Dimensionsstabile Approximation für Verteilungen von zufälligen quadratischen Formen im Repeated-Measures-Design

Carola Werner

Abteilung Medizinische Statistik
Georg-August-Universität Göttingen

1 Einleitung

Die vorliegende Arbeit beschäftigt sich mit der Auswertung von Datensätzen, die *verbundene Messungen* (sogenannte *repeated measures*) enthalten. Diese tauchen zum Beispiel in klinischen Studien auf, wenn an wenigen Individuen über lange Zeiträume wiederholt Messungen durchgeführt werden. Ein anderes, immer stärker an Bedeutung gewinnendes Gebiet, in dem *repeated measures* auftreten, ist das der Microarrays. Hier hat man wenige Datenträger, auf denen sich jeweils sehr viele Messungen zu einer Person befinden. Datensätze dieser Art heißen hochdimensional, da ihre Dimension d (Anzahl der Messwiederholungen) den Stichprobenumfang n (Zahl der Individuen) weit übersteigt. Die für viele Tests wichtige Voraussetzung, dass der Stichprobenumfang größer als die Zahl der Messwiederholungen sein soll, ist hier verletzt. Und so wird sich zeigen, dass alle Verfahren, die bisher bei *repeated measures* angewandt wurden, für $d > n$ degenerieren oder nicht einmal mehr berechenbar sind.
Während man zum Beispiel die Wald-Typ-Statistik nicht ausrechnen kann, wenn die Dimension größer ist als der Stichprobenumfang, so ist die von Brunner, Dette, Munk in [6] beschriebene ANOVA-Typ-Statistik noch berechenbar. Simulationen haben aber gezeigt, dass sie kein vorgegebenes Niveau einhält und extrem konservativ wird. Die Motivation für die vorliegende Arbeit war deshalb, die ANOVA-Typ-Statistik so zu modifizieren, dass im Falle $d > n$ das Niveau eingehalten wird und man somit sinnvolle Ergebnisse erhält.
Andere moderne Verfahren, mit denen hochdimensionale Daten ausgewertet werden können, basieren auf Dimensionsreduktion (Kropf [13]) oder Permutationsverfahren (Westfall, Young [20]). Hier geht es zumeist nicht um globale sondern um multiple Tests, welche ein bestimmtes, multiples Niveau einhalten sollen (Benjamini, Hochberg [2]). Bei der Literaturrecherche zu globalen Verfahren für hochdimensionale Datensätze findet man bisher nur wenige Arbeiten (Donoho et. al. [9]).

2 ANOVA-Typ-Statistik

2.1 Box-Approximation

Um Hypothesen der Form $\mathbf{T}\boldsymbol{\mu} = \mathbf{0}$ zu testen, kann die sogenannte ANOVA-Typ-Statistik verwendet werden. Diese wird in [6], [7] und [8] genau beschrieben.
Man betrachtet zunächst die quadratische Form $\mathbf{X}'\mathbf{T}\mathbf{X}$ und untersucht deren Verteilung. Die kann approximativ bestimmt werden, wenn man die Tatsache ausnutzt, dass quadratische Formen in normalverteilten Zufallsvariablen als gewichtete Summen von χ^2_1-verteilten Zufallsvariablen dargestellt werden können. Man approximiert die Verteilung dieser Linearkombination $\sum_i \lambda_i C_i$ durch eine mit g gestreckte χ^2_f-Verteilung derart, dass die beiden ersten Momente von $\sum_i \lambda_i C_i$ und $g\,\chi^2_f$ übereinstimmen. Dieser Weg, eine Verteilung zu approximieren, wird in dieser Arbeit "Box-Approximation" (siehe [3],[4]) genannt, obwohl es bereits 1949 von Patnaik (siehe [17]) eine Veröffentlichung zu dieser Art der Approximation gab. Man erhält die folgenden Beziehungen:

$$g\,f = E(g\,\chi_f^2) \quad \stackrel{!}{=} \quad E(\sum_i \lambda_i C_i) = \sum_i \lambda_i = \mathrm{Sp}(\mathbf{TV})$$
$$2\,g^2 f = \mathrm{Var}(g\,\chi_f^2) \quad \stackrel{!}{=} \quad \mathrm{Var}(\sum_i \lambda_i C_i) = 2\sum_i \lambda_i^2 = 2\,\mathrm{Sp}(\mathbf{TVTV})$$

Durch Umformen erhält man f und g und somit hat die quadratische Form $\mathbf{X'TX}$ die folgende approximative Verteilung:

$$\frac{\mathbf{X'TX}}{\mathrm{Sp}(\mathbf{TV})} \stackrel{\cdot}{\sim} \chi_f^2/f = F(f,\infty)$$

2.2 Konsistenz von Schätzern

Definition 2.1 (Konsistenz 1) *Eine Folge von Schätzern $\hat{\theta}_n$ heißt* **konsistent** *für θ, falls $\hat{\theta}_n - \theta \xrightarrow{p} 0$ für festes θ (d.h. $\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0 \;\forall\, \varepsilon > 0$).*

Die Konsistenz ist eine Minimal-Eigenschaft, die vernünftige Schätzer mindestens erfüllen sollten. Sie impliziert, dass sich der Schätzer mit wachsendem Stichprobenumfang n immer besser dem Parameter annähert.
Betrachten wir als Beispiel zunächst die Stichprobenkovarianzmatrix

$$\widehat{\mathbf{V}}_n = \frac{1}{n-1}\sum_{k=1}^{n}(\mathbf{X}_k - \overline{\mathbf{X}}_\cdot)(\mathbf{X}_k - \overline{\mathbf{X}}_\cdot)' \tag{2.1}$$

Da $\widehat{\mathbf{V}}_n$ ein konsistenter Schätzer für $\mathbf{V}$ ist, gilt nach dem Satz von Slutzky, dass auch $[\mathrm{Sp}(\mathbf{T}\widehat{\mathbf{V}}_n)]^2$ und $\mathrm{Sp}[(\mathbf{T}\widehat{\mathbf{V}}_n)^2]$ als stetige Funktionen von konsistenten Schätzern konsistente Schätzer sind. Aus dem Korollar zum Satz von Slutzky folgt dann (vorausgesetzt, der Nenner ist ungleich null), dass auch der Quotient der beiden Schätzer konsistenter Schätzer für den Quotienten ist.
Betrachtet man nun anstatt Folgen von Schätzern ein Feld von Schätzern $\hat{\theta}_{n,d}$ die von zwei Parametern abhängen, so benötigt man eine modifizierte Definition der Konsistenz:

Definition 2.2 (Konsistenz 2) *Ein Feld von Schätzern $\hat{\theta}_{n,d}$ heißt* **konsistent** *für θ_d, falls $\hat{\theta}_{n,d} - \theta_d \xrightarrow{p} 0$ für festes d, also für festes θ_d (d.h. $\lim_{n\to\infty} P(|\hat{\theta}_{n,d} - \theta_d| > \varepsilon) = 0 \;\forall\, \varepsilon > 0$).*

Man kann die klassischen Verfahren für Schätzer $\hat{\theta}_n$, die gegen θ streben, dahingehend verändern, dass statt des Schätzers selbst nun der Quotient $\frac{\hat{\theta}_{n,d}}{\theta_d}$ bei festem d gegen 1 streben soll. Das folgende Lemma bietet eine einfache Möglichkeit, mit Hilfe der Tschebycheff-Ungleichung die Konsistenz zu überprüfen.

Lemma 2.3 *Ein Feld von asymptotisch erwartungstreuen Schätzern $\hat{\theta}_{n,d}$ ist konsistent für $\theta_d > 0$, falls gilt:*

$$\lim_{n\to\infty}\left\{\frac{\mathrm{Var}(\hat{\theta}_{n,d})}{\theta_d^2}\right\} = 0 \qquad \forall d < \infty,\; d \textit{ fest}$$

Beweis: Mit Hilfe der Tschebycheff-Ungleichung $P(|X| > \varepsilon) \le \frac{EX^2}{\varepsilon^2}$ mit $X = \frac{\hat{\theta}_{n,d}}{\theta_d} - 1$ erhält man:

$$\begin{aligned}
&P\left(\left|\tfrac{\hat{\theta}_{n,d}}{\theta_d} - 1\right| > \varepsilon\right) \\
&\qquad \le \frac{1}{\varepsilon^2}E\left(\left|\tfrac{\hat{\theta}_{n,d}}{\theta_d} - 1\right|^2\right) = \frac{1}{\varepsilon^2}E\left(\left|\tfrac{\hat{\theta}_{n,d}}{\theta_d} - E(\tfrac{\hat{\theta}_{n,d}}{\theta_d}) + E(\tfrac{\hat{\theta}_{n,d}}{\theta_d}) - 1\right|^2\right) \\
&\qquad \le \frac{1}{\varepsilon^2}\left\{E\left[\left(\tfrac{\hat{\theta}_{n,d}}{\theta_d} - E(\tfrac{\hat{\theta}_{n,d}}{\theta_d})\right)^2\right] + E\left[\left(E(\tfrac{\hat{\theta}_{n,d}}{\theta_d}) - 1\right)^2\right]\right\} \\
&\qquad = \frac{1}{\varepsilon^2}\left\{\mathrm{Var}(\tfrac{\hat{\theta}_{n,d}}{\theta_d}) + \left(E(\tfrac{\hat{\theta}_{n,d}}{\theta_d}) - 1\right)^2\right\}
\end{aligned}$$

$$\Rightarrow \lim_{n\to\infty} P\left(\left|\frac{\hat{\theta}_{n,d}}{\theta_d} - 1\right| > \varepsilon\right) \quad \leq \quad \frac{1}{\varepsilon^2} \lim_{n\to\infty} \left\{ \mathrm{Var}(\frac{\hat{\theta}_{n,d}}{\theta_d}) + \left(E(\frac{\hat{\theta}_{n,d}}{\theta_d}) - 1\right)^2 \right\} \overset{\text{Vor.}}{=} 0$$

□

2.3 Dimensionsstabilität

Betrachtet man nun derartige Felder von Schätzern, so benötigt man ein neues Stabilitätskriterium zur Beschreibung des Verhaltens der Schätzer, wenn sich einer der Parameter verändert.

Definition 2.4 *Ein Feld von Schätzern $\hat{\theta}_{n,d}$ heißt* **dimensionsstabil**, *wenn $\forall d \geq 1, n \geq 1$ gilt:*

1. $|E\left(\frac{\hat{\theta}_{n,d}}{\theta_d} - 1\right)| \leq A(n) < \infty$, *wobei A nicht von d abhängt.*
2. $\mathrm{Var}\left(\frac{\hat{\theta}_{n,d}}{\theta_d}\right) = \frac{1}{\theta_d^2}\mathrm{Var}(\hat{\theta}_{n,d}) \leq B(n) < \infty$, *wobei B nicht von d abhängt.*

Anschaulich bedeutet die Definition, dass die Approximation mit zunehmender Anzahl von Zeitpunkten bzw. Messwiederholungen nicht schlechter wird.
In dieser Arbeit soll nun gezeigt werden, dass die Standardschätzverfahren, die bei der ANOVA-Typ-Statistik angewendet werden, nicht dimensionsstabil sind, d.h. für hochdimensionale Daten nicht geeignet sind. Da aber im Zuge der Technisierung von Datenerhebungen immer größere Datensätze auf Statistiker zukommen und im Zuge von Kostenreduktion immer weniger Patienten (Individuen) untersucht werden, ist die Dimensionsstabilität eine immer wichtiger werdende Eigenschaft von Approximationen. Deshalb wird ein dimensionsstabiles Verfahren entwickelt, mit dessen Hilfe Datensätze ausgewertet werden können, bei denen die Anzahl der Messwiederholungen den Stichprobenumfang übersteigt. Zum Beispiel ist $\mathrm{Sp}(\mathbf{T}\widehat{\mathbf{V}}_n\mathbf{T}\widehat{\mathbf{V}}_n)$, der Schätzer den man für $\mathrm{Sp}[(\mathbf{TV})^2]$ erhält, wenn man die Stichprobenkovarianzmatrix einsetzt, im Allgemeinen nicht dimensionsstabil.

2.4 Verzerrung der klassischen Schätzer

Eine wichtige Voraussetzung für die Dimensionsstabilität ist die asymptotische Erwartungstreue eines Feldes von Schätzern für festes d. Die Schätzer für $\mathrm{Sp}(\mathbf{TV})$, $[\mathrm{Sp}(\mathbf{TV})]^2$ und $\mathrm{Sp}[(\mathbf{TV})^2]$, die man durch Einsetzen der Stichprobenkovarianzmatrix $\widehat{\mathbf{V}}_n$ für $\mathbf{V}$ erhält, sind für kleines $d < n$ nur asymptotisch erwartungstreu. Hält man also n fest, so fällt die Verzerrung mit wachsendem d immer stärker ins Gewicht (für $\mathrm{Sp}[(\mathbf{TV})^2]$ wurde dies oben analytisch gezeigt).
Um diese Verzerrung durch Simulationen zu demonstrieren, werden normalverteilte Zufallsvariablen mit CS-Struktur erzeugt. Der Stichprobenumfang n ist konstant gleich 10, die Dimension d variiert zwischen 5 und 200. Anhand einer vorgegebenen Kovarianzstruktur können die wahren Werte der zu schätzenden Größen berechnet werden. Mit Hilfe der simulierten Zufallsvariablen werden dann die einzelnen Schätzer berechnet und somit der Erwartungswert und die empirische Varianz nach 1000 Simulationsdurchläufen bestimmt.
Zunächst wird der Schätzer für $\mathrm{Sp}(\mathbf{TV})$ betrachtet, den man erhält, indem $\mathbf{V}$ durch $\widehat{\mathbf{V}}_n$ ersetzt wird. Die Simulationen zeigen das Verhalten des Schätzers in Relation zu seinem wahren Wert für größer werdendes d. Es ist hier angebracht, den relativen Erwartungswert zu betrachten, da auch der wahre Wert mit der Dimension wächst. Es ist zu sehen, dass der Schätzer für beliebiges d erwartungstreu ist und die Varianz mit wachsendem d immer kleiner wird, d.h. dass der Schätzer dimensionsstabil ist.

Weitere Simulationen haben gezeigt, dass der Schätzer für $[\mathrm{Sp}(\mathbf{TV})]^2$ bei Verwendung der Stichproben-Kovarianzmatrix nahezu erwartungstreu ist; die Verzerrung wird bei wachsendem d zumindest nicht größer.
Der Schätzer für $\mathrm{Sp}[(\mathbf{TV})^2]$ ist dagegen ziemlich stark verzerrt. Diese Verzerrung nimmt mit wachsendem d zu. Außerdem wächst die Varianz auch mit steigender Dimension.
Aus den bisherigen Simulationen folgt, dass der Schätzer für den Freiheitsgrad, also der Quotient

der beiden Schätzer, für wachsendes d immer stärker unterschätzt wird. Dementsprechend zeigen die Simulationensergebnisse, dass die Verzerrung des Schätzers mit der Anzahl der Zeitpunkte d zunimmt. Der wahre Wert wird immer stärker unterschätzt.
Wenn nun aber der erste Freiheitsgrad der F-Verteilung bzw. der χ^2_f/f-Verteilung kleiner wird, so wird das dazugehörige Quantil größer. Das hat zur Folge, dass der Test mit wachsender Anzahl von Zeitpunkten konservativ wird. Außerdem sinkt die Wahrscheinlichkeit, dass eine Statistik größer als ein beliebig gewähltes $(1-\alpha)$-Quantil ist. Schließlich wird diese Wahrscheinlichkeit irgendwann so gering, dass der Test gar nicht mehr ablehnt. Diese Tatsache wird am Ende dieser Arbeit anhand von Simulationen demonstriert.
Die Verzerrung der Schätzer ist besonders schwerwiegend, wenn man kleine Stichproben hat. Noch gravierender wird es, wenn dazu noch die Anzahl der Zeitpunkte steigt. Der Test wird zunehmend konservativ, wie durch zahlreiche Simulationen in vergangenen Arbeiten bereits gezeigt wurde (siehe BRUNNER, DOMHOF, LANGER [7]). In dem Buch wird auch darauf hingewiesen, dass die unverzerrte Schätzung von f weiterhin ein offenes Problem ist.

2.5 Literaturvergleich

In der Literatur wurde bereits häufig darauf hingewiesen, dass der Schätzer für das sogenannte Box'sche Epsilon (den 1. Freiheitsgrad der Verteilung der F-Statistik)

$$\varepsilon = \frac{[\mathrm{Sp}(\mathbf{TV})]^2}{(d-1)\mathrm{Sp}[(\mathbf{TV})^2]} = \frac{f}{d-1}$$

nicht erwartungstreu ist. Setzt man die Stichprobenkovarianzmatrix $\widehat{\mathbf{V}}_n$ ein, so tritt bei Verletzung der Annahmen an die Kovarianzstruktur (*sphericity* bzw. *compound symmetry*) eine starke Verzerrung des Schätzers auf.

Bereits 1976 haben zum Beispiel HUYHN, FELDT ([12]) eine Korrektur des Schätzers für das Box'sche ε veröffentlicht. Dem Artikel im *Journal of Educational Statistics* folgte ein weiterer in *JASA*. Die Korrektur des Schätzers basiert darauf, den Zähler und Nenner getrennt erwartungstreu zu schätzen. Daraus resultiert der Schätzer $\tilde{\varepsilon}$:

$$\tilde{\varepsilon} = \frac{n(d-1)\hat{\varepsilon} - 2}{(d-1)(n-1-(d-1)\hat{\varepsilon})} \quad \text{mit} \quad \hat{\varepsilon} = \frac{[\mathrm{Sp}(\mathbf{T}\widehat{\mathbf{V}}_n)]^2}{(d-1)\mathrm{Sp}[(\mathbf{T}\widehat{\mathbf{V}}_n)^2]}$$

HUYNH, FELDT haben in ihrer Arbeit Monte-Carlo-Simulationen durchgeführt, die zeigen, dass ANOVA-Tests mit $\tilde{\varepsilon}$ statt $\hat{\varepsilon}$ das Niveau weitaus besser einhalten und dass $\tilde{\varepsilon}$ viel weniger verzerrt ist als $\hat{\varepsilon}$. Diese sogenannte "Huynh-Feldt-Korrektur" ist seitdem auch in einigen Softwarepaketen (z.B. SPSS) implementiert. Die zahlreichen Veröffentlichungen zu Verbesserungen dieser Korrektur in Journals wie *Social Psychology Quarterly, Journal of Higher Education* zeigen, dass die "Huynh-Feldt-Korrektur" hauptsächlich im Bereich der Psychologie verwendet wird.

Wie man leicht sieht, wird der Schätzer für $d > n$ negativ. Er ist also für die $d > n$-Situation, die hier betrachtet werden soll, nicht anwendbar.

Festzuhalten bleibt aber trotzdem die Vorgehensweise bei der "Huynh-Feldt-Korrektur" des Schätzers: die Verzerrung eines Quotienten aus abhängigen Zufallsvariablen wird berechnet, indem man die Erwartungswerte von Zähler und Nenner getrennt berechnet. Diese Verzerrung wird jeweils abgezogen und anschließend werden die so konstruierten erwartungstreuen Schätzer wieder zusammensetzt. Die Erwartungstreue dieses neuen Schätzers wird dann nur anhand von Simulationen gezeigt.
Die Autoren begründen ihre Vorgehensweise mit Beispielen aus der Literatur und führen HÁJEK ([10]) an, der den Varianzschätzer beim Behrens-Fisher-Problem genau so konstruiert hat. Dieser wiederum begründet sein Vorgehen kaum, verweist aber auf WELCH ([19]), der dies vor ihm schon so gemacht hat. Allerdings begründet auch Welch sein Vorgehen nur mit ein paar kurzen Worten:

> "[...] It may be shown that the numerator of (29) has, in repeated samples an average value [*echter Wert*], and the denominator has average value [*echter Wert*]. In a certain sense, therefore, (29) is a fair estimate of (28)."

Analog zu HUYNH, FELDT sollen im Folgenden mit Hilfe des Verfahrens der getrennten Schätzung von Zähler und Nenner andere erwartungstreue Schätzer konstruiert werden, die auch in der $d > n$-Situation anwendbar sind. Die so konstruierten Schätzer werden verwendet, um die Verteilung der ANOVA-Typ-Statistik neu zu approximieren. Hierbei erhält man Quotienten von Schätzern, deren Eigenschaften betrachtet werden müssen. Zur Beurteilung dieser Eigenschaften der neuen Schätzer werden aber nicht wie bei HUYNH, FELDT nur Simulationen herangezogen. Wann immer es möglich ist, werden die Eigenschaften auch analytisch hergeleitet. Durch die Anwendung von Taylor-Approximationen ist man in der Lage, Simulationsergebnisse durch Rechnungen zu bestätigen.

2.6 F-Approximation

Wenn Komponenten der Teststatistik geschätzt werden, so ändert sich deren Verteilung. Also muss auch die Approximation angepasst werden.

Da die wahre Kovarianzmatrix im Allgemeinen unbekannt ist, muss sie aus den Daten geschätzt werden. Im Folgenden bezeichnen:

- B_0 einen Schätzer für $\mathrm{Sp}(\mathbf{TV})$,
- B_1 einen Schätzer für $[\mathrm{Sp}(\mathbf{TV})]^2$,
- B_2 einen Schätzer für $\mathrm{Sp}[(\mathbf{TV})^2]$.

Setzt man nun die Schätzer sowohl in die Statistik als auch in den Freiheitsgrad ein, so erhält man folgende Quotienten:

$$\widetilde{F} = \frac{Q_n}{B_0} \qquad \widetilde{f}_1 = \frac{B_1}{B_2}$$

Die Statistik $\widetilde{F}$ hat somit keine $\chi^2_{f_1}/f_1$-Verteilung mehr, da der Schätzer B_0 für $\mathrm{Sp}(\mathbf{TV})$ eine Zufallsvariable ist. Also muss die Verteilung der Statistik neu approximiert werden. Sie sollte im Grenzfall $n \to \infty$ gegen die $\chi^2_{f_1}/f_1$-Verteilung konvergieren. Betrachtet man diese als eine $F(f_1, \infty)$-Verteilung, so liegt es nahe, $\widetilde{F}$ mit einer $F(f_1, f_2)$-Verteilung zu approximieren. Außerdem sollte sichergestellt sein, dass $f_2 \to \infty$ für $n \to \infty$. Es wird wieder die Box-Approximation (d.h. Gleichsetzen der ersten zwei Momente) verwendet werden.
Zur Berechnung der Momente der neuen Statistik werden allerdings die konkreten Schätzer B_0, B_1, B_2 benötigt, die erst im folgenden Kapitel hergeleitet werden. Diese Thematik wird deshalb am Ende des nächsten Kapitels wieder aufgegriffen.

3 Erwartungstreue Schätzer

3.1 Konstruktion von erwartungstreuen Schätzern

Im Folgenden werden erwartungstreue Schätzer für $[\mathrm{Sp}(\mathbf{TV})]^2$ und $\mathrm{Sp}[(\mathbf{TV})^2]$ konstruiert. Die einzelnen Schätzer werden getrennt erwartungstreu und konsistent geschätzt und anschließend zusammengesetzt (siehe Abschnitt 2.5). Von dem so konstruierten Schätzer für den Freiheitsgrad lässt sich dann mit Hilfe einer Taylor-Entwicklung die Verzerrung bestimmen.
Sei $\mathbf{X}_k$ ein multivariat normalverteilter Vektor mit Erwartungswertvektor $\boldsymbol{\mu}$ und Kovarianzmatrix $\mathbf{V}$. Da die Hypothesen, die getestet werden sollen, immer in der Standardform $\mathbf{T}\boldsymbol{\mu} = \mathbf{0}$ - wobei $\mathbf{T}$ ein Projektor ist - dargestellt werden können, wird im Folgenden zur besseren Übersichtlichkeit der Zufallsvektor $\mathbf{TX}_k$ mit $\mathbf{Y}_k$ bezeichnet. Unter $H_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$ gilt dann für die $\mathbf{Y}_k$:

$$E(\mathbf{Y}_k) = \mathbf{0} \qquad \mathrm{Cov}(\mathbf{Y}_k) = \mathrm{Cov}(\mathbf{TX}_k) = \mathbf{TVT} =: \mathbf{S}$$

Deshalb wird als Kovarianzschätzer der folgende, unter Hypothese erwartungstreue Schätzer $\widehat{\mathbf{S}}_n$ verwendet.

Lemma 3.1 *Sei*

$$\widehat{\mathbf{S}}_n := \frac{1}{n}\sum_{k=1}^{n}\mathbf{S}_k = \frac{1}{n}\sum_{k=1}^{n}\mathbf{Y}_k\mathbf{Y}_k'. \tag{3.2}$$

Es gelte die Hypothese $H_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$. *Dann ist* $\widehat{\mathbf{S}}_n$ *erwartungstreuer Schätzer für* $\mathbf{TVT}$.

Beweis:

$$E(\widehat{\mathbf{S}}_n) = \frac{1}{n}\sum_{k=1}^{n}E(\mathbf{Y}_k\mathbf{Y}_k') = \mathrm{Var}(\mathbf{Y}_k) + [E(\mathbf{Y}_k)]^2 = \mathbf{TVT} + \mathbf{0} \qquad \square$$

Der kanonische Schätzer B_0 für $\mathrm{Sp}(\mathbf{TV}) = \mathrm{Sp}(\mathbf{S})$, den man durch Einsetzen des Kovarianzschätzers $\widehat{\mathbf{S}}_n$ aus (3.2) erhält, ist erwartungstreu, da die Spur eine linearer Abbildung ist und Erwartungswert und Spur somit vertauschbar sind.
Wird B_0 allerdings quadriert, um den Schätzer B_1 für $[\mathrm{Sp}(\mathbf{TV})]^2$ zu erhalten, so kommt eine Verzerrung τ^2 hinzu:

$$E\left([\mathrm{Sp}(\widehat{\mathbf{S}}_n)]^2\right) = \mathrm{Var}(\mathrm{Sp}(\widehat{\mathbf{S}}_n)) + E\left([\mathrm{Sp}(\widehat{\mathbf{S}}_n)]\right)^2 = \underbrace{\mathrm{Var}(\mathrm{Sp}(\widehat{\mathbf{S}}_n))}_{:=\tau^2} + [\mathrm{Sp}(\mathbf{S})]^2 \tag{3.3}$$

Für diese Varianz kann man einen erwartungstreuen Schätzer angeben:

$$\begin{aligned}
\tau^2 &= \mathrm{Var}\left(\mathrm{Sp}(\widehat{\mathbf{S}}_n)\right) = \mathrm{Var}\left(\mathrm{Sp}(\frac{1}{n}\sum_{k=1}^{n}\mathbf{Y}_k\mathbf{Y}_k')\right) \\
&= \mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^{n}\underbrace{\mathbf{X}_k'\mathbf{T}\mathbf{X}_k}_{:=A_k \text{ u.i.v.}}\right) = \frac{1}{n}\mathrm{Var}(A_k) \\
\Longrightarrow \widehat{\tau}^2 &= \frac{1}{n(n-1)}\sum_{k=1}^{n}(A_k - \overline{A}_\cdot)^2
\end{aligned}$$

Hier wurde die empirische Varianz der A_k eingesetzt, wobei $\overline{A}_\cdot := 1/n\sum A_k$ den empirischen Mittelwert der A_k bezeichnet. Ein erwartungstreuer Schätzer für $[\mathrm{Sp}(\mathbf{TV})]^2$ hat also folgende Gestalt:

$$B_1 = [\mathrm{Sp}(\widehat{\mathbf{S}}_n)]^2 - \frac{1}{n(n-1)}\sum_{k=1}^{n}(A_k - \overline{A}_\cdot)^2. \tag{3.4}$$

Bei der Entwicklung des Schätzers für $\mathrm{Sp}[(\mathbf{TV})^2] = \mathrm{Sp}(\mathbf{S}^2)$, der durch doppeltes Einsetzen des Kovarianzschätzers aus (3.2) entsteht, wird ein Resultat aus der Matrizenrechnung verwendet:

Lemma 3.2 *Für zwei quadratische Matrizen* $\mathbf{A}, \mathbf{B}$ *der gleichen Dimension* n *gilt:*

$$\mathrm{Sp}(\mathbf{AB}') = \mathbf{1}_n'(\mathbf{A}\#\mathbf{B})\mathbf{1}_n,$$

wobei $(\mathbf{A}\#\mathbf{B})$ *das Hadamard-Produkt der zwei Matrizen, also die komponentenweise Multiplikation, bezeichnet.*

Mit Hilfe dieser Gleichung lässt sich die Verzerrung π^2 des Schätzers direkt herleiten, wobei $\mathbf{S}_k = \mathbf{Y}_k\mathbf{Y}_k'$ als Abkürzung verwendet wird:

$$\begin{aligned}
E[\mathrm{Sp}(\widehat{\mathbf{S}}_n^{\,2})] &= \mathbf{1}_d'E[\widehat{\mathbf{S}}_n\#\widehat{\mathbf{S}}_n]\mathbf{1}_d = \mathbf{1}_d'E\left[\left(\frac{1}{n}\sum_{k=1}^{n}\mathbf{S}_k\right)\#\left(\frac{1}{n}\sum_{k=1}^{n}\mathbf{S}_k\right)\right]\mathbf{1}_d \\
&= \mathbf{1}_d'E\left[\frac{1}{n^2}\sum_{k=1}^{n}\sum_{k'=1}^{n}\mathbf{S}_k\#\mathbf{S}_{k'}\right]\mathbf{1}_d = \mathbf{1}_d'\Big[\frac{n-1}{n}\mathbf{S}\#\mathbf{S} + \underbrace{\frac{1}{n^2}\sum_{k=1}^{n}E(\mathbf{S}_k\#\mathbf{S}_k)}_{:=\pi^2}\Big]\mathbf{1}_d
\end{aligned}$$

Also muss π^2 geschätzt werden. Es wird wieder das Resultat in Lemma 3.2 aus der Matrizenrechnung verwendet. Dann lautet die Gleichung:

$$\mathbf{1}_d'(\mathbf{S}_k \# \mathbf{S}_k)\mathbf{1}_d = \mathrm{Sp}(\mathbf{S}_k^2).$$

Die Überlegungen führen schließlich zu einem erwartungstreuen Schätzer B_2 für $\mathrm{Sp}[(\mathbf{TV})^2]$:

$$B_2 = \mathrm{Sp}\left(\frac{n}{n-1}\widehat{\mathbf{S}}_n^{\,2} - \frac{1}{n(n-1)}\sum_{k=1}^{n}\mathbf{S}_k^2\right). \tag{3.5}$$

□

Im Folgenden wird gezeigt, dass die nun konstruierten Schätzer B_1 und B_2 durch Umformungen auf eine einfachere und übersichtlichere Form gebracht werden können:

Lemma 3.3
Seien

$$\mathbf{X}_k = (X_{k1}, ..., X_{kd})' \qquad k, l = 1, ..., n$$

u.i.v. Zufallsvektoren, $\mathbf{T}$ *ein Projektor und* $\mathbf{Y}_k := \mathbf{T}\mathbf{X}_k$. *Sei*

$$\widehat{\mathbf{S}}_n := \frac{1}{n}\sum_{k=1}^{n}\mathbf{X}_k\mathbf{T}\mathbf{X}_k'$$

der nichtzentrierte Schätzer für die Kovarianzmatrix $\mathbf{V}$. *Bezeichne weiter* $A_{kl} = \mathbf{X}_k'\mathbf{T}\mathbf{X}_l$ *eine symmetrische Bilinearform in* $\mathbf{X}_k$ *und* $\mathbf{X}_l$, *für* $k = l$ *sei* $A_k := A_{kk}$ *die entsprechende quadratische Form. Dann gilt:*

$$1.\ B_1 = \left[\mathrm{Sp}(\widehat{\mathbf{S}}_n)\right]^2 - \frac{1}{n(n-1)}\sum_{k=1}^{n}(A_k - \overline{A}_\cdot)^2 = \frac{1}{n(n-1)}\underbrace{\sum_{k=1}^{n}\sum_{l=1}^{n}}_{k\neq l}A_kA_l$$

$$2.\ B_2 = \mathrm{Sp}\left(\frac{n}{n-1}\widehat{\mathbf{S}}_n^{\,2} - \frac{1}{n(n-1)}\sum_{k=1}^{n}\mathbf{S}_k^2\right) = \frac{1}{n(n-1)}\underbrace{\sum_{k=1}^{n}\sum_{l=1}^{n}}_{k\neq l}A_{kl}^2$$

Beweis:

1. Man beachte zunächst, dass mit $\mathbf{Y}_k = \mathbf{T}\mathbf{X}_k$ und $\mathbf{S}_k = \mathbf{Y}_k\mathbf{Y}_k'$ gilt:

$$\begin{aligned}
\mathrm{Sp}(\widehat{\mathbf{S}}_n) &= \mathrm{Sp}\left(\frac{1}{n}\sum_{k=1}^{n}\mathbf{S}_k\right) = \frac{1}{n}\sum_{k=1}^{n}\mathrm{Sp}\left(\mathbf{S}_k\right)\\
&= \frac{1}{n}\sum_{k=1}^{n}\mathrm{Sp}\left(\mathbf{Y}_k\mathbf{Y}_k'\right) = \frac{1}{n}\sum_{k=1}^{n}\mathrm{Sp}\left(\mathbf{Y}_k'\mathbf{Y}_k\right)\\
&= \frac{1}{n}\sum_{k=1}^{n}\mathrm{Sp}\left(\underbrace{\mathbf{X}_k'\mathbf{T}\mathbf{T}\mathbf{X}_k}_{=A_k}\right)\\
&= \frac{1}{n}\sum_{k=1}^{n}A_k = \overline{A}_\cdot
\end{aligned}$$

Dies folgt aus der Invarianz der Spur unter zyklischen Vertauschungen und den Projektoreigenschaften von $\mathbf{T}$. Dann kann man folgende Umformungen durchführen:

$$B_1 = \left[\mathrm{Sp}(\widehat{\mathbf{S}}_n)\right]^2 - \frac{1}{n(n-1)}\sum_{k=1}^{n}(A_k - \overline{A}_\cdot)^2 \quad =$$

$$
\begin{aligned}
\overline{A}_{\cdot}^{\,2} - \frac{1}{n(n-1)}\left[\sum_{k=1}^{n} A_k^2 - n\overline{A}_{\cdot}^{\,2}\right] &= \\
\frac{n}{n-1}\overline{A}_{\cdot}^{\,2} - \frac{1}{n(n-1)}\sum_{k=1}^{n} A_k^2 &= \\
\frac{1}{n(n-1)}\left[\left(\sum_{k=1}^{n} A_k\right)^2 - \sum_{k=1}^{n} A_k^2\right] &= \frac{1}{n(n-1)}\underbrace{\sum_{k=1}^{n}\sum_{l=1}^{n}}_{k\neq l} A_k A_l
\end{aligned}
$$

2. Es gilt - zunächst ohne die Spuren:

$$
\begin{aligned}
\frac{n}{n-1}\left(\frac{1}{n}\sum_{k=1}^{n}\mathbf{S}_k\right)^2 - \frac{1}{n(n-1)}\sum_{k=1}^{n}\mathbf{S}_k^2 &= \\
\frac{1}{n(n-1)}\left[\sum_{k=1}^{n}\mathbf{S}_k\mathbf{S}_l - \sum_{k=1}^{n}\mathbf{S}_k^2\right] &= \frac{1}{n(n-1)}\sum_{k\neq l}\mathbf{S}_k\mathbf{S}_l
\end{aligned}
$$

Jetzt erhält man mit der Invarianz der Spur unter zyklischen Vertauschungen die folgende Darstellung ($k \neq l$):

$$
\begin{aligned}
\mathrm{Sp}(\mathbf{S}_k\mathbf{S}_l) &= \mathrm{Sp}(\mathbf{Y}_k\mathbf{Y}_k'\mathbf{Y}_l\mathbf{Y}_l') \\
&= \mathbf{Y}_l'\mathbf{Y}_k\mathbf{Y}_k'\mathbf{Y}_l = A_{lk}A_{kl} \overset{Sym.}{=} A_{kl}^2
\end{aligned}
$$

Schließlich gilt für den Schätzer B_2 dann:

$$
B_2 = \frac{1}{n(n-1)}\underbrace{\sum_{k=1}^{n}\sum_{l=1}^{n}}_{k\neq l} A_{kl}^2
$$

□

Durch Zusammensetzen der obigen Ergebnisse erhält man schließlich einen neuen Schätzer $\widetilde{f}$:

$$
\widetilde{f} = \frac{B_1}{B_2} = \frac{\sum_{k\neq l} A_k A_l}{\sum_{k\neq l} A_{kl}^2}.
$$

3.2 Quadratische und Bilinearformen

Quadratische Formen sind Spezialfälle der Bilinearformen: für $k = l$ wird aus der Bilinearform A_{kl} die quadratische Form A_k. Deshalb werden zunächst die Ergebnisse über quadratische Formen beschrieben und dann die der Bilinearformen daraus abgeleitet. Im Folgenden soll gelten, dass immer dann eine quadratische Form gemeint ist, wenn die Indizes gleich sind und die Bilinearform nur dann Bilinearform genannt wird, wenn die Indizes verschieden sind:

Definition 3.4 *Es gelte folgende Unterscheidung:*

$$
A_{kl} := \mathbf{X}_k'\mathbf{T}\mathbf{X}_l \mathrel{\widehat{=}} \begin{cases} \text{Quadratform} & k = l \\ \text{Bilinearform} & k \neq l \end{cases}
$$

Zunächst also die Eigenschaften der quadratischen Formen: das Darstellungslemma zeigt, dass sich eine quadratische Form als Summe von unkorrelierten Zufallsvariablen darstellen lässt.

Lemma 3.5 (Darstellung einer quadratischen Form)
Sei $\mathbf{X} = (X_1, ..., X_n)'$ *ein Zufallsvektor mit* $E(\mathbf{X}) = \boldsymbol{\mu}$ *und* $\mathrm{Cov}(\mathbf{X}) = \mathbf{V} > 0$ *sowie* $\mathbf{T}$ *ein Projektor. Dann gilt für die quadratische Form*

$$A = \mathbf{X}'\mathbf{T}\mathbf{X} = \begin{cases} \sum_{i=1}^{n} \lambda_i (U_i + b_i)^2 & \text{für } E(\mathbf{X}) = \boldsymbol{\mu} \neq \mathbf{0} \\ \sum_{i=1}^{n} \lambda_i U_i^2 & \text{für } E(\mathbf{X}) = \mathbf{0} \end{cases}$$

wobei die λ_i *die Eigenwerte von* $\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}$ *sind.*
Außerdem ist $\mathbf{U} = (U_1, ..., U_d)$ *ein Zufallsvektor mit* $E(\mathbf{U}) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{U}) = \mathbf{I}$ *und* $(b_1, ..., b_d) = (\mathbf{P}'\mathbf{V}^{\frac{1}{2}}\boldsymbol{\mu})'$ *mit* $\mathbf{P}\mathbf{P}' = \mathbf{I}$.

Beweis: siehe MATHAI, PROVOST [15] (S.28 f)

Wenn man zusätzlich fordert, dass die $\mathbf{X}$ normalverteilt sind, dann kann man folgendes Resultat über die Verteilung der entsprechenden quadratischen Form herleiten.

Korollar 3.6 (Verteilung einer quadratischen Form)
Seien $\mathbf{X} = (X_1, ..., X_n)' \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$, $|\mathbf{V}| \neq 0$, $\mathbf{V} = \mathbf{V}'$ *und* $r(\mathbf{V}) = n$ *sowie* $\mathbf{T}$ *ein Projektor. Dann gilt für die quadratische Form*

$$A = \mathbf{X}'\mathbf{T}\mathbf{X} = \sum_{i=1}^{n} \lambda_i C_i$$

wobei $C_i \sim \chi_1^2$ *u.i.v. und die* λ_i *die Eigenwerte von* $\mathbf{T}\mathbf{V}$ *sind.*

Beweis: Es gilt: $|\mathbf{V}| \neq 0$ und $\mathbf{V} = \mathbf{V}' \Rightarrow \exists \mathbf{V}^{1/2}$ und$\mathbf{V}^{-1/2}$, symmetrisch und invertierbar $\Rightarrow \mathbf{V}^{1/2}\mathbf{V}^{-1/2} = \mathbf{I}$.

$$\begin{aligned} \mathbf{X}'\mathbf{T}\mathbf{X} &= \mathbf{X}'\mathbf{V}^{-1/2}\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}\mathbf{V}^{-1/2}\mathbf{X} \\ &= \mathbf{X}'\mathbf{V}^{-1/2}\mathbf{P}'\mathbf{P}\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}\mathbf{P}'\mathbf{P}\mathbf{V}^{-1/2}\mathbf{X} \end{aligned}$$

Da $\mathbf{V}^{1/2}$ und $\mathbf{T}$ symmetrisch sind, gilt auch, dass $\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}$ symmetrisch ist. Damit folgt aus dem Satz über die Hauptachsentransformation: es existiert eine orthogonale Matrix $\mathbf{P}$, sodass $\mathbf{P}\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}\mathbf{P}' = diag\{\lambda_1', ..., \lambda_n'\} = \Delta$. Die λ_i' sind hierbei die Eigenwerte von $\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}$. Also folgt weiter:

$$\mathbf{X}'\mathbf{T}\mathbf{X} = (\mathbf{P}\mathbf{V}^{-1/2}\mathbf{X})'\underbrace{\mathbf{P}\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}\mathbf{P}'}_{\Delta}(\mathbf{P}\mathbf{V}^{-1/2}\mathbf{X}) = \mathbf{Z}'\Delta\mathbf{Z} = \sum_{i=1}^{n} \lambda_i' Z_i^2$$

mit $\mathbf{Z} = \mathbf{P}\mathbf{V}^{-1/2}\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}\mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2}\mathbf{P}') = \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$.

Es bleibt also zu zeigen, dass $\lambda_i = \lambda_i'$ gilt.
Sei deshalb λ ein beliebiger Eigenwert von $\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}$. Dann gilt:

$$\begin{aligned} \mathbf{0} = |\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2} - \lambda\mathbf{I}| &= |\mathbf{I}||\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2} - \lambda\mathbf{I}| \\ &= |\mathbf{V}^{-1/2}||\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2} - \lambda\mathbf{I}||\mathbf{V}^{1/2}| \\ &= |\mathbf{V}^{-1/2}\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}\mathbf{V}^{1/2} - \lambda\mathbf{V}^{-1/2}\mathbf{I}\mathbf{V}^{1/2}| \\ &= |\mathbf{T}\mathbf{V} - \lambda\mathbf{I}| \end{aligned}$$

Also ist λ auch Eigenwert von $\mathbf{T}\mathbf{V}$. □

Das Korollar zeigt, dass eine quadratische Form wie eine Summe von unkorrelierten (bzw. unabhängigen bei Normalverteilung) χ_1^2-verteilten Zufallsvariablen verteilt ist. Mit Hilfe dieser Information ist es möglich, die ersten zentralen Momente von quadratischen Formen auszurechnen.

Lemma 3.7 (Momente von quadratischen Formen)
Seien $\mathbf{X}_k = (X_{k1}, ..., X_{kn}) \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$, *u.i.v.*, $(k = 1, ..., N)$ *sowie* $\mathbf{T}$ *ein Projektor. Dann gilt für die quadratischen Formen* $A_k = \mathbf{X}_k' \mathbf{T} \mathbf{X}_k$ *zunächst allgemein:*

1. $E(A_k) = \mathrm{Sp}(\mathbf{TV}) + \boldsymbol{\mu}' \mathbf{T} \boldsymbol{\mu}$
2. $\mathrm{Var}(A_k) = 2\,\mathrm{Sp}[(\mathbf{TV})^2] + 4\boldsymbol{\mu}' \mathbf{TVT} \boldsymbol{\mu}$

Falls $\mathbf{T}\boldsymbol{\mu} = \mathbf{0}$ *ist, so gilt außerdem:*

3. $E(A_k) = \mathrm{Sp}(\mathbf{TV}) =: \nu$
4. $\mathrm{Var}(A_k) = 2\,\mathrm{Sp}[(\mathbf{TV})^2] =: \tau^2$
5. *Für die Varianzen bzw. Kovarianzen der gemischten Paare* $A_k A_l$ *und* $A_m A_n$ *(es sei immer* $k \neq l$ *bzw.* $m \neq n$*) gilt weiterhin:*

$$\mathrm{Cov}(A_k A_l, A_m A_n) = \begin{cases} \tau^4 + 2\,\tau^2 \nu^2 & : \quad (k,l) = (m,n) \\ & \quad bzw.\ (k,l) = (n,m) \\ \tau^2 \nu^2 & : \quad k = m \text{ und } l \neq n \\ & \quad bzw.\ k \neq m \text{ und } l = n \\ 0 & : \quad sonst \end{cases}$$

Beweis:

1. (Satz von Lancaster)

$$\begin{aligned} E(A_k) = E(\mathbf{X}'\mathbf{TX})) &= E(\mathrm{Sp}(\mathbf{X}'\mathbf{TX})) = E(\mathrm{Sp}(\mathbf{TXX}')) \\ &= \mathrm{Sp}(\mathbf{T} \cdot E(\mathbf{XX}')) \\ &= \mathrm{Sp}(\mathbf{TV}) + \mathrm{Sp}(\mathbf{T}\boldsymbol{\mu}\boldsymbol{\mu}') \\ &= \mathrm{Sp}(\mathbf{TV}) + \boldsymbol{\mu}' \mathbf{T} \boldsymbol{\mu} \end{aligned}$$

2. siehe Mathai, Provost [15] (S.53)
3. folgt unmittelbar aus (1)
4. folgt unmittelbar aus (2)
5. - $\mathrm{Var}(A_k A_l) = E(A_k A_l A_k A_l) - [E(A_k A_l)]^2$
     $= [E(A_k^2)]^2 - [E(A_k)]^4 = (\tau^2 + \nu^2)^2 - \nu^4$
     $= \tau^4 + 2\,\tau^2 \nu^2$
   - $\mathrm{Cov}(A_k A_l, A_k A_m) = E(A_k A_l A_k A_m) - E(A_k A_l) E(A_k A_m)$
     $= (\tau^2 + \nu^2)\nu^2 - \nu^4 = \tau^2 \nu^2$
   - $\mathrm{Cov}(A_k A_l, A_m A_n) = \nu^4 - \nu^4 = 0$

□

Nachdem nun alle erforderlichen Eigenschaften der quadratischen Formen hergeleitet wurden, sollen jetzt die Bilinearformen betrachtet werden. Zunächst also wieder das Darstellungslemma in leicht abgewandelter Form (das Darstellungslemma oben ist offensichtlich nur ein Spezialfall hiervon).

Lemma 3.8 (Darstellung einer Bilinearform)
Seien $\mathbf{X} = (X_1, ..., X_n)'$ *und* $\mathbf{Y} = (Y_1, ..., Y_n)'$ *unabhängige, identisch verteilte Zufallsvektoren mit*

$E(\mathbf{X}) = E(\mathbf{Y}) = \boldsymbol{\mu}$ *und* $\mathrm{Cov}(\mathbf{X}) = \mathrm{Cov}(\mathbf{Y}) = \mathbf{V} > 0$ *sowie* $\mathbf{T}$ *ein Projektor. Dann gilt für die Bilinearform*

$$A = \mathbf{X}'\mathbf{T}\mathbf{Y} = \begin{cases} \sum_{i=1}^{n} \lambda_i (U_i + b_i)(W_i + b_i) & \text{für } E(\mathbf{X}) = E(\mathbf{Y}) = \boldsymbol{\mu} \neq \mathbf{0} \\ \sum_{i=1}^{n} \lambda_i U_i W_i & \text{für } E(\mathbf{X}) = E(\mathbf{Y}) = \mathbf{0} \end{cases}$$

wobei die λ_i *die Eigenwerte von* $\mathbf{TV}$ *sind.*
Außerdem sind $\mathbf{U} = (U_1, ..., U_d), \mathbf{W} = (W_1, ..., W_d)$ *Zufallsvektoren mit* $E(\mathbf{U}) = E(\mathbf{W}) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{U}, \mathbf{W}) = \mathbf{0}$ *und* $\mathrm{Cov}(\mathbf{U}) = \mathrm{Cov}(\mathbf{W}) = \mathbf{I}$ *und* $(b_1, ..., b_d) = (\mathbf{P}'\mathbf{V}^{\frac{1}{2}}\boldsymbol{\mu})'$ *mit* $\mathbf{P}\mathbf{P}' = \mathbf{I}$.

Beweis: Setze $\mathbf{Z}_X = (\mathbf{V}^{-1/2}\mathbf{X} - \mathbf{V}^{-1/2}\boldsymbol{\mu})$ und analog $\mathbf{Z}_Y = (\mathbf{V}^{-1/2}\mathbf{Y} - \mathbf{V}^{-1/2}\boldsymbol{\mu})$. Nach dem Satz über die Hauptachsentransformation existiert eine orthogonale Matrix $\mathbf{P}$, so dass gilt:

$$\mathbf{P}\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}\mathbf{P}' = diag\{\lambda_1', ..., \lambda_n'\} =: \Delta, \qquad \mathbf{P}\mathbf{P}' = \mathbf{I}$$

Setze dann $\mathbf{U} = \mathbf{P}'\mathbf{Z}_X$ und $\mathbf{W} = \mathbf{P}'\mathbf{Z}_Y$. Damit gilt :

$$E(\mathbf{U}) = E(\mathbf{W}) = \mathbf{0} \qquad \mathrm{Cov}(\mathbf{U}) = \mathrm{Cov}(\mathbf{W}) = \mathbf{I}$$

Dann folgt für die Bilinearform:

$$\begin{aligned} \mathbf{X}'\mathbf{T}\mathbf{Y} &= \mathbf{X}'\mathbf{V}^{-1/2}\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}\mathbf{V}^{-1/2}\mathbf{Y} \\ &= (\mathbf{Z}_X + \mathbf{V}^{-1/2}\boldsymbol{\mu})'\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}(\mathbf{Z}_Y + \mathbf{V}^{-1/2}\boldsymbol{\mu}) \\ &= (\mathbf{Z}_X + \mathbf{V}^{-1/2}\boldsymbol{\mu})'\mathbf{P}'\underbrace{\mathbf{P}\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}\mathbf{P}'}_{\Delta}\mathbf{P}(\mathbf{Z}_Y + \mathbf{V}^{-1/2}\boldsymbol{\mu}) \\ &= (\mathbf{U} + \mathbf{b})'\Delta(\mathbf{W} + \mathbf{b}) \end{aligned}$$

□

Wenn man nun zusätzlich fordert, dass die Vektoren $\mathbf{X}$ und $\mathbf{Y}$ normalverteilt sind, so kann die Aussage über die Verteilung einer Bilinearform präzisiert werden.

Korollar 3.9 (Verteilung einer Bilinearform)
Seien $\mathbf{X} = (X_1, ..., X_n)'$ *und* $\mathbf{Y} = (Y_1, ..., Y_n)' \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$ *und unabhängig,* $|\mathbf{V}| \neq 0$, $\mathbf{V} = \mathbf{V}'$ *und* $r(\mathbf{V}) = d$ *sowie* $\mathbf{T}$ *ein Projektor. Dann gilt für die Bilinearform*

$$A = \mathbf{X}'\mathbf{T}\mathbf{Y} \sim \sum_{i=1}^{n} \lambda_i C_i D_i$$

wobei $C_i, D_i \sim \mathcal{N}(0, 1)$ *u.i.v. und die* λ_i *die Eigenwerte von* $\mathbf{TV}$ *sind.*

Beweis: Es gilt: $|\mathbf{V}| \neq 0$ und $\mathbf{V} = \mathbf{V}' \Rightarrow \exists \mathbf{V}^{1/2}$ und$\mathbf{V}^{-1/2}$, symmetrisch und invertierbar $\Rightarrow \mathbf{V}^{1/2}\mathbf{V}^{-1/2} = \mathbf{I}$.

$$\begin{aligned} \mathbf{X}'\mathbf{T}\mathbf{Y} &= \mathbf{X}'\mathbf{V}^{-1/2}\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}\mathbf{V}^{-1/2}\mathbf{Y} \\ &= \mathbf{X}'\mathbf{V}^{-1/2}\mathbf{P}'\mathbf{P}\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}\mathbf{P}'\mathbf{P}\mathbf{V}^{-1/2}\mathbf{Y} \end{aligned}$$

Da $\mathbf{V}^{1/2}$ und $\mathbf{T}$ symmetrisch sind ist auch $\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}$ symmetrisch. Damit folgt aus dem Satz über die Hauptachsentransformation: es existiert ein orthogonales $\mathbf{P}$, sodass $\mathbf{P}\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}\mathbf{P}' = diag\{\lambda_1', ..., \lambda_n'\} = \Delta$. Die λ_i' sind hierbei die Eigenwerte von $\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}$. Es folgt weiter:

$$\mathbf{X}'\mathbf{T}\mathbf{Y} = (\mathbf{P}\mathbf{V}^{-1/2}\mathbf{X})'\underbrace{\mathbf{P}\mathbf{V}^{1/2}\mathbf{T}\mathbf{V}^{1/2}\mathbf{P}'}_{\Delta}(\mathbf{P}\mathbf{V}^{-1/2}\mathbf{Y}) = \mathbf{C}'\Delta\mathbf{D} = \sum_{i=1}^{n} \lambda_i' C_i D_i$$

mit $\mathbf{C} = \mathbf{P}\mathbf{V}^{-1/2}\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}\mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2}\mathbf{P}') = \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ und analog
$\mathbf{D} = \mathbf{P}\mathbf{V}^{-1/2}\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}\mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2}\mathbf{P}') = \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$.
Es bleibt noch zu zeigen, dass $\lambda_i = \lambda_i'$ gilt. Das wurde aber bereits im Beweis zu Korollar 3.6 gezeigt.

□

Mit Hilfe der obigen Betrachtungen kann man nun die ersten zentralen Momente der Bilinearformen berechnen.

Lemma 3.10 (Momente von Bilinearformen)
Seien $\mathbf{X}_k = (X_{k1}, ..., X_{kn})'$, $\mathbf{X}_l = (X_{l1}, ..., X_{ln})' \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$, *u.i.v.*, $(k \neq l \in \{1, ..., n\})$ *sowie* $\mathbf{T}$ *ein Projektor. Dann gilt für die Bilinearformen* $A_{kl} = \mathbf{X}_k'\mathbf{T}\mathbf{X}_l$ *:*

1. $E(A_{kl}) = 0$
2. $\mathrm{Var}(A_{kl}) = E(A_{kl}^2) = \mathrm{Sp}[(\mathbf{TV})^2]$
3. $E(A_{kl}^4) = 6\,\mathrm{Sp}[(\mathbf{TV})^4] + 3\,(\mathrm{Sp}[(\mathbf{TV})^2])^2$
4. $E(A_{kl}^2 A_{kn}^2) = 2\,\mathrm{Sp}[(\mathbf{TV})^4] + (\mathrm{Sp}[(\mathbf{TV})^2])^2$

Beweis:

1. $E(A_{kl}) = E(\sum_{i=1}^d \lambda_i C_i D_i) = \sum_{i=1}^d \lambda_i \underbrace{E(C_i)E(D_i)}_{=0} = 0$

2. $\mathrm{Var}(A_{kl}) = E(A_{kl}^2) = E(\sum_{i=1}^d \lambda_i C_i D_i \sum_{j=1}^d \lambda_j C_j D_j)$

$$= E(\textstyle\sum_{i=1}^d \lambda_i^2 C_i^2 D_i^2) = \sum_{i=1}^d \lambda_i^2 \underbrace{E(C_i^2)}_{=1}\underbrace{E(D_i^2)}_{=1}$$

$$= \textstyle\sum_{i=1}^d \lambda_i^2 = \mathrm{Sp}[(\mathbf{TV})^2]$$

3. $E(A_{kl}^4) = E\left((\sum \lambda_i C_i D_i)^4\right)$

$$= \sum_i \sum_j \sum_r \sum_s \lambda_i\lambda_j\lambda_r\lambda_s E(C_iD_iC_jD_jC_rD_rC_sD_s)$$

$$= \sum_i \lambda_i^4 \underbrace{E(C_i^4 D_i^4)}_{9} + 3\sum_{i\neq j} \lambda_i^2\lambda_j^2 \underbrace{E(C_i^2D_i^2C_j^2D_j^2)}_{1}$$

$$= 6\sum_i \lambda_i^4 + 3\sum_{i,j}\lambda_i^2\lambda_j^2 = 6\,\mathrm{Sp}[(\mathbf{TV})^4] + 3\,(\mathrm{Sp}[(\mathbf{TV})^2])^2$$

4. $E(A_{kl}^2 A_{kn}^2) = E((\sum \lambda_i C_i D_i)^2(\sum \lambda_i C_i E_i)^2)$

$$= \sum_i \sum_j \sum_r \sum_s \lambda_i\lambda_j\lambda_r\lambda_s E(C_iD_iC_jD_jC_rE_rC_sE_s)$$

$$= \sum_i \lambda_i^4 \underbrace{E(C_i^4 D_i^2 E_i^2)}_{3} + \sum_{i\neq j} \lambda_i^2\lambda_j^2 \underbrace{E(C_i^2D_i^2C_j^2E_j^2)}_{1}$$

$$= 2\sum_i \lambda_i^4 + \sum_{i,j}\lambda_i^2\lambda_j^2 = 2\,\mathrm{Sp}[(\mathbf{TV})^4] + (\mathrm{Sp}[(\mathbf{TV})^2])^2$$

□

3.3 Konsistenz und Dimensionsstabilität der Schätzer

Folgendes Lemma liefert eine Regularitätsbedingung für die Eigenwerte der Kovarianzmatrix, die die Konsistenz und Dimensionsstabilität der neuen Schätzer sichert.

Lemma 3.11 *Seien* $\lambda_i,\ (\lambda_i \geq \lambda_0 > 0)$ *die Eigenwerte von* $\mathbf{TVT}$ *und* $m := \max_i \lambda_i < \infty$.

1.
$$\lim_{d\to\infty} \frac{m}{\sum_{i=1}^{d} \lambda_i} = 0 \qquad \Rightarrow \qquad \lim_{d\to\infty} \frac{\mathrm{Sp}[(\mathbf{TV})^2]}{[\mathrm{Sp}(\mathbf{TV})]^2} = 0$$

2.
$$\lim_{d\to\infty} \frac{m^2}{\sum_{i=1}^{d} \lambda_i^2} = 0 \qquad \Rightarrow \qquad \lim_{d\to\infty} \frac{\mathrm{Sp}[(\mathbf{TV})^4]}{(\mathrm{Sp}[(\mathbf{TV})^2])^2} = 0$$

Beweis:

1.
$$\frac{\mathrm{Sp}[(\mathbf{TV})^2]}{[\mathrm{Sp}(\mathbf{TV})]^2} = \frac{\sum \lambda_i^2}{(\sum \lambda_i)^2} \leq \frac{m \sum \lambda_i}{(\sum \lambda_i)^2} = \frac{m}{\sum \lambda_i} \overset{d\to\infty}{\longrightarrow} 0$$

2.
$$\frac{\mathrm{Sp}[(\mathbf{TV})^4]}{(\mathrm{Sp}[(\mathbf{TV})^2])^2} = \frac{\sum \lambda_i^4}{(\sum \lambda_i^2)^2} \leq \frac{m^2 \sum \lambda_i^2}{(\sum \lambda_i^2)^2} = \frac{m^2}{\sum \lambda_i^2} \overset{d\to\infty}{\longrightarrow} 0$$

Bemerkung 3.12 *Aus der zweiten Bedingung folgt die erste:*

$$\frac{m^2}{(\sum \lambda_i)^2} \leq \frac{m^2}{\sum \lambda_i^2} \text{ denn } \left(\sum \lambda_i\right)^2 \geq \sum \lambda_i^2$$

da die $\lambda_i \geq 0$ *sind.*

Mit Hilfe von Lemma 2.3 wird nun die Konsistenz bzw. Dimensionsstabilität der beiden erwartungstreuen Schätzer B_1 und B_2 gezeigt, vorausgesetzt, sie erfüllen die Bedingungen aus Lemma 3.11. Es muß also gezeigt werden, dass der Quotient aus der Varianz des Schätzers und dem Parameter zum Quadrat für hinreichend großes d gleichmäßig beschränkt ist und diese Schranke nur noch von n abhängt.
Die folgenden Aussagen über die Varianz einer Summe von abhängigen Zufallsvariablen X_i werden anschließend verwendet:

$$\begin{aligned} \mathrm{Var}\sum_i X_i &= \sum_i \mathrm{Var}(X_i) + \sum_{i\neq j} \mathrm{Cov}(X_i, X_j) \\ \mathrm{Var}\sum_i X_i &= E(\sum_{i,j} X_i X_j) - [E(\sum_i X_i)]^2 \end{aligned}$$

Daraus folgt für $V_1 = \mathrm{Var}(B_1)$:

$$n^2(n-1)^2 V_1 = \mathrm{Var}(\sum_{k\neq l} A_k A_l) = \sum_{k\neq l} \mathrm{Var}(A_k A_l) + \underbrace{\sum_{k\neq l}\sum_{r\neq s}}_{(k,l)\neq(r,s)} \mathrm{Cov}(A_k A_l, A_r A_s)$$

Weiter folgt mit Lemma 3.7:

$$\begin{aligned} n^2(n-1)^2 V_1 &= n(n-1)\mathrm{Var}(A_k A_l) \\ &\quad +n(n-1)(n-2)(n-3)\mathrm{Cov}(A_k A_l, A_r A_s) \\ &\quad +4n(n-1)(n-2)\mathrm{Cov}(A_k A_l, A_k A_r) \\ &\quad +n(n-1)\underbrace{\mathrm{Cov}(A_k A_l, A_l A_k)}_{=\mathrm{Var}(A_k A_l)} \\ &= 2n(n-1)(\sigma^4 + 2\,\sigma^2\mu^2) + 4\,n(n-1)(n-2)\sigma^2\mu^2 \\ &= 2n(n-1)\sigma^4 + (4n^3 - 8\,n^2 + 4n)\sigma^2\mu^2 \end{aligned}$$

Man sieht, dass der Term $\mu^4 = [\mathrm{Sp}(\mathbf{TV})]^4$ in der Varianz nicht mehr vorkommt. Mit Hilfe von Lemma 3.11 kann man dann zeigen, dass $V_1/[\mathrm{Sp}(\mathbf{TV})]^4$ für $d \to \infty$ verschwindet:

$$\frac{V_1}{[\mathrm{Sp}(\mathbf{TV})]^4} = \frac{2}{n(n-1)} \underbrace{\frac{(\mathrm{Sp}[(\mathbf{TV})^2])^2}{[\mathrm{Sp}(\mathbf{TV})]^4}}_{\to 0} + \frac{(4n^3 - 8\,n^2 + 4n)}{n^2(n-1)^2} \underbrace{\frac{\mathrm{Sp}[(\mathbf{TV})^2]}{[\mathrm{Sp}(\mathbf{TV})]^2}}_{\to 0} \tag{3.6}$$

Außerdem ist der Quotient für alle $d, n > 1$ beschränkt durch $\frac{4n}{(n-1)^2}$:

$$\begin{aligned} \frac{V_1}{[\mathrm{Sp}(\mathbf{TV})]^4} &= \frac{2}{n(n-1)} \underbrace{\frac{(\mathrm{Sp}[(\mathbf{TV})^2])^2}{[\mathrm{Sp}(\mathbf{TV})]^4}}_{\leq 1} + \frac{(4n^3 - 8\,n^2 + 4n)}{n^2(n-1)^2} \underbrace{\frac{\mathrm{Sp}[(\mathbf{TV})^2]}{[\mathrm{Sp}(\mathbf{TV})]^2}}_{\leq 1} \\ &\leq \frac{2}{n(n-1)} + \frac{(4n^2 - 8\,n + 4)}{n(n-1)^2} \leq \frac{4n^2 - 6n + 2}{n(n-1)^2} \leq \frac{4n}{(n-1)^2} \end{aligned}$$

Für den zweiten Quotienten $V_2 = \mathrm{Var}(B_2)$ gilt:

$$n^2(n-1)^2 V_2 = \mathrm{Var}(\sum_{k \neq l} A_{kl}^2) = E(\sum_{k \neq l} \sum_{r \neq s} A_{kl}^2 A_{rs}^2) - (E(\sum_{k \neq l} A_{kl}^2))^2$$

Weiter folgt mit Lemma 3.10:

$$\begin{aligned} n^2(n-1)^2 V_2 &= 2n(n-1)E(A_{kl}^4) \\ &\quad +4n(n-1)(n-2)E(A_{kl}^2 A_{kn}^2) \\ &\quad +n(n-1)(n-2)(n-3)E(A_{kl}^2)^2 \\ &\quad -n^2(n-1)^2 \mathrm{Var}(A_{kl})^2 \\ &= 2n(n-1)(6\mathrm{Sp}[(\mathbf{TV})^4] + 3(\mathrm{Sp}[(\mathbf{TV})^2])^2) \\ &\quad +4n(n-1)(n-2)(2\mathrm{Sp}[(\mathbf{TV})^4] + (\mathrm{Sp}[(\mathbf{TV})^2])^2)) \\ &\quad +n(n-1)(n-2)(n-3)(\mathrm{Sp}[(\mathbf{TV})^2])^2) \\ &\quad -n^2(n-1)^2(\mathrm{Sp}[(\mathbf{TV})^2])^2) \\ &= n(n-1)[(8n-4)\mathrm{Sp}[(\mathbf{TV})^4] + 4(\mathrm{Sp}[(\mathbf{TV})^2])^2)] \end{aligned}$$

Leider erhält man hier nicht so ein schönes Ergebnis wie im Zähler. Der Term $(\mathrm{Sp}[(\mathbf{TV})^2])^2$ verschwindet im Ausdruck V_2 nicht, deshalb bleibt im Quotienten der zweite Summand für festes n stehen:

$$\frac{V_2}{(\mathrm{Sp}[(\mathbf{TV})^2])^2} = \frac{8n-4}{n(n-1)} \underbrace{\frac{\mathrm{Sp}[(\mathbf{TV})^4]}{(\mathrm{Sp}[(\mathbf{TV})^2])^2}}_{\to 0} + \frac{4}{n(n-1)} \tag{3.7}$$

Dieser Quotient ist für alle $d, n > 1$ beschränkt durch $\frac{8}{n-1}$:

$$\frac{V_2}{(\mathrm{Sp}[(\mathbf{TV})^2])^2} = \frac{8n-4}{n(n-1)} \underbrace{\frac{\mathrm{Sp}[(\mathbf{TV})^4]}{(\mathrm{Sp}[(\mathbf{TV})^2])^2}}_{\leq 1} + \frac{4}{n(n-1)} \leq \frac{8n-4+4}{n(n-1)} \leq \frac{8}{(n-1)}$$

Es konnte also gezeigt werden, dass B_1 für festes n und $d \to \infty$ konsistent im Sinne der $\mathcal{L}_2$-Norm und somit dimensionsstabil ist. Ausserdem ist B_1 auch konsistent für $n \to \infty$ und beliebiges d. B_2 dagegen ist nur für $n \to \infty$ und beliebiges d konsistent. Für festes n und $d \to \infty$ ist die Varianz von B_2 aber immerhin beschränkt.

3.4 Verzerrung des Quotienten

Nachdem die dimensionsstabilen Schätzer für $[\mathrm{Sp}(\mathbf{TV})]^2$ und $\mathrm{Sp}[(\mathbf{TV})^2]$ konstruiert wurden, sollte man die Verzerrung des Quotienten B_1/B_2 untersuchen. Ohne Kenntnis der gemeinsamen Verteilung von B_1 und B_2 ist es jedoch im Allgemeinen nicht möglich, den Erwartungswert des Quotienten exakt auszurechnen. In STANGE [18] wird eine Taylor-Approximation für solch einen Quotienten angegeben:

$$E\left(\frac{B_1}{B_2}\right) \doteqdot \frac{E(B_1)}{E(B_2)}\left(1+\frac{\mathrm{Var}(B_2)}{[E(B_2)]^2}-\frac{\mathrm{Cov}(B_1,B_2)}{E(B_1)E(B_2)}\right), \tag{3.8}$$

dabei bedeutet das Zeichen $\doteqdot$ "ist ungefähr gleich".
Der größte Teil der Arbeit für die Berechnung dieses Erwartungswertes ist bereits getan:

$$E\left(\frac{B_1}{B_2}\right) \doteqdot \frac{[\mathrm{Sp}(\mathbf{TV})]^2}{\mathrm{Sp}[(\mathbf{TV})^2]}\left(1+\frac{4}{n(n-1)}-\frac{\mathrm{Cov}(B_1,B_2)}{[\mathrm{Sp}(\mathbf{TV})]^2\mathrm{Sp}[(\mathbf{TV})^2]}\right).$$

Was hier noch fehlt, ist die Kovarianz $\mathrm{Cov}(B_1,B_2)$. In diesem Fall handelt es sich um eine gemischte Summe von quadratischen und Bilinearformen.

Lemma 3.13 *Seien B_1 und B_2 definiert wie in Lemma 3.3. Dann gilt:*

$$\mathrm{Cov}(B_1,B_2)=\frac{2}{n(n-1)}\sum_{i\neq j}\lambda_i^2\lambda_j^2+\frac{8n-4}{n(n-1)}\sum_{i\neq j}\lambda_i\lambda_j^3+\frac{8n+2}{n(n-1)}\sum_i\lambda_i^4.$$

Beweis: Der Beweis des Lemmas ist sehr technisch. Zur Vereinfachung des Leseflusses stehen diese Berechnungen im Anhang.

Mit diesem Resultat kann man leicht zeigen, dass die Kovarianz für die Berechnung der Verzerrung nicht relevant ist:

Korollar 3.14 *Der Quotient*

$$\frac{\mathrm{Cov}(B_1,B_2)}{[\mathrm{Sp}(\mathbf{TV})]^2\mathrm{Sp}[(\mathbf{TV})^2]}$$

verschwindet für großes d, wenn die Regularitätsbedingungen von Lemma 3.11 erfüllt sind.

Beweis: Da die Summanden alle positiv sind, reicht es, für jeden einzelnen Summanden zu zeigen, dass er für hinreichend großes d verschwindet.
Sei $m = max\ \lambda_i$. Dann gilt:

1. $\dfrac{\sum_{i\neq j}\lambda_i^2\lambda_j^2}{(\sum\lambda_i)^2\sum\lambda_i^2} \leq \dfrac{(\sum\lambda_i^2)^2}{(\sum\lambda_i)^2\sum\lambda_i^2}=\dfrac{\sum\lambda_i^2}{(\sum\lambda_i)^2}\to 0$

2. $\dfrac{\sum_{i\neq j}\lambda_i\lambda_j^3}{(\sum\lambda_i)^2\sum\lambda_i^2} \leq \dfrac{\sum\lambda_i\sum\lambda_i^3}{(\sum\lambda_i)^2\sum\lambda_i^2}=\dfrac{\sum\lambda_i^3}{(\sum\lambda_i)\sum\lambda_i^2}\leq\dfrac{m^3}{(\sum\lambda_i)\sum\lambda_i^2}\to 0$

3. $\underbrace{\dfrac{\sum\lambda_i^4}{(\sum\lambda_i)^2}}_{\geq\sum\lambda_i^2}\dfrac{}{\sum\lambda_i^2} \leq \dfrac{\sum\lambda_i^4}{(\sum\lambda_i^2)^2}\to 0$

Also verschwindet der ganze Quotient für hinreichend großes d. □

Setzt man schließlich dieses Ergebnis in (3.8) ein, so erhält man folgende Verzerrung für den neuen Schätzer:

$$E(\tilde{f}) \doteqdot \frac{[\mathrm{Sp}(\mathbf{TV})]^2}{\mathrm{Sp}[(\mathbf{TV})^2]}\left(1+\frac{4}{n(n-1)}\right) \tag{3.9}$$

Wie man sieht, ist der neue Schätzer nicht vollkommen unverzerrt, aber die Verzerrung verschwindet mit wachsendem n. Weiterhin sind die Bestandteile des neuen Schätzers dimensionsstabil, d.h. sowohl für grosses n als auch für großes d werden die Schätzer nicht schlechter oder degenerieren. Außerdem ist der neue Schätzer für $n \to \infty$ bei beliebigem d asymptotisch unverzerrt.
Simulationen zeigen, dass die Verzerrung bei den neuen Schätzern B_1 und B_2 viel kleiner ist als bei den alten. Die stärkste Verbesserung ist beim Schätzer B_2 zu beobachten, dessen Momente mit der Stichprobenkovarianzmatrix sehr weit vom wahren Wert abwichen.
Simulationen zum Verhalten des Quotienten B_1/B_2 zeigen, dass die Verzerrung mit wachsendem d nicht stark zunimmt und auch der Quotient der Varianzen nicht mit der Dimension wächst. Die Schwankungen der Werte sind auf Simulationsungenauigkeiten und numerische Probleme bei der Approximation der Verteilung zurückzuführen. Bei großem d (ab 100) ist die Varianz bereits so klein, dass sich Rundungsfehler sogar schon auf die dritte Nachkommastelle auswirken. Um die Schwankungen so gering wie möglich zu halten, wurden für die folgenden Ergebnisse 100 000 Wiederholungen durchgeführt.

3.5 F-Approximation (Fortsetzung)

Da jetzt alle notwendigen Schätzer hergeleitet sind und ihr Verhalten beschrieben wurde, kann die F-Approximation aus Kapitel 2.6 wieder aufgenommen werden. Die Verteilung der Statistik $\widetilde{F}$ soll folgendermaßen approximiert werden:

$$\widetilde{F} = \frac{Q_n}{B_0} \mathrel{\dot{\sim}} F(f_1, f_2)$$

Dazu werden die folgenden Momente benötigt:

Lemma 3.15 *Sei $B_0 = \frac{1}{n}\sum A_k$ und $Q_n = \frac{1}{n}\sum_k \sum_l A_{kl}$, wobei die $A_{kl} = \mathbf{X}_k\mathbf{T}\mathbf{X}_l$ symmetrische Bilinearformen sind. Dann gilt:*

1. $E(B_0) = \mathrm{Sp}(\mathbf{TV})$
2. $\mathrm{Var}(B_0) = \frac{2}{n}\mathrm{Sp}[(\mathbf{TV})^2]$
3. $\mathrm{Cov}(Q_n, B_0) = \mathrm{Var}(B_0)$

Beweis: Unter Verwendung der Ergebnisse aus Lemma 3.7 über die Momente von quadratischen Formen und Lemma 3.10 über die Momente von Bilinearformen folgen die Behauptungen:

$$\begin{aligned}
1. \quad E(\frac{1}{n}\sum A_k) &= \frac{1}{n}\sum E(A_{kk}) = \mathrm{Sp}(\mathbf{TV}) \\
2.\ \mathrm{Var}(\frac{1}{n}\sum A_k) &= \frac{1}{n^2}\sum \mathrm{Var}(A_k) = \frac{2}{n}\mathrm{Sp}[(\mathbf{TV})^2] \\
3. \quad \mathrm{Cov}(Q_n, B_0) &= E(Q_nB_0) - E(Q_n)E(B_0) \\
&= E\left(\frac{1}{n^2}\sum A_{kl}A_m\right) - E\left(\frac{1}{n}\sum A_{kl}\right)\mathrm{Sp}(\mathbf{TV}) \\
&= \frac{1}{n^2}\sum E(A_kA_l) - [\mathrm{Sp}(\mathbf{TV})]^2 \\
&= \frac{1}{n}E(A_k^2) + \frac{n-1}{n}[E(A_k)]^2 - [\mathrm{Sp}(\mathbf{TV})]^2 = \frac{2}{n}\mathrm{Sp}[(\mathbf{TV})^2]
\end{aligned}$$

□

Also ist $B_0 = \frac{1}{n}\sum A_k$ ein erwartungstreuer, konsistenter und dimensionsstabiler Schätzer für $\mathrm{Sp}(\mathbf{TV})$.
Nun werden erneut mit Hilfe einer Taylor-Entwicklung (siehe 3.8) die ersten beiden zentralen Momente der Verteilung von $\widetilde{F}$ berechnet. Es gilt nämlich:

$$E\left(\frac{X}{Y}\right) \doteq \frac{E(X)}{E(Y)}\left(1 + \frac{\mathrm{Var}(Y)}{[E(Y)]^2} - \frac{\mathrm{Cov}(X,Y)}{E(X)E(Y)}\right)$$

$$\mathrm{Var}\left(\frac{X}{Y}\right) \doteqdot \frac{E(X)^2}{E(Y)^2}\left(\frac{\mathrm{Var}(X)}{[E(X)]^2} + \frac{\mathrm{Var}(Y)}{[E(Y)]^2} - 2\frac{\mathrm{Cov}(X,Y)}{E(X)E(Y)}\right)$$

Mit Hilfe der Ergebnisse des voranstehenden Lemmas und

$$E(Q_n) = \mathrm{Sp}(\mathbf{TV}) \qquad \mathrm{Var}(Q_n) = 2\,\mathrm{Sp}[(\mathbf{TV})^2]$$

erhält man folgende Beziehungen:

$$\begin{aligned}
E(\widetilde{F}) &\doteqdot \frac{EQ_n}{EB_0}\left[1 + \frac{\mathrm{Var}(B_0)}{(EB_0)^2} - \frac{\mathrm{Cov}(Q_n, B_0)}{EQ_nEB_0}\right] \\
&\doteqdot \frac{\mathrm{Sp}(\mathbf{TV})}{\mathrm{Sp}(\mathbf{TV})}\left[1 + \frac{\mathrm{Var}(B_0)}{[\mathrm{Sp}(\mathbf{TV})]^2} - \frac{\mathrm{Var}(B_0)}{[\mathrm{Sp}(\mathbf{TV})]^2}\right] \doteqdot 1 \\
\mathrm{Var}(\widetilde{F}) &\doteqdot \frac{(EQ_n)^2}{(EB_0)^2}\left[\frac{\mathrm{Var}(Q_n)}{(EQ_n)^2} + \frac{\mathrm{Var}(B_0)}{(EB_0)^2} - 2\frac{\mathrm{Cov}(Q_n, B_0)}{EQ_nEB_0}\right] \\
&\doteqdot \frac{[\mathrm{Sp}(\mathbf{TV})]^2}{[\mathrm{Sp}(\mathbf{TV})]^2}\left[\frac{\mathrm{Var}(Q_n) - \mathrm{Var}(B_0)}{[\mathrm{Sp}(\mathbf{TV})]^2}\right] \\
&\doteqdot \frac{(2 - \frac{2}{n})\mathrm{Sp}[(\mathbf{TV})^2]}{[\mathrm{Sp}(\mathbf{TV})]^2}
\end{aligned}$$

Zusammen mit den Formeln für die ersten zwei Momente einer $F(f_1, f_2)$-Verteilung erhält man diese Tabelle:

	E	Var
$F(f_1, f_2)$	$\frac{f_2 - 2}{f_2}$	$\frac{f_2^3 + 2f_2^2 f_1 - 4f_2^2}{(f_1 f_2^2 - 4f_2 f_1 + 4f_1)(f_2 - 4)}$
$\widetilde{F}$	1	$\frac{(2 - \frac{2}{n})\mathrm{Sp}[(\mathbf{TV})^2]}{[\mathrm{Sp}(\mathbf{TV})]^2}$

Durch Gleichsetzen der beiden Spalten erhält man folgende Lösungen für f_1 und f_2:

$$\widetilde{f} = f_1 = \frac{[\mathrm{Sp}(\mathbf{TV})]^2}{(1 - \frac{1}{n})\mathrm{Sp}[(\mathbf{TV})^2]}, \quad f_2 = \infty.$$

Die Verteilung von $\tilde{F}$ kann mit einer $\chi^2_{\widetilde{f}}/\widetilde{f}$-Verteilung approximiert werden, bei der die Komponenten $[\mathrm{Sp}(\mathbf{TV})]^2$ und $\mathrm{Sp}[(\mathbf{TV})^2]$ erwartungstreu, konsistent und dimensionsstabil geschätzt werden. Es ergibt sich folgender Schätzer $\widetilde{f}$:

$$\widetilde{f} = \frac{B_1}{(1 - \frac{1}{n})B_2} = \frac{n}{n-1}\frac{B_1}{B_2}.$$

Dieser Freiheitsgrad konvergiert für $n \to \infty$ gegen den Freiheitsgrad der ersten Approximation, die unter Annahme einer bekannten Kovarianzmatrix gemacht wurde. Leider ist er weiterhin nicht vollständig unverzerrt, die Verzerrung aus 3.9 wird durch den Faktor $n/(n-1)$ nicht verändert. Aber es gilt, dass die Verzerrung dieses neuen Schätzers für den Freiheitsgrad der Verteilung der quadratischen Form weitaus geringer ausfällt als beim alten Schätzer mit der Stichprobenkovarianzmatrix.

4 Zusammenfassung und Ausblick

In der vorliegenden Arbeit wurde ein globaler Test für den Fall $d > n$ bei *repeated measures* hergeleitet.
Nachdem die Nachteile alter Schätzer aufgedeckt und analysiert worden sind, wurden neue konstruiert, die diese Mängel nicht mehr aufweisen. Der Begriff der Dimensionsstabilität wurde definiert, um das Verhalten von Schätzern bei hohen Dimensionen zu beschreiben. Mit Hilfe der Verteilung quadratischer und Bilinearformen konnten Eigenschaften der neuen Schätzer wie Erwartungstreue und Dimensionsstabilität nachgewiesen werden. Es wurde sowohl analytisch als auch durch Simulationen gezeigt, dass die neu konstruierten Schätzer in allen Kategorien besser abschneiden als die alten. Zur Abrundung der Theorie wurde eine F-Approximation der neuen Statistik für kleine Stichprobenumfänge angegeben.
Zur einfachen Anwendung der neuen Tests in der Praxis wurden Makros entwickelt, die an vorhandenen und konstruierten Datensätzen getestet wurden.
In zahlreichen Simulationen konnte schließlich gezeigt werden, dass der neue Test im Gegensatz zur normalen ANOVA-Typ-Statistik das Niveau einhält und dass die Power bei wachsender Dimension steigt. Somit ist der modifizierte Test bei hochdimensionalen Daten einsetzbar und liefert sinnvolle und gut interpretierbare Ergebnisse.
Ein anderer Aspekt ist die Übertragung der Theorie auf gemischte Modelle. Wäre man in der Lage, erwartungstreue Schätzer für das Split-Plot-Design (ein unverbundener, ein verbundener Faktor) herzuleiten, so ließe sich das vorliegende Testverfahren auch auf beliebige multivariate Modelle ausdehnen. Hiermit hätte man einen Ersatz für Hotellings T^2 Test, der nicht mehr anwendbar ist, sobald die Dimension den Stichprobenumfang übersteigt.
In einem anderen Schritt kann man versuchen, den Test auf die nichtparametrische Statistik auszuweiten. Hier muss man zunächst der Frage nachgehen, inwieweit die Normalverteilung für die Herleitung des Tests und seiner Verteilung notwendig ist. Es wurde ja bereits durch Simulationen gezeigt, dass der Test für exponential- oder gleichverteilte Zufallsvariablen das Niveau sehr gut einhält. Es gilt also herauszufinden, welche Bedingungen an die Verteilung der Zufallsvektoren $\mathbf{X}_k$ gestellt werden müssen, damit das Niveau eingehalten wird.
Schließlich stellt sich die Frage, wie man die Verteilung der Statistik approximieren kann, wenn die Dimension gegen unendlich strebt, ein Fall, der in vorliegenden Arbeit nicht untersucht wurde. Für $d \to \infty$ strebt die Varianz der Verteilung gegen null und man erhält eine total degenerierte Verteilung. Bei den Simulationen war dieses Verhalten der Teststatistik bzw. des Schätzers für den Freiheitsgrad schon ab $d = 100$ zu sehen. Die Simulationsergebnisse variierten trotz einer hohen Anzahl von Wiederholungen sehr stark.
Anhand dieser Beispiele sieht man, dass der hier hergeleitete globale Test für hochdimensionale Daten vielfältige Möglichkeiten der Ausweitung und Übertragung auf andere Gebiete liefert.

Literatur

[1] Arnold, S.F. (1981). *The Theory of Linear Models and Multivariate Analysis.* Wiley, New York.

[2] Benjamini, Y. und Hochberg, Y. (1995). Controlling the False Discovery Rate. *Journal of the Royal Statistical Society. Series B* **57**, 289-300.

[3] Box, G.E.P. (1954). Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems I. Effect of the Inequality of Variance in the One-Way Classification. *The Annals of Mathematical Statistics* **25**, 290-302.

[4] Box, G.E.P. (1954). Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, II. Effects of Inequality of Variance and of Correlation Between Errors in the Two-Way Classification. *The Annals of Mathematical Statistics* **25**, 484-498.

[5] Boysen, L. (2002). *Analyse von intra-individuellen Effekten bei longitudinalen Daten.* Diplomarbeit, Institut für Mathematische Stochastik Göttingen.

[6] Brunner, E., Dette, H. und Munk A. (1997). Box-Type Approximations in Nonparametric Factorial Designs. *Journal of the American Statistical Association* **92**, 1494-1502.

[7] Brunner, E., Domhof, S. und Langer F. (2001). *Nonparametric Analysis of Longitudinal Data in Factorial Experiments.* Wiley, New York.

[8] Brunner, E., Munzel, U. und Puri M.L. (1999). Rank-Score Tests in Factorial Designs with Repeated Measures. *Journal of Multivariate Analysis* **70**, 286-317.

[9] Candes, E.J. und Donoho, D. (1999). Ridgelets: A key to higher-dimensional intermittency? *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* **357**, 2495-2509.

[10] Hájek, J. (1962). Inequalities for the generalized students distribution and their applications, *Selected Translations in Mathematical Statistics and Probabilities* **2**, 62-74.

[11] Hochberg, Y. und Tamhane, A.C. (1987). *Multiple Comparison Procedures*, Wiley, New York.

[12] Huynh, H. und Feldt, L.S. (1976). Estimation of the Box-Correction for degrees of freedom from sample data in randomized block and split-plot-design. *Journal of Educational Statistics* **1**, 69-82.

[13] Kropf, S. (1999). *Hochdimensionale multivariate Verfahren in der medizinischen Statistik.* Habilitationsschrift, Universität Magdeburg.

[14] Lindsay, B.G. und Basak, P. (2000). Moments determine the Tail of a Distribution (But Not Much Else). *The American Statistician***54**, 248-251.

[15] Mathai, A.M. und Provost, S.B. (1992). *Quadratic Forms in Random Variables, Theory and Applications.* Marcel Dekker, Inc.

[16] Okamoto, M. (1960). An Inequality for the Weighted Sum of χ^2-Variables. *Bulletin of Mathematical Statistics* **9**, 69-70.

[17] Patnaik, P.B. (1949). The non-central χ^2-and F-distributions and their Applications. *Biometrika*, 202-232.

[18] Stange, K. (1970). *Angewandte Statistik.* Springer, Berlin.

[19] Welch, B.L. (1947). The generalization of Student's problem when several different population variances are involved. *Biometrika*, 28-35.

[20] Westfall, P.H. und Young, S.S. (1993). *Resampling-Based Multiple Testing.* Wiley, New York.