

– LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN –
INSTITUT FÜR STATISTIK



Regularisierte Schätzverfahren für Bradley-Terry-Luce Modelle

Masterarbeit

im Fachgebiet Statistik mit wirtschafts- und sozialwissenschaftlicher
Ausrichtung

vorgelegt von:

Manuela Bitterlich

Gutachter:

Prof. Dr. Gerhard Tutz

Betreuer:

Gunther Schauberger, M. Sc.

Abgabetermin:

18. März 2013

Zusammenfassung

Die statistische Modellierung von Paarvergleichsdaten spielt in der Wissenschaft eine große Rolle: Präferenzmodelle sind speziell in der Psychologie eine häufig angewendete Methode um, über ein oder mehrere Paarvergleiche, subjektive Kriterien in eine eindeutige Reihenfolge zu bringen. Oft wird davon ausgegangen, dass diese Präferenzentscheidungen sowohl von den Eigenschaften der zu beurteilenden Objekten als auch von den Charaktereigenschaften der beurteilenden Personen selbst abhängen. Für die Aufnahme subjekt-spezifischer Informationen in Bradley-Terry-Luce Modelle gibt es bislang keinen allgemein gültigen Modellierungsansatz. Die Einbindung subjekt-spezifischer Information erfolgt deshalb über verschiedenste Ansätze wie etwa über modellbasiertes rekursives Partitionieren oder über loglineare Modelle. Der Schwerpunkt dieser Arbeit liegt auf der Untersuchung weiterer alternativer Schätzverfahren für Bradley-Terry-Luce Modelle, über die eine direkte Einbindung subjekt-spezifischer Informationen ermöglicht wird. Dabei wird zum einen die Schätzung über Generalisierte Lineare Modelle mittels Logistischer Regression betrachtet, die ohne zusätzliche Informationen identische Ergebnisse liefert wie die Schätzung über einfache Bradley-Terry-Luce Modelle. Zum anderen wird das Group-LASSO Verfahren als regularisierte Alternative für Paarvergleichsmodelle analysiert. Anhand von Simulationsstudien wird gezeigt, dass Group-LASSO Verfahren mittels Modellregularisierung und Variablenselektion zu guten bis sehr guten Modellanpassungen führen. Die Anwendung erfolgt anhand zweier Datenbeispiele, die zeigen, dass subjekt-spezifische Kovariablen einen wesentlichen Einfluss bei Präferenzentscheidungen haben.

Inhaltsverzeichnis

Inhaltsverzeichnis	I
Abbildungsverzeichnis	III
Tabellenverzeichnis	IV
Abkürzungsverzeichnis	V
1. Einleitung	1
2. Die Methodik des paarweisen Vergleichs	4
2.1. Komparative Skalierung	5
2.2. Rating, Ranking, Paarvergleich	7
2.3. Datenstruktur bei Paarvergleichen	9
3. Das Bradley-Terry-Luce (BTL-) Modell	10
3.1. Luce's Choice Axiom	10
3.2. Mathematisches Modell	13
3.2.1. Annahmen und Voraussetzungen	13
3.2.2. Likelihood-Inferenz	14
3.3. Methoden für Bradley-Terry-Luce Modelle	15
4. BTL-Modell über Generalisierte Lineare Modelle	17
4.1. Das Generalisierte Lineare Modell	17
4.1.1. Allgemeine Modelldefinition	18
4.1.2. Likelihood-Inferenz	21
4.1.3. Iterative Algorithmen	23
4.1.4. Modellwahl	24
4.2. Binäre Regression	25
4.2.1. Logistisches Regressionsmodell	26
4.2.2. Weitere Linkfunktionen	28

4.3.	Datentransformation	28
4.3.1.	Keine Berücksichtigung von Kovariablen	30
4.3.2.	Berücksichtigung von Kovariablen	32
4.4.	Umsetzung und Problematik	34
5.	BTL-Modell über Group-LASSO	36
5.1.	Das LASSO Verfahren	38
5.2.	Das Group-LASSO Verfahren	39
5.2.1.	Group-LASSO bei Logistischer Regression	42
5.2.2.	Datentransformation	44
5.2.3.	Variablenselektion	46
5.3.	Simulationsergebnisse	48
6.	Anwendungsbeispiele	65
6.1.	Germany's Next Topmodel 2007	65
6.1.1.	Anwendung BTL-Modell über GLM	67
6.1.2.	Anwendung BTL-Modell über Group-LASSO	72
6.1.3.	Vergleich mit modellbasierten rekursiven Partitionieren	75
6.2.	CEMS-Universitäten	79
6.2.1.	Anwendung BTL-Modell über GLM	82
6.2.2.	Anwendung BTL-Modell über Group-LASSO	86
6.2.3.	Vergleich mit Verfahren über Log-Lineare Modelle	87
7.	Zusammenfassung und Ausblick	94
A.	Anhang	VI
A.1.	Beweis von <i>Lemma 1</i>	VI
A.2.	R-Code	VII
A.2.1.	Simulationsmaske: BTL-Modell über GLM	VII
A.2.2.	Simulationsmaske: BTL-Modell über Group-LASSO	IX
A.3.	Weitere Simulationsergebnisse	XII
A.4.	Elektronischer Anhang	XIII
	Literaturverzeichnis	XV

Abbildungsverzeichnis

2.1. Paarvergleich zwischen zwei Untersuchungsobjekten	5
2.2. Darstellung eines Präferenzschemas einer Person	6
2.3. Direktes Rating eines Untersuchungsobjektes	8
2.4. Ausgangssituation beim Ranking	9
3.1. Darstellung des Luce's Choice Axiom	11
4.1. Die Logit-Funktion	27
5.1. Übersicht über Variablenselektionsverfahren	37
5.2. Darstellung l_1 -Bestrafung, Group-LASSO Bestrafung und l_2 -Bestrafung . .	40
5.3. Plot der Simulationsergebnisse mit 6 Items und 10 Kovariablen	52
5.4. AIC/BIC in Abhängigkeit der Höhe der Bestrafung Lambda	53
5.5. Koeffizientenpfade von Group-LASSO für Simulation mit 6 Items	56
5.6. Vergleich der Simulationsergebnisse mit unterschiedlicher Anzahl an Items	59
5.7. Vergleich der Simulationsergebnisse mit unterschiedlicher Anzahl informa- tiver Kovariablen	62
5.8. Plot der Simulationsergebnisse mit 10 Items und 20 Kovariablen	64
6.1. Beobachtete Häufigkeiten der Paarvergleiche für den Topmodel-Datensatz .	66
6.2. Koeffizientenschätzer des GLMs ohne Kovariablen für Topmodel-Datensatz	68
6.3. Schätzer der wahren Präferenzen für die Topmodel-Kandidatinnen	69
6.4. Koeffizientenschätzer des GLMs mit Kovariablen für Topmodel-Datensatz .	70
6.5. Vergleich der Koeffizientenschätzer zwischen Group-LASSO und GLM . . .	74
6.6. Partitioniertes Paarvergleichsmodell für Topmodel-Datensatz	77
6.7. Beobachtete Häufigkeiten der Paarvergleiche für CEMSChoice-Datensatz .	80
6.8. Koeffizientenschätzer des GLMs ohne Kovariablen für CEMSChoice-Datensatz	83
6.9. Koeffizientenschätzer des GLMs mit Kovariablen für CEMSChoice-Datensatz	85
6.10. Präferenzskala der Universitäten bei Anpassung eines LLBT ohne Kova- riablen	91
6.11. Vergleich der Koeffizientenschätzer zwischen LLBT und GLM	93

Tabellenverzeichnis

4.1. Einfache Exponentialfamilien	20
4.2. Anordnung von Paarvergleichen	29
5.1. Simulationsergebnisse mit 6 Items, 10 Kovariablen, 6 ohne Information . .	51
5.2. BIC vs. Lambda bei Modellselektion	54
5.3. Simulationsergebnisse mit 10 Items, 10 Kovariablen, 6 ohne Information . .	57
5.4. Simulationsergebnisse mit 15 Items, 10 Kovariablen, 6 ohne Information . .	58
5.5. Simulationsergebnisse mit 10 Items, 10 Kovariablen, 2 ohne Information . .	60
5.6. Simulationsergebnisse mit 10 Items, 10 Kovariablen, 4 ohne Information . .	61
5.7. Simulationsergebnisse mit 10 Items, 10 Kovariablen, 8 ohne Information . .	61
5.8. Simulationsergebnisse mit 10 Items, 20 Kovariablen, 12 ohne Information .	63
6.1. Koeffizientenschätzer des GLMs mit Kovariablen für Topmodel-Datensatz .	70
6.2. Koeffizientenschätzer über Group-LASSO für Topmodel-Datensatz	74
6.3. Schätzer der Worth-Parameter der jeweiligen Teilstichproben	78
6.4. Eigenschaften der CEMS-Universitäten	80
6.5. Koeffizientenschätzer des GLMs mit Kovariablen für CEMSChoice-Datensatz	84
6.6. Designstruktur eines LLBT mit Unentschieden	89
6.7. Koeffizientenschätzer des LLBTs ohne Kovariablen	90
6.8. Koeffizientenschätzer des LLBTs mit Kovariablen	92
A.1. Ergebnisse des Simulationssettings gemäß Topmodel-Datensatz	XII
A.2. Simulationsergebnisse mit 20 Items, 10 Kovariablen, 6 ohne Information . .	XIII

Abkürzungsverzeichnis

AIC	Akaikes Informationskriterium
AICc	korrigiertes Akaike Informationskriterium
AS	Active Set
BIC	Bayesianisches Informationskriterium
BT	Bradley–Terry
BTL	Bradley–Terry–Luce
FG	Freiheitsgrad
GLM	Generalisiertes Lineares Modell
IIA	Independence of Irrelevant Alternatives
iid	unabhängig und identisch verteilt (independent and identical distributet)
KQ	Methode der kleinsten Quadrate
LASSO	Least Absolute Shrinkage and Selection Operator
LCA	Luce’s Choice Axiom
LLBT	loglineares Bradley-Terry Modell
ML	Maximum–Likelihood
RIC	Residuen Informationskriterium
Std.Abw.	Standardabweichung
YL	Approximation der Freiheitsgrade nach Yuan and Lin (2006) für logistische Group-LASSO Modelle

1. Einleitung

Ranglisten sind heutzutage allgemein verbreitet: die Liste der Dienstgrade beim Militär, die Mitgliederliste eines Vereins, die Ergebnislisten im Sport, wie die aktuelle Tabelle in der Fußball Bundesliga oder die Liste der Top-Manager Gehälter. In beinahe allen Lebensbereichen werden vergleichbare Objekte oder deren Eigenschaften bewertet und in Reihenfolgen gebracht.

Darüber hinaus existieren neben messbaren Kriterien auch subjektive Kriterien, wie Lebensqualität, Geschmack oder Schönheit, die nicht direkt messbar sind und somit nicht direkt in eine Reihenfolge gebracht werden können. Diese Variablen werden in diesem Zusammenhang als latente Variablen bezeichnet [David, 1988]. Doch was ist „schön“? Blond oder brünett, dick oder dünn, groß oder klein? Jeder glaubt das zu wissen, doch wie eine Redewendung besagt, liegt Schönheit im Auge des Betrachters. Es ist kaum vorstellbar, sie objektiv messen zu können [Novafeel, 2013].

Die Wahl eines geeigneten statistischen Modells zur Bewertung subjektiver Kriterien ist demnach von großer Bedeutung. Eine Möglichkeit, nicht direkt messbare Variablen in eine eindeutige Rangordnung zu bringen, ist der Paarvergleich. Hierbei werden Items systematisch über Paarkombinationen miteinander verglichen, um kleine Unterschiede bei der Präferenzentscheidung aufzuzeigen [David, 1988]. Dabei versteht man unter einem Item die zu bewertende Zielgröße, wie im Fall von Schönheit etwa eine Frau. Die Einstufung subjektiver Kriterien ist über den direkten Paarvergleich zweier Items für die Menschen oft einfacher, schneller und verlässlicher zu beantworten [Zeileis, 2011].

Aber bereits in den 1940er Jahren haben Kendall and Babington Smith [1940] auf den hohen Aufwand bei der Durchführung von Paarvergleichen hingewiesen: Bei m zu vergleichenden Items ergeben sich $\binom{m}{2}$ Paarvergleiche, bei denen die befragten Personen ihre Präferenz aussprechen müssen. Bei 20 zu vergleichenden Items sind 190 Paarvergleiche durchzuführen, bei 30 sind es schon 435 und bei 50 Items geht es mit 1225 Paarvergleichen bereits in den vierstelligen Bereich. Zur statistischen Modellierung werden deswegen Methoden benötigt, die sich für diese spezielle Datensituation eignen.

Breiman [2001] unterscheidet zwischen zwei Kulturen bei der statistischen Modellierung:

Datenmodelle und Algorithmische Modelle. Bei Ersterem wird angenommen, dass die Daten über ein Stochastisches Modell erzeugt werden und häufig parametrischer Natur sind. Darunter fallen etwa (verallgemeinerte) lineare Modelle, zu denen das Bradley-Terry-Luce (BTL-) Modell [Bradley and Terry, 1952; Luce, 1959] zur Schätzung von Paarvergleichsdaten zählt. Bei Algorithmischen Modellen hingegen bleibt der datengenerierende Prozess unbekannt um flexible Modelle, wie zum Beispiel Rekursives Partitionieren/Entscheidungsbäume oder neuronale Netze, zu ermöglichen [Zeileis, 2011]. Der Schwerpunkt der hier vorliegenden Arbeit liegt auf der Kultur der Datenmodelle, speziell auf dem Paarvergleichsmodell von Ralph A. Bradley, Milton E. Terry und Robert D. Luce [Bradley and Terry, 1952; Luce, 1959], welches hauptsächlich in der Psychologie und in verwandten Themengebieten Anwendung findet. Anwendungsbeispiele sind etwa Untersuchungen bei Wahrnehmungen von Schmerz, Geräuschen oder Geschmack (siehe z.B., Matthews and Morris [1995], Choisel and Wickelmaier [2007], Oberfeld et al. [2009]) sowie Studien im Bereich des Gesundheitswesens, der Bildung oder der Politik (siehe z.B., Dittrich et al. [1998], Dittrich et al. [2006]).

In zahlreichen Anwendungen wird davon ausgegangen, dass die Präferenzentscheidungen der Personen nicht allein von den Eigenschaften der zu beurteilenden Stimuli abhängt, sondern auch von den Charaktereigenschaften der Personen selbst [Strobl et al., 2009]. So kann etwa das Geschlecht der befragten Person ebenso einen Einfluss auf die Entscheidung haben, wie die Haarfarbe einer Frau, die nach ihrer Attraktivität beurteilt werden soll. Um derartige Informationen bei der Anpassung von Paarvergleichsmodellen berücksichtigen zu können, ist es üblich, separate Anpassungen, z.B. für weibliche und männliche Teilnehmer, des Bradley-Terry-Luce Modells durchzuführen. In erweiterten Anwendungen, wie in Dittrich et al. [1998], Böckenholt [2001], werden die Kovariablen direkt ins Modell mit eingebunden.

Eine Alternative ist die Modellierung von Paarvergleichen über das Generalisierte Lineare Modell (GLM) [Nelder and Wedderburn, 1972], dessen Vorteil in der einfachen Ausführung über Standardsoftware liegt. Die Struktur des Paarvergleichs ist in den Kontext des Generalisierten Linearen Modells zu übertragen und die Einbindung von Kovariablen ist darauf abzustimmen [Critchlow and Fligner, 1991]. Die Problematik der hohen Dimensionalität kommt jedoch erneut zum Tragen. Bei einer großen Anzahl m zu vergleichender Items wächst die Dimension der Designmatrix enorm, insbesondere da eine Kovariable bereits $(m - 1)$ Spalten belegt. Eine relativ hohe Anzahl p an Kovariablen kann deshalb bei einem niedrigen Stichprobenumfang n schnell zu einem hochdimensionalen Problem führen. In diesen Fällen stößt das Generalisierte Lineare Modell an seine Grenzen, und liefert keine zufriedenstellenden Ergebnisse bezüglich Vorhersagegenauigkeit und Inter-

pretierbarkeit [Tibshirani, 1996].

Um trotzdem eine adäquate Vorhersagegenauigkeit und sinnvoll interpretierbare Modelle zu erhalten muss der klassische GLM-Ansatz erweitert werden. Die Idee liegt in der Anpassung eines Regressionsmodells, bei dem nur die wichtigsten Prädiktoren berücksichtigt werden, indem komplizierte Modelle stark bestraft werden und der Effekt unwichtiger Parameter auf Null geschrumpft wird. Somit wird ein möglichst sparsames Modell generiert, welches aus einer großen Anzahl an Kovariablen jeweils nur die Informativen bezüglich der Präferenzentscheidung enthält. Die bekannteste Methode in diesem Sinne ist das LASSO Verfahren [Tibshirani, 1996] (Least Absolute Shrinkage and Selection Operator), das sowohl genauere Vorhersagen ermöglicht als auch gut interpretierbare Modelle liefert und somit beide Aspekte in sich vereint [Tibshirani, 1996].

Durch das Vorwissen über die spezielle Datenstruktur bei Paarvergleichsmodellen stellt die Gruppierung von Variablen einen wesentlichen Vorteil dar. Im Gegensatz zur LASSO Methode, bei der lediglich metrische Variablen oder individuelle Dummy-Variablen im Modell berücksichtigt werden können, wurde das Group-LASSO Verfahren [Yuan and Lin, 2006] speziell zur Einbindung kategorialer Variablen entwickelt. Faktorvariablen können somit einheitlich als komplette Gruppe wahrgenommen werden und entweder als Ganzes im Modell berücksichtigt oder geschlossen als Gruppe aus dem Modell genommen werden. Die Variablenselektion und -schrumpfung wird nunmehr auf Faktorebene durchgeführt. Diese besondere Eigenschaft des Group-LASSO kann zur Verarbeitung der Paarvergleichsinformation eingesetzt werden. Die einzelnen Kovariablen können somit über die Paarvergleiche gruppiert werden. Ziel dieser Arbeit ist es, die Funktionsweise der Selektion von Kovariablen bei der Anpassung von Paarvergleichsdaten über Group-LASSO zu untersuchen und zu überprüfen, ob dieser Ansatz eine Alternative für BTL-Modelle darstellt.

Im anschließenden Kapitel 2 wird zunächst die Methodik des paarweisen Vergleichs ausführlich vorgestellt und mit Ranking und Rating Verfahren verglichen. Das Bradley-Terry-Luce Modell zur Analyse von Paarvergleichen wird in Kapitel 3 eingeführt. Das alternative Schätzverfahren über Generalisierte Lineare Modelle wird in Kapitel 4 näher erläutert. Die notwendigen Datentransformationen zur Übertragung von Paarvergleichsdaten in den Kontext parametrischer Modelle werden in Abschnitt 4.3 durchgeführt. Mit der regularisierten Schätzung des Bradley-Terry-Luce Modells über Group-LASSO beschäftigt sich Kapitel 5. Hierbei werden in Abschnitt 5.3 die Simulationsergebnisse vorgestellt. Anhand zweier Beispieldatensätze werden in Kapitel 6 die regularisierten Schätzverfahren für Bradley-Terry-Luce Modelle angewandt und auf ihre Eignung als alternative Verfahren zur Schätzung von Paarvergleichen überprüft. Abschließend wird in Kapitel 7 eine kurze Zusammenfassung und ein Ausblick auf weitere Analysemöglichkeiten gegeben.

2. Die Methodik des paarweisen Vergleichs

Der Paarvergleich ist ein psychometrisches Skalierungsverfahren, dessen Grundprinzip durch [Thurstone \[1927b\]](#) entwickelt wurde. Eine ausführliche Einführung in die Methodik des paarweisen Vergleichs, den Aufbau von Experimenten bei Paarvergleichen und der Schätzung von Paarvergleichsmodellen ist in [David \[1988\]](#) gegeben. Frühe Anwendungen und Entwicklungen von Paarvergleichsdaten sind in einem umfangreichen Literaturverzeichnis von [Davidson and Farquhar \[1976\]](#) zusammengefasst.

Bei einem Paarvergleich werden Objekte/Items oder deren Eigenschaften durch das systematische Gegenüberstellen in eine Rangordnung bzw. Präferenzliste gebracht. Um die gemessenen Eigenschaften von Objekten oder die Objekte selbst positionieren zu können ist eine Skalierung notwendig. Zu unterscheiden sind Verfahren komparativer (vergleichender) Skalierung und Verfahren nichtkomparativer Skalierung. Dabei beinhalten Techniken komparativer Skalierung den direkten Vergleich von Objekten und man spricht von einer sogenannten Selbsteinstufung. Da diese Skalierung nur Ordinalskalenniveau zulässt nennt man diese auch nichtmetrische Skalierung. Bei der nichtkomparativen Skalierung hingegen, erfolgt die Bewertung von Objekten isoliert, d.h. unabhängig von anderen Untersuchungsobjekten und wird somit auch als Fremdeinstufung bezeichnet. Die Skalierung kann als metrisch angenommen werden, denn obwohl hier ebenfalls nur ordinale Daten vorliegen, ist dies aufgrund der Annahme gleicher Abstände der Skalenpunkte möglich [[Altobelli, 2007](#)].

Da in dieser Arbeit lediglich die Methodik des paarweisen Vergleichs von Interesse ist, wird ausschließlich auf das Verfahren komparativer Skalierung eingegangen. Die Beispiele beziehen sich auf eine Studie, in der 192 Personen die Attraktivität der sechs Finalistinnen der TV-Show Germany's Next Topmodel 2007 über Paarvergleiche beurteilen sollten. Für dieses Beispiel werden demnach $m = 6$ Items, entsprechend den sechs Finalistinnen, über Paarvergleiche miteinander verglichen. Die Fotos der Kandidatinnen und Teile der Beispiele sind aus [Strobl et al. \[2009\]](#) entnommen. Die genauen Begebenheiten der Datenerhebung werden in Abschnitt [6.1](#) erläutert.

2.1. Komparative Skalierung

Der Paarvergleich ist das häufigste Verfahren, wenn Objekte über den direkten Vergleich von Paarkombinationen in eine Rangordnung gebracht werden sollen. Hauptsächlich wird die Methode des paarweisen Vergleichs in Situationen angewendet, bei denen die Objekte, nur subjektiv bewertet werden können [David, 1988].

Beispielsweise möchte man anhand vorgegebener subjektiver Kriterien, wie etwa „Geschmack“ oder „Schönheit“ eine Reihenfolge aufstellen. Da diese Kriterien nicht direkt messbar sind, werden sie als latente Variablen bezeichnet. In Abbildung 2.1 ist ein direkter Paarvergleich zwischen zwei Topmodel-Kandidatinnen aufgezeigt, bei dem die Personen über die Frage „Welche Frau finden sie attraktiver?“ entscheiden und ihre Präferenz nach vorgegebenem Kriterium aussprechen sollen.

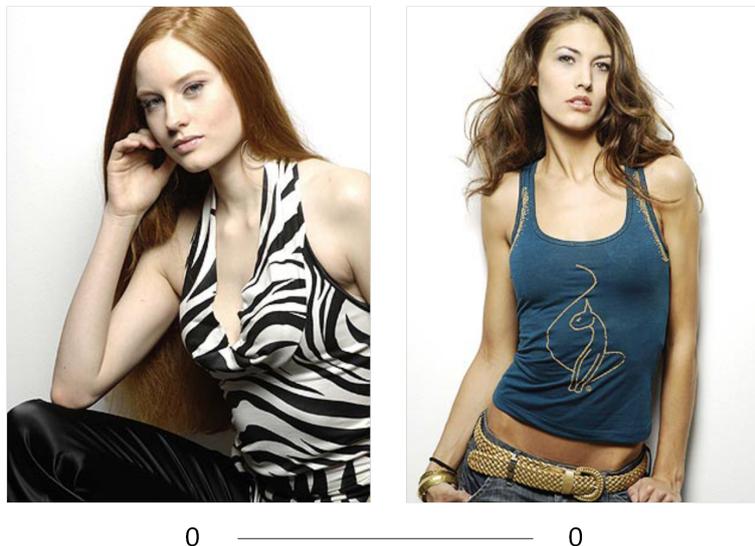


Abbildung 2.1.: Paarvergleich zwischen den Topmodel-Kandidatinnen Barbara und Fiona.

Ein wesentlicher Vorteil dieses Vorgehens ist, dass die Personen bei der systematischen Gegenüberstellung zweier Objekte zu einer Entscheidung gezwungen werden, da kein Unentschieden zwischen den Objektpaaren zugelassen ist und somit eine gerade Anzahl an Skalenstufen vorliegt. Man spricht von der sogenannten Forced-Choice Befragungstechnik, durch welche die „Tendenz zur Mitte“ vermieden wird. Durch das Verhindern eines Unentschiedens können jedoch auch wichtige Informationen verloren gehen [SDI-Research, 2012].

Bei m zu vergleichenden Items ergeben sich $\binom{m}{2} = \frac{m(m-1)}{2}$ Paarvergleiche, bei denen die Personen, welche unabhängig voneinander agieren, nach einem vorgegebenem Kriterium ihre Präferenz angeben. Bei n Personen sind das $n \cdot \binom{m}{2}$ durchzuführende Paarvergleiche.

2. Die Methodik des paarweisen Vergleichs

Die Entscheidung für eine Option im Paarvergleich kann für eine einzelne Person in einer Tabelle, wie in Abbildung 2.2, dargestellt werden.

Es werden $m = 6$ Items (A, B, C, D, E, F) paarweise miteinander verglichen. Die Präferenz von Item A gegenüber Item B kann über die Formulierungen $A \succ B$ bzw. $B \prec A$ ausgedrückt werden. Die Anordnung der Items A bis F erfolgt beliebig, so dass es insgesamt $(m!)^2$ Möglichkeiten gibt, die Präferenzen einer Person in einer Tabelle, wie in Abbildung 2.2, zu übertragen. Eine einheitliche Anordnung der Items in Spalten und Zeilen reduziert die Anzahl auf $(m!)$ Möglichkeiten [Kendall and Babington Smith, 1940].

	A	B	C	D	E	F
A	–	1	1	0	1	1
B	0	–	0	1	1	0
C	0	1	–	1	1	1
D	1	0	0	–	0	0
E	0	0	0	1	–	1
F	0	1	0	1	0	–

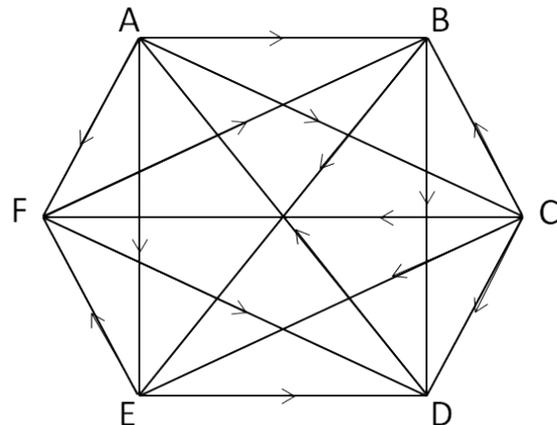


Abbildung 2.2.: Links: Beispielhafte Darstellung der Präferenz einer Person beim Vergleich von 6 Items; Rechts: Geometrische Darstellung nebenstehenden Präferenzschemas. Quelle: Kendall and Babington Smith [1940]

Das Item, das am häufigsten bevorzugt wird erhält Rang 1 und das Item, welches am seltensten präferiert wird erhält Rang n . Dieses Vorgehen erweist sich als sinnvoll, wenn die Anzahl der zu beurteilenden Objekte begrenzt bzw. nicht sehr groß ist, da es sonst sehr unübersichtlich wird [Altobelli, 2007].

Beim Vergleich von 10 Items müssen 45 Paarvergleiche durchgeführt werden, bei 20 Items sind es 190 Paarvergleiche und bereits ab 46 Items kommt man in den vierstelligen Bereich durchzuführender Paarvergleiche. Dies kann bei Befragungen dazu führen, dass das Konzentrations- und Durchhaltevermögen der Testpersonen zu sehr beansprucht wird und dadurch die Ergebnisse verzerrt werden [Bortz and Döring, 2006].

Die Geduld der Personen wird schnell auf eine harte Probe gestellt, mit dem Ergebnis, dass die Reihenfolge der abzuarbeitenden Paarvergleiche einen Einfluss haben kann. Außerdem ist keine Aussage darüber möglich, ob das präferierte Item der Person im absoluten Sinne gefällt oder lediglich ausgewählt wird um das Verfahren abzukürzen [Altobelli, 2007].

Demnach eignen sich Komparative Skalierungsverfahren insbesondere, wenn Präferen-

zen bzw. Wichtigkeitsbewertungen, anhand einer überschaubaren Anzahl an Items, erhoben werden sollen [Altobelli, 2007]. Über die einzelnen Bewertungen der Items gemäß dem Gesamteindruck, kann eine eindeutige Rang- bzw. Präferenzordnung aufgestellt werden. Durch das Erzwingen einer Entscheidung infolge der Forced-Choice Befragungstechnik, wird das Urteil spontan ohne Hinterfragung möglicher Aspekte abgegeben [Müller-Herbers, 2012].

2.2. Rating, Ranking, Paarvergleich

Die subjektive Beurteilung einer Person gegenüber einem Objekt kann neben dem Paarvergleich auch über Rating- oder Ranking-Verfahren erfolgen. Für alle drei Verfahren genügt, zur Erreichung einer Rangfolge, als Input ein diskretes Entscheidungsproblem. Im Gegensatz zum paarweisen Vergleich können bei der Bewertung der Objekte über das Rating-Verfahren Unentschieden auftreten. Das Ranking-Verfahren schließt Unentschieden ebenso aus, wie Inkonsistenzen bei den Präferenzentscheidungen, welche beim Paarvergleich nicht ausgeschlossen werden können und so manch eine Entscheidung einer Anderen widerspricht [David, 1988].

Beim Rating geben die Personen bezüglich der einzelnen Items direkt ein quantitatives Urteil auf einer numerischen Skala ab, wobei eine entsprechende Abstufung für den betrachteten Untersuchungsgegenstand vorgegeben ist [Zeileis, 2011]. Bei der Bewertung der Attraktivität einer Topmodel-Kandidatin ist die Abstufung von 1 = „überhaupt nicht attraktiv“ über 5 = „mittelmäßig attraktiv“ bis 10 = „sehr attraktiv“ denkbar. Abbildung 2.3 stellt gegenüber Abbildung 2.1 das direkte Rating einer Topmodel-Kandidatin auf einer Skala von 1-10 dar.

Die größte Schwierigkeit bei diesem Vorgehen ist die unterschiedliche Vorstellung bei einer gleichen Bewertung der Testpersonen. Wie lassen sich die Bewertungen von 1 bis 10 genau definieren und welche Vorstellungen verbinden die Personen damit? Wie sollte eine Topmodel-Kandidatin aussehen um mit einer 1 bewertet zu werden? Genanntes zeigt, dass das Rating in einigen Fragestellungen unflexibel ist. Anstelle der Fragestellung „Wie attraktiv finden Sie diese Person?“ interessiert im Rating-Verfahren vielmehr die Fragestellung „Um wieviel attraktiver finden Sie die Person gegenüber den Anderen?“ [Tsukida and Gupta, 2011].

Es zeigt sich, dass der Paarvergleich bei der Bewertung von Attraktivität gegenüber einer direkten Bewertung auf einer Skala einfacher zu bewerten ist, vor allem da man bei jedem Vergleich ein eindeutiges Urteil bezüglich der Attraktivität erhält, wohingegen man beim

Rating mit gleichen Bewertungen unterschiedlicher Bedeutung arbeiten muss. Aufgrund der Einstufung auf einer Skala spricht man auch von Einstufungsverfahren.



1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10

Abbildung 2.3.: Direktes Rating einer Topmodel-Kandidatin.

Das Vorgehen beim Ranking unterscheidet sich vom Paarvergleich oder Rating dahingehend, dass bereits zu Beginn alle Objekte betrachtet werden können. Diese werden dem Einstufungskriterium entsprechend von den Testpersonen in eine Rangfolge einsortiert [Zeileis, 2011]. Über das Zuordnen eindeutiger Ränge wird das Unentschieden mehrerer Items ausgeschlossen. Zum Beispiel sollen den Topmodel-Kandidatinnen, wie in Abbildung 2.4, nach ihrer Attraktivität Ränge zugeordnet werden.

Eine wichtige Eigenschaft von Ranking-Verfahren ist die Folgerung, dass beim Vergleich dreier Items A , B und C mit den Präferenzen $A \succ B$ und $B \succ C$ automatisch $A \succ C$ gilt [Kendall and Babington Smith, 1940]. Beim Paarvergleich hingegen, können sogenannte Inkonsistenzen auftreten, d.h. es besteht die Möglichkeit sich selbst bei einer Präferenzentscheidung zu widersprechen [David, 1988]. In diesen Fällen kann der Fall $A \succ B$, $B \succ C$ und $C \succ A$ auftreten. Liegen inkonsistente Beurteilungen bei Paarvergleichen vor, erhält man als Folge verzerrte Ergebnisse. Ein einwandfreies Resultat ist nur bei konsistenten Beurteilungen möglich [Medialexikon, 2012]. Für das Beispiel in Abbildung 2.4 müssen demnach die Ränge 1 bis 6 den einzelnen Topmodels zugeordnet werden. Die Zuweisung von Rängen verleiht dem Ranking-Verfahren auch die Bezeichnung Rangordnungsverfahren.



Abbildung 2.4.: Ausgangssituation beim Ranking: Die Topmodel-Kandidatinnen Barbara, Anni, Hana, Fiona, Mandy und Anja (von links nach rechts, oben nach unten) sollen in eine Rangordnung gebracht werden. Quelle: [Strobl et al., 2009]

2.3. Datenstruktur bei Paarvergleichen

Die eindeutige Rangordnung bei Paarvergleichen resultiert aus der Schätzung von Skalierungsparametern, genannt *Item Parameter*. Als Beobachtungen werden die Präferenzen der n Personen verwendet, die sich jeweils aus den systematischen Gegenüberstellungen zweier Items ergeben. Für alle Personen wird die gleiche Präferenzskala vorausgesetzt, so dass die Beobachtungen als unabhängige Ergebnisse der selben Verteilung angenommen werden können.

Sei Y_{rs} eine stochastische Variable, welche die Präferenz zwischen zwei Items beschreibt, so dass

$$Y_{rs} = \begin{cases} 1, & \text{falls Testperson Item } r \text{ gegenüber Item } s \text{ bevorzugt} \\ 0, & \text{sonst} \end{cases}$$

für alle $r, s = 1, \dots, m$ gilt. Für die Wahrscheinlichkeit, dass eine Person Item r gegenüber Item s bevorzugt, gilt $p_{rs} = P(Y_{rs} = 1)$ und die Verteilung von Y_{rs} ist nach Frisinfeldt Thuesen [2007] gegeben durch

$$Y_{rs} \sim \text{Bin}(p_{rs}, 1) \quad \forall r, s = 1, \dots, m. \quad (2.1)$$

Die zwei am häufigsten verwendeten statistischen Verfahren zur Schätzung von Paarvergleichsdaten gemäß komparativer Skalierung sind das Thurstone [1927a]-Mosteller [1951] und das Bradley-Terry [1952] Modell. Im nächsten Kapitel 3 wird das Bradley-Terry-Luce Modell vorgestellt, welches eine Verallgemeinerung des Bradley-Terry Modells darstellt.

3. Das Bradley-Terry-Luce Modell

Das statistische Modell von [Bradley and Terry \[1952\]](#) und [Luce \[1959\]](#), genannt Bradley-Terry-Luce Modell (BTL-Modell), wird häufig zur Analyse von Präferenzentscheidungen verwendet. Das BTL-Modell basiert auf dem Verfahren komparativer Skalierung (siehe Abschnitt 2.1) und lässt beim direkten Vergleich zweier Items kein Unentschieden zu.

Letztlich entstand das BTL-Modell durch die Einführung des Wahlaxioms von Robert D. Luce im Jahre 1959, welches das Bradley-Terry Modell (BT-Modell) aus dem Jahr 1952 entscheidend erweiterte. Die Grundidee des BTL-Modells geht auf [Bradley and Terry \[1952\]](#) zurück, die in ihrem Artikel *Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons* bereits die Analyse von Paarvergleichsdaten für den binären Fall untersuchten. Eine Verallgemeinerung dieses Modells führte [Luce \[1959\]](#) in seinem Buch *Individual Choice Behavior: A Theoretical Analysis* durch das sogenannte *Luce's Choice Axiom* ein, welches in Abschnitt 3.1 genauer vorgestellt wird. Die Erweiterung gegenüber dem Bradley-Terry Modell aus dem Jahr 1952 ermöglicht die Beantwortung komplexerer Fragestellungen.

3.1. Luce's Choice Axiom

Als Modell zur Bewertung individueller Entscheidungen, hat [Luce \[1959\]](#) das Bradley-Terry Modell [[Bradley and Terry, 1952](#)] gründlich in seinem Einzelwerk *Individual Choice Behavior: A Theoretical Analysis* untersucht. Dabei zeigt Luce, wie das Bradley-Terry-Luce Modell aus Axiomen über binäre Wahlentscheidungen abgeleitet werden kann. Das sogenannte Luce's Choice Axiom (LCA, Luce'sches Wahlaxiom) macht die Annahme, dass das Verhältnis der Wahrscheinlichkeiten zweier Alternativen gleich bleibt, unabhängig von der Anzahl an Alternativen, die in der Grundmenge enthalten sind.

Als Beispiel für Luce's Choice Axiom kann etwa die Wahl eines Menüs im Restaurant gesehen werden. Ein Restaurant bietet vier Menüs an, mit der Wahl für den Hauptgang zwischen Schwein, Huhn, Rind und einer vegetarischen Hauptspeise. 20% entscheiden sich für das Menü mit Schwein und 30% wählen das Menü mit Huhn. Nun wird ein neues Menü angeboten, bei dem nur mehr die Wahl zwischen Schwein und Huhn als Hauptgang

besteht. Das Menü mit Schwein wird nun von 40% gewählt und das Menü mit Huhn von 60% [Cohen, 2004].

In diesem Fall spricht man von Independence of Irrelevant Alternatives (IIA), welche für dieses Beispiel wie folgt geschrieben werden kann.

$$\frac{P(\text{Schwein}, \{\text{Schwein}, \text{Huhn}\})}{P(\text{Huhn}, \{\text{Schwein}, \text{Huhn}\})} = \frac{P(\text{Schwein}, \{\text{Schwein}, \text{Huhn}, \text{Rind}, \text{Vegetarisch}\})}{P(\text{Huhn}, \{\text{Schwein}, \text{Huhn}, \text{Rind}, \text{Vegetarisch}\})}$$

Die Annahme der IIA ist eine Schlussfolgerung bezüglich der allgemeinen Definition des Luce's Choice Axioms, welches im Folgenden genauer vorgestellt wird.

Luce's Choice Axiom

Es werde mit $P_S(R)$ die Wahrscheinlichkeit bezeichnet, die sich beim direkten Paarvergleich ergibt, wenn die Wahl einer Person über eine abgeschlossene Menge von S Alternativen auf eine Alternative aus der Untermenge $R \subset S$ fällt. Betrachtet man $S = \{x, y\}$ so beschreibt $P(x, y) := P_S = P_{\{x,y\}}(x)$ die Wahrscheinlichkeit, dass Alternative x gegenüber Alternative y vorgezogen wird. Unter diesen Annahmen definiert Luce [1959] das sogenannte Luce'sche Wahlaxiom:

Sei T eine abgeschlossene Menge, sodass für jedes $S \subset T$, P_S definiert ist.

1. Falls $P(x, y) \neq 0, 1$ für alle $x, y \in T$, dann gilt für $R \subset S \subset T$

$$P_T(R) = P_S(R)P_T(S) \tag{1}$$

2. Falls $P(x, y) = 0$, für einige $x, y \in T$, dann gilt für $S \subset T$

$$P_T(S) = P_{T-\{x\}}(S - \{x\}) \tag{2}$$

Dabei steht die Notation $S - \{x\}$ für $S \setminus \{x\}$, in Worten bedeutet dies S ohne x . Graphisch kann dies wie in Abbildung 3.1 dargestellt werden.

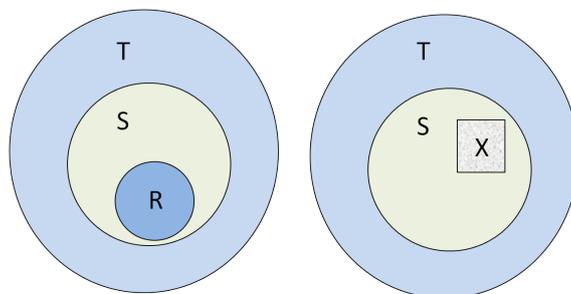


Abbildung 3.1.: Darstellung des Luce's Choice Axioms: Links: 1. Teil der Definition; Rechts: 2. Teil der Definition. Quelle: [Cohen, 2004]

Annahme (2) der Definition wird häufig vernachlässigt, so dass auch hier nur auf Folgerungen bezüglich dem ersten Teil (1) der Definition eingegangen wird:

- (1) kann umformuliert werden über

$$P_S(R) = \frac{P_T(R)}{P_T(S)}. \quad (3)$$

Es ist zu beachten, dass es sich hierbei formal nicht um die bedingte Wahrscheinlichkeit handelt, da Luce's Choice Axiom für T keinen allgemein gültigen Ereignisraum annimmt [Luce, 1959].

- (1) ist äquivalent zur Darstellung

$$P_S(x) = \frac{v(x)}{\sum_{y \in S} v(y)}. \quad (4)$$

Diese Annahme wurde bereits durch Bradley and Terry [1952] für den binären Fall untersucht.

- Annahme (4) impliziert für $T = \{x, y, z\}$ die Produktregel

$$\frac{P(x, y)}{P(y, x)} \frac{P(y, z)}{P(z, y)} \frac{P(z, x)}{P(x, z)} = 1. \quad (5)$$

- Die Annahme der Independence of Irrelevant Alternatives (IIA), welche bereits im Beispiel angesprochen wurde, lässt sich formal wie folgt darstellen.

$$\frac{P(x, y)}{P(y, x)} = \frac{P_S(x)}{P_S(y)} \quad (6)$$

Der Unterschied zwischen den beiden Wahrscheinlichkeiten $P(x, y)$ und $P(y, x)$ liegt darin, dass bei Ersterem x und bei Zweiterem y beim Paarvergleich vorgezogen wird. In einigen Zusammenhängen ist die Annahme der Independence of Irrelevant Alternatives (IIA) des Axioms unbefriedigend [Luce, 1959]. Betrachtet man zum Beispiel eine Person bei der Wahl zwischen einem Fahrrad (a), einem roten Bus (b) und einem blauen Bus (c). Wird eine Entscheidung zwischen den drei Möglichkeiten von der Person verlangt, wählt diese mit einer Wahrscheinlichkeit von 50:50 zwischen dem Fahrrad und dem Bus, da ihr die Farbe des Busses egal ist, sodass man die Wahrscheinlichkeiten $\frac{1}{2}, \frac{1}{4}, \frac{1}{4}$ erhält. Über den ersten Teil (1) des Wahlaxioms ergeben sich für die Entscheidung der Person die Wahrscheinlichkeiten $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$, und kann somit eine gegebene Information nicht verarbeiten [Debreu, 1960].

3.2. Mathematisches Modell

Bei der Betrachtung von Paarvergleichen stellt sich die Frage, wie diese Informationen über ein mathematisches Modell formuliert werden können. [Guttman \[1946\]](#) entwickelte eine Methode zur Quantifizierung von Paarvergleichen, stand jedoch vor dem Problem, repräsentative, numerische Werte für die einzelnen Items mit bestimmten Eigenschaften zu bestimmen und auf eine Population von n Individuen zu übertragen.

[Bradley and Terry \[1952\]](#) lösen dieses Problem durch die Annahme einer verallgemeinerten Form des Binomialmodells. Mit dem von ihnen formulierte mathematische Modell und den Maximum-Likelihood-Schätzern für die Bewertung der einzelnen Items gelang es ihnen eine einfache Lösung für Guttman's Problem zu finden.

3.2.1. Annahmen und Voraussetzungen

Für folgende Ausführungen wird ein Experiment angenommen, bei dem m Items systematisch paarweise miteinander verglichen werden. Zunächst wird vorausgesetzt, dass jedes einzelne dieser Items wahre Präferenzen oder Ratings, π_1, \dots, π_m , auf einer subjektiven Skala besitzt. Hierfür gelten die Bedingungen $\pi_i \geq 0$ und $\sum \pi_i = 1$, wobei Letzteres aus Gründen der Identifizierbarkeit notwendig ist. Die Wahrscheinlichkeit, dass beim Vergleich der Items i und j das i -te Item präferiert wird, ist definiert durch

$$p_{ij} = \frac{\pi_i}{\pi_i + \pi_j}; \quad (3.1)$$

Äquivalent steht π_i/π_j für die Chance, dass Item i gegenüber Item j vorgezogen wird. Daher ist das BTL-Modell ein „proportional-odds“ Modell [[Bradley and Terry, 1952](#)].

Für alle Vergleiche gilt $p_{ij} + p_{ji} = 1$, da eine Entscheidung erzwungen wird und kein Unentschieden zugelassen ist. Über die Transformation $\pi_i = \exp(\theta_i) \Leftrightarrow \theta_i = \ln(\pi_i)$ lässt sich Gleichung (3.1) äquivalent darstellen über

$$\log \frac{p_{ij}}{p_{ji}} = \log \frac{p_{ij}}{1 - p_{ij}} = \theta_i - \theta_j. \quad (3.2)$$

Alternativ gilt

$$p_{ij} = \frac{\exp(\theta_i)}{\exp(\theta_i) + \exp(\theta_j)}.$$

Demnach gilt bei $p_{ij} = \frac{1}{2}$, dass $\theta_i = \theta_j$ und bei $p_{ij} > 0$, dass $\theta_i > \theta_j$. Im BTL-Modell ist daher die Wahrscheinlichkeit $p_{ij} = g^{-1}(\theta_i - \theta_j)$ über die Funktion g spezifiziert, für die die Logit-Funktion angenommen wird [[Critchlow and Fligner, 1991](#)].

3.2.2. Likelihood-Inferenz

Um die Likelihood Funktion für Bradley-Terry-Luce Modelle zu erhalten, wird zwischen den einzelnen Vergleichen der Item-Paare Unabhängigkeit angenommen. Zunächst wird noch einmal die Wahrscheinlichkeit aus Gleichung (3.1) betrachtet, die die Präferenz von Item i gegenüber Item j beschreibt. Diese kann ebenso über die Formulierung

$$\left(\frac{\pi_i}{\pi_i + \pi_j} \right)^{2-r_{ijk}} \left(\frac{\pi_j}{\pi_i + \pi_j} \right)^{2-r_{jik}} = \frac{\pi_i^{2-r_{ijk}} \pi_j^{2-r_{jik}}}{(\pi_i + \pi_j)}$$

dargestellt werden. Wählt man dabei $r_{ijk} = 1$ und $r_{jik} = 2$, so vereinfacht sich der Ausdruck auf $\pi_i/(\pi_i + \pi_j)$ gemäß der Wahrscheinlichkeit, dass Item i in der k -ten Wiederholung gegenüber Item j vorgezogen wird. Entsprechend erhält man für $r_{ijk} = 2$ und $r_{jik} = 1$ die Wahrscheinlichkeit $\pi_j/(\pi_i + \pi_j)$, dass Item j bei der k -ten Wiederholung gegenüber Item i präferiert wird [Bradley and Terry, 1952].

Nun werden für alle n Wiederholungen die entsprechenden Ausdrücke multipliziert und unter der Annahme der Unabhängigkeit der einzelnen Vergleiche, erhält man als Likelihood Funktion den Ausdruck

$$L(\boldsymbol{\pi}) = \prod_{i < j} \frac{\pi_i^{y_{ij}} \pi_j^{n_{ij} - y_{ij}}}{(\pi_i + \pi_j)^{n_{ij}}}. \quad (3.3)$$

Dabei steht n_{ij} für die Anzahl der Vergleiche, die zwischen den Objekten i und j durchgeführt werden und y_{ij} bezeichnet die Anzahl an Vergleichen, bei denen Item i gegenüber Item j präferiert wird.

Durch Logarithmieren der Likelihood Funktion erhält man die Log-Likelihood Funktion

$$l(\boldsymbol{\pi}) = \sum_{i < j} \left(y_{ij} \log \frac{\pi_i}{\pi_j} + n_{ij} \log \frac{\pi_j}{\pi_i + \pi_j} \right). \quad (3.4)$$

Zur Berechnung des Maximum-Likelihood (ML-) Schätzers wird üblicherweise die Log-Likelihood Funktion maximiert, da diese Funktion, aufgrund der streng monotonen Funktion des Logarithmus, das selbe Maximum aufweist wie die Likelihood Funktion. Es handelt sich demnach bei BTL-Modellen um Maximierungsprobleme mit Nebenbedingungen. Der ML-Schätzer $\hat{\pi}_i$ berechnet sich nach Huang et al. [2006] über

$$\begin{aligned} & \max_{\boldsymbol{\pi}} \quad l(\boldsymbol{\pi}) \\ & \text{in Abhängigkeit von} \quad \forall \pi_j \geq 0, \quad \sum_j \pi_j = 1. \end{aligned}$$

Dieser Ansatz geht auf [Zermelo \[1929\]](#) zurück und wurde für allgemeinere Situationen vielfach erweitert oder abgeändert. [Bradley and Terry \[1952\]](#) lösen das Maximierungsproblem mit Hilfe der Lagrange-Multiplikatorenregel. Eine einfache, iterative Lösung des Problems ist in [Huang et al. \[2006\]](#) gegeben.

Neben der direkten (iterativen) Schätzung der Likelihood kann die Schätzung des Bradley-Terry-Luce Modells auch über alternative Verfahren erfolgen. Eine Variante ist über ein Generalisiertes Lineares Modell (GLM) mittels logistischer Regression gegeben, dessen Vorgehen in Kapitel 4 aufgezeigt wird.

3.3. Methoden für Bradley-Terry-Luce Modelle

Das Bradley-Terry-Luce Modell schließt, basierend auf der Methode der komparativen Skalierung, ein Unentschieden aus. Dennoch ist die Fülle an Daten enorm, da bereits bei wenigen zu vergleichenden Items etliche Paarvergleiche durchzuführen sind. Paarvergleichsmodelle sind ausschließlich auf diese spezielle Datensituation ausgerichtet, um neben messbaren Kriterien auch subjektive Kriterien sinnvoll in Ranglisten einstufen zu können. Für BTL-Modelle existieren zahlreiche Erweiterungen, wie das Einbinden von Unentschieden [[Davidson, 1970](#)], die Berücksichtigung geordneter Response [[Tutz, 1986](#)], der Vergleich mehrerer Items [[Mallovs, 1957](#)] oder die Berücksichtigung der Reihenfolge bei den Paarvergleichen [[Davidson and Beaver, 1977](#); [Fienberg, 1979](#)].

In vielen Anwendungen wird davon ausgegangen, dass die Präferenzentscheidungen der Personen nicht allein von den Eigenschaften der zu beurteilenden Stimuli abhängt, sondern auch von den Charaktereigenschaften der Personen selbst [[Strobl et al., 2009](#)]. Die direkte Einbindung derartiger Informationen ist in der allgemeinen Definition des Bradley-Terry-Luce Modells und in seinen genannten Erweiterungen nicht gegeben. Überlicherweise werden dafür separate Modellanpassungen für die einzelnen Charaktereigenschaften durchgeführt, etwa für weibliche und männliche Personen getrennt. Ansätze für die direkte Einbindung von Kovariablen sind in [Dittrich et al. \[1998\]](#) und [Böckenholt \[2001\]](#) gegeben. Alternativ können für BTL-Modelle aber auch regularisierte Schätzverfahren angepasst werden.

Zunächst bietet sich die Modellierung von Paarvergleichen über Generalisierte Lineare Modelle (GLM) [[Nelder and Wedderburn, 1972](#)] an, siehe Kapitel 4. Ein wesentlicher Vorteil des parametrischen Verfahrens ist durch die einfache Anwendung statistischer Inferenz über Standardsoftware gegeben. Die besondere Datenstruktur, resultierend aus den direkten Paarvergleichen, ist in den Kontext Generalisierter Linearer Modelle zu übertragen

und die Einbindung von Kovariablen (siehe Unterabschnitt 4.3.2) ist daran anzupassen [Critchlow and Fligner, 1991]. Werden keine zusätzlichen Kovariablen berücksichtigt, so wird lediglich der Effekt des Paarvergleichs geschätzt, wie in Unterabschnitt 4.3.1 vorgestellt wird. Bei einer großen Anzahl p an Kovariablen und einer relativ kleinen Anzahl n an Personen kann der Spezialfall $p > n$ eintreten. Die Schätzung über ein Generalisiertes Lineares Modell stößt in solchen Fällen an seine Grenzen und liefert keine zufriedenstellenden Ergebnisse bezüglich Vorhersagegenauigkeit und Interpretierbarkeit [Tibshirani, 1996].

Ein weiteres alternatives Schätzverfahren für Bradley-Terry-Luce Modelle ist über das Group-LASSO Verfahren [Yuan and Lin, 2006] gegeben, das in Kapitel 5 untersucht wird. Ziel dieser statistischen Methode ist es, die Vorhersagegenauigkeit zu verbessern und gut interpretierbare Modelle zu liefern. Die Idee liegt in der Anpassung eines Regressionsmodells, bei der nur die wichtigsten Prädiktoren berücksichtigt werden, die über Methoden der Modellregularisierung, speziell der Schrumpfung von Parametern, sowie der Variablenselektion als einflussreich eingestuft werden. Durch dieses Vorgehen wird ein möglichst sparsames Modell generiert, welches aus einer großen Anzahl an Kovariablen jeweils nur die Informativen bezüglich der Präferenzentscheidung enthält. Das Group-LASSO Verfahren wurde speziell zur Einbindung kategorialer Kovariablen konzipiert, um Faktorvariablen als Gruppen behandeln zu können. Diese Eigenschaft kann zur Verarbeitung der speziellen Datenstruktur bei Paarvergleichsdaten verwendet werden, um die einzelnen Kovariablen einer Person gruppiert über die Paarvergleiche betrachten zu können.

In den folgenden beiden Kapiteln 4 und 5 werden beide genannten alternativen Schätzverfahren für Bradley-Terry-Luce Modelle analysiert. Zunächst wird jeweils auf die allgemeine Definition der statistischen Verfahren eingegangen und auf den Spezialfall Logistischer Regressionsmodelle erweitert, da es sich bei der Responsevariable „Präferenz“ um eine binäre Größe handelt. Die notwendigen Datentransformationen zur Übernahme der Paarvergleichsstruktur werden in Abschnitt 4.3 genauer erläutert. Zur Überprüfung der Modellregularisierung und Variablenselektion beim Group-LASSO Verfahren werden in Abschnitt 5.3 Simulationsstudien vorgestellt.

4. BTL-Modell über Generalisierte Lineare Modelle

Regression ist wohl die am häufigsten eingesetzte und am besten untersuchte Methodik in der Statistik. Über ein Regressionsmodell wird der Zusammenhang zwischen einer Zielgröße und einer oder mehreren erklärenden Größen untersucht [Fahrmeir et al., 2009]. Dieser Zusammenhang lässt eine einfache Einbindung von Kovariablen bei Paarvergleichen vermuten. Die Datenstruktur (siehe Abschnitt 2.3) der stochastische Variable Y_{rs} mit den Ausprägungen $\{0, 1\}$ für die Präferenz hat in diesem Kontext eine binäre Zielgröße zur Folge. Daher kann die parametrische Methode der Generalisierten Linearen Modelle als alternatives Schätzverfahren für Bradley-Terry-Luce Modelle herangezogen werden, insbesondere der Spezialfall binärer Regression.

Vorab wird in Abschnitt 4.1 auf die allgemeine Definition Generalisierter Linearer Modelle eingegangen. Der Spezialfall der binären Zielgröße wird in Abschnitt 4.2 vorgestellt. Abgeschlossen wird dieses Kapitel mit der Erläuterung der notwendigen Datentransformationen (Abschnitt 4.3) zur parametrischen Schätzung von Paarvergleichsdaten über Generalisierte Lineare Modelle.

4.1. Das Generalisierte Lineare Modell

Generalisierte Lineare Modelle (GLMs) [Nelder and Wedderburn, 1972] repräsentieren eine Klasse von Regressionsmodellen für feste Effekte für verschiedene Typen abhängiger Variablen, d.h. stetig, dichotom oder zählend. McCullagh and Nelder [1989] beschreiben diese im Detail, eine umfassende Einführung liefert Dobson [2002]. Zusätzlich sei auf Agresti [2002] verwiesen.

Regressionsmodelle beschreiben den Einfluss von Kovariablen x_1, \dots, x_k auf die Eigenschaften einer Zielvariablen y . Die Zielvariable wird auch als abhängige Variable bzw. als Response und die Kovariablen werden als erklärende Variablen oder als Regressoren bezeichnet. Die Regressionsmodelle unterscheiden sich durch verschiedene Typen abhängiger Variablen (stetig, binär, kategorial oder zählend) und unterschiedliche Typen von

Kovariablen, welche metrisch, binär oder kategorial sein können.

Sei $\mathbf{y} = (y_1, \dots, y_n)^\top$ die Realisation der Stichprobe $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ und $E(\mathbf{Y}) = \boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^\top$ der Vektor der Erwartungswerte.

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix}$$

ist die Designmatrix und $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)^\top$ sind die Parameter, durch welche das Modell beschrieben werden soll, wobei $\text{rg}(\mathbf{X}) = k + 1 = p$ gilt.

Das klassische Lineare Regressionsmodell hat die Form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}; \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}; \sigma^2 \mathbf{I}),$$

mit dem Fehlervektor $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^\top$ und der Einheitsmatrix \mathbf{I} . Für den Erwartungswert $E(\mathbf{y})$ gilt der lineare Zusammenhang

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}.$$

Generalisierte Lineare Modelle (GLMs) erweitern einfache Lineare Regressionsmodelle für nicht normal-verteilte Zielvariablen. Dabei wird die Annahme, dass der Response einer Normalverteilung folgt dahingehend erweitert, dass eine Verteilung aus der Exponentialfamilie zugelassen wird. Die Annahme der strikten Linearität wird abgeschwächt, indem man für den Response zulässt, dass dieser über eine monotone Funktion vom linearen Prädiktor abhängt [Fahrmeir et al., 2009].

4.1.1. Allgemeine Modelldefinition

Generalisierte Lineare Modelle werden über drei Komponenten näher beschrieben: Einer *zufälligen Komponente*, welche die **Verteilungsannahme** der Zielvariablen beschreibt; einer *systematischen Komponente*, welche die Kovariablen im linearen Prädiktor bestimmt; und einer *Linkfunktion*, die den Erwartungswert der Zielvariablen $E(\mathbf{y})$ mit der systematischen Komponente verknüpft. Die beiden Letzteren ergeben die **Strukturannahme** [Agresti, 2002].

Somit besitzen lineare Modelle und Regressionsansätze für nicht normalverteilte Zielvariablen gemeinsame Eigenschaften, die sich in einheitlicher Darstellung schreiben lassen:

1. **Verteilungsannahme:** Die *zufällige Komponente* in einem GLM nimmt für die Ziel-

variable \mathbf{y} mit unabhängigen Beobachtungen (y_1, \dots, y_n) eine Verteilung aus der einparametrischen Exponentialfamilie an. Die Dichte einer einparametrischen Exponentialfamilie hat die Form

$$f(y_i|\theta_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right). \quad (4.1)$$

Einige wichtige Verteilungen sind Spezialfälle, wie etwa die Poisson- bzw. Binomialverteilung. Eine Übersicht über Verteilungen aus der einfachen Exponentialfamilie ist in Tabelle 4.1 gegeben.

Für die Funktion $b(\theta_i)$ müssen die erste und zweite Ableitung $b'(\theta_i)$ und $b''(\theta_i)$ existieren. Man kann zeigen, dass $E(y_i|\mathbf{x}_i) = \mu_i = b'(\theta_i)$ und $\text{Var}(y_i|\mathbf{x}_i) = b''(\theta_i)a(\phi)$ gilt [Fahrmeir et al., 2009]. Bei der Komponente $a(\phi)$ kann für a eine beliebige Funktion von ϕ angenommen werden, falls ϕ bekannt ist. Ist ϕ unbekannt, so kann man schreiben $a(\phi) = \phi/\omega_i$, wobei es sich bei ω_i um eine bekannte Konstante handelt, die in den meisten Fällen gleich 1 ist [Wood, 2006].

2. **Strukturannahme:** Die Strukturannahme setzt sich aus der *systematischen Komponente* und der *Linkfunktion* zusammen.

Die systematische Komponente eines GLMs verknüpft die erklärenden Variablen mit dem Vektor (η_1, \dots, η_n) über ein lineares Modell. Sei x_{ij} der Wert des Prädiktors j ($j = 1, 2, \dots, p$) für Subjekt i , dann gilt

$$\eta_i = \sum_j \beta_j x_{ij}, \quad i = 1, \dots, n.$$

Diese lineare Kombination der Kovariablen wird als *linearer Prädiktor* bezeichnet. Zur Berücksichtigung eines Intercepts im Modell wird üblicherweise für ein $x_{ij} = 1$ für alle i angenommen.

Die Linkfunktion eines GLMs verbindet die zufällige Komponente mit der systematischen Komponente. Der Erwartungswert $\mu_i = E(Y_i), i = 1, \dots, n$ ist mit dem linearen Prädiktor durch $\eta_i = g(\mu_i)$ verknüpft, wobei g eine monotone differenzierbare Funktion ist.

Somit verknüpft g den Erwartungswert der Zielvariablen mit den Kovariablen durch

$$g(\mu_i) = \sum_j \beta_j x_{ij}, \quad i = 1, \dots, n. \quad (4.2)$$

Eine geeignete Linkfunktion hängt vom Typ der Zielvariablen ab.

	Normal	Binomial	Poisson	Gamma	Inverse Gauß
Verteilung	$N(\mu, \sigma^2)$	$B(n, \pi)/n$	$Po(\mu)$	$\Gamma(\mu, \nu)$	$IG(\mu, \sigma^2)$
$f(y)$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$	$\binom{n}{y} \pi^y (1-\pi)^{n-y}$	$\frac{\mu^y \exp(-\mu)}{y!}$	$\frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right)^\nu \exp\left(-\frac{\nu y}{\mu}\right)$	$\sqrt{\frac{\sigma^2}{2\pi y^3}} \exp\left(-\frac{\sigma^2(y-\mu)^2}{2\mu^2 y}\right)$
Wertebereich	$-\infty < y < \infty$	$y = 0, 1, \dots, n$	$y = 0, 1, 2, \dots$	$y > 0$	$y > 0$
$\theta(\mu)$	μ	$\log\left(\frac{\pi}{1-\pi}\right)$	$\log(\mu)$	$-\frac{1}{\mu}$	$\frac{1}{\mu^2}$
$b(\theta)$	$\frac{\theta^2}{2}$	$\log(1 + \exp(\theta))$	$\exp(\theta)$	$-\log(-\theta)$	$-\sqrt{-2\theta}$
ϕ	σ^2	$\frac{1}{n}$	1	$\frac{1}{\nu}$	σ^2
$a(\phi)$	$\phi(= \sigma^2)$	$\phi(= \frac{1}{n})$	$\phi(= 1)$	$\phi(= \frac{1}{\nu})$	$\phi(= \sigma^2)$
$c(y; \phi)$	$-\frac{1}{2} \left(\frac{y^2}{\phi} + \log(2\pi\phi)\right)$	$\log\left(\binom{n}{ny}\right)$	$-\log(y!)$	$\nu \log(\nu y) - \log\{y\Gamma(\nu)\}$	$-\frac{1}{2} \left(\log(2\pi\phi y^3) + \frac{1}{\phi y}\right)$
$\mu(\theta) = E(Y; \theta)$	θ	$\frac{\exp(\theta)}{1+\exp(\theta)}$	$\exp(\theta)$	$-\frac{1}{\theta}$	$-\sqrt{-2\theta}$
$V(\mu)$	1	$\mu(1-\mu)$	μ	μ^2	μ^3

Tabelle 4.1.: Charakteristiken einiger Verteilungen aus der Exponentialfamilie. Quelle: In Anlehnung an [Fahrmeir et al. \[2009, Seite 219\]](#); [Wood \[2006, Seite 61\]](#); [McCullagh and Nelder \[1989, Seite 30\]](#).

g heißt kanonische Linkfunktion, wenn die strukturelle Komponente die Form

$$\theta_i = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$$

hat, das heißt der kanonische Parameter entspricht dem linearen Prädiktor. Gemäß Tabelle 4.1 entspricht bei einer Normalverteilung das lineare Modell $\mu_i = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ der natürlichen Linkfunktion, bei binären Regressionsmodellen das Logit-Modell und bei Poisson-Modellen das loglineare Modell [Agresti, 2002].

Ein klassisches Generalisiertes Lineares Modell ist letztlich vollständig durch den Typ der Exponentialfamilie (Normal-, Binomial-, Poisson-, Gammaverteilung usw.), durch die Wahl der Linkfunktion und die Definition und Auswahl der Kovariablen charakterisiert [Agresti, 2002; Fahrmeir et al., 2009].

4.1.2. Likelihood-Inferenz

Zur Schätzung unbekannter Parameter in GLMs wird die Maximum-Likelihood (ML)-Methode herangezogen. Bei gegebenen Daten $(y_i, \mathbf{x}_i), i = 1, \dots, n$ erfolgt, gemäß dem *Maximum-Likelihood Prinzip*, die Schätzung der Parameter über die Maximierung der Likelihood.

Das ML-Prinzip lässt sich wie folgt ausdrücken. Die Wahrscheinlichkeit für das Auftreten der beobachteten Stichprobe y_1, \dots, y_n , wird, unter Anwendung eines geeignet bestimmten ML-Schätzers $\hat{\boldsymbol{\beta}}$ für $\boldsymbol{\beta}$, maximiert, so dass die Stichprobe möglichst plausibel (likely) ist.

Unter Berücksichtigung der (bedingten) Unabhängigkeit der Zielvariablen ergibt sich die Likelihood $L(\boldsymbol{\beta})$ als Produkt

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n f(y_i | \boldsymbol{\beta}; \mathbf{x}_i) \quad (4.3)$$

der einzelnen Dichten $f(y_i | \boldsymbol{\beta})$. Im Allgemeinen ist die Likelihood eine differenzierbare Funktion bezüglich $\boldsymbol{\beta}$, deren Maximum durch differenzieren, Null setzen und anschließend dem Auflösen des resultierenden Gleichungssystems bezüglich $\boldsymbol{\beta}$ berechnet werden kann. Üblicherweise wird jedoch die Log-Likelihood $l(\boldsymbol{\beta}) = \log(L(\boldsymbol{\beta}))$ maximiert. Hierbei besitzen $l(\boldsymbol{\beta})$ und $L(\boldsymbol{\beta})$ das selbe argmax, da es sich bei dem Logarithmus um eine streng monotone Funktion handelt. Das bedeutet, dass die Likelihood und die Log-Likelihood unterschiedliche Maxima aufweisen, aber beide Maxima bei $\hat{\boldsymbol{\beta}}$ liegen.

In allgemeiner Form lässt sich die Log-Likelihood für n unabhängige Beobachtungen durch

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n l_i(\boldsymbol{\beta}) = \sum_{i=1}^n \log(f(y_i | \boldsymbol{\beta})) = \sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + \sum_{i=1}^n c(y_i, \phi) \quad (4.4)$$

darstellen, wobei die Abhängigkeit von $\boldsymbol{\beta}$ über den natürlichen Parameter θ_i der Exponentialfamilie

$$\mu_i = b'(\theta_i) = h(\mathbf{x}_i^T \boldsymbol{\beta})$$

gegeben ist, mit h als Responsefunktion, für die die Beziehung $g = h^{-1}$ gilt [Fahrmeir et al., 2009].

Zur Berechnung des ML-Schätzers ist die 1. Ableitung der Log-Likelihood nach $\boldsymbol{\beta}$ zu bilden, die Score-Funktion

$$s(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \left(\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_1}, \dots, \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_k} \right)^T. \quad (4.5)$$

Der Erwartungswert der Scorefunktion ist gleich Null, d.h. $E(s(\boldsymbol{\beta})) = \mathbf{0}$, da $E(y_i | \mathbf{x}_i) = \mu_i$ gilt. Die *ML-Gleichungen* ergeben sich durch Nullsetzen der Score-Funktion

$$s(\hat{\boldsymbol{\beta}}) = \mathbf{0}.$$

Aus der 2. Ableitung erhält man die beobachtete Fisher-Information, welche definiert ist durch:

$$\mathbf{H}(\boldsymbol{\beta}) = -\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \begin{pmatrix} \frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_1} & \dots & \frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_k} \\ \vdots & & \vdots \\ \frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_1} & \dots & \frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_k} \end{pmatrix} \quad (4.6)$$

Die beobachtete Fisher-Informationsmatrix dient als Maß für die Information, welche die Likelihood bezüglich der unbekannt Parameter liefert. Der Erwartungswert der Fisher-Information

$$\mathbf{F}(\boldsymbol{\beta}) = E \left(-\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right) = \text{Cov}(s(\boldsymbol{\beta})) = E(s(\boldsymbol{\beta})s^T(\boldsymbol{\beta}))$$

wird als erwartete Fisher-Information bezeichnet. Für das Logit-Modell entsprechen sich die erwartete und die beobachtete Informationsmatrix, d.h. es gilt $\mathbf{H}(\boldsymbol{\beta}) = \mathbf{F}(\boldsymbol{\beta})$. Dies ist jedoch nicht auf andere Modelle übertragbar, wie etwa das Probit- oder das komplementäre log-log-Modell, welche im Unterabschnitt 4.2.2 genauer vorgestellt werden [Fahrmeir et al., 2009].

Zur Bestimmung des ML-Schätzers $\hat{\boldsymbol{\beta}}$ sind in den meisten Anwendungen verschiedene iterative numerische Verfahren notwendig, welche im Folgenden beschrieben werden.

4.1.3. Iterative Algorithmen

Oft ist ein numerisches Verfahren zur Berechnung des ML-Schätzers erforderlich, um die Nullstellen der Score-Funktion bestimmen zu können, die auf Grund eines nichtlinearen Gleichungssystems nach dem Differenzieren der Score-Funktion nicht analytisch berechnet werden können. Ein Überblick numerischer Verfahren ist etwa in Lange [2000] gegeben. Hier wird lediglich auf das Fisher-Scoring-Verfahren genauer eingegangen.

Der Fisher-Scoring-Algorithmus kann auf mehrdimensionale Parameter $\beta = (\beta_0, \dots, \beta_k)^T$ angewendet werden. Dabei werden die folgenden Iterationsschritte durchlaufen.

Fisher-Scoring-Algorithmus

1. Setze Startwerte $\hat{\beta}^{(0)}$. Setze $k = 0$.
2. Erhalte verbesserte Schätzungen $\hat{\beta}^{(k+1)}$ durch

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + \mathbf{F}^{-1}(\hat{\beta}^{(k)})s(\hat{\beta}^{(k)}), \quad k = 0, 1, 2, \dots \quad (4.7)$$

3. Stoppe, falls für einen sehr kleinen Wert $\varepsilon > 0$

$$\frac{\|\hat{\beta}^{(k+1)} - \hat{\beta}^{(k)}\|}{\|\hat{\beta}^{(k)}\|} \leq \varepsilon$$

gilt. $\hat{\beta} \equiv \hat{\beta}^{(k)}$ ist der ML-Schätzer. Andernfalls setze $k = k + 1$ und fahre mit 2. fort.

Setzt man anstelle der erwarteten Informationsmatrix die beobachtete Informationsmatrix $\mathbf{H}(\hat{\beta}^{(k)})$ ein, so erhält man das Newton- (bzw. Newton-Raphson-) Verfahren. Wegen der Gleichheit der erwarteten und der beobachteten Informationsmatrix beim Logit-Modell entsprechen sich hierbei das Fisher-Scoring- und das Newton-Raphson-Verfahren.

Voraussetzung, damit die Iterationen aus Gleichung (4.7) gegen eine Lösung $\hat{\beta}$ der ML-Gleichungen konvergieren, ist die Invertierbarkeit der Fisher-Matrix $\mathbf{F}(\beta)$ für alle β . Hierfür muss die Designmatrix vollen Rang p besitzen, $\text{rg}(\mathbf{X}) = p$. Der Algorithmus konvergiert in der Regel und stoppt nach einigen Iterationsschritten gemäß dem Abbruchkriterium in der Nähe eines Maximums. Es kann gezeigt werden, dass das erreichte Maximum für natürliche Linkfunktionen eindeutig ist. Im Allgemeinen ist dies jedoch nicht der Fall, so

dass verschiedene Startwerte verwendet werden sollten.

Tritt jedoch keine Konvergenz ein, d.h. die Iterationen divergieren, liegt das daran, dass die (Log-)Likelihood kein Maximum für endliches β besitzt. Mindestens eine Komponente von $\hat{\beta}^{(k)}$ geht gegen $\pm\infty$, so dass die sukzessiven Differenzen $\|\hat{\beta}^{(k+1)} - \hat{\beta}^{(k)}\|$ anwachsen statt gegen Null zu konvergieren, und somit kein endlicher ML-Schätzer existiert. Dies kann bei sehr ungünstigen Datenkonstellationen auftreten, speziell wenn der Stichprobenumfang n im Vergleich zur Dimension p klein ist [Agresti, 2002; Fahrmeir et al., 2009], wie es bei Paarvergleichsdaten vorkommen kann.

4.1.4. Modellwahl

Um zwischen Modellen mit verschiedenen Prädiktoren und Parametern auszuwählen, ist ein Kompromiss zwischen guter Datenanpassung und zu großer Modellkomplexität durch eine hohe Parameterzahl erforderlich [Fahrmeir et al., 2009]. Die Modellwahl durch den direkten Vergleich der Likelihoods ist nicht geeignet, da das Modell mit mehr Variablen immer die höhere Likelihood besitzt. Auch wenn ein überflüssiger Parameter zum richtigen Modell mit aufgenommen wird, erhöht sich die Likelihood, da der zusätzliche Parameter das Modell näher an die beobachteten Daten bringt, obwohl das nur als „Modellierung des Rauschens“ bezeichnet werden kann [Wood, 2006].

Um diese Überanpassung (overfitting) an den Datensatz durch weitere Hinzunahme von Kovariablen, Interaktionen usw., also einer zu hohen Parameterzahl zu vermeiden, müssen Kriterien zur reinen Modellanpassung modifiziert werden.

Das bekannteste Kriterium zur Modellwahl ist das Akaike Informationskriterium (AIC) [Akaike, 1973], das allgemein definiert ist durch

$$\text{AIC} = -2l(\hat{\beta}) + 2p. \quad (4.8)$$

Hierbei bestraft der Modellterm $2p$ ein überparametrisiertes Modell. Beim Vergleich verschiedener Modelle, wird dasjenige Modell mit dem kleinsten AIC-Wert (und nicht mit der größten Log-Likelihood $l(\hat{\beta})$!) gewählt. Oftmals wird anstelle des AICs der mit dem Stichprobenumfang n dividierte Wert AIC/n verwendet [Fahrmeir et al., 2009].

Ein weiteres Kriterium, welches bei der Modellwahl Anwendung findet, ist das Bayesianische Informationskriterium (BIC) [Schwartz, 1965]. Die allgemeine Definition des BICs lautet

$$\text{BIC} = -2l(\hat{\beta}) + \log(n)p. \quad (4.9)$$

Aus einer Menge konkurrierender Modelle ist dasjenige mit dem kleinsten BIC-Wert auszuwählen.

Formal erkennt man, dass das BIC eine ähnliche Gestalt wie das AIC aufweist, jedoch sind beide Informationskriterien unterschiedlich motiviert. Bei der formalen Betrachtung unterscheidet sich lediglich der Faktor vor der Parameterzahl p , der beim AIC durch 2 gegeben ist und sich auf $\log(n)$ beim BIC ändert. Dies impliziert, dass bei der Verwendung des BICs komplexe Modelle deutlich stärker bestraft werden als mit dem AIC, d.h. in der Regel werden Modelle mit weniger Parametern selektiert und es resultieren sparsamere Modelle [Fahrmeir et al., 2009].

4.2. Binäre Regression

Binäre Regression ist notwendig, falls die Responsevariable y dichotom ist und nur zwei Ausprägungen annehmen kann. Zum Beispiel steht y für die Wahl eines Autos (einheimisches Produkt/ Import), für die Kreditwürdigkeit eines Kunden (ja/nein) oder für die Diagnose einer Krankheit (vorhanden/nicht vorhanden) [Agresti, 2002]. Betrachtet man hingegen Paarvergleiche so steht y für das Ergebnis aus der Präferenzentscheidung. Allgemein nimmt jede Beobachtung eine von zwei möglichen Ausprägungen an, welche mit 0 und 1 bezeichnet werden. Bei der Wahl eines Autos gilt:

$$\text{Wahl}_i = \begin{cases} 1, & \text{falls eine einheimische Automarke bei Wahl } i \text{ gewählt wird,} \\ 0, & \text{sonst.} \end{cases}$$

Für die Anwendung auf Paarvergleiche lässt sich dies darstellen über

$$\text{Präferenz}_{ij} = \begin{cases} 1, & \text{falls Item } i \text{ beim Vergleich der Items } i \text{ und } j \text{ präferiert wird,} \\ 0, & \text{sonst.} \end{cases}$$

Der Erwartungswert einer binären Variable $y \in \{0, 1\}$ lautet

$$E(y) = P(y = 0) \cdot 0 + P(y = 1) \cdot 1 = P(y = 1).$$

Weiter kann der Erwartungswert in Abhängigkeit von den Kovariablen durch die Wahrscheinlichkeit

$$P(y = 1|x_1, \dots, x_k) = \pi(\mathbf{x})$$

dargestellt werden. Die Varianz von y

$$\text{Var}(y|\mathbf{x}) = \pi(\mathbf{x})(1 - \pi(\mathbf{x}))$$

hängt somit ebenfalls von den Kovariablen und den Parametern ab [Agresti, 2002].

Ein klassisches lineares Regressionsmodell

$$y_i = P(y_i = 1) + \varepsilon_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i; \quad \varepsilon_i \sim N(0; \sigma^2)$$

ist bei einer binären Zielgröße aus verschiedenen Gründen ungeeignet:

- Lediglich die linke Seite der Formel y_i ist binär.
- Die Fehlervarianz $\text{Var}(\varepsilon_i) = \text{Var}(y_i|x_i)$ ist nicht homoskedastisch, d.h. sie entspricht nicht für alle i dem gleichen Wert σ^2 , auch wenn auf die Normalverteilungsannahme für ε_i verzichtet wird [Fahrmeir et al., 2009].
- Im linearen Modell kann $P(y_i = 1)$ Werte von $\pi_i < 0$ und $\pi_i > 1$ annehmen, welche für Wahrscheinlichkeiten unzulässig sind [Fahrmeir et al., 2009].

4.2.1. Logistisches Regressionsmodell

Im Allgemeinen ergeben sich binäre Daten durch einen *nichtlinearen* Zusammenhang zwischen $\pi(x)$ und x . Eine Veränderung in x hat meist einen geringeren Einfluss wenn $\pi(x)$ nahe 0 oder 1 ist, als wenn $\pi(x)$ nahe 0.5 liegt. Diese nichtlinearen Zusammenhänge sind meist gleichbleibend, das bedeutet, dass $\pi(x)$ mit zunehmenden x entweder stetig ansteigt oder stetig fällt [Agresti, 2002].

Angenommen man hat n unabhängige Beobachtungen y_1, \dots, y_n und die i -te Beobachtung kann als Realisierung der Zufallsvariable Y_i gesehen werden. Für Y_i wird eine Binomialverteilung

$$Y_i \sim B(n_i, \pi_i)$$

angenommen, mit Wahrscheinlichkeit π_i . Für eine Beobachtung gilt $n_i = 1$, für alle i . Hierdurch wird die stochastische Struktur des Modells definiert [Fahrmeir et al., 2009].

Um die Probleme bei der Anwendung eines linearen Modells für binäre Zielgrößen zu umgehen, nimmt man für das Modell

$$\pi_i = P(y_i = 1) = F(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})$$

an. Der Wertebereich für die Funktion F ist auf das Intervall $[0, 1]$ beschränkt. Für die Wahl von F bieten sich Verteilungsfunktionen an, da F aus interpretatorischen Gründen monoton wachsen sollte. Beim logistischen Regressionsmodell wählt man daher die logistische Verteilungsfunktion

$$F(\eta) = \frac{\exp(\eta)}{1 + \exp(\eta)},$$

und erhält somit das *Logit-Modell*

$$\pi_i = P(y_i = 1) = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \quad (4.10)$$

mit dem linearen Prädiktor

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}.$$

Hierbei kommen die *odds (Chance)* bzw. *log-odds (logarithmierte Chance)*

$$\text{odds}_i = \frac{\pi_i}{1 - \pi_i} \quad \text{bzw.} \quad \eta_i = \text{logit}(\pi_i) = \log \frac{\pi_i}{1 - \pi_i} \quad (4.11)$$

zur Anwendung [[Fahrmeir et al., 2009](#)]. Eine graphische Darstellung der Logit-Transformation ist in [Abbildung 4.1](#) gegeben.

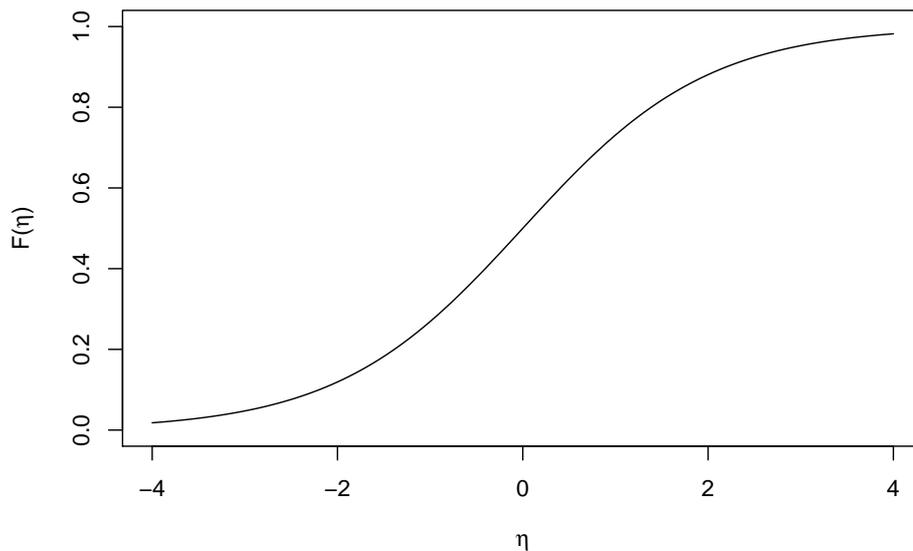


Abbildung 4.1.: Die Logit-Funktion.

Logistische Regressionsmodelle sind demnach GLMs mit binomialverteilter zufälliger Komponente und dem Logit-Link. Die log-odds sind der natürliche Parameter der Binomialverteilung, so dass der Logit-Link dem kanonischen Link entspricht. Während $\pi(x)$ im Wertebereich $[0, 1]$ liegen muss, können die log-odds Werte reeller Zahlen annehmen. Die reellen Zahlen bilden ebenfalls den Bereich für den linearen Prädiktor, welcher die systematische Komponente eines GLMs bildet [Agresti, 2002].

4.2.2. Weitere Linkfunktionen

Anstelle der Logit-Funktion können auch andere Linkfunktionen verwendet werden, die aber nicht so anschaulich interpretiert werden können.

Probit-Modell

Sei $\Phi(\cdot)$ die Verteilungsfunktion der Standardnormalverteilung. Das Modell lässt sich wie folgt schreiben:

$$\pi = \Phi(\eta) = \Phi(\mathbf{x}^T \beta) \quad (4.12)$$

Für Φ existiert die Inverse Funktion $\Phi^{-1}(\cdot)$, sodass sich Gleichung (4.12) äquivalent darstellen lässt über

$$\Phi^{-1}(\pi) = \eta. \quad (4.13)$$

Das Probit-Modell hat den kleinen Nachteil, dass bei der Likelihood-Schätzung der Parameter β die numerische Auswertung von Φ notwendig wird [Fahrmeir et al., 2009].

Komplementäres log-log-Modell

Bei diesem Modell wird die Extremwert-Verteilungsfunktion angenommen

$$\pi = 1 - \exp(-\exp(\eta)) \quad (4.14)$$

und man erhält somit als Linkfunktion $g(\pi) = \log(-\log(1 - \pi))$ [Fahrmeir et al., 2009].

4.3. Datentransformation

Zur Schätzung von Paarvergleichsmodellen mittels Generalisierter Linearer Modelle muss die spezielle Datenstruktur in den Kontext parametrischer Modelle übertragen werden. Insbesondere kommt der Anordnung der Paarvergleiche eine wichtige Rolle zu. Die Reihenfolge wurde an den Datensatz der Anwendungsbeispiele aus Kapitel 6 angepasst. Die Paarvergleichsstruktur lässt sich wie folgt anordnen:

$$1:2, 1:3, 2:3, 1:4, 2:4, 3:4, 1:5, 2:5, 3:5, 4:5, \dots, 1:m, 2:m, \dots, (m-1):m$$

Zur besseren Übersicht kann diese in eine Tabelle, wie in Tabelle 4.2 dargestellt, übertragen werden.

Vergleich mit i -tem Item				
Vergleich mit dem 2. Item	1:2			
Vergleich mit dem 3. Item	1:3	2:3		
Vergleich mit dem 4. Item	1:4	2:4	3:4	
Vergleich mit dem 5. Item	1:5	2:5	3:5	4:5
\vdots				
Vergleich mit dem m . Item	1:m	2:m	...	(m-1):m

Tabelle 4.2.: Anordnungsschema bei Paarvergleichen: m gibt die Anzahl der zu vergleichenden Items an.

Bevor die Paarvergleichtsstruktur in eine entsprechende Designmatrix übertragen werden kann, wird zunächst noch einmal kurz auf das Bradley-Terry-Luce Modell für binären Response eingegangen.

Es werden m Items $\{a_1, \dots, a_m\}$ miteinander verglichen, wobei γ_m gleich Null gesetzt wird. Dabei steht γ_m für die Präferenz bezüglich Item m . Betrachtet man den Vergleich von Item r mit Item s erhält man

$$\begin{aligned} \log \frac{P(r \succ s | (r, s))}{P(s \succ r | (r, s))} &= \gamma_r - \gamma_s \\ &= x_1^{(r,s)} \gamma_1 + \dots + x_{m-1}^{(r,s)} \gamma_{m-1} \\ &= (\mathbf{x}^{(r,s)})^T \boldsymbol{\gamma} \end{aligned}$$

mit $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{m-1})^T$, $\gamma_m = 0$. Dabei steht etwa die Wahrscheinlichkeit $P(r \succ s | (r, s))$ für die Präferenz von Item r gegenüber Item s bei der Paarkombination (r, s) (siehe Kapitel 2). Die Einträge für die Designmatrix ergeben sich durch

$$x_j^{(r,s)} = \begin{cases} 1, & j = r \\ -1, & j = s \\ 0, & \text{sonst.} \end{cases} \quad (4.15)$$

Diese Darstellung entspricht der in der Statistik bekannten Effektkodierung. Analog lässt sich Gleichung (4.15) über

$$\mathbf{x}(r, s) = \mathbf{1}_r - \mathbf{1}_s$$

darstellen, wobei $\mathbf{1}_r = (0, \dots, 0, 1, 0, \dots, 0)$ die Länge $m - 1$ hat mit einer 1 an Position r . Das bedeutet, dass die Designmatrix in jeder Zeile Einträge der Werte 1, -1 und 0 besitzt, repräsentierend für einen Paarvergleich für eine betrachtete Person. Da $\gamma_m = 0$ gilt, muss nicht in jeder Zeile der Wert -1 vorkommen.

Über die Darstellung mit Zufallsvariablen gilt analog

$$\mathbf{Y}_{(r,s)} = \begin{cases} 1, & r \succ s \\ 0, & s \succ r \end{cases}$$

und der Vergleich von Item r mit Item s lässt sich wie folgt schreiben [Tutz, 2013].

$$\log \frac{P(r \succ s | (r, s))}{P(s \succ r | (r, s))} = \log \frac{P(\mathbf{Y}_{(r,s)} = 1)}{P(\mathbf{Y}_{(r,s)} = 0)}$$

Weiter muss die Anzahl der Personen, mit denen der Paarvergleich durchgeführt wird berücksichtigt werden. Zunächst wird der 1. Paarvergleich, Item 1 mit Item 2 (1:2), für alle betrachteten Personen durchgeführt, danach folgt der 2. Paarvergleich zwischen Item 1 und Item 3 (1:3) für alle Personen usw. Wie dies entsprechend in der Designmatrix berücksichtigt wird, wird in Unterabschnitt 4.3.1 genau erläutert.

Es sind verschiedene Designmatrizen möglich, je nachdem ob Kovariablen bei der Schätzung der Parameter für die jeweiligen Paarvergleiche berücksichtigt werden sollen. Einerseits besteht die Möglichkeit, keine zusätzlichen Informationen über Kovariablen in das Modell mit aufzunehmen, so dass lediglich die Effekte der einzelnen Paarvergleiche modelliert werden (siehe Unterabschnitt 4.3.1). Dieses Vorgehen entspricht sozusagen der Schätzung über das allgemeine Bradley-Terry-Luce Modell. Andererseits können durch die Aufnahme von Kovariablen (siehe Unterabschnitt 4.3.2) mögliche weitere Effekte aufgezeigt werden, die Aussagen über Einflussvariablen in Paarvergleichsmodellen ermöglichen. Eine Unterscheidung zwischen binären/kategorialen und metrischen Kovariablen ist von Relevanz. Dieses Vorgehen erleichtert die Einbindung von Kovariablen gegenüber dem Bradley-Terry-Luce Modell.

4.3.1. Keine Berücksichtigung von Kovariablen

Interessiert nur der Effekt der einzelnen Paarvergleiche, so werden keine Kovariablen im Modell berücksichtigt. Es wird lediglich über einen Intercept der Effekt über alle Paarvergleiche der n Personen geschätzt. Bei m zu vergleichenden Items erhält man $m - 1$ Schätzungen für die Präferenzen, da der Effekt der Paarvergleiche mit dem m -ten Item aus Gründen der Identifizierbarkeit auf Null gesetzt wird, $\gamma_m = 0$.

Im Folgenden ist beispielhaft eine Designmatrix einer einzelnen Person dargestellt. Die Datenstruktur für die einzelnen Paarvergleiche entspricht den Formulierungen aus Gleichung (4.15). Das Anordnungsschema der Paarvergleiche kann aus Tabelle 4.2 entnommen werden.

$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

In dieser Darstellung werden $m = 6$ Items miteinander verglichen. Dies ergibt $I = \binom{m}{2} = 15$ Paarvergleiche, die sich in der Anzahl der Zeilen der Designmatrix wiederfinden. Die Anzahl der Spalten der Designmatrix beträgt $m - 1$. Somit steht jede Zeile für einen Paarvergleich einer Person und jede Spalte für das i -te Item, ausgenommen Item m .

Die 4. Zeile der Designmatrix steht für den Paarvergleich zwischen Item 1 und Item 4. Entsprechend der Formulierung aus Gleichung (4.15) gilt:

$$x_j^{(1,4)} = \begin{cases} 1, & j = 1 \\ -1, & j = 4 \\ 0, & \text{sonst} \end{cases}$$

Da die Spalten für die einzelnen Items stehen, enthalten die Zeilen genau an den Positionen Werte ungleich Null, die miteinander verglichen werden. Ist nur ein Wert ungleich Null vorhanden, handelt es sich um einen Paarvergleich mit dem m -ten Item, die aus Gründen der Identifizierbarkeit weggelassen werden.

Allerdings ist nicht nur der Effekt einer Person von Interesse, sondern die Effekte über alle n Personen, so dass über die Struktur der Paarvergleiche eine eindeutige Rangordnung

folgende Struktur

$$\gamma_{ir} = \beta_{r0} + x_i \beta_r.$$

Die Kovariablen können sowohl binär/kategorial $x_i \in \{0, 1\}$, $x_i \in \{-1, 1\}$ oder metrisch sein. Der Parameter β_r ändert den Effekt der Kovariable bezüglich Item r . Hingegen steht der Parameter β_{r0} wieder für den Intercept hinsichtlich der Effekte der einzelnen Paarvergleiche.

Ein Beispiel, wie β_r den Effekt der Kovariable von Item r betreffend verändert, ist bei der binären Kovariable $x_i \in \{-1, 1\}$ folgendermaßen gegeben:

$$\begin{aligned} x_i = 1 : \quad \gamma_{ir} &= \beta_{r0} + \beta_r \\ x_i = -1 : \quad \gamma_{ir} &= \beta_{r0} - \beta_r \end{aligned}$$

Je nach Ausprägung der binären Variable verändert sich der Einfluss von β_r auf die Präferenz und hat in diesem Beispiel entweder einen positiven oder einen negativen Einfluss bei der Präferenzwahl im Paarvergleich.

Beim Paarvergleich von Item r mit Item s erhält man unter Berücksichtigung einer Kovariable x_i als resultierendes Modell für Person i

$$\begin{aligned} \log \frac{\mathrm{P}(r \succ s | (r, s), x_i)}{\mathrm{P}(s \succ r | (r, s), x_i)} &= \beta_{r0} - \beta_{s0} + x_i(\beta_r - \beta_s) \\ &= x_1^{(r,s)} \beta_{1,0} + \dots + x_{m-1}^{(r,s)} \beta_{m-1,0} + x_i x_1^{(r,s)} \beta_1 + \dots + x_i x_{m-1}^{(r,s)} \beta_{m-1} \end{aligned} \quad (4.16)$$

mit $\beta_{m,0} = \beta_m = 0$. Die Parameter $\beta_{i,0}$ für $i = 1, \dots, m - 1$ repräsentieren die Intercepts mit den Effekten der Paarvergleiche, während die Parameter β_i , $i = 1, \dots, m - 1$ den Effekt der Kovariablen ausdrücken.

Die Verallgemeinerung auf einen Kovariablenvektor $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ik})$ ergibt

$$\begin{aligned} \gamma_{ir} &= \beta_{r0} + x_{i1} \beta_{r1} + \dots + x_{ik} \beta_{rk} \\ &= \beta_{r0} + \mathbf{x}_i^T \boldsymbol{\beta}_r. \end{aligned} \quad (4.17)$$

Im allgemeinen Fall verändert β_{rj} den Einfluss der Kovariable x_j bei der Betrachtung von Item r und es ergibt sich mit $x_{i0} = 1$ das Modell [Tutz, 2013]

$$\log \frac{\mathrm{P}(r \succ s | (r, s), i)}{\mathrm{P}(s \succ r | (r, s), i)} = \sum_{t=1}^m \sum_{j=0}^k x_{ij} x_t^{(r,s)} \beta_{tj}. \quad (4.18)$$

In der Anwendung bedeutet dies, dass sich die Designmatrix neben den Effekten der Paarvergleiche über die Intercepts aus Unterabschnitt 4.3.1 um weitere $m - 1$ Spalten pro zusätzliche Kovariable vergrößert, wie in Gleichung (4.16) für eine Kovariable dargestellt ist. Der Einfluss der Kovariable wird, wie aus der allgemeinen Gleichung (4.18) zu entnehmen ist, mit den Effekten aus den Paarvergleichen multipliziert, wobei zu beachten ist, dass hier keine Matrixmultiplikation durchgeführt wird.

Die sich daraus ergebenden Dimensionen der Designmatrix wachsen bei der Betrachtung von Paarvergleichsdaten stark an. Werden zum Beispiel von $n = 200$ Personen $m = 10$ Items über die Methodik des paarweisen Vergleichs (45 Vergleiche) miteinander verglichen und weitere 10 Kovariablen berücksichtigt, so erhält man für die Dimension der Designmatrix (9.000×99). Bei $m = 20$ Items im Vergleich erhöht sich diese auf (38.000×99). Bildlich kann man sich also den Aufbau der Designmatrix wie folgt vorstellen:

$$nI \begin{pmatrix} 1 \dots (m-1) & \left| \right. & 1 \dots (m-1) \\ \vdots & & & & & & & & \\ \text{Intercept} & & \text{1. Kovariable} & & \text{2. Kovariable} & & \text{3. Kovariable} & & \text{usw.} \end{pmatrix}$$

Ein Beispiel für eine binäre Kovariable ist etwa durch die Variable Geschlecht weiblich/männlich $x_i \in \{0, 1\}$ gegeben. Beim ersten Paarvergleich von Item 1 mit Item 2 bei insgesamt $m = 6$ Items erhält man $(1 \quad -1 \quad 0 \quad 0 \quad 0)$ für den Effekt des Paarvergleichs, entsprechend der 1. Zeile der Designmatrix aus 4.3.1. Durch Multiplikation erhält man

$$x_i \in \{0, 1\} : \begin{array}{l} 0 : \quad 1 \rightarrow 0, \quad -1 \rightarrow 0, \quad 0 \rightarrow 0 \\ 1 : \quad 1 \rightarrow 1, \quad -1 \rightarrow -1, \quad 0 \rightarrow 0 \end{array}$$

Entsprechend gilt dies für weitere binäre bzw. kategoriale oder metrische Kovariablen.

4.4. Umsetzung und Problematik

Der wesentliche Vorteil bei der Schätzung von Bradley-Terry-Luce Modellen über Generalisierte Lineare Modelle ist die einfache Einbindung zusätzlicher Kovariablen im Kontext paarweiser Vergleiche. Diese können über die Designmatrix bei der Schätzung über Standardsoftware eingebunden werden. In diesem Fall erfolgt die Modellanpassung über die Statistik Software R, Version 2.15.0 [R Development Core Team, 2012], speziell über das

`stats`-Paket [R Core Team and contributors worldwide, 2012] für grundlegende Statistik-Funktionalitäten.

Zur Anpassung eines Generalisierten Linearen Modells mittels binärer Regression als Alternative für Bradley-Terry-Luce Modelle wird die R-Funktion `glm.fit` verwendet, bei der die Designmatrix und der Responsevektor zu definieren sind. Die wichtigsten Komponenten der R-Funktion zur Anwendung auf Paarvergleichsdaten lauten:

$$\text{glm.fit}(x, y, \text{family}=\text{binomial}(\text{link}=\text{"logit"}), \text{intercept}=\text{FALSE}) \quad (4.19)$$

Bei x handelt es sich um die Designmatrix, die wie in Unterabschnitt 4.3.1 entweder nur die Effekte aus den Paarvergleichen enthält oder wie in Unterabschnitt 4.3.2 zusätzliche Kovariablen berücksichtigt. Für y wird die Responsevariable Y eingesetzt, die bei Paarvergleichen der Präferenz entspricht und binärer Natur ist. Zur Übergabe dieser Information wird über die Bezeichnung `family=binomial(link="logit")` das Logistische Regressionsmodell angegeben. Da der Intercept bei Paarvergleichen bereits in der Designmatrix enthalten ist, ist dieser in der Funktion `glm.fit` auf `FALSE` zu setzen. Als Ergebnis erhält man die Koeffizientenschätzer gemäß den Paarvergleichen und weitere Schätzer für den Fall zusätzlicher Kovariablen.

Der R-Code zur Simulation der Schätzung von Paarvergleichsmodellen über GLMs ist in Anhang Abschnitt A.2.1 beigefügt. Über die Simulationsmaske kann also die Vorhersagegenauigkeit von Paarvergleichsdaten über Generalisierte Lineare Modelle überprüft werden.

Bei der Anpassung Generalisierter Linearer Modelle auf Paarvergleichsdaten zeigt sich bei verschiedenen Simulationsdurchläufen, dass GLMs eine gute Alternative zur Schätzung von Bradley-Terry-Luce Modellen sind. Die Anpassung eines GLMs liefert sinnvolle Schätzer und interpretierbare Ergebnisse. Einzig die Problematik der hohen Dimensionen bei Paarvergleichen schränkt die Anwendung ein. Für den Fall $p > n$ kann bei der Anwendung iterativer Algorithmen (siehe Unterabschnitt 4.1.3) zur Berechnung des ML-Schätzers keine Konvergenz eintreten. In solchen Fällen existiert kein endlicher ML-Schätzer und die Anpassung eines GLMs für BTL-Modelle stößt an seine Grenzen.

Abhilfe verschafft bei solch ungünstigen Datenkonstellationen die regularisierte Schätzung von Bradley-Terry-Luce Modellen über Group-LASSO [Yuan and Lin, 2006], das in Kapitel 5 vorgestellt wird. Das Verfahren der Modellregularisierung, speziell der Modellschrumpfung, und Variablenselektion kann auf logistische Regressionsmodelle (siehe Unterabschnitt 5.2.1) übertragen werden und somit als weiteres alternatives Schätzverfahren von Paarvergleichsdaten angesehen werden.

5. BTL-Modell über Group-LASSO

Eine weitere Möglichkeit Bradley-Terry-Luce Modelle zu schätzen, ist durch die Anwendung des Group-LASSO Verfahrens gegeben, das durch Modellregularisierung und Variablenselektion möglichst sparsame, interpretierbare Modelle generiert. Die Abkürzung LASSO steht für „**L**east **A**bsolute **S**hrinkage and **S**election **O**perator“. Beim Group-LASSO Verfahren [Yuan and Lin, 2006] handelt es sich um ein sogenanntes Shrinkage-Verfahren, das auch bei hochdimensionalen Problemen ohne Einschränkungen angewendet werden kann. Der Vorteil gegenüber LASSO [Tibshirani, 1996] liegt in der ganzheitlichen Betrachtung kategorialer Variablen als Gruppen. Diese Eigenschaft kann bei der Schätzung von Paarvergleichsdaten in dem Sinne genutzt werden, dass die einzelnen Kovariablen über die Paarvergleiche gruppiert werden. Einzige Voraussetzung sind standardisierte Kovariablen. Über Modellregularisierung und Variablenselektion werden demnach die zugehörigen Modellparameter der einzelnen Gruppen geschrumpft bzw. mit Effekt Null geschätzt, so dass die einzelnen Gruppen entweder im Modell verbleiben oder vollständig aus diesem fallen. Dadurch erhält man sparsamere und interpretierbare Modelle.

Variablenselektionsverfahren sind ein wesentlicher Bestandteil bei der Modellbildung, um die Variablen zu finden bzw. auszuwählen, die informativ ins finale Modell eingehen. Aufgrund der hohen Dimensionalität bei Paarvergleichsdaten, sollen nur Kovariablen im Modell berücksichtigt werden, die einen wesentlichen Einfluss auf die Zielgröße haben. Daher ist ein Kompromiss zwischen Modellanpassung und Modellkomplexität erforderlich. Ein Überblick über Variablenselektionsverfahren ist in Abbildung 5.1 gegeben.

Es wird zwischen Subset-Selection und Shrinkage Verfahren unterschieden. Bei Ersterem wird eine Teilmenge an Variablen beibehalten, die Restlichen finden im Modell keine Berücksichtigung. Es werden dem Regressionsmodell schrittweise Kovariablen zugefügt oder entfernt bis die Modellanpassung keine wesentlichen Verbesserungen aufweist. Da es sich beim Subset-Selection um einen diskreten Prozess handelt, bei dem die Variablen entweder aufgenommen oder verworfen werden, kommt es zu Schwierigkeiten bei hoher Variabilität, welche bei Shrinkage Methoden besser verarbeitet werden kann [Hastie et al., 2008]. Bei Shrinkage Verfahren werden über Strafterme Bedingungen an die Regressionskoeffizienten gestellt, die bei der Schätzung berücksichtigt werden müssen. Die Koeffizi-

entschätzer ergeben sich durch die Maximierung der penalisierten Log-Likelihood. Der optimale Penalisierungsparameter, meist mit λ bezeichnet, wird über Optimalitätskriterien wie AIC oder BIC ermittelt. Gilt $\lambda = 0$ erhält man die unpenalisierte Schätzung, die mit der ML-Schätzung übereinstimmt [Hastie et al., 2008]. Bei der Anwendung von Group-LASSO [Yuan and Lin, 2006] erfolgt die Bestrafung über eine Mischung aus l_1 - und l_2 -Norm, die einige nicht informative Kovariablen aus dem Modell selektieren soll. Somit handelt es sich beim Group-LASSO Verfahren um einen simultanen Schätz- und Modellbildungs-Algorithmus.

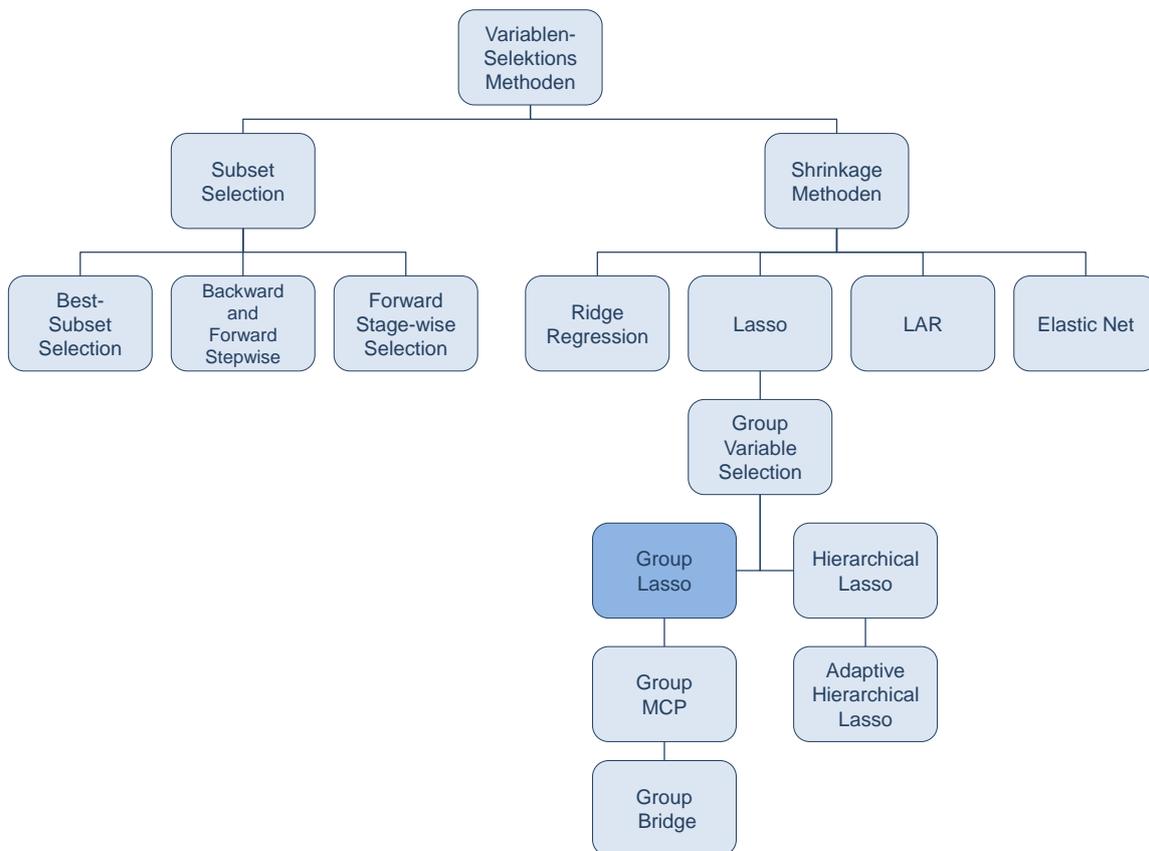


Abbildung 5.1.: Übersicht über verschiedene Variablenselektionsverfahren. Quelle: Yang [2011]

Da sowohl beim LASSO als auch beim Group-LASSO Verfahren Normen bei der Penalisierung verwendet werden, werden diese im Folgenden kurz eingeführt. Das LASSO Verfahren wendet die l_1 -Norm im Bestrafungsterm an, die über

$$|\mathbf{v}| = |v_1| + |v_2| + \dots + |v_n|$$

für einen Vektor $\mathbf{v} = (v_1, v_2, \dots, v_n)$ definiert ist, wobei $|\cdot|$ für den Betrag steht [Elad,

2010]. Der Bestrafungsterm bei Group-LASSO Verfahren verwendet ein Zwischenprodukt aus l_1 - und l_2 -Norm. Darunter ist zu verstehen, dass die l_1 -Norm als Bestrafung zwischen den Gruppen und die l_2 -Norm als Bestrafung innerhalb der Gruppen erfolgt [Yuan and Lin, 2006]. Die Schreibweise $\|\cdot\|$ entspricht der l_2 -Norm, auch euklidische Norm genannt, die für einen Vektor $\mathbf{v} = (v_1, v_2, \dots, v_n)$ über

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}$$

berechnet werden kann [Elad, 2010].

5.1. Das LASSO Verfahren

Das LASSO Verfahren [Tibshirani, 1996] wurde ursprünglich für lineare Regressionsmodelle aufgestellt und wird zur Modellregularisierung und Variablenselektion angewendet. Beim üblichen Ansatz über lineare Regressionsmodelle liegen eine stetige Zielgröße $\mathbf{Y} \in \mathbb{R}^n$, eine $(n \times p)$ Designmatrix \mathbf{X} und ein Parametervektor $\boldsymbol{\beta} \in \mathbb{R}^p$ vor. Der LASSO Schätzer ist definiert durch

$$\hat{\boldsymbol{\beta}}_\lambda = \arg \min_{\boldsymbol{\beta}} (\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \sum_{j=1}^p |\beta_j|) \quad (5.1)$$

mit $\|\mathbf{u}\|_2^2 = \sum_{i=1}^n \mathbf{u}_i^2$ für einen Vektor $\mathbf{u} \in \mathbb{R}^n$ [Meier et al., 2008]. Folglich wird im Fall Linearer Regression Gleichung (5.1), unter Berücksichtigung eines zusätzlichen Bestrafungsterms, über die Methode der kleinsten Quadrate minimiert. Bei steigendem λ schrumpfen die Werte aller Koeffizienten, β_λ , oder werden exakt auf Null geschätzt. Demnach sollen nur relevante/informative Koeffizienten ins Modell aufgenommen werden. Bei der LASSO Bestrafung sind die Lösungen nicht eindeutig und variieren für jedes λ [Tibshirani, 1996].

Die hier angewendete l_1 -Typ Bestrafung beim LASSO Verfahren kann ebenso auf Logistische Regressionsmodelle [Lokhorst, 1999; Shevade and Keerthi, 2003; Roth, 2004; Genkin et al., 2007] angewendet werden. Dabei hat die LASSO Bestrafung die selbe Struktur wie bei Linearen Regressionsmodellen in dem Sinne, dass man eine Likelihood Funktion mit zusätzlicher Bestrafung hat. Im Fall von Logistischen Regressionsmodellen wird die penalisierte Log-Likelihood Funktion verwendet [Meier et al., 2008]. Die Log-Likelihood Funktion bei logistischer Regression lautet für $\boldsymbol{\beta} = (\beta_0^T, \beta_1^T, \beta_2^T, \dots, \beta_k^T)^T$, mit $k = p - 1$:

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n [y_i \mathbf{x}_i^T \boldsymbol{\beta} - \log(1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta}))] \quad (5.2)$$

Um sparsame Lösungen der Modelle sicher zu stellen, wird die LASSO Bestrafung zur Log-Likelihood hinzugefügt:

$$\hat{\beta}_\lambda = \arg \min_{\beta} (l(\beta) + \lambda \sum_{j=1}^p |\beta_j|) \quad (5.3)$$

Auch für den Fall logistischer Regression bei der neben stetigen auch kategoriale Prädiktoren (Faktoren) vorkommen, liefert das LASSO Verfahren keine zufriedenstellenden Ergebnisse, da einzelne Dummy-Variablen anstelle ganzer Faktoren selektiert werden. Die Ergebnisse hängen davon ob, wie die Dummy-Variablen kodiert sind. So führen im Allgemeinen unterschiedliche Anordnungen bei kategorialen Variablen zu unterschiedlichen Schätzungen.

Bei Paarvergleichsdaten würde das nicht Erkennen einer Gruppe ebenfalls falsche Schätzungen zur Folge haben. Daher wird für diese spezielle Datenstruktur das Group-LASSO Verfahren angewendet, welches zunächst allgemein in Abschnitt 5.2 vorgestellt wird. Der Spezialfall einer binären Zielgröße wird in Unterabschnitt 5.2.1 behandelt, bei dem eine Erweiterung von Group-LASSO auf Logistische Regressionsmodelle erfolgt. Abschließend wird in Abschnitt 5.3 anhand von Simulationsstudien die Vorhersagegenauigkeit von Group-LASSO Modellen an Paarvergleichsdaten überprüft.

5.2. Das Group-LASSO Verfahren

Das Group-LASSO Verfahren [Bakin, 1999; Cai, 2001; Antoniadis and Fan, 2001; Yuan and Lin, 2006] wurde als eine Erweiterung des LASSO Verfahrens für lineare Regressionsmodelle eingeführt und ursprünglich zur Einbindung kategorialer Variablen entwickelt. Das LASSO Verfahren ist darauf beschränkt, Faktorvariablen als einzelne dummykodierte Variablen ins Modell aufzunehmen und bietet somit keine Möglichkeit Faktorvariablen als Ganzes zu betrachten. Das Group-LASSO Verfahren umgeht dieses Problem indem Faktorvariablen zu Gruppen von Variablen zusammengefasst werden, die entweder als Ganzes im Modell berücksichtigt oder geschlossen als diese aus dem Modell genommen werden. Bei der Anwendung auf Paarvergleichsdaten wird diese Eigenschaft insbesondere zur Verwertung der Paarvergleichsstruktur eingesetzt, sodass einzelne Kovariablen als Gruppe betrachtet werden können.

Der Group-LASSO Schätzer bei Linearer Regression ist definiert durch

$$\hat{\beta}_\lambda = \arg \min_{\beta} (\|\mathbf{Y} - \mathbf{X}\beta\|_2^2 + \lambda \sum_{g=1}^G \|\beta_{\mathcal{I}_g}\|_2) \quad (5.4)$$

5. BTL-Modell über Group-LASSO

wobei \mathcal{I}_g die Indexmenge der zugehörigen g -ten Gruppe an Variablen ist, $g = 1, \dots, G$ [Meier et al., 2008]. Gleichung (5.4) entspricht bei einer einheitlichen Gruppengröße von 1 dem normalen LASSO Verfahren. Wie bereits erwähnt ist die Bestrafungsfunktion ein Zwischenprodukt der l_1 - und l_2 -Norm. Somit nutzt Group-LASSO alle Vorteile aus LASSO-Verfahren durch l_1 -Typ Bestrafung und Ridge-Regression durch l_2 -Typ Bestrafung.

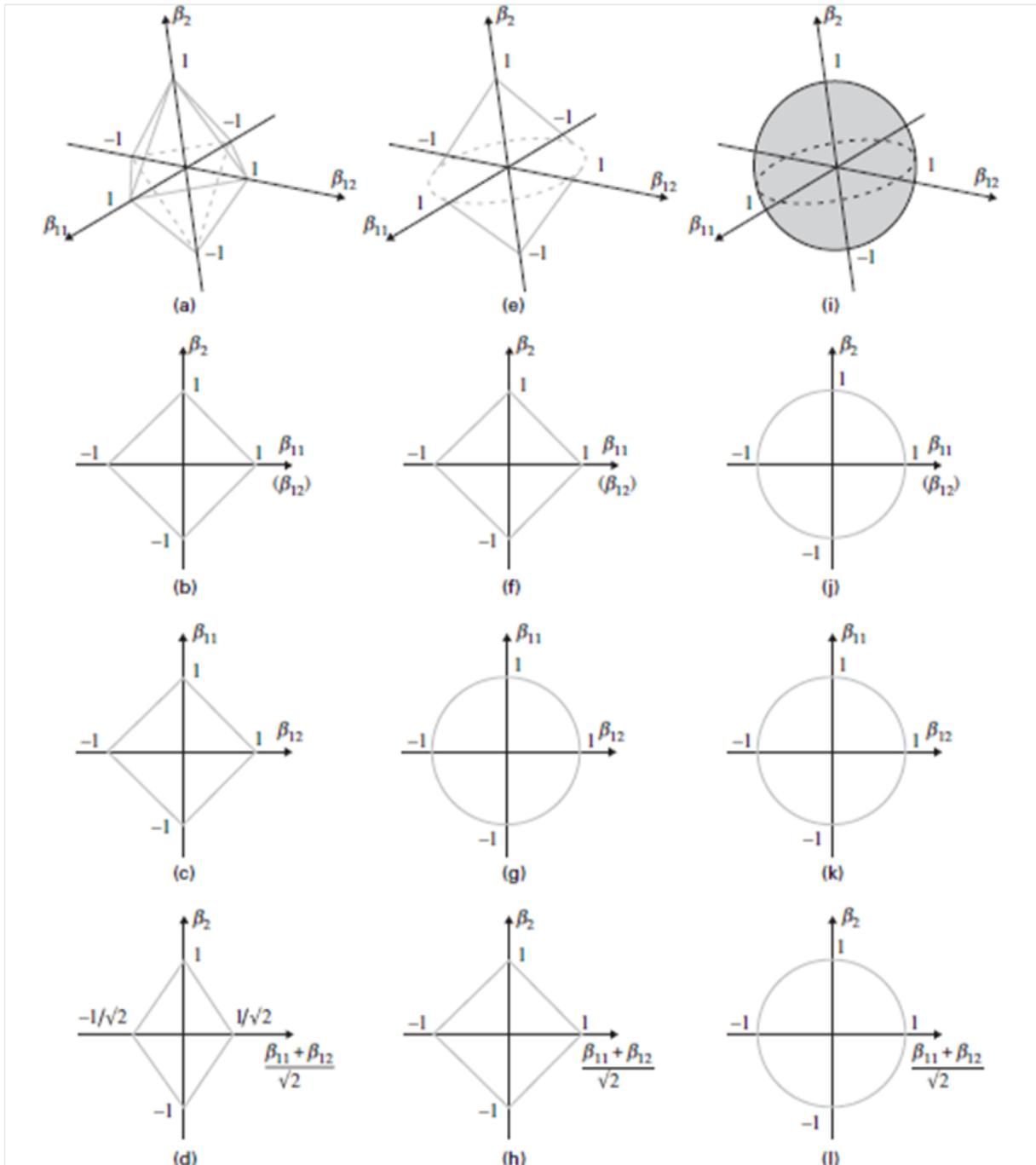


Abbildung 5.2.: (a)-(d) l_1 -Bestrafung, (e)-(h) Group-LASSO Bestrafung und (i)-(l) l_2 -Bestrafung.

Die Mischung beider Bestrafungsfunktionen hat den Vorteil, dass die Variablenselektion auf Gruppen-Ebene durchgeführt wird und, wie bei der Ridge-Regression, unter (gruppenweisen) orthogonalen Transformationen invariant ist [Yuan and Lin, 2006]. Die Mischung nutzt die l_1 -Bestrafung zwischen den Gruppen und die l_2 -Bestrafung innerhalb der Gruppen. Eine graphische Darstellung dieses Zusammenhangs ist in Abbildung 5.2 gegeben.

Zur Erläuterung von Abbildung 5.2 sei im Folgenden der Fall angenommen, dass es zwei Faktorvariablen gibt, einen 2-fachen Vektor $\beta_1 = (\beta_{11}, \beta_{12})^T$ und ein Skalar β_2 . Die Darstellungen 1(a), 1(e) und 1(i) zeigen die Konturen der Bestrafungsfunktionen. In 1(a) ist die l_1 -Bestrafung $|\beta_{11}| + |\beta_{12}| + |\beta_2| = 1$ dargestellt, die zwischen den Gruppen bei Group-LASSO verwendet wird, 1(e) beschreibt $\|\beta_1\| + |\beta_2| = 1$ die Group-LASSO Bestrafung und 1(i) zeigt die l_2 -Bestrafung $\|(\beta_1^T, \beta_2)^T\| = 1$, die innerhalb der Gruppen vorliegt. Die Schnittpunkte mit den Ebenen $\beta_{12} = 0$ (oder $\beta_{11} = 0$), $\beta_2 = 0$ und $\beta_{11} = \beta_{12}$ sind in den Abbildungen 1(b)-1(d), 1(f)-1(h) und 1(j)-1(l) dargestellt. Wie aus Abbildung 5.2 ersichtlich ist, werden bei der l_1 -Bestrafung die drei Achsen-Richtungen differenziert betrachtet, mit dem Ergebnis, dass individuelle Koeffizienten seltener vorkommen. Bei der l_2 -Bestrafung werden hingegen alle Richtungen gleich behandelt und die Anzahl an Koeffizienten reduziert sich nicht. Die Group-LASSO Bestrafung unterstützt somit durch die Mischung der beiden Bestrafungen die Selektion an Koeffizienten auf Faktorebene [Yuan and Lin, 2006].

Yuan and Lin [2006] wählen die l_2 -Norm zur Penalisierung der Koeffizienten innerhalb der Gruppen, mit

$$\|\beta_a\| = \sqrt{\beta_{a1}^2 + \dots + \beta_{ap_a}^2}$$

wobei p_a für die einzelnen Gruppengrößen steht. Obwohl die euklidische Norm eines Vektors β_a nur Null ist, wenn alle seine Komponenten Null sind, führt dieses Vorgehen zu einer Selektion an Gruppen und an Variablen auf individueller Ebene. Dadurch können einige Werte von λ den vollständigen Vektor $\beta_a = \mathbf{0}$ auf Null setzen und die a -te Gruppe wird bei der Schätzung des Modells nicht berücksichtigt, so dass als Ergebnis ein sparsames Modell resultiert.

Dieses Vorgehen wurde von Bakin [1999] als eine Erweiterung des LASSO Verfahrens zur Selektion von Gruppen von Variablen vorgeschlagen und von Yuan and Lin [2006] verallgemeinert. Meier et al. [2008] erweiterten das Konzept des Group-LASSO Verfahrens auf Logistische Regressionsmodelle, welches im Folgenden Unterabschnitt 5.2.1 vorgestellt wird.

5.2.1. Group-LASSO bei Logistischer Regression

Das allgemeine Group-LASSO Verfahren für lineare Regressionsmodelle kann auf Logistische Regressionsmodelle übertragen werden. Diese Eigenschaft führt dazu, dass auch Paarvergleichsmodelle wie das Bradley-Terry-Luce Modell über Group-LASSO geschätzt werden können. Folgende Ausführungen beziehen sich auf [Meier et al. \[2008\]](#).

Zuerst wurde das Group-LASSO Verfahren für Logistische Regressionsmodelle von [Kim et al. \[2006\]](#) untersucht, jedoch hängt bei dem von [Kim et al. \[2006\]](#) vorgeschlagenen Algorithmus die Konvergenzeigenschaft von unbekanntem Konstanten ab. In [Meier et al. \[2008\]](#) hingegen werden Methoden eingeführt, welche die direkte Arbeit am Penalisierungsproblem unterstützen. Die vorgeschlagenen Algorithmen sind effizient, in dem Sinne, dass sie mit Problemsituationen wie große Anzahl Kovariablen/Parameter und kleinen Stichprobenumfang umgehen können.

Zur allgemein gültigen Definition von logistischen Group-LASSO Modellen seien (\mathbf{x}_i, y_i) , $i = 1, \dots, n$ unabhängig und identisch verteilte (iid (independent and identical distributed)) Beobachtungen eines p -dimensionalen Vektors $\mathbf{x}_i \in \mathbb{R}^p$ mit G Gruppen von Variablen und einer binären Responsevariable $y_i \in \{0, 1\}$ gegeben. Für die erklärenden Variablen können sowohl kategoriale als auch metrische Variablen angenommen werden. Man bezeichnet mit df_g die Freiheitsgrade des g -ten Prädiktors, damit man \mathbf{x}_i als $\mathbf{x}_i = (\mathbf{x}_{i,1}^T, \dots, \mathbf{x}_{i,G}^T)^T$, mit der Gruppe der Variablen $\mathbf{x}_{i,g} \in \mathbb{R}^{df_g}$ über alle Gruppen $g = 1, \dots, G$ schreiben kann. Zum Beispiel werden metrische Variablen als eigene Gruppe angesehen und es gilt $df_g = 1$. Bei kategorialen Variablen sind die Freiheitsgrade über $df_g = n_g - 1$, mit n_g entsprechend der Anzahl an Levels der Faktorvariable, definiert. So ergibt sich für eine Faktorvariable mit vier Levels $df_g = 3$.

Bei der Anwendung auf Paarvergleichsdaten kann diese Berechnung der Freiheitsgrade nicht übertragen werden, da die Gruppierung anders erfolgt. Bei der Schätzung von Präferenzen über Group-LASSO mittels Logistischer Regression werden sowohl metrische als auch binäre und kategoriale Kovariablen gruppiert. Die genaue Berechnung der Freiheitsgrade für Paarvergleichsdaten wird in Unterabschnitt [5.2.3](#) definiert.

Beim linearen logistischen Regressionsmodell wird die bedingte Wahrscheinlichkeit $p_\beta(\mathbf{x}_i) = P_\beta(\mathbf{Y} = 1|\mathbf{x}_i)$ folgendermaßen modelliert

$$\eta_\beta(\mathbf{x}_i) = \log \left(\frac{p_\beta(\mathbf{x}_i)}{1 - p_\beta(\mathbf{x}_i)} \right) = \beta_0 + \sum_{g=1}^G \mathbf{x}_{i,g}^T \boldsymbol{\beta}_g. \quad (5.5)$$

Hierbei entspricht β_0 dem Intercept und $\beta_g \in \mathbb{R}^{df_g}$ dem Parametervektor des g -ten Prädiktors. Die Bezeichnung $\beta \in \mathbb{R}^p$ steht für den gesamten Parametervektor, z.B. $\beta = (\beta_0^T, \beta_1^T, \dots, \beta_k^T)^T$.

Der Group-LASSO Schätzer $\hat{\beta}_\lambda$ bei Logistischer Regression ergibt sich durch die Minimierung der konvexen Funktion

$$S_\lambda(\beta) = -l(\beta) + \lambda \sum_{g=1}^G s(df_g) \|\beta_g\|. \quad (5.6)$$

Die Funktion $s(\cdot)$ wird durch die Freiheitsgrade der einzelnen Gruppen bestimmt und zur Umskalierung der Bestrafung, hinsichtlich der Dimensionalität des Parametervektors β_g , verwendet. Um sicher zu gehen, dass der Bestrafungsterm in der Größenordnung der Anzahl der Parameter df_g liegt, wird in Meier et al. [2008] $s(df_g) = \sqrt{df_g}$ definiert. Selbige Umskalierung wird auch in Yuan and Lin [2006] verwendet.

Die Funktion $l(\cdot)$ wird als Log-Likelihood bei Logistischen Regressionsmodellen, wie in Gleichung (5.2), definiert:

$$l(\beta) = \sum_{i=1}^n [y_i(\beta_0 + \sum_{g=1}^G x_{i,g}^T \beta_g) - \log(1 + \exp(\beta_0 + \sum_{g=1}^G x_{i,g}^T \beta_g))] \quad (5.7)$$

Der Tuning-Parameter $\lambda \geq 0$ legt die Stärke der Bestrafung der Gruppen fest. Wichtig hierbei ist, dass der Intercept nicht bestraft/penalisiert wird. Lemma 1 legt die Bedingungen für die Daten fest, so dass $S_\lambda(\cdot)$ minimiert wird. Für den Fall dass die Designmatrix vollen Rang hat, ist die Lösung eindeutig, ansonsten ist die Lösung eine Menge von Minimierern, deren Elemente das gleiche Minimum bei $S_\lambda(\cdot)$ haben.

Lemma 1. Angenommen es gilt $0 < \sum_{i=1}^n y_i < n$. Für $\lambda > 0$ und $s(d) > 0$ für alle $d \in \mathbb{N}$ wird das Minimum im Optimierungsproblem (5.6) erreicht.

Lemma 1 besagt, dass bei einer binären Responsevariable mindestens ein Wert ungleich Null sein muss. Bei der Berechnung der gruppenweisen l_2 -Norm in Gleichung (5.6) erhält man eine Zwischensumme von der Bestrafung bei LASSO (l_1) und der Bestrafung bei Ridge-Regression (l_2), wie oben bereits erwähnt. Dies unterstützt, dass allgemein entweder $\hat{\beta}_g = \mathbf{0}$ oder $\hat{\beta}_{g,j} \neq 0$ für alle $j \in \{1, \dots, df_g\}$ gilt. Anders ausgedrückt können bei Group-LASSO Verfahren alle Koeffizienten einer Gruppe auf Null gesetzt werden. Eine geometrische Interpretation der speziellen Eigenschaft an wenigen Parametern im Modell ist in Abbildung 5.2 aus Yuan and Lin [2006] gegeben. Der Beweis, dass Lemma 1 gilt, ist im Anhang Abschnitt A.1 gegeben.

Die $(n \times df_g)$ Matrix \mathbf{X}_g entspricht den Spalten der Designmatrix entsprechend dem g -ten Prädiktor. Wenn für jede Block-Matrix \mathbf{X}_g voller Rang angenommen wird, kann eine (blockweise) Orthonormalisierung $\mathbf{X}_g^T \mathbf{X}_g = \mathbf{I}_{df_g}, g = 1, \dots, G$ erreicht werden. Demnach wird vor der Minimierung von $S_\lambda(\cdot)$ nach den Gruppen orthonormalisiert, so dass die neue Kodierung der Dummy-Variablen keinen Einfluss auf den Group-LASSO Schätzer hat. Nachdem die Parameter geschätzt wurden ist eine Rücktransformation der Koeffizienten notwendig, um richtig interpretieren zu können. Aufbauend kann nun der Minimierungsprozess begonnen werden.

Meier et al. [2008] stellen zwei Algorithmen zur Schätzung mit Group-LASSO vor: *Block Co-ordinate Descent* und *Block Co-ordinate Gradient Descent*.

Der *Block Co-ordinate Descent* ist ein Minimierungsalgorithmus, bei dem alle Parametergruppen durchlaufen werden und die Funktion $S_\lambda(\cdot)$ minimiert wird, unter der Bedingung, dass alle Parametergruppen festgehalten werden außer der aktuellen Gruppe. Im ersten Schritt des Algorithmus wird überprüft ob das Minimum am nicht-differenzierbaren Punkt von $\beta_g = \mathbf{0}$ ist. Falls dies nicht der Fall sein sollte, kann ein numerisches Minimierungsverfahren, wie etwa ein Newton-Algorithmus, verwendet werden um β_g zu minimieren. Der Nachteil dieser Methode ist, dass die blockweise Minimierung der aktiven Gruppen numerisch ausgeführt werden muss. Für kleine bis mittlere Anwendungen fällt der genannte Nachteil nicht ins Gewicht. Für umfangreiche Anwendungen scheint der Algorithmus nicht effizient zu sein. Daher wird im Folgenden die Block Co-ordinate Gradient Descent Methode von Tseng and Yun [2009] genauer vorgestellt.

Die zentrale Idee des *Block Co-ordinate Gradient Descent* ist die Kombination einer quadratischen Approximation der Log-Likelihood mit einem zusätzlichen Line-Search-Verfahren. Beim Gradient Descent Verfahren werden die Schritte genommen, welche proportional zur negativen Steigung des Punktes der aktuellen Gleichung sind. Diese Methode wird auf die Blockkoordinaten der Gruppen angewendet und ist die Basis des R-Pakets `grplasso` [Meier, 2009], das in Unterabschnitt 5.2.3 diskutiert wird [Meier et al., 2008].

5.2.2. Datentransformation

Wendet man Group-LASSO als alternatives Schätzverfahren für Bradley-Terry-Luce Modelle an, so ist, wie bei der Anwendung über Generalisierte Lineare Modelle, die spezielle Struktur von Paarvergleichsdaten zu berücksichtigen. Da man bei Paarvergleichsdaten bereits eine natürliche Gruppierung vorliegen hat, ändert sich die allgemeine Definition bezüglich der Gruppenbildung.

Zunächst wird noch einmal kurz auf das Bradley-Terry-Luce Modell für binären Respon-

se unter Berücksichtigung von Kovariablen eingegangen. Die Präferenz berechnet sich für den allgemeinen Fall mehrerer Kovariablen $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ik})$, wie in Gleichung (4.17), über

$$\begin{aligned}\gamma_{ir} &= \beta_{r0} + x_{i1}\beta_{r1} + \dots + x_{ik}\beta_{rk} \\ &= \beta_{r0} + \mathbf{x}_i^T \boldsymbol{\beta}_r.\end{aligned}$$

Wird Item r betrachtet, so verändert sich der Einfluss der Kovariablen x_j durch β_{rj} . Somit erhält man das Modell

$$\log \frac{\mathbb{P}(r \succ s | (r, s), i)}{\mathbb{P}(s \succ r | (r, s), i)} = \sum_{t=1}^m \sum_{j=0}^k x_{ij} x_t^{(r,s)} \beta_{tj} \quad (5.8)$$

mit $x_{i0} = 1$ [Tutz, 2013]. Diese Darstellungen können aus Abschnitt 4.3 übernommen werden, da es sich, wie bei der Betrachtung über Generalisierte Lineare Modelle, ebenso um ein parametrisches Modell handelt. Bei der Schätzung über Group-LASSO wird nur der Fall mit zusätzlichen Kovariablen betrachtet, da die Schätzung eines Modells ohne Kovariablen fehlschlägt. Grund hierfür ist die Eigenschaft, dass beim Intercept keine Penalisierung durchgeführt wird. Die Einbindung von Kovariablen ermöglicht über Modellregularisierung und Variablenselektion sparsame Modelle zu erhalten. Somit können die Ausführungen aus Unterabschnitt 4.3.2 ohne Einschränkungen auf das Group-LASSO Verfahren übertragen werden.

Sollen über verschiedene Betrachtungen von LASSO Kovariablen selektiert und ausgewählt werden, so sind zusätzliche Bestrafungsterme notwendig, die sich für Paarvergleichsdaten wie folgt darstellen lassen:

$$\begin{aligned}\text{LASSO} \quad J(\beta) &= \sum_{t=1}^m \sum_{j=1}^k |\beta_{tj}| \\ \text{Group-LASSO} \quad J(\beta) &= \sum_{j=1}^k \|\boldsymbol{\beta}_{\cdot j}\| \\ \boldsymbol{\beta}_{\cdot j}^T &= (\beta_{1,j}, \dots, \beta_{m-1,j})\end{aligned}$$

Die Penalisierung erfolgt lediglich für die betrachteten Kovariablen $\boldsymbol{\beta}_{\cdot 1}, \dots, \boldsymbol{\beta}_{\cdot k}$ und nicht für die Intercepts, die den Effekt der Paarvergleiche enthalten. Somit werden $(m-1)k$ Parameter bei der Penalisierung berücksichtigt, da Item m aus Gründen der Identifizierbarkeit auf Null gesetzt wird und keine Interaktionen zwischen den Kovariablen angenommen werden [Tutz, 2013].

Da die Parameter noch nicht identifizierbar sind, muss für das Modell folgende Parametrisierung gelten

$$\gamma_{ir} = \beta_{r0} - c + \mathbf{x}_i^T(\boldsymbol{\beta}_r - \boldsymbol{\gamma}) = \tilde{\beta}_{r0} + \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}_r, \quad (5.9)$$

wobei c eine Konstante und $\boldsymbol{\gamma}$ ein fixer Vektor ist. Um eindeutige Parameter $\tilde{\beta}_{r0}, \tilde{\boldsymbol{\beta}}_r$ zu erhalten, wird in diesem Fall $c = \beta_{10}$ und $\boldsymbol{\gamma} = \boldsymbol{\beta}_1$ gewählt [Tutz, 2013].

Bei der Anwendung von Group-LASSO auf Paarvergleichsdaten mit der binären Zielgröße Präferenz sind also die einzelnen Kovariablen gruppiert über die Paarvergleiche zu betrachten und mit entsprechender Penalisierung zu schätzen. Die Einbindung der Kovariablen und die Berücksichtigung des Effekts der einzelnen Paarvergleiche erfolgt, wie in Unterabschnitt 4.3.2, über die dort beschriebene Designmatrix \mathbf{X} . Wie die Modelle über Group-LASSO bei Logistischer Regression über die Statistiksoftware R geschätzt werden können und wie das optimale Modell gefunden werden kann, wird in Unterabschnitt 5.2.3 vorgestellt.

5.2.3. Variablenselektion

Als Basis zur Modellselektion wird das Group-LASSO Verfahren angewendet. Nach der Variablengruppierung wird jede Gruppe über einen Bestrafungsterm penalisiert, und entweder als ganze Gruppe im Modell aufgenommen oder komplett als diese aus dem Modell genommen, indem alle Koeffizienten einer Gruppe auf Null geschätzt werden.

Zur Anwendung des Group-LASSO Verfahrens für logistische Regressionsmodelle wird das R-Paket `grplasso` [Meier, 2009] verwendet. Hierbei müssen die Designmatrix \mathbf{X} , die Response-Variablen, ein Vektor von Indizes, welcher die zu gruppierenden Variablen definiert, und eine Sequenz von λ 's zur Bestrafung definiert werden. Im Folgenden werden die wichtigsten Komponenten der R-Funktion `grplasso` dargestellt:

$$\text{grplasso}(x, y, \text{model} = \text{LogReg}(), \text{lambda}=\text{lambda}, \text{index}=\text{index}) \quad (5.10)$$

Hier entspricht x der Designmatrix unter Einbeziehung von Kovariablen aus Unterabschnitt 4.3.2, y ist die Responsevariable Y und `lambda` ist ein Vektor von Werten, die vom Anwender definiert sind um die Bestrafung im Group-LASSO Verfahren festzulegen. Im R-Paket `grplasso` kann der maximale Wert von λ über die Funktion `lambda_max` bestimmt werden. Diese Funktion berechnet den Wert des Bestrafungsparameters λ , wenn die erste Gruppe von penalisierten Parametern im Modell aufgenommen wird. Ausgehend von dem maximalen Wert für λ kann eine beliebig lange Sequenz bis zum Wert 0 für die einzelnen λ 's gewählt werden. Wird für λ der Wert 0 gewählt, so entsprechen die

geschätzten Koeffizienten den ML-Schätzern. Über den `index` werden die Variablen der Designmatrix Gruppen zugewiesen, so dass die Länge des Vektors der Anzahl der Spalten in \mathbf{X} entspricht. Gibt man beim Index den Wert `NA` an, bedeutet dies, dass die dadurch festgelegte Gruppe nicht penalisiert wird. Da beim Group-LASSO Verfahren, wie bereits erwähnt der Intercept nicht penalisiert wird, ist hierbei als Gruppenindex `NA` anzugeben. Als Ergebnis von `grplasso` erhält man für jeden Wert von λ ein geschätztes Modell, aus denen das Beste über Modell-Selektionskriterien zu bestimmen ist.

Viele Selektionskriterien/Optimalitätskriterien können als gewichtetes Mittel zwischen einem Maß für die Modellanpassung und einem Maß für die Modellkomplexität aufgefasst werden. Die Devianz $D = -2 * \log(\text{Likelihood})$ wird als Maß für die Modellanpassung herangezogen. Der Koeffizient λ bestimmt das relative Gewicht der Modellkomplexität zur Modellanpassung, sodass für ein Selektionskriterium Folgendes gilt:

$$\text{Kriterium} = -2 * \log(\text{Likelihood}) + \lambda * \text{Modellkomplexität} \quad (5.11)$$

Die verschiedenen Selektionsmethoden unterscheiden sich erstens bei der Wahl von λ und zweitens bei der Wahl des Maßes für die Modellkomplexität. Für Letzteres wird oft die Anzahl an Freiheitsgraden des Modells gewählt. Beim logistischen Regressionsmodell entspricht dies der Anzahl geschätzter Parameter [Reineking and Schröder, 2003].

Für logistische Group-LASSO Modelle ist in Yuan and Lin [2006] eine Approximation für die Freiheitsgrade definiert, die im Folgenden mit (YL) für Yuan und Lin abgekürzt wird:

$$df_{\beta}(\lambda) = \sum_{j=1}^k \mathbf{I}(\|\beta_j(\lambda)\| > \mathbf{0}) + \sum_{j=1}^k \frac{\|\beta_j(\lambda)\|}{\|\beta_j(\lambda)^{ML}\|} (m - 2) \quad (5.12)$$

Gleichung (5.12) stellt sicher, dass Gruppen, die im Modell berücksichtigt werden, mindestens einen Freiheitsgrad und höchstens so viele Freiheitsgrade bekommen wie die Gruppe Ausprägungen hat, und Gruppen, welche nicht im Modell sind 0 Freiheitsgrade zugewiesen bekommen. Da sich die β 's aus genannter Gleichung nur auf Gruppen beim Group-LASSO Verfahren beziehen, gilt die Approximation der Freiheitsgrade nicht für unpenalisierte Variablen wie etwa für die Intercepts. Diese erhalten per Definition automatisch einen Freiheitsgrad, weil diese nicht geschrumpft werden.

Alternativ kann die Definition der Freiheitsgrade einfacher LASSO Modelle übernommen werden. Hierbei wird die Anzahl geschätzter Parameter ungleich Null mit den Freiheitsgraden gleichgesetzt. Man spricht von einem sogenannten Active Set (AS). Diese zwei unterschiedlichen Betrachtungen für Freiheitsgrade können beide über die Simulationsmaske, siehe Anhang A.2.2, betrachtet werden.

Bei der Wahl von λ verwendet das AIC, wie in Gleichung (4.8), $\lambda = 2$ während das BIC, siehe Gleichung (4.9), $\lambda = \log(n)$ mit n Anzahl der Beobachtungen verwendet. Neben dem AIC und dem BIC gibt es weitere Selektionskriterien wie etwa das korrigierte Akaike Informationskriterium (AICc) oder das Residuen Informationskriterium (RIC), die speziell bei kleineren Stichproben und hoch korrelierten Daten zur Anwendung kommen. Aber auch LASSO und Group-LASSO selbst sind Modell-Selektionskriterien, wie in Abbildung 5.1 dargestellt ist. Bei diesen Verfahren werden anstatt der Anzahl der Modellparameter die Parameterwerte als Komplexitätsmaß angesehen [Reineking and Schröder, 2003].

In den folgenden Simulationsbeispielen in Abschnitt 5.3 werden ausgewählte simulierte Paarvergleichsmodelle unter Einbeziehung von Kovariablen über das Group-LASSO Verfahren geschätzt. Zur Bestimmung des besten Group-LASSO Modells werden die Selektionskriterien AIC und BIC angewendet. Das gewählte Modell gibt Aufschluss darüber, welche und wie viele Kovariablen im Modell berücksichtigt werden.

5.3. Simulationsergebnisse

In diesem Abschnitt wird anhand von Simulationsstudien die Vorhersagegenauigkeit von Group-LASSO Modellen bei Paarvergleichsdaten untersucht. Der zugehörige R-Code der zugrundeliegenden Simulationsmaske ist im Anhang, Abschnitt A.2.2, beigefügt. Anhand folgender Beispiele wird überprüft, ob sich das Group-LASSO Verfahren als alternatives Schätzverfahren für Bradley-Terry-Luce Modelle eignet und inwieweit die Simulationen gute Ergebnisse liefern.

Simulation allgemein

Bei der Simulation von Paarvergleichsdaten über Group-LASSO wird lediglich der Fall mit Kovariablen betrachtet, um über Modellregularisierung und Variablenselektion sparsame Modelle zu erhalten, die nur einflussreiche Variablen beinhalten. Dabei können sowohl binäre als auch metrische Kovariablen einbezogen werden.

Zur Simulation der metrischen Kovariablen wird aus einer Standardnormalverteilung gezogen und für die binären Kovariablen aus einer Gleichverteilung auf $\{0, 1\}$. Der entsprechende Effekt wird aus einer Normalverteilung mit $N(0, \sigma^2 I)$ gezogen oder gleich Null gesetzt, falls es sich um eine uninformative Variable handeln soll.

Um den Unterschied zwischen binären/metrischen Kovariablen und simulierten Kovariablenschätzern aufzuzeigen, sei noch einmal auf Gleichung (4.17) verwiesen, für die bei

simulierten Daten die entsprechenden Verteilungsannahmen gelten.

$$\gamma_{ir} = \beta_{r0} + \underset{\substack{\uparrow \\ N(0,1)}}{\mathbf{x}_{i,metr}^T} \underset{\substack{\uparrow \\ N(0,\sigma^2 I)}}{\boldsymbol{\beta}_{r,metr}^{inf}} + \underset{\substack{\uparrow \\ N(0,1)}}{\mathbf{x}_{i,metr}^T} \underset{=0}{\boldsymbol{\beta}_{r,metr}^{uninf}} + \underset{\substack{\uparrow \\ U\{0,1\}}}{\mathbf{x}_{i,bin}^T} \underset{\substack{\uparrow \\ N(0,\sigma^2 I)}}{\boldsymbol{\beta}_{r,bin}^{inf}} + \underset{\substack{\uparrow \\ U\{0,1\}}}{\mathbf{x}_{i,bin}^T} \underset{=0}{\boldsymbol{\beta}_{r,bin}^{uninf}}$$

Über die Positionen `true.pos` bzw. `false.pos` im Simulationscode können die Anteile korrekt informativ geschätzter bzw. fälschlicherweise informativ geschätzter Variablen berechnet werden. Diese Anteile werden für die einzelnen Simulationsszenarien in Tabellen dargestellt.

Die Informationen der simulierten normalverteilten Schätzer gehen über den linearen Prädiktor in das Logit-Modell mit ein. Zur Erzeugung eines Responsevektors \mathbf{y} wird zufällig aus der Binomialverteilung mit Wahrscheinlichkeit π_i aus dem Logit-Modell gezogen. Nun kann unter Angabe der gewünschten Gruppierung über den Index (NA für den Intercept) und einer Sequenz von λ 's zur Bestrafung des Group-LASSO Modells für Paarvergleichsdaten simuliert werden. Wie in Unterabschnitt 5.2.3 beschrieben, wird über die Modellselektionskriterien AIC bzw. BIC das beste Modell bestimmt.

Simulation konkret

Zur Überprüfung der Vorhersagegenauigkeit von Group-LASSO Modellen auf Paarvergleichsdaten werden verschiedene Simulationssettings untersucht.

Zunächst unterscheiden sich drei Simulationsstudien lediglich durch die Anzahl an Items, der Anteil an informativen Kovariablen ist konstant. Das erste Simulationssetting vergleicht $m = 6$ Items, sodass $I = 15$ Paarvergleiche resultieren. Im nächsten Simulationssetting erhöht sich die Anzahl zu vergleichender Items auf zehn, sodass 45 Paarvergleiche durchgeführt werden. Beim dritten Setting werden $m = 15$ Items über 105 Paarvergleiche miteinander verglichen. Für alle Settings werden insgesamt zehn Kovariablen und ein Intercept, für den Effekt der Paarvergleiche, im Modell berücksichtigt. Die Kovariablen teilen sich in jeweils fünf binäre und fünf metrische Kovariablen auf, wobei jeweils drei Variablen als uninformativ angenommen werden. Die Simulationen erfolgen an 100 fiktiven Personen und werden 100-mal wiederholt. Ziel dieser Untersuchung ist herauszufinden, ob eine Erhöhung der Items zu einer besseren Vorhersagegenauigkeit von Group-LASSO Modellen führt.

Bei weiteren vier Simulationsstudien variiert der Anteil informativer Kovariablen und die Anzahl an Items bleibt konstant. Die Simulationen betrachten Settings mit insgesamt zehn Kovariablen bei zehn zu vergleichenden Items, wovon entweder zwei, vier, sechs oder acht der Kovariablen als uninformativ vorausgesetzt werden. Simuliert wird für 100 Personen

und es erfolgen 100 Durchläufe, mit dem Ziel, Verbesserungen bzw. Verschlechterungen bei der Vorhersagegenauigkeit zwischen den einzelnen Modellen zu erkennen.

Das letzte hier vorgestellte Simulationssetting untersucht die Präferenzentscheidungen bei 45 Paarvergleichen resultierend aus zehn systematisch miteinander verglichenen Items. Hierbei erhöht sich die Anzahl an Kovariablen auf 20, wovon zwölf keine Informationen enthalten. Aufgrund der großen Anzahl betrachteter Kovariablen wird die Simulation an 250 Personen durchgeführt und 100 mal wiederholt. Das Interesse bei diesem Setting liegt auf der Untersuchung der Vorhersagegenauigkeit von Group-LASSO Modellen bei einer erhöhten Anzahl von Kovariablen.

Die Ergebnisse der vorgestellten Simulationssettings werden im Folgenden vorgestellt. Weitere Simulationsergebnisse sind im Anhang Abschnitt A.3 gegeben. Zudem befinden sich weitere Ergebnisse auf der beigefügten CD.

Ergebnisse

Zunächst wird das Simulationsergebnis des Settings mit $m = 6$ Items und zehn Kovariablen ausführlich vorgestellt. Danach erfolgt die Interpretation der weiteren Ergebnisse und Vergleiche zwischen Simulationssettings bezüglich Vorhersagegenauigkeit von Group-LASSO Modellen.

Die Ergebnisse des Simulationssettings mit sechs zu vergleichenden Items sind in Tabelle 5.1 dargestellt. Bei optimaler Anpassung über Group-LASSO wäre der Anteil korrekt informativ geschätzter Kovariablen 100% und der Anteil fälschlicherweise informativ geschätzter Kovariablen 0%. Die Berechnung der Freiheitsgrade unterscheidet, wie bereits in Unterabschnitt 5.2.3 erläutert, zwischen der Approximation bei logistischen Group-LASSO Modellen (YL), siehe Gleichung (5.12), und dem Active Set (AS), welches bei einfachen LASSO Modellen Anwendung findet.

In der ersten Spalte sind die unterschiedlichen Standardabweichungen (Std.Abw.) aufgelistet, mit denen aus der Normalverteilung für die simulierten informativen Schätzer gezogen wird. Zieht man für dieses Beispiel aus der Standardnormalverteilung, Std.Abw. = 1, und berechnet die Freiheitsgrade über die Approximation nach Yuan and Lin [2006] (YL), werden alle informativen Kovariablen richtig erkannt, allerdings liegt die Rate der fälschlicherweise informativ geschätzten Kovariablen bei 23%. Erfolgt die Berechnung der Freiheitsgrade über das Active Set (AS), so werden 99.75% der informativen Kovariablen korrekt als informativ erkannt und nur 4% der uninformativen Kovariablen fälschlicherweise als informativ geschätzt.

Die Anteile berechnen sich bei zehn Kovariablen, mit sechs uninformativen und vier in-

formativen Kovariablen bei 100 Simulationsdurchläufen wie folgt. Für den Anteil korrekt informativ geschätzter Kovariablen sind bei 100 Simulationen jeweils die vier als informativ vorausgesetzten Kovariablen zu erkennen, also müssten im Idealfall 400 Variablen informativ geschätzt werden. Bei einer Rate von 100% werden alle 400 Variablen erkannt, bei einer Rate von 99.75% erfolgt in 398 Fällen die richtige Zuordnung. Bei den sechs uninformativen Kovariablen erhält man, falls alle richtig erkannt werden, einen Anteil fälschlicherweise informativ geschätzter Kovariablen von 0% und falls alle als informativ eingestuft werden von 100%.

Std.Abw.	Group-LASSO (YL)		Group-LASSO (AS)	
	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)
0.20	44.00	0.17	3.50	0.00
0.25	71.75	1.33	19.00	0.00
0.3	86.00	2.67	46.00	0.83
0.4	96.75	5.33	81.50	1.50
0.5	99.50	10.17	93.50	2.50
1	100.00	23.00	99.75	4.00

Tabelle 5.1.: Ergebnisse des Simulationssettings: 6 Items, 10 Kovariablen, davon 6 ohne Information, 100 Personen, 100 Simulationen.

Reduziert man die Standardabweichung gegenüber der Standardnormalverteilung auf 0.5 so werden, über die Freiheitsgrade der Approximation für logistische Group-LASSO Modelle (YL), noch 99.5% informativer Kovariablen richtig spezifiziert, mit einem verringerten Anteil fälschlicherweise informativ geschätzter Kovariablen von 10.17%. Auch bei einer Standardabweichung von 0.4 ist die Anpassung mit 96.75% noch sehr gut bei nur 5.33% inkorrekt informativ geschätzter Kovariablen. Eine weitere Verringerung der Standardabweichung führt zu schlechteren Anpassungen, obwohl der Anteil fälschlicherweise informativ geschätzter Kovariablen gegen Null geht.

Werden die Freiheitsgrade über das Active Set berechnet, liegt der Anteil korrekt spezifizierter informativer Kovariablen bei einer Standardabweichung von 0.5 nur mehr bei 93.5%, bei einer Standardabweichung von 0.4 sind es noch 81.5% und ab einer Standardabweichung kleiner als 0.3 liegt die korrekte Anpassung bereits unter 50%. Der Anteil fälschlicherweise informativ geschätzter Kovariablen reduziert sich kontinuierlich gegen Null.

Zur graphischen Darstellung der einzelnen Ergebnisse aus Tabelle 5.1 bietet sich eine Art Boxplot an, bei der statt dem Median der Mittelwert abgetragen ist und die Ränder der Box durch \pm Standardabweichung beschränkt sind. Zur einheitlichen Darstellung wird der Anteil korrekt informativ geschätzter Kovariablen in den Anteil fälschlicherweise nicht informativ geschätzter Kovariablen umgewandelt. Zum Beispiel erhält man im Fall der Standardnormalverteilung, bei Freiheitsgradberechnung (AS), statt 99.75% korrekt spezifizierter Schätzer, 0.25% fälschlicherweise nicht korrekt spezifizierte Schätzer. Dadurch ist eine bessere Vergleichsmöglichkeit mit dem Anteil fälschlicherweise informativ geschätzter Kovariablen gegeben.

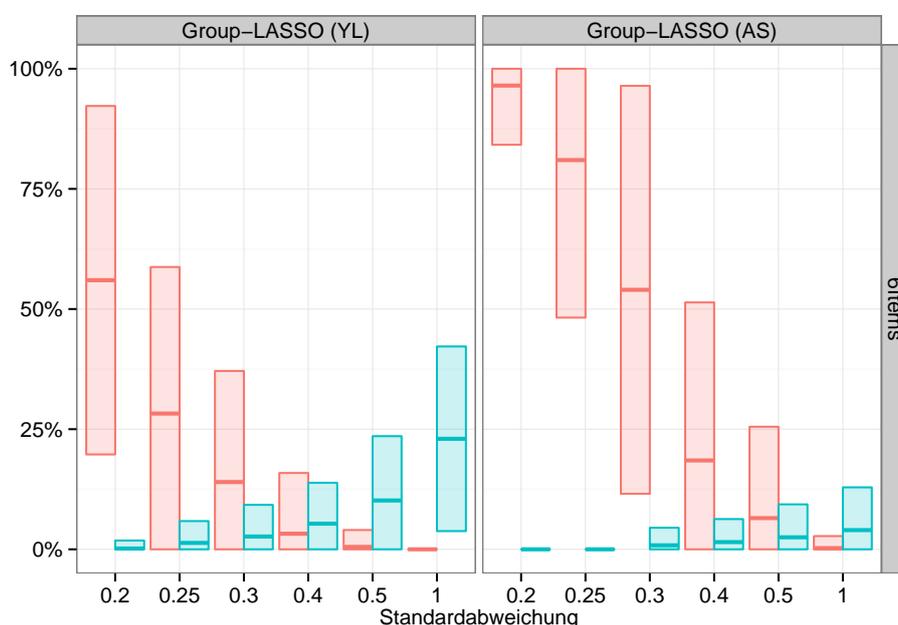


Abbildung 5.3.: Boxen der fälschlicherweise informativ bzw. nicht informativ geschätzten Parameter des Simulationssettings mit 6 Items und 10 Kovariablen, 6 ohne Information.

Die roten Boxen stellen den Anteil fälschlicherweise nicht informativ geschätzter Kovariablen dar, die blauen Boxen beschreiben den Anteil fälschlicherweise informativ geschätzter Kovariablen. In Abbildung 5.3 ist zu erkennen, dass bei größer werdender Standardabweichung der Anteil korrekt informativ geschätzter Kovariablen steigt, da der Anteil fälschlicherweise nicht informativ geschätzter Kovariablen sinkt (rote Boxen). Hingegen steigt der Anteil fälschlicherweise informativ geschätzter Kovariablen (blaue Boxen), wie ebenfalls aus Tabelle 5.1 ersichtlich ist. Es ist jedoch schwierig zu beurteilen, welche der beiden Freiheitsgradberechnungen bei sechs zu vergleichenden Items zu bevorzugen ist.

Zur Bestimmung des besten Modells über Modell-Selektionskriterien wird als Beispiel die

Anpassung über Group-LASSO für Standardabweichung 1 und Freiheitsgradberechnung (YL) verwendet. Zur Reproduktion der Darstellungen ist in R `set.seed(1)` festzulegen, d.h. es wird das erste der $i = 100$ simulierten Schätzungen betrachtet. Für λ wird in der Simulation eine Sequenz von 40 Werten, absteigend vom maximalen Wert 94.999 bis 0.001, angenommen. Entsprechend der verschiedenen Werte für λ und der damit verbundenen unterschiedlichen Bestrafungen werden 40 Modelle angepasst.

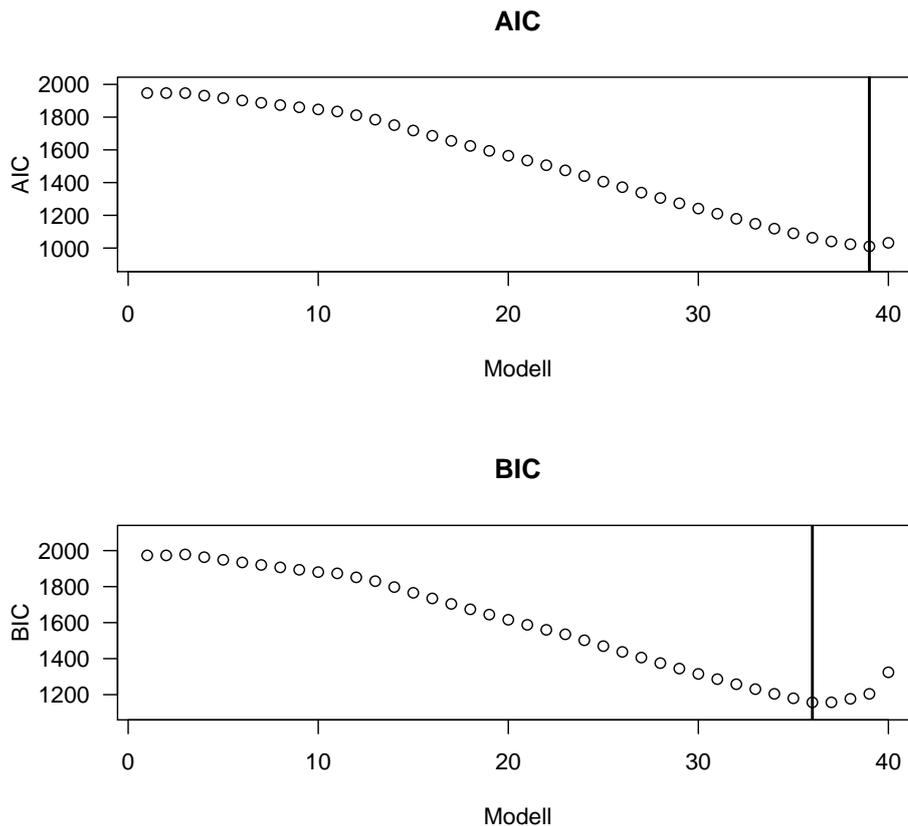


Abbildung 5.4.: Oben: AIC in Abhängigkeit der Höhe der Bestrafung Lambda für Modell mit 6 Items; Unten: BIC in Abhängigkeit der Höhe der Bestrafung Lambda für Modell mit 6 Items.

Das beste Modell kann über die Kriterien AIC, siehe Gleichung (4.8), oder BIC, siehe Gleichung (4.9), selektiert werden, wie in Abbildung 5.4 dargestellt ist. Bei beiden Kriterien wird das Modell mit dem jeweils kleinsten Wert gewählt.

Basierend auf Abbildung 5.4 wird beim AIC das 39. Modell gewählt mit $\lambda = 2.43$ und dem Minimum $AIC = 1010.203$. Betrachtet man das optimale Modell gemäß AIC, zeigt sich, dass nur eine der sechs uninformativen Kovariablen als diese erkannt wird und die entsprechenden Koeffizienten mit 0 geschätzt werden. Die fünf weiteren uninformativen Kovariablen werden als informativ geschätzt und haben Schätzer ungleich Null. Ein mögli-

cher Grund, dass nicht alle uninformativen Kovariablen erkannt werden, ist die schwache Bestrafung beim AIC. Um Modelle mit weniger Parametern zu erhalten, müssen diese stärker bestraft werden, wie es beim Selektionskriterium BIC der Fall ist.

Zur Wahl eines Modells über das Selektionskriterium BIC wird für jeden Wert von λ ein logistisches Modell angepasst und das BIC berechnet. Die Ergebnisse sind in Tabelle 5.2 aufgelistet.

Modell	Lambda	Anzahl Koeffizienten ungleich Null	BIC	ausgeschlossene Variablen
40	0.00	55	1324.260	Keine
39	2.43	50	1204.368	binär3
38	4.87	50	1177.005	Keine
37	7.30	35	1157.360	binär5, metr4, metr5
36	9.74	25	1157.334	binär4, metr3
35	12.18	25	1179.863	Keine
34	14.61	25	1204.518	Keine
33	17.05	25	1230.682	Keine
32	19.48	25	1258.006	Keine
31	21.92	25	1286.254	Keine
30	24.35	25	1315.271	Keine
29	26.79	25	1344.960	Keine
28	29.23	25	1375.251	Keine
27	31.66	25	1406.122	Keine
26	34.10	25	1437.552	Keine
25	36.53	25	1469.536	Keine
24	38.97	25	1502.095	Keine
23	41.41	25	1535.239	Keine
22	43.84	20	1560.088	binär2
21	46.28	20	1587.791	Keine
20	48.71	20	1616.006	Keine
19	51.15	20	1644.754	Keine
18	53.59	20	1674.065	Keine
17	56.02	20	1703.969	Keine
16	58.46	20	1734.503	Keine
15	60.89	20	1765.707	Keine
14	63.33	20	1797.633	Keine
13	65.76	20	1830.329	Keine
12	68.20	15	1851.675	metr2
11	70.64	15	1873.823	Keine
10	73.07	10	1880.791	metr1
9	75.51	10	1893.641	Keine
8	77.94	10	1906.858	Keine
7	80.38	10	1920.452	Keine
6	82.82	10	1934.421	Keine
5	85.25	10	1948.780	Keine
4	87.69	10	1963.535	Keine
3	90.12	10	1978.700	Keine
2	92.56	5	1973.582	binär1
1	94.99	5	1973.582	Keine

Tabelle 5.2.: BIC vs. Lambda bei Modellselektion. In Anlehnung an Condon [2012].

Wie aus Abbildung 5.4 und aus Tabelle 5.2 zu entnehmen ist, wird das 36. Modell gewählt, bei dem $\lambda = 9.74$ und $\text{BIC} = 1157.334$ sind und das BIC sein Minimum hat.

Insgesamt werden über das Paarvergleichsmodell mit $m = 6$ zu vergleichenden Items, einem Intercept und zehn Kovariablen 55 Koeffizienten geschätzt. Diese Anzahl berechnet sich bei der Einbindung von elf Variablen über $11(m - 1) = 55$. Die Dimension der Designmatrix beträgt somit bei $n = 100$ Personen (1500×55).

In Tabelle 5.2 werden mit binär1 und binär2 die informativen binären Kovariablen und mit binär3, binär4 und binär5 die uninformativen binären Kovariablen bezeichnet. Analog stehen metr1 und metr2 für die informativen metrischen Kovariablen und metr3, metr4 und metr5 für die uninformativen metrischen Kovariablen. Das 36. Modell enthält 25 Koeffizienten für den Intercept, binär1, binär2, metr1 und metr2. Somit werden in diesem Fall alle uninformativen Kovariablen erkannt und aus dem Modell ausgeschlossen, indem die Koeffizientenschätzer den Wert Null annehmen.

Der Aufbau von Tabelle 5.2 ist so zu verstehen, dass das 1. geschätzte Modell über Group-LASSO mit dem maximalen Wert von λ penalisiert wird, das 2. Modell mit einem etwas kleineren λ , usw. bis schließlich im letzten Modell, hier dem 40. Modell, mit $\lambda = 0$ keine Penalisierung mehr erfolgt und man die ML-Schätzer erhält. Ausgehend vom unpenalisierten Group-LASSO Modell, das alle Kovariablen enthält, fallen bei einem größer werdenden Bestrafungsparameter zunehmend uninformativ Kovariablen aus dem Modell. Dabei ist zu beachten, dass alle Kovariablen die einmal aus dem Modell ausgeschlossen wurden dies auch bleiben.

Die graphische Darstellung der Simulationsergebnisse mit $m = 6$ Items und zehn Kovariablen, wovon sechs als uninformativ angenommen werden, ist in Abbildung 5.5 gegeben. Abbildung 5.5 zeigt die Pfade der Koeffizienten, ohne Intercepts, der β -Parameter, die sich durch kleiner werdenden Tuning-Parameter λ ergeben. Auf der x-Achse wird eine Transformation von λ verwendet, da diese bessere Skalierungseigenschaften besitzt. Somit wird anstelle der λ -Werte die Skalierung $\|\beta\|/\max(\|\beta\|)$ auf der x-Achse abgetragen, mit $\max(\|\beta\|)$ entsprechend der l_2 -Norm der maximal erhältlichen Schätzern, den ML-Schätzern. Am rechten Rand von Abbildung 5.5 sieht man die Schätzer für $\lambda = 0$ ($\|\beta\|/\max(\|\beta\|) = 1$), die den ML-Schätzern des Group-LASSO Modells entsprechen. Auf der linken Seite sind alle Parameter über den maximalen Bestrafungsparameter λ auf Null geschrumpft. Daher zeigt Abbildung 5.5, wie sich die Schätzer über kleiner werdende Regularisierung entwickeln. Am rechten Ende wird keine Regularisierung angewendet, am linken Ende ist die Regularisierung so stark, dass alle β -Parameter auf Null gesetzt werden. Die vertikal hervorgehobene Linie zeigt den Tuning-Parameter, der nach dem

Selektionskriterium BIC zum besten Modell führt. In diesem Fall wird über das BIC, wie bereits erwähnt, das 36. Modell gewählt. Entsprechend der Darstellung in Tabelle 5.2, werden alle informativen Kovariablen erkannt und mit Koeffizienten ungleich Null geschätzt, wohingegen die uninformativen Kovariablen Schätzer gleich Null haben. Somit ist die Anpassung für dieses Beispiel optimal [Tutz and Schauburger, 2012].

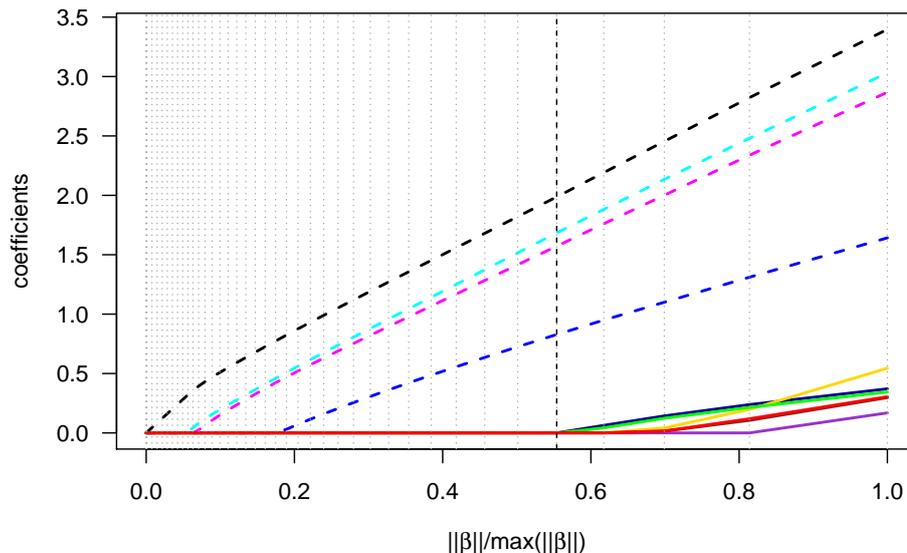


Abbildung 5.5.: Koeffizientenpfade von Group-LASSO für Simulation mit 6 Items und 10 Kovariablen; gestrichelte Linien: informative Kovariablen; durchgezogene Linien: uninformativen Kovariablen.

Vergleicht man die Selektionskriterien AIC und BIC miteinander, so zeigt sich, dass über das BIC eindeutig sparsamere Modelle mit weniger Parametern gewählt werden, als über das AIC. Da an weiteren Beispielen Gleiches beobachtet werden kann, eignet sich bei der Wahl des optimalen Group-LASSO Modells das BIC besser.

Die ausführliche Darstellung der Wahl des besten Modells über Group-LASSO wurde beispielhaft für das Simulationssetting mit sechs Items und zehn Kovariablen vorgestellt, wobei die Koeffizientenschätzer aus der Standardnormalverteilung simuliert und die Freiheitsgrade über die Approximation für logistische Group-LASSO Modelle nach Yuan and Lin [2006] berechnet wurden. Die Anwendung kann auf beliebige Simulationsszenarien oder reale Datensituationen erfolgen.

Zur Untersuchung, ob eine Erhöhung der Items zu einer besseren Vorhersagegenauigkeit von Group-LASSO Modellen führt, werden nun die Simulationsergebnisse der Settings

mit $m = 10$ Items und $m = 15$ Items vorgestellt und mit den Ergebnissen aus Tabelle 5.1 mit $m = 6$ Items verglichen.

Die Ergebnisse der Simulationen mit $m = 10$ Items im Vergleich sind in Tabelle 5.3 dargestellt. Die graphische Darstellung ist in Abbildung 5.6 (Mitte) gegeben.

Std.Abw.	Group-LASSO (YL)		Group-LASSO (AS)	
	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)
0.2	94.25	0.33	26.00	0.00
0.25	99.00	0.83	76.00	0.00
0.3	99.50	2.17	97.00	0.33
0.4	100.00	7.17	99.50	0.33
0.5	100.00	9.50	100.00	0.17
1	100.00	27.67	100.00	3.17

Tabelle 5.3.: Ergebnisse des Simulationssettings: 10 Items, 10 Kovariablen, davon 6 ohne Information, 100 Personen, 100 Simulationen.

Im Gegensatz zum Simulationssetting mit $m = 6$ Items und den daraus resultierenden 15 Paarvergleichen ist die Anpassung bei $m = 10$ Items über Group-LASSO deutlich besser. Beginnend mit der Simulation der informativen Kovariablenschätzer aus der Standardnormalverteilung und der Freiheitsgradberechnung über das Active Set werden alle informativen Kovariablen erkannt und nur 3.17% der uninformativen Kovariablen fälschlicherweise als informativ geschätzt. In diesem Beispiel stellt die Reduzierung der Varianz sogar eine Verbesserung dar. Bis zu einer Standardabweichung von 0.3 werden 97% der informativen Kovariablen erkannt und der Anteil fälschlicherweise informativ geschätzter Kovariablen liegt mit 0.33% nahe bei der optimalen Anpassung. Für Standardabweichungen kleiner als 0.3 kann beim Anteil korrekt informativ geschätzter Kovariablen nicht mehr von guter Anpassung gesprochen werden.

Wird hingegen bei der Berechnung der Freiheitsgrade die Approximation für logistische Group-LASSO Modelle (YL) angewendet, so liegt der Anteil korrekt informativ geschätzter Kovariablen sogar bei einer sehr kleinen Standardabweichung von 0.2 noch bei 94.25%. Allerdings liegt der Anteil der fälschlicherweise als informativ eingestufteten Kovariablen, im Fall der Standardnormalverteilung, bei nahezu 30%. Mit abnehmender Varianz sinkt dieser erheblich in Richtung 0%.

Beim Vergleich der Anpassungen der Group-LASSO Modelle mit unterschiedlichen Berechnungen an Freiheitsgraden zeigt sich, dass bei großer Standardabweichung die Berechnung über das Active Set (AS) bessere Ergebnisse liefert und bei kleineren Standardabweichungen die Berechnung über die Approximation für logistische Group-LASSO Modelle nach Yuan and Lin [2006] (YL) von Vorteil ist.

In Tabelle 5.4 sind die Ergebnisse des Simulationssettings mit 15 Items dargestellt.

Std.Abw.	Group-LASSO (YL)		Group-LASSO (AS)	
	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)
0.20	99.75	0.33	81.00	0.00
0.25	100.00	1.17	99.00	0.00
0.3	100.00	3.00	100.00	0.00
0.4	100.00	8.17	100.00	0.00
0.5	100.00	12.33	100.00	0.17
1	100.00	36.00	100.00	1.83

Tabelle 5.4.: Ergebnisse des Simulationssettings: 15 Items, 10 Kovariablen, davon 6 ohne Information, 100 Personen, 100 Simulationen.

Auf den ersten Blick ist sofort zu erkennen, dass die Vorhersagegenauigkeit bei 15 Items steigt bzw. nahezu perfekt ist. Für die Standardabweichungen von 0.3 und 0.4 bei der Freiheitsgradberechnung über das Active Set erhält man die perfekte Anpassung. Erfolgt die Berechnung der Freiheitsgrade über die Approximation (YL) aus Gleichung (5.12) wird der Anteil korrekt informativ geschätzter Kovariablen mit 100% richtig spezifiziert. Die beste Vorhersage wird bei einer Standardabweichung von 0.25 getroffen. Dabei werden lediglich 1.17% der uninformativen Variablen fälschlicherweise als informativ eingestuft. Für dieses Beispiel zeigt sich, dass kleine Varianzen deutliche Verbesserungen bei der Schätzgenauigkeit gegenüber der Standardnormalverteilung aufweisen. Dies lässt vermuten, dass bei einer Erhöhung der Items die Varianz verringert werden sollte um bessere Vorhersagen zu erhalten.

Der graphische Vergleich der Simulationsergebnisse mit variierender Anzahl an Items und konstanter Anzahl an Kovariablen ist in Abbildung 5.6 gegeben. Betrachtet man zunächst die Freiheitsgradberechnung (YL) zeigt sich, dass der Anteil fälschlicherweise nicht informativ geschätzter Kovariablen mit steigender Anzahl an Items sinkt (rote Boxen). D.h. der Anteil der korrekt informativ geschätzten Kovariablen nimmt bei steigender Anzahl an

Items zu, so dass man bessere Vorhersagen erhält. Hingegen steigt der Anteil fälschlicherweise informativ geschätzter Kovariablen mit steigender Standardabweichung bei einer Erhöhung der Items (blaue Boxen), so dass die besten Vorhersagen bei einer erhöhten Anzahl an Items bei kleiner werdender Standardabweichung erreicht werden.

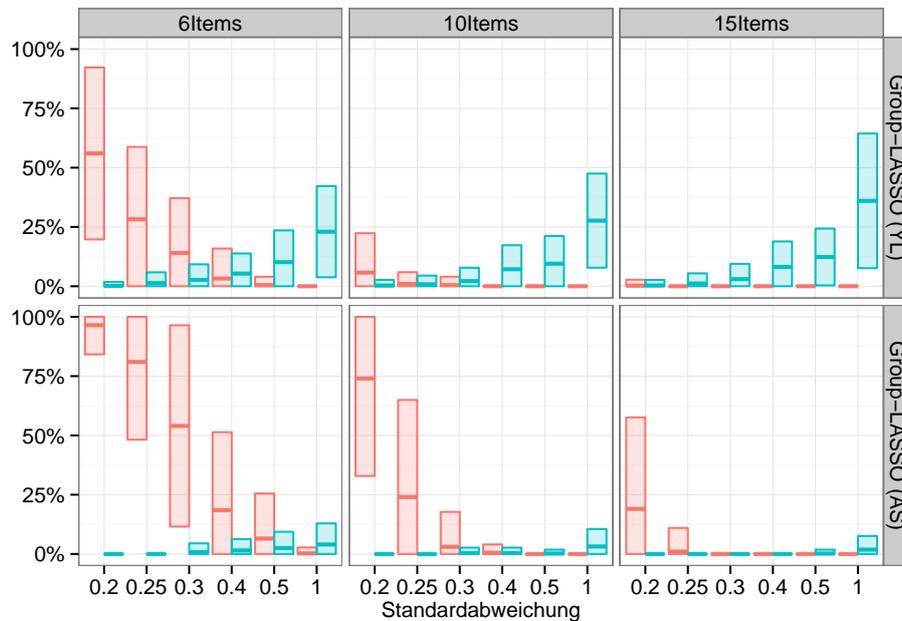


Abbildung 5.6.: Vergleich der Simulationsergebnisse mit unterschiedlicher Anzahl an Items bei konstanten 10 Kovariablen, wovon 6 uninformativ sind.

Der Vergleich der Group-LASSO Modelle mit Freiheitsgradberechnung über das Active Set (AS) zeigt bei einer Erhöhung der Items deutlich bessere Anpassungen. Bei 15 Items ist bei den Standardabweichungen 0.3 und 0.4 die perfekte Anpassung zu erkennen. Selbst beim Anteil fälschlicherweise informativ geschätzter Kovariablen (blaue Boxen) ist eine Verbesserung gegenüber der Anpassung mit sechs bzw. zehn Items auszumachen.

Somit konnte gezeigt werden, dass eine Erhöhung der zu vergleichenden Items die Vorhersagegenauigkeit der Group-LASSO Modelle verbessert. Im Folgenden wird untersucht, wie sich Veränderungen beim Anteil informativer Kovariablen bei konstanter Anzahl an Items auf die Anpassung von Group-LASSO Modellen auswirken. Dafür werden zunächst kurz die Simulationsergebnisse der einzelnen Settings vorgestellt und miteinander verglichen.

Das Simulationssetting variiert lediglich in der Anzahl informativer/uninformativer Kovariablen, ansonsten werden bei allen Simulationen zehn Items von 100 Personen in 45 Paarvergleichen miteinander verglichen. Zunächst werden zwei der zehn Kovariablen als

uninformativ vorausgesetzt, danach vier, sechs (Simulationsergebnisse siehe Tabelle 5.3) und schließlich acht. Untersucht wird, ob eine Veränderung bei der Anzahl informativer Kovariablen eine Auswirkung auf die Vorhersagegenauigkeit von Group-LASSO Modellen hat und welches Setting die besten Ergebnisse liefert.

Die Simulationsergebnisse des Settings mit zwei uninformativen Kovariablen sind in Tabelle 5.5 aufgeführt.

Std.Abw.	Group-LASSO (YL)		Group-LASSO (AS)	
	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)
0.2	95.50	0.50	10.25	10.00
0.25	99.50	5.50	80.50	31.00
0.3	99.625	10.00	97.75	42.50
0.4	100.00	33.00	100.00	70.50
0.5	100.00	45.50	100.00	78.00
1	100.00	89.00	100.00	100.00

Tabelle 5.5.: Ergebnisse des Simulationssettings: 10 Items, 10 Kovariablen, davon 2 ohne Information, 100 Personen, 100 Simulationen.

Die Simulationsergebnisse zeigen, dass allgemein eine schlechte Vorhersage über Group-LASSO Modelle bei einer geringen Anzahl uninformativer Kovariablen erfolgt. Die Group-LASSO Modelle mit Freiheitsgradberechnung (YL) zeigen gegenüber Modellen mit Freiheitsgradberechnung (AS) eine leicht bessere Anpassung. Der Anteil korrekt informativ geschätzter Kovariablen ist über alle Standardabweichungen mit einem Minimum von 95.50% sehr gut. Der Anteil fälschlicherweise informativ geschätzter Kovariablen steigt mit zunehmender Standardabweichung sehr stark an, so dass bei der Simulation mit Standardnormalverteilung 89% uninformativer Kovariablen nicht erkannt wird. Bei der Berechnung der Freiheitsgrade über das Active Set werden bei der Simulation aus der Standardnormalverteilung alle Kovariablen als Informativ geschätzt. Eine kleinere Varianz kann dieser Fehleinstufung leicht entgegenwirken, dennoch ist diese mit einem Anteil von 70% sehr hoch. Dies lässt vermuten, dass bei Schätzungen mit einer geringen Anzahl uninformativer Kovariablen besser die Approximation für logistische Group-LASSO Modelle verwendet werden sollte, um bessere Vorhersagen zu erhalten.

Die Ergebnisse des Simulationssettings mit vier Kovariablen ohne Information können in Tabelle 5.6 nachgesehen werden. Auch hier ist die Fehlerquote beim Anteil fälschlicher-

weise informativ geschätzter Kovariablen mit Abweichungen über 50% bei der Schätzung über die Standardnormalverteilung sehr hoch. Der Anteil korrekt informativ geschätzter Kovariablen ist ab einer Standardabweichung von 0.3 über beide Freiheitsgradberechnungen sehr gut, bei einer kleineren Standardabweichung ist insbesondere die Anpassung der Modelle mit Freiheitsgradberechnung (AS) schlecht.

Std.Abw.	Group-LASSO (YL)		Group-LASSO (AS)	
	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)
0.20	93.83	1.25	16.33	0.00
0.25	98.83	3.75	77.50	1.00
0.3	99.83	6.00	97.00	3.25
0.4	100.00	13.25	99.83	3.50
0.5	100.00	22.25	100.00	4.75
1	100.00	64.00	100.00	52.75

Tabelle 5.6.: Ergebnisse des Simulationssettings: 10 Items, 10 Kovariablen, davon 4 ohne Information, 100 Personen, 100 Simulationen.

Erhöht sich die Anzahl uninformativer Kovariablen auf acht, siehe Tabelle 5.7, erhält man gute Vorhersagen bei Standardabweichungen von 1, 0.5 oder 0.4.

Std.Abw.	Group-LASSO (YL)		Group-LASSO (AS)	
	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)
0.2	81.50	0.00	31.50	0.00
0.25	96.50	0.125	66.00	0.00
0.3	99.5	0.375	90.00	0.00
0.4	100.00	1.00	99.00	0.00
0.5	100.00	2.75	100.00	0.00
1	100.00	8.375	100.00	0.125

Tabelle 5.7.: Ergebnisse des Simulationssettings: 10 Items, 10 Kovariablen, davon 8 ohne Information, 100 Personen, 100 Simulationen.

Die Berechnung der Freiheitsgrade über das Active Set führt, bei einer Standardabwei-

chung von 0.5 zum Idealfall, da alle informativen und uninformativen Kovariablen richtig spezifiziert werden. Auch bei der Approximation der Freiheitsgrade werden gute Anpassungen erreicht, insbesondere der Anteil der fälschlicherweise informativ geschätzten Kovariablen ist nicht mehr so hoch.

Der graphische Vergleich der Simulationsergebnisse mit unterschiedlicher Anzahl informativer Kovariablen ist in Abbildung 5.7 dargestellt.

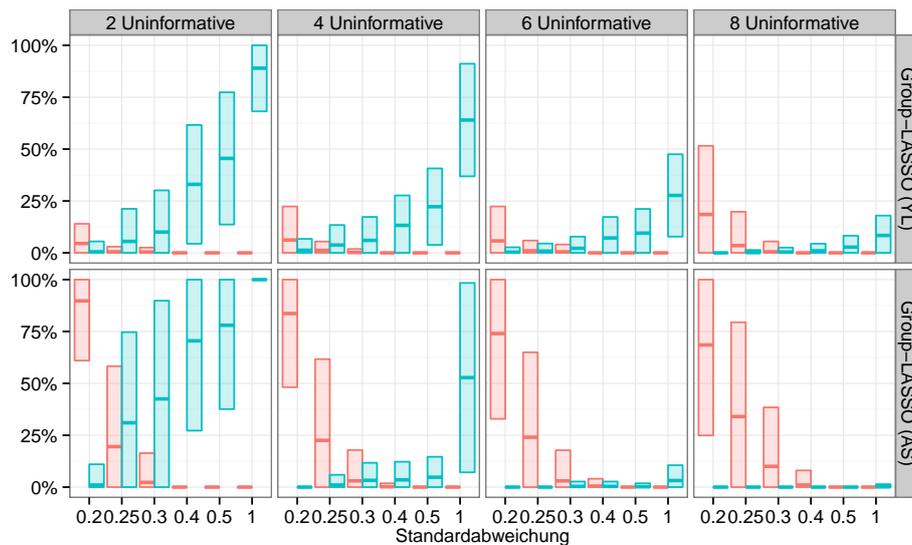


Abbildung 5.7.: Vergleich der Simulationsergebnisse mit unterschiedlicher Anzahl informativer Kovariablen von insgesamt 10 Kovariablen beim Vergleich von 10 Items.

Vergleicht man zunächst die Modellanpassungen mit Freiheitsgradberechnung nach [Yuan and Lin \[2006\]](#) (YL) ist zu erkennen, dass der Anteil der fälschlicherweise informativ geschätzten Kovariablen (blaue Boxen) bei der Schätzung mit zwei Uninformativen mit steigender Standardabweichung stark ansteigt, sich dieser Effekt jedoch durch eine Erhöhung des Anteils uninformativer Kovariablen abschwächt. Der Anteil fälschlicherweise nicht informativ geschätzter Kovariablen (rote Boxen) wird am besten bei einer kleinen Anzahl uninformativer Kovariablen angepasst.

Beim Vergleich der Group-LASSO Modelle mit Freiheitsgradberechnung (AS) ist eine wesentliche Verbesserung beim Anteil fälschlicherweise informativ geschätzter Kovariablen (blaue Boxen) bei einer hohen Anzahl uninformativer Kovariablen zu erkennen. Im Modell mit acht uninformativen Variablen wird bei einer Standardabweichung von 0.5 optimale Anpassung erreicht. Dafür ist der Anteil fälschlicherweise nicht informativ geschätzter Kovariablen höher.

Sowohl Group-LASSO (AS) als auch Group-LASSO (YL) zeigen, dass eine erhöhte Anzahl

uninformativer Kovariablen bessere Vorhersagen ermöglicht. Ein Vergleich zwischen den Modellen mit unterschiedlicher Freiheitsgradberechnung macht deutlich, dass bei einer hohen Anzahl informativer Kovariablen (zwei bzw. vier Uninformative) die Berechnung über Group-LASSO (YL) von Vorteil ist, während bei einer kleinen Anzahl informativer Kovariablen (sechs bzw. acht Uninformative) die Berechnung über Group-LASSO (AS) bessere Vorhersagen ermöglicht.

Letztlich wird noch ein Simulationssetting vorgestellt, bei dem untersucht wird, wie sich die Vorhersagegenauigkeit von Group-LASSO Modellen bei einer erhöhten Anzahl an Kovariablen verhält. Hierfür wurden $m = 10$ Items von 250 Personen in 45 Paarvergleichen miteinander verglichen. Insgesamt werden 20 Kovariablen, davon jeweils zehn binäre und zehn metrische mit jeweils sechs uninformativen simuliert. Die zugehörigen Simulationsergebnisse sind in Tabelle 5.8 dargestellt.

Std.Abw.	Group-LASSO (YL)		Group-LASSO (AS)	
	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)
0.2	99.875	4.67	99.375	1.00
0.25	100.00	7.33	99.875	1.083
0.3	100.00	10.67	100.00	0.917
0.4	100.00	14.33	100.00	2.75
0.5	100.00	17.33	100.00	3.583
1	100.00	39.50	100.00	8.417

Tabelle 5.8.: Ergebnisse des Simulationssettings: 10 Items, 20 Kovariablen, davon 12 ohne Information, 250 Personen, 100 Simulationen.

Auffallend ist, dass der Anteil korrekt informativ geschätzter Kovariablen für alle Berechnungen nahezu optimal ist. Der Anteil fälschlicherweise informativ geschätzter Kovariablen liegt bei der Berechnung der Freiheitsgrade über das Active Set stets unter 10%. Die beste Vorhersage wird bei einer Standardabweichung von 0.3 getroffen, wie es zudem in Abbildung 5.8 ersichtlich ist. Bei den Modellen Group-LASSO (YL) erfolgt die beste Anpassung bei einer Standardabweichung von 0.2, obwohl der Anteil korrekt informativ geschätzter Kovariablen nicht bei 100% liegt. Mit steigender Standardabweichung steigt der Anteil der fälschlicherweise informativ geschätzten Kovariablen zunehmend an.

Vergleicht man die Anpassungen über die Freiheitsgradberechnungen Group-LASSO (YL) und Group-LASSO (AS) wie in Abbildung 5.8 dargestellt, so ist eindeutig zu erkennen,

dass bei einer erhöhten Anzahl an Kovariablen die Anpassung über Group-LASSO (AS) bessere Modellanpassungen liefert.

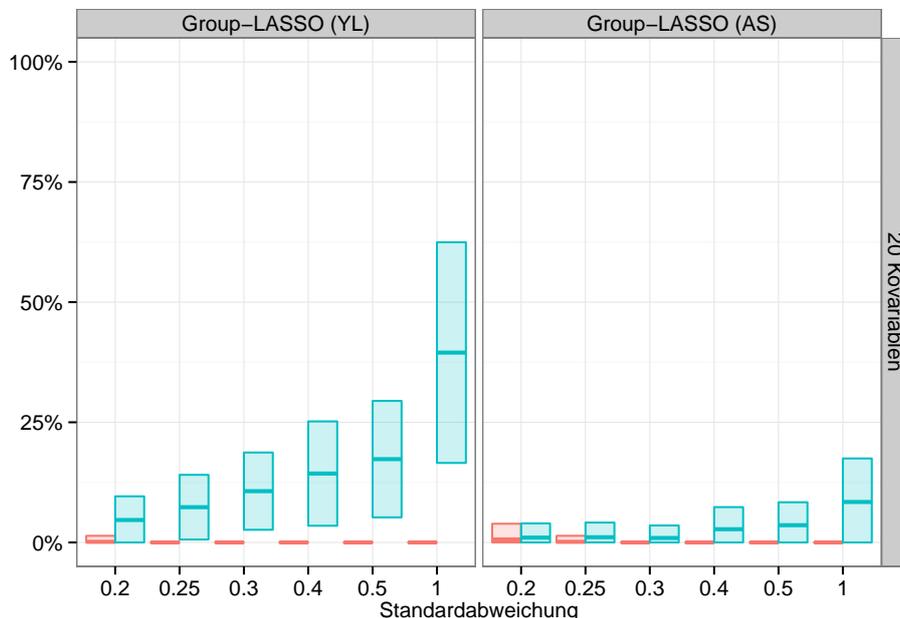


Abbildung 5.8.: Boxen der fälschlicherweise informativ bzw. nicht informativ geschätzten Parameter des Simulationssettings mit 10 Items und 20 Kovariablen, 12 ohne Information.

Im Allgemeinen kann in diesem Zusammenhang von einer guten Vorhersagegenauigkeit gesprochen werden. Durch die Einbindung von 20 Kovariablen, gegenüber zehn Kovariablen in den vorherigen Ausführungen, steigt der Anteil der fälschlicherweise als informativ geschätzten Kovariablen, allerdings ist der Anteil korrekt informativ geschätzter Kovariablen präziser. Die höhere Fehlerrate könnte möglicherweise eine Folge der erhöhten Personenzahl sein.

Fazit

Die betrachteten Simulationsstudien haben gezeigt, dass die Schätzung von Paarvergleichsdaten über Group-LASSO Modelle möglich ist und dass die Anpassung der einzelnen Modelle zu guten Ergebnissen führen. Fasst man die Ergebnisse der einzelnen Simulationen zusammen, kommt man zu dem Schluss, dass sowohl eine hohe Zahl an Kovariablen als auch an Items zu besseren Vorhersagen führen. Ebenso gilt dies für Modelle mit einem hohen Anteil uninformativer Kovariablen. Aufgrund der Ergebnisse kann allerdings keine allgemein gültige Aussage darüber getroffen werden, ob Modelle besser über Group-LASSO (AS) oder über Group-LASSO (YL) angepasst werden sollen, dies hängt jeweils von der Problemstellung ab.

6. Anwendungsbeispiele

Zur illustrativen Darstellung der alternativen Schätzverfahren für Bradley-Terry-Luce Modelle über Generalisierte Lineare Modelle bzw. über Group-LASSO werden im Folgenden zwei Anwendungsbeispiele durchgeführt. Zunächst erfolgt die Anwendung auf den Topmodel-Datensatz, der bereits Grundlage für die Beispiele aus Kapitel 2 war. Basierend auf den Informationen aus den Paarvergleichen werden zusätzliche Kovariablen bei der Schätzung mit eingebunden. Dabei wird bei der Anpassung über Group-LASSO untersucht, welche Kovariablen als Uninformativ eingestuft werden und aus dem Modell fallen. Desweiteren werden die beiden alternativen Schätzmethoden auf einen bekannten Datensatz aus dem Bereich der Bildung angewendet: Nach [Dittrich et al. \[1998\]](#) und [Böckenholt \[2001\]](#) wird beim CEMSChoice-Datensatz der Einfluss von Kovariablen bei der Präferenzentscheidung über die Wahl einer Universität für ein Auslandssemester von BWL-Studenten untersucht und mit den Ergebnissen von [Dittrich et al. \[1998\]](#) verglichen.

6.1. Germany's Next Topmodel 2007

Als Datenbasis für das erste Anwendungsbeispiel dient der Datensatz `Topmodel2007` aus dem R-Paket `psychotree` [[Zeileis et al., 2012b](#)]. Die Daten resultieren aus einer durchgeführten Befragung des Instituts für Psychologie der Universität Tübingen aus dem Jahr 2007, bei der 192 Personen ihre Präferenz bezüglich der Attraktivität der sechs Finalteilnehmerinnen der zweiten Staffel der Castingshow *Germany's Next Topmodel*, welche zwischen März-Mai 2007 gesendet wurde, angeben sollten. Die Befragung fand kurz nach der Ausstrahlung des Finales der TV-Sendung statt. *Germany's Next Topmodel* ist eine deutsche Castingshow für Topmodel-Anwärterinnen im deutschen Fernsehen auf dem Sender ProSieben. Die Sendung basiert auf dem Konzept von *Next Topmodel* aus den USA und wird von Heidi Klum moderiert [siehe z.B. [Wikipedia, 2012](#)].

Basierend auf Fotos der Top 6 Kandidatinnen, siehe Abbildung 2.4, die zur Zeit der Befragung auf der Homepage von ProSieben (www.prosieben.de) zur Verfügung standen, beurteilten die Studienteilnehmer die Attraktivität dieser $m = 6$ Teilnehmerinnen (Barbara Meier, Anni Wendler, Hana Nitsche, Fiona Erdmann, Mandy Graff und Anja

Platzer, absteigend nach den Platzierungen von oben links nach unten rechts angeordnet, beginnend mit der Siegerin Barbara Meier) in einem Forced-Choice Experiment, dessen Vorgehen bereits in Abschnitt 2.1 erläutert wurde. Die Fotos wurden aus Gründen der Vergleichbarkeit so gewählt, dass jeweils das Gesicht und der Oberkörper sichtbar ist und alle Frauen leger gekleidet sind. Die beobachteten Präferenzen der befragten Personen bezüglich der Attraktivität der Topmodels sind in Abbildung 6.1 dargestellt.

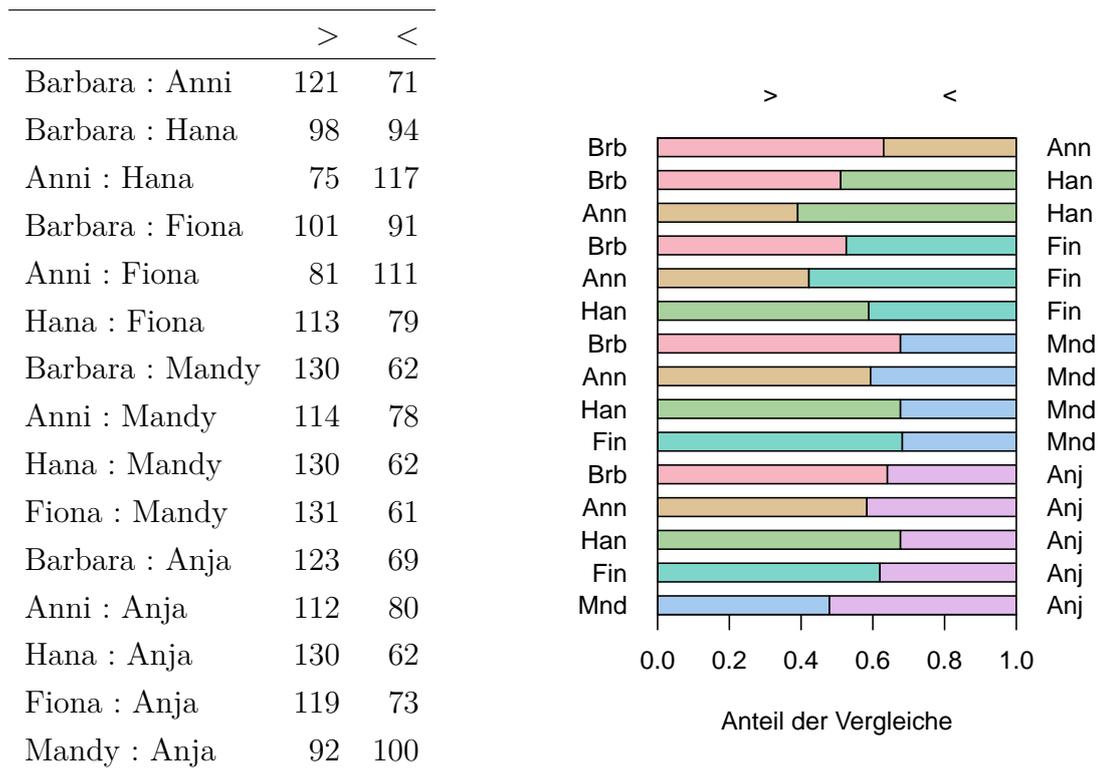


Abbildung 6.1.: Beobachtete Häufigkeiten der Paarvergleiche für den Topmodel-Datensatz.

Der Datensatz `Topmodel12007` enthält 192 Beobachtungen und 6 Variablen:

- **preference**: Präferenz aus den 15 Paarvergleichen über die 6 Topmodels
- **gender**: Geschlecht der befragten Personen (Faktorvariable: male(1)/female(2))
- **age**: Alter der befragten Personen in Jahren (Ganzzahlig)
- **q1**: Kennen Sie die Frauen auf den Fotos? / Kennen Sie die Fernsehsendung Germany's Next Topmodel? (Faktorvariable: yes(1)/no(2))
- **q2**: Sehen Sie die Sendung Germany's Next Topmodel regelmäßig? (Faktorvariable: yes(1)/no(2))

- q3: Haben Sie das Finale von Germany's Next Topmodel gesehen? / Wissen Sie wer in der Sendung gewonnen hat? (Faktorvariable: yes(1)/no(2))

Wurde eine der Fragen q1 und q3 positiv beantwortet, erfolgte für die andere Frage ebenfalls eine positive Einstufung. Insgesamt wurden 96 Frauen und 96 Männer im Alter zwischen 15 und 77 Jahren befragt, in beiden Gruppen jeweils 48 Personen unter 30 Jahre und 48 Personen über 30 Jahre.

Die Befragten sollten jeweils in den sich ergebenden 15 Paarvergleichen ihre Präferenz für eine der beiden Kandidatinnen entsprechend ihrer Attraktivität abgeben. Die Fragen q1–q3 wurden nach den Paarvergleichen gestellt um die Bewertungen der Teilnehmer einstufen zu können [Strobl et al., 2009; Zeileis et al., 2012b].

6.1.1. Anwendung BTL-Modell über GLM

Zunächst wird als alternatives Schätzverfahren für Bradley-Terry-Luce Modelle das Generalisierte Lineare Modell angewendet. Hierbei ist wie in Kapitel 4 zu unterscheiden, ob bei der Modellierung von Paarvergleichsdaten, wie in Unterabschnitt 4.3.2, zusätzliche Kovariablen berücksichtigt werden sollen oder nicht (siehe Unterabschnitt 4.3.1). Je nachdem, ob und wieviele Kovariablen bei der Schätzung eingebunden werden, verändert sich die Dimension der Designmatrix. Für den Datensatz `Topmodel2007`, mit $m = 6$ zu vergleichenden Items, $I = 15$ Paarvergleichen und $n = 192$ beurteilenden Personen, ergibt sich für die Grundstruktur der Designmatrix \mathbf{X} die Dimension (2880×5) . Die Designmatrix enthält aus Gründen der Identifizierbarkeit für jeden Effekt $m - 1$ Spalten, so dass der letzte Koeffizient, hier der Koeffizient für Topmodel-Kandidatin Anja, immer auf Null geschätzt wird.

Keine Berücksichtigung von Kovariablen

Bei der Schätzung von Paarvergleichsdaten über Generalisierte Lineare Modelle ohne Berücksichtigung von Kovariablen erhält man die Effekte aus den einzelnen Paarvergleichen. Daraus ergibt sich für die untersuchte latente Variable, hier für die Attraktivität der sechs Topmodel-Kandidatinnen, eine eindeutige Rangordnung. Das bedeutet, dass anhand der Präferenzentscheidungen der zu beurteilenden Personen die Models nach ihrer Attraktivität in eine Reihenfolge gebracht werden können. Über das GLM erhält man die Schätzer:

	Barbara	Anni	Hana	Fiona	Mandy	Anja
Schätzer:	0.6792	0.2489	0.7195	0.5272	-0.1072	0
Rangzuordnung:	2	4	1	3	6	5

Die graphische Darstellung der Ergebnisse ist in Abbildung 6.2 gegeben. Da die Topmodels nach ihren Platzierungen in der Fernsehsendung Germany's Next Topmodel angeordnet sind, zeigen sich in dieser Studie abweichende Ergebnisse von der Juryentscheidung. Barbara, die spätere Siegerin der TV-Show, reiht sich hinter Hana, der Drittplatzierten der Show, auf Rang zwei ein. Anni, die Zweitplatzierte der Sendung, wird ihrer Attraktivität nach nur auf Platz 4 hinter der eigentlich Viertplatzierten Fiona eingestuft.

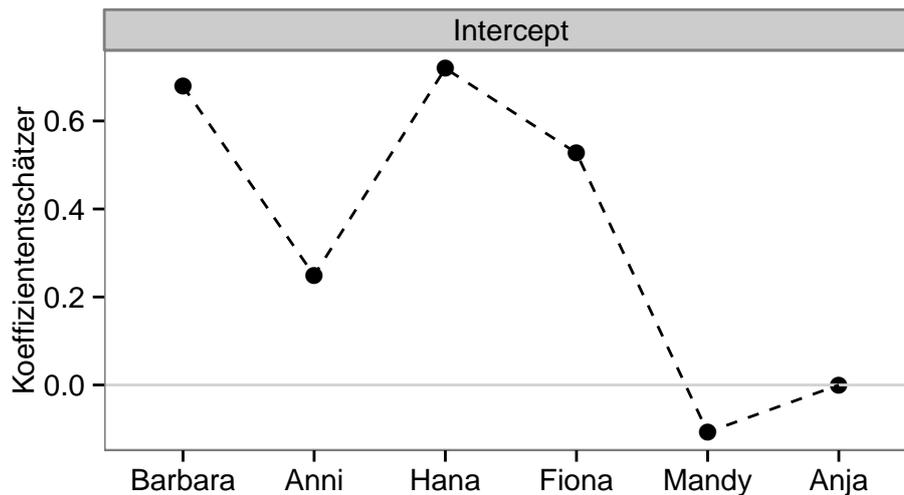


Abbildung 6.2.: Koeffizientenschätzer des Generalisierten Linearen Modells ohne Kovariablen für Topmodel-Datensatz.

Die Ergebnisse der Schätzung von Paarvergleichsdaten über Generalisierte Lineare Modelle können, da keine zusätzlichen Kovariablen berücksichtigt werden, mit den Ergebnissen der Schätzung über Bradley-Terry-Luce Modelle verglichen werden. Hierfür wird die R-Funktion `btReg.fit` aus dem Paket `psychotools` [Zeileis et al., 2012a] angewendet. Dabei resultieren die selben Schätzer wie bei der Modellierung über Generalisierte Lineare Modelle. Da es sich hierbei um ein spezielles Paarvergleichsmodell handelt kann man sich die wahren Präferenzen/Ratings $\pi_i \geq 0$ ($i = 1, \dots, m$) der einzelnen Items ausgeben lassen:

	Barbara	Anni	Hana	Fiona	Mandy	Anja
π_i	0.22	0.14	0.23	0.19	0.10	0.11

Werden die Schätzer der wahren Präferenzen, die auch mit „Worth-Parameter“ bezeichnet werden, wie in Abbildung 6.3 graphisch betrachtet, erhält annähernd den selben Verlauf wie bei der Betrachtung der geschätzten Koeffizienten.

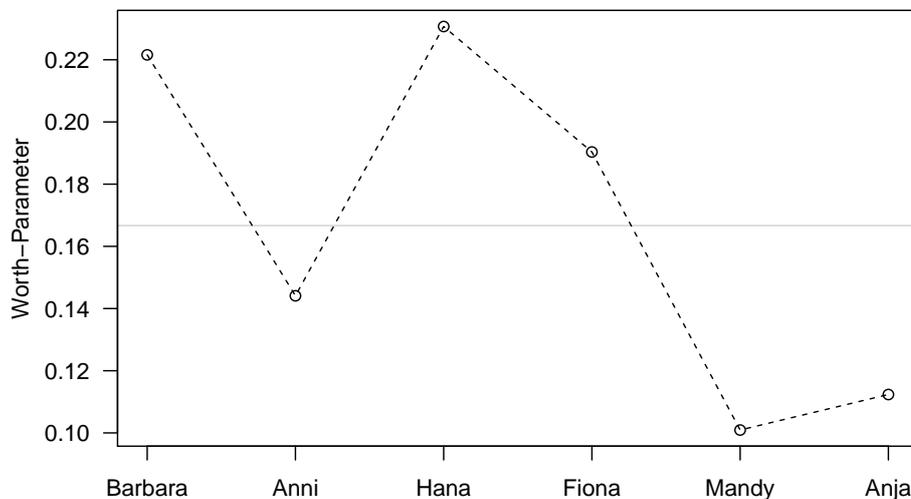


Abbildung 6.3.: Schätzer der wahren Präferenzen für die Topmodel-Kandidatinnen.

Über die wahren Präferenzen können nun die Wahrscheinlichkeiten, dass beim Vergleich der Items i und j das i -te Item präferiert wird, siehe Gleichung (3.1), berechnet werden. Vergleicht man zum Beispiel Item 1, Barbara, mit Item 2, Anni, berechnet sich die Wahrscheinlichkeit, dass Topmodel-Kandidatin Barbara gegenüber Topmodel-Kandidatin Anni präferiert wird, über

$$p_{B,Ann} = \frac{0.22}{0.22 + 0.14} = 0.61.$$

Barbara wird demnach mit einer Wahrscheinlichkeit von 61% gegenüber Anni präferiert. Wird hingegen der Paarvergleich zwischen Barbara und Hana betrachtet, erhält man für die Wahrscheinlichkeit $p_{B,H}$ den Wert 0.49 und für die Wahrscheinlichkeit $p_{H,B}$ den Wert 0.51.

Der Vorteil des alternativen Schätzverfahrens für BTL-Modelle über das GLM ist die einfache Einbindung von Kovariablen, die im Folgenden für das Topmodel-Beispiel erfolgt.

Berücksichtigung von Kovariablen

Bei der Schätzung von Paarvergleichsdaten interessieren neben den Effekten aus den einzelnen Paarvergleichen Effekte von Kovariablen, wie etwa der Einfluss des Geschlechts oder des Alters. Zu Beachten ist dabei, dass die Kovariablen standardisiert werden müssen. Bei der Anwendung auf den Topmodel-Datensatz werden als zusätzliche Kovariablen die binären Größen Geschlecht (**gender**), Kenntnis der Fernsehendung (**q1**), regelmäßiger

Konsum der Sendung (q2) und Kenntnis der Siegerin der TV-Show (q3) und die metrische Größe Alter (age) betrachtet. Somit ergibt sich für die Designmatrix \mathbf{X} die Dimension (2880×30) .

Die Schätzer der einzelnen Kovariablen lauten wie folgt:

	Barbara	Anni	Hana	Fiona	Mandy	Anja
Intercept	0.7319	0.2798	0.7847	0.5629	-0.0936	0
gender	0.3511	0.3521	0.0033	0.0989	-0.0049	0
q1	-0.2751	-0.0886	-0.5429	-0.2150	-0.4035	0
q2	-0.1968	-0.3662	-0.6066	-0.1008	-0.4192	0
q3	0.1025	0.2539	0.4780	0.2856	0.4197	0
age	0.0598	-0.4948	-0.1593	-0.3059	0.0529	0

Tabelle 6.1.: Koeffizientenschätzer des Generalisierten Linearen Modells mit Kovariablen für Topmodel-Datensatz.

Die Referenzkategorien der einzelnen binären Kovariablen lauten wie folgt. Beim Geschlecht (**gender**) fallen die Männer in die Referenzkategorie, bei den Fragen q1, q2 und q3 jeweils die Personen, die die Fragen mit ja beantworteten.

Die graphische Darstellung dieser Schätzergebnisse ist in Abbildung 6.4 gegeben.

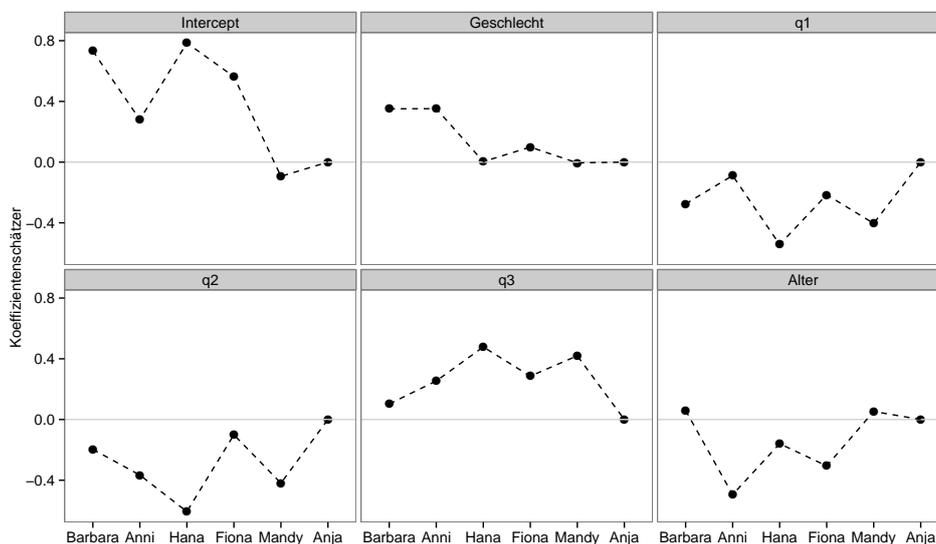


Abbildung 6.4.: Koeffizientenschätzer des Generalisierten Linearen Modells mit Kovariablen für Topmodel-Datensatz.

Bei der Schätzung mit Kovariablen ergeben sich für die Effekte aus den einzelnen Paarvergleichen (Intercept) leicht veränderte Schätzer, die daraus resultierende Rangordnung

bleibt unverändert. Bei der Interpretation der Koeffizientenschätzer der Kovariablen sei auf Gleichung (4.16) und Gleichung (4.17) verwiesen. Vergleicht man Item r mit Item s erhält man für das Chancenverhältnis, die Odds-Ratio

$$\frac{P(r \succ s | (r, s), x_i + 1)}{P(s \succ r | (r, s), x_i + 1)} \bigg/ \frac{P(r \succ s | (r, s), x_i)}{P(s \succ r | (r, s), x_i)} = \exp(\beta_r - \beta_s). \quad (6.1)$$

Es ergibt sich ein multