schwache Konvergenz und Konvergenz der zweiten Momente äquivalent zur Konvergenz bezüglich d_2 ist (vgl. Bickel und Freedman (1981)).

Unser Ziel ist es nun, zu zeigen, dass F_n bezüglich d_2 für P_F - fast alle Realisierungen von $X_1, \ldots, X_n$ gegen F konvergiert. Dafür noch eine Definition:

Für eine ($\mathcal{B}^n$-$\mathcal{B}$)-messbare Funktion $t\colon \mathbb{R}^n \to \mathbb{R}$ sei die Halbmetrik $d_{2,t}$ definiert durch:

$$\begin{aligned} d_{2,t}(G,H) \quad := \quad \inf\big\{E_{G,H}[t(Y_1,\ldots,Y_n) - t(Z_1,\ldots,Z_n)]^2\colon & \\ Y_1,\ldots,Y_n \text{ i.i.d. mit } \mathscr{L}(Y_1) = G & \\ \text{und} & \\ Z_1,\ldots,Z_n \text{ i.i.d. mit } \mathscr{L}(Z_1) = H\ \big\} & \end{aligned}$$

für $G, H \in \mathscr{F}_{2,t}$ mit

$$\begin{aligned} \mathscr{F}_{2,t} \quad := \quad & \big\{F\colon F \text{ Verteilungsfunktion auf } \mathbb{R} \text{ mit } E_F t^2(X_1,\ldots,X_n) < \infty, \\ & \quad \text{falls } X_1,\ldots,X_n \text{ i.i.d. mit } \mathscr{L}(X_1) = F\ \big\}, \end{aligned}$$

wobei „i.i.d." für independent and identically distributed, also unabhängig und identisch verteilt, steht.

Auf die Überprüfung der Eigenschaften einer Halbmetrik werden wir ebenfalls verzichten (vgl. Bickel und Freedman (1981) - analog zum Nachweis der Metrikeigenschaften der Mallows-Metrik).

Ab dem nächsten Abschnitt werden wir als konkretes t den Kern h betrachten. Um dabei in Übereinstimmung mit unseren Resultaten aus den bisherigen Kapiteln zu bleiben, schränken wir die Familie von Verteilungsfunktionen $\mathscr{F}$ auf $\mathscr{F}^*$ ein mit

$$\mathscr{F}^* := \mathscr{F} \cap (\mathscr{F}_2 \cup \mathscr{F}_{2,h}).$$

Dabei sei sinnvollerweise $\mathscr{F} \supseteq \mathscr{F}_2 \cup \mathscr{F}_{2,h}$ und damit insgesamt $\mathscr{F}^* \neq \emptyset$.

3.3 Asymptotik

3.3.1 Degenerierte U-Statistiken

Zuerst stellen wir eine Verbindung zwischen der Mallows-Metrik d_2 und der Halbmetrik $d_{2,h}$ her.

Lemma 5 (vgl. Dehling und Mikosch (1994)). *Es sei* $\theta\colon \mathscr{F}^* \to \mathbb{R}$ *ein reguläres statistisches Funktional vom Grad 2 mit Kern* h. *Weiterhin seien* $Y_1, \ldots, Y_n$ *und* $Z_1, \ldots, Z_n$ *mit* $n \geq 2$ *jeweils unabhängig und identisch verteilt mit* $\mathscr{L}(Y_1) = G$ *und* $\mathscr{L}(Z_1) = H$ *mit* $G, H \in \mathscr{F}^*$. *Es gelte:*

(i) $E_G h^2(Y_1, Y_2) < \infty$,

(ii) $E_H h^2(Z_1, Z_2) < \infty$ *und*

(iii) h *sei unter* G *und* H *vollständig degeneriert.*

Dann ist

$$d_2^2 \left(\mathscr{L}_G \left(\frac{2}{n} \sum_{\binom{n}{2}} h_G(Y_i, Y_j) \right), \mathscr{L}_H \left(\frac{2}{n} \sum_{\binom{n}{2}} h_H(Z_i, Z_j) \right) \right) \leq 2 d_{2,h}^2(G, H).$$

Beweis. Es sei $(Y, Z)^T$ ein Zufallsvektor mit $\mathscr{L}(Y) = G$ und $\mathscr{L}(Z) = H$. Weiter seien $(Y_1, Z_1)^T, \ldots, (Y_n, Z_n)^T$ unabhängig und identisch verteilte Replikationen von $(Y, Z)^T$.

Dass dies keine zusätzliche Einschränkung bezüglich der Voraussetzungen des Lemmas ist, wird folgendermaßen klar: Nach Konstruktion haben $(Y_i, Z_i)^T$ eine offen gelassene Abhängigkeit voneinander, die für $i = 1, \ldots, n$ gleich ist. Da aber gilt, dass $Y_1, \ldots, Y_n$ und $Z_1, \ldots, Z_n$ jeweils unabhängig und identisch verteilt sind, kann jedes Y_i entweder mit keinem Z_j in Verbindung stehen oder aber zu maximal einem Z_j eine Beziehung aufweisen, wobei diese Zugehörigkeit eineindeutig und die Abhängigkeit für jedes Paar $(Y_i, Z_j)^T$ gleich sein muss. Und genau das entspricht unserer Konstruktion: Falls $Y_1, \ldots, Y_n$ und $Z_1, \ldots, Z_n$ einen Zusammenhang aufweisen, so nehmen wir lediglich an, dass diese Abhängigkeit zwischen Y_i und Z_i besteht. Dies entspricht also gegebenfalls einfach einer Umindizierung.

Dann gilt:

$$\begin{aligned}
& d_2^2\left(\mathscr{L}_G\left(\frac{2}{n}\sum_{\binom{n}{2}} h_G(Y_i,Y_j)\right), \mathscr{L}_H\left(\frac{2}{n}\sum_{\binom{n}{2}} h_H(Z_i,Z_j)\right)\right) \\
\leq \;& E_{G,H}\left[\frac{2}{n}\sum_{\binom{n}{2}}[h_G(Y_i,Y_j)-h_H(Z_i,Z_j)]\right]^2 \\
= \;& E_{G,H}\left[\frac{4}{n^2}\sum_{\binom{n}{2}}\sum_{\binom{n}{2}}[h_G(Y_i,Y_j)-h_H(Z_i,Z_j)][h_G(Y_k,Y_l)-h_H(Z_k,Z_l)]\right] \\
= \;& \frac{4}{n^2}\sum_{\binom{n}{2}}\sum_{\binom{n}{2}} E_{G,H}[[h_G(Y_i,Y_j)-h_H(Z_i,Z_j)][h_G(Y_k,Y_l)-h_H(Z_k,Z_l)]] \\
= \;& \frac{4}{n^2}\sum_{\binom{n}{2}}\sum_{\binom{n}{2}}[E_{G,H}[h_G(Y_i,Y_j)h_G(Y_k,Y_l)] \\
& \qquad\qquad - 2E_{G,H}[h_G(Y_i,Y_j)h_H(Z_k,Z_l)] \\
& \qquad\qquad + E_{G,H}[h_H(Z_i,Z_j)h_H(Z_k,Z_l)]].
\end{aligned}$$

(a) Nun ist:

$$\begin{aligned}
& E_{G,H}[h_G(Y_i,Y_j)h_G(Y_k,Y_l)] \\
= \;& E_{G,H}[[h(Y_i,Y_j)-h_1(Y_j)-h_1(Y_i)+E_{G,H}h(Y_i,Y_j)] \\
& \qquad \cdot [h(Y_k,Y_l)-h_1(Y_l)-h_1(Y_k)+E_{G,H}h(Y_k,Y_l)]] \\
= \;& E_{G,H}[[h(Y_i,Y_j)-\theta_G-\theta_G+\theta_G][h(Y_k,Y_l)-\theta_G-\theta_G+\theta_G]] \\
= \;& E_{G,H}[[h(Y_i,Y_j)-E_{G,H}h(Y_i,Y_j)][h(Y_k,Y_l)-E_{G,H}h(Y_k,Y_l)]] \\
= \;& Cov_{G,H}(h(Y_i,Y_j),h(Y_k,Y_l)),
\end{aligned}$$

Anmerkung: Wegen Degeneration gilt: $0=\sigma_1^2=Var_G h_1(Y_1)$. Da aber unter Verwendung von Lemma 1 (ii) $E_G h_1(Y_1)=E_G h(Y_1,Y_2)=\theta_G$ ist, folgt:

$$h_1(Y_1)=\theta_G \; P_G \text{ - fast überall.}$$

Falls $|\{i,j,k,l\}|=4$, so gilt wegen der Unabhängigkeit von $Y_1,\dots,Y_n$:

$$Cov_{G,H}(h(Y_i,Y_j),h(Y_k,Y_l))=0.$$

Falls $|\{i,j,k,l\}|=3$, so folgt mit Lemma 2 und da bei der Summation $i\neq j$ und $k\neq l$ gilt:

$$Cov_{G,H}(h(Y_i,Y_j),h(Y_k,Y_l))=\sigma_1^2=0.$$

Somit brauchen wir bei der Summenbildung für diese Kovarianzen nur den Fall $i = k$ und $j = l$ betrachten.

Eine analoge Rechnung ergibt sich für $E_{G,H}[h_H(Z_i, Z_j)h_H(Z_k, Z_l)]$.

(b) Weiterhin ist wegen Degeneration:

$$
\begin{aligned}
&E_{G,H}[h_G(Y_i, Y_j)h_H(Z_k, Z_l)] \\
&\quad= E_{G,H}[[h(Y_i, Y_j) - h_1(Y_j) - h_1(Y_i) + E_{G,H}h(Y_i, Y_j)] \\
&\qquad\qquad \cdot [h(Z_k, Z_l) - h_1(Z_l) - h_1(Z_k) + E_{G,H}h(Z_k, Z_l)]] \\
&\quad= E_{G,H}[[h(Y_i, Y_j) - \theta_G - \theta_G + \theta_G][h(Z_k, Z_l) - \theta_H - \theta_H + \theta_H]] \\
&\quad= E_{G,H}[[h(Y_i, Y_j) - E_{G,H}h(Y_i, Y_j)][h(Z_k, Z_l) - E_{G,H}h(Z_k, Z_l)]] \\
&\quad= Cov_{G,H}(h(Y_i, Y_j), h(Z_k, Z_l)).
\end{aligned}
$$

Falls $|\{i, j, k, l\}| = 4$, so gilt:

$$Cov_{G,H}(h(Y_i, Y_j), h(Z_k, Z_l)) = 0,$$

da mit der Unabhängigkeit von $(Y_1, Z_1)^T, \ldots, (Y_n, Z_n)^T$ in diesem Fall die Unabhängigkeit von Y_i, Y_j, Z_k, Z_l folgt.

Falls $|\{i, j, k, l\}| = 3$ und wegen $i \neq j$ und $k \neq l$, so ist:

$$
\begin{aligned}
&Cov_{G,H}(h(Y_i, Y_j), h(Z_k, Z_l)) \\
&\quad= E_{G,H}[h(Y_i, Y_j)h(Z_k, Z_l)] - [E_{G,H}h(Y_i, Y_j)][E_{G,H}h(Z_k, Z_l)] \\
&\quad= E_{G,H}[h(Y_i, Y_j)h(Z_k, Z_l)] - \theta_G\theta_H
\end{aligned}
$$

und mit der Unabhängigkeit von $(Y_1, Z_1)^T, \ldots, (Y_n, Z_n)^T$ und dem Satz von Fubini (vgl. Dunford und Schwartz (1988a), S. 193):

$$
\begin{aligned}
&E_{G,H}[h(Y_1, Y_2)h(Z_2, Z_3)] \\
&\quad= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(y_1, y_2)h(z_2, z_3)dG(y_1)dG(y_2)dH(z_2)dH(z_3) \\
&\quad= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(y_2)h_1(z_2)dG(y_2)dH(z_2) \\
&\quad= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \theta_G\theta_H dG(y_2)dH(z_2) \\
&\quad= \theta_G\theta_H.
\end{aligned}
$$

Also:

$$Cov_{G,H}(h(Y_i, Y_j), h(Z_k, Z_l)) = 0.$$

Somit müssen wir auch hier bei der Summation lediglich den Fall $i = k$ und $j = l$ beachten.

Insgesamt haben wir also, da $(Y_1, Z_1)^T, \ldots, (Y_n, Z_n)^T$ allesamt unabhängig und identisch verteilte Replikationen von $(Y, Z)^T$ sind:

$$\begin{aligned}
&\frac{4}{n^2} \sum_{\binom{n}{2}} \sum_{\binom{n}{2}} [E_{G,H}[h_G(Y_i, Y_j) h_G(Y_k, Y_l)] \\
&\qquad\qquad - 2E_{G,H}[h_G(Y_i, Y_j) h_H(Z_k, Z_l)] \\
&\qquad\qquad + E_{G,H}[h_H(Z_i, Z_j) h_H(Z_k, Z_l)]] \\
&= \frac{4}{n^2} \sum_{\binom{n}{2}} E_{G,H}[h_G(Y_i, Y_j) - h_H(Z_i, Z_j)]^2 \\
&= \frac{4}{n^2} \frac{n(n-1)}{2} E_{G,H}[h_G(Y_1, Y_2) - h_H(Z_1, Z_2)]^2 \\
&\leq 2E_{G,H}[h_G(Y_1, Y_2) - h_H(Z_1, Z_2)]^2.
\end{aligned}$$

Schließlich werden wir noch diesen Erwartungswert abschätzen:

$$\begin{aligned}
&E_{G,H}[h_G(Y_1, Y_2) - h_H(Z_1, Z_2)]^2 \\
&= E_{G,H} h_G^2(Y_1, Y_2) - 2E_{G,H}[h_G(Y_1, Y_2) h_H(Z_1, Z_2)] + E_{G,H} h_H^2(Z_1, Z_2).
\end{aligned}$$

(c) Es ist:

$$\begin{aligned}
E_{G,H} h_G^2(Y_1, Y_2) &= E_{G,H}[h(Y_1, Y_2) - h_1(Y_2) - h_1(Y_1) + E_{G,H} h(Y_1, Y_2)]^2 \\
&= E_{G,H}[h(Y_1, Y_2) - \theta_G - \theta_G + \theta_G]^2 \\
&= E_{G,H}[h(Y_1, Y_2) - \theta_G]^2.
\end{aligned}$$

Auf die gleiche Weise erhalten wir:

$$E_{G,H} h_H^2(Z_1, Z_2) = E_{G,H}[h(Z_1, Z_2) - \theta_H]^2.$$

(d) Außerdem gilt:

$$\begin{aligned}
&E_{G,H}[h_G(Y_1, Y_2) h_H(Z_1, Z_2)] \\
&= E_{G,H}[[h(Y_1, Y_2) - h_1(Y_2) - h_1(Y_1) + \theta_G][h(Z_1, Z_2) - h_1(Z_2) - h_1(Z_1) + \theta_H]] \\
&= E_{G,H}[[h(Y_1, Y_2) - \theta_G - \theta_G + \theta_G][h(Z_1, Z_2) - \theta_H - \theta_H + \theta_H]] \\
&= E_{G,H}[[h(Y_1, Y_2) - \theta_G][h(Z_1, Z_2) - \theta_H]]
\end{aligned}$$

Somit können wir nun zusammenfassen:

$$\begin{aligned}
&E_{G,H}h_G^2(Y_1,Y_2) - 2E_{G,H}[h_G(Y_1,Y_2)h_H(Z_1,Z_2)] + E_{G,H}h_H^2(Z_1,Z_2)\\
&= E_{G,H}h^2(Y_1,Y_2) - 2\theta_G E_{G,H}h(Y_1,Y_2) + \theta_G^2\\
&\quad - 2[E_{G,H}[h(Y_1,Y_2)h(Z_1,Z_2)] - \theta_G E_{G,H}h(Z_1,Z_2) - \theta_H E_{G,H}h(Y_1,Y_2) + \theta_G\theta_H]\\
&\quad + E_{G,H}h^2(Z_1,Z_2) - 2\theta_H E_{G,H}h(Z_1,Z_2) + \theta_H^2\\
&= E_{G,H}h^2(Y_1,Y_2) - \theta_G^2\\
&\quad - 2E_{G,H}[h(Y_1,Y_2)h(Z_1,Z_2)] + 2\theta_G\theta_H\\
&\quad + E_{G,H}h^2(Z_1,Z_2) - \theta_H^2\\
&= E_{G,H}[h(Y_1,Y_2) - h(Z_1,Z_2)]^2 - (\theta_G - \theta_H)^2\\
&\leq E_{G,H}[h(Y_1,Y_2) - h(Z_1,Z_2)]^2.
\end{aligned}$$

Wenn wir nun das Infimum über alle gemeinsamen Verteilungen von $(Y_1, Z_1)^T$ bilden, erhalten wir das zu Zeigende. □

Dieser Zusammenhang zwischen d_2 und $d_{2,h}$ wird uns bei der Konvergenz bezüglich d_2 helfen:

Lemma 6 (vgl. Dehling und Mikosch (1994)). *Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängig und identisch verteilter Zufallsvariablen mit $\mathscr{L}(X_1) = F \in \mathscr{F}^*$. F_n sei die empirische Verteilungsfunktion von $X_1, \dots, X_n$ und $X_1^*, \dots, X_n^*$ unabhängig und identisch verteilt, mit $\mathscr{L}(X_1^*) = F_n$. Für ein reguläres statistisches Funktional $\theta\colon \mathscr{F}^* \to \mathbb{R}$ vom Grad 2 sei h der Kern dieses Funktionals. Es gelte:*

(i) $E_F h^2(X_1, X_2) < \infty$,

(ii) $E_F|h(X_1, X_1)| < \infty$ und

(iii) h sei unter F und F_n, für alle $n \in \mathbb{N}$, vollständig degeneriert.

Dann gilt:

$$d_2\left(\mathscr{L}\left(\frac{2}{n}\sum_{\binom{n}{2}} h_F(X_i, X_j)\right), \mathscr{L}\left(\frac{2}{n}\sum_{\binom{n}{2}} h_{F_n}\left(X_i^*, X_j^*\right)\right)\right) \xrightarrow[n\to\infty]{} 0$$

für P_F-fast alle Realisierungen von (X_n).

Beweis. Mit Lemma 5 genügt es zu zeigen, dass

$$d_{2,h}(F, F_n) \xrightarrow[n\to\infty]{} 0.$$

Dafür sei zunächst $\widehat{h}\colon \mathbb{R}^2 \to \mathbb{R}$ eine stetige, beschränkte, $(\mathcal{B}^2\text{-}\mathcal{B})$-messbare Funktion, die h approximiert, sodass

$$E_F \left[h(X_1, X_2) - \widehat{h}(X_1, X_2)\right]^2 \le \varepsilon$$

für ein $\varepsilon > 0$.

Mit dem starken Gesetz der großen Zahlen für U-Statistiken (vgl. Serfling (1980), S. 190) haben wir:

$$\begin{aligned}
&E_{F_n} \left[h(X_1^*, X_2^*) - \widehat{h}(X_1^*, X_2^*)\right]^2 \\
= \quad & \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[h(X_i, X_j) - \widehat{h}(X_i, X_j)\right]^2 \\
= \quad & \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} h^2(X_i, X_j) \\
& \quad - 2 \left[\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} h(X_i, X_j) \widehat{h}(X_i, X_j)\right] \\
& \quad + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \widehat{h}^2(X_i, X_j) \\
\xrightarrow{P_F\text{-f.s.}} \quad & E_F h^2(X_1, X_2) \\
& \quad - 2E_F \left[h(X_1, X_2) \widehat{h}(X_1, X_2)\right] \\
& \quad + E_F \widehat{h}^2(X_1, X_2) \\
= \quad & E_F \left[h(X_1, X_2) - \widehat{h}(X_1, X_2)\right]^2.
\end{aligned}$$

Außerdem gilt:

$$X_1^* \xrightarrow{P_F\text{-f.s.}} X_1,$$

weil nach dem starken Gesetz der großen Zahlen von Etemadi (vgl. Billingsley (1995), S. 282)

$$F_n \xrightarrow{P_F\text{-f.s.}} F.$$

Da $\widehat{h}$ nach Voraussetzung beschränkt ist, können wir den Satz von Lebesgue über majorisierte Konvergenz (vgl. Billingsley (1995), S. 209) anwenden und haben:

$$E_{F,F_n} \left[\widehat{h}(X_1, X_2) - \widehat{h}(X_1^*, X_2^*)\right]^2 \xrightarrow[n\to\infty]{} 0.$$

Nun sind wir fertig, denn wir haben insgesamt unter Zuhilfenahme der Minkowski-

Ungleichung (vgl. Dunford und Schwartz (1988a), S. 120):

$$\begin{aligned}
d_{2,h}(F,F_n) &\leq \sqrt{E_{F,F_n}\left[h(X_1,X_2)-h(X_1^*,X_2^*)\right]^2} \\
&\leq \sqrt{E_{F,F_n}\left[h(X_1,X_2)-\widehat{h}(X_1,X_2)\right]^2} \\
&\quad + \sqrt{E_{F,F_n}\left[\widehat{h}(X_1,X_2)-\widehat{h}(X_1^*,X_2^*)\right]^2} \\
&\quad + \sqrt{E_{F,F_n}\left[\widehat{h}(X_1^*,X_2^*)-h(X_1^*,X_2^*)\right]^2}
\end{aligned}$$

und damit:

$$\lim_{n\to\infty} d_{2,h}(F,F_n) \leq 2\sqrt{\varepsilon}$$

für ein beliebiges $\varepsilon > 0$. □

Mit den beiden Lemmas sind wir nun am Ziel und halten als Satz fest:

Satz 8 (vgl. Dehling und Mikosch (1994))**.** *Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängig und identisch verteilter Zufallsvariablen mit $\mathscr{L}(X_1) = F \in \mathscr{F}^*$. F_n sei die empirische Verteilungsfunktion von $X_1,\ldots,X_n$ und $X_1^*,\ldots,X_n^*$ unabhängig und identisch verteilt, mit $\mathscr{L}(X_1^*) = F_n$. Für ein reguläres statistisches Funktional $\theta\colon \mathscr{F}^* \to \mathbb{R}$ vom Grad 2 sei h der Kern dieses Funktionals. Es gelte:*

(i) $E_F h^2(X_1,X_2) < \infty$,

(ii) $E_F|h(X_1,X_1)| < \infty$ *und*

(iii) h sei unter F und F_n, für alle $n \in \mathbb{N}$, vollständig degeneriert.

Außerdem bezeichnen wir mit Q die Grenzverteilung von $n(U_n-\theta)$ aus Satz 6, also:

$$\mathscr{L}(n(U_n-\theta)) \xrightarrow{w} Q,$$

wobei „$\xrightarrow{w}$“ die schwache Konvergenz bezeichnet.
Dann gilt:

$$\mathscr{L}(nU_n^*) \xrightarrow{w} Q$$

für P_F-fast alle Realisierungen von (X_n).

Beweis. Die Aussage folgt mit Lemma 6 und wegen der bereits genannten Tatsache, dass schwache Konvergenz und Konvergenz der zweiten Momente äquivalent zur Konvergenz bezüglich d_2 ist (vgl. Bickel und Freedman (1981)). □

3.3.2 Degenerierte V-Statistiken

Die Bootstrap-Grenzverteilung degenerierter V-Statistiken werden wir hier wie angekündigt nicht weiter untersuchen. Allerdings kann man auch für vollständig degenerierte V-Statistiken vom Grad 2 zeigen, dass deren Bootstrap-Verteilung die selbe Grenzverteilung wie die ursprüngliche V-Statistik aus Satz 7 besitzt. Dafür verweisen wir auf eine Arbeit von Arcones und Giné aus dem Jahr 1992. Die Idee dort besteht darin, eine V-Statistik in die Summe von V-Statistiken, ähnlich der Hoeffding-Zerlegung von U-Statistiken, zu zerlegen. Aufbauend darauf kommt man dann schließlich zu dem erwähnten Resultat.

Literaturverzeichnis

[1] Arcones, M. A. und Giné, E. (1992). On the bootstrap of U and V statistics. *The Annals of Statistics* **20**, 655-674.

[2] Bickel, P. J. und Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics* **9**, 1196-1217.

[3] Billingsley, P. P. (1995). *Probability and Measure.* Wiley, New York.

[4] Billingsley, P. P. (1999). *Convergence of Probability Measures.* Wiley, New York.

[5] Comtet, L. (1974). *Advanced Combinatorics. The Art of Finite and Infinite Expansions.* Reidel, Dordrecht.

[6] Dehling, H. G. und Mikosch, T. (1994). Random quadratic forms and the bootstrap for U-statistics. *Journal of Multivariate Analysis* **51**, 392-413.

[7] Dunford, N. und Schwartz, J. T. (1988a). *Linear Operators. Part I. General Theory.* Wiley, New York.

[8] Dunford, N. und Schwartz, J. T. (1988b). *Linear Operators. Part II. Spectral Theory.* Wiley, New York.

[9] Efron, B. (1979). Bootstrap methods. Another look at the jackknife. *The Annals of Statistics* **7**, 1-26.

[10] Georgii, H.-O. (2009). *Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik.* de Gruyter, Berlin.

[11] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* **19**, 293-325.

[12] Lee, A. J. (1990). *U-Statistics. Theory and Practice.* CRC Press, New York.

[13] Mallows, C. L. (1972). A note on asymptotic joint normality. *The Annals of Mathematical Statistics* **43**, 508-515.

[14] Meintrup, D. und Schäffler, S. (2005). *Stochastik. Theorie und Anwendungen.* Springer, Leipzig.

[15] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

[16] von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *The Annals of Mathematical Statistics* **18**, 309-348.

Selbstständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe.

Seitens des Verfassers bestehen keine Einwände, die vorliegende Bachelorarbeit für die öffentliche Benutzung im Universitätsarchiv zur Verfügung zu stellen.

Jena, 1.9.2011,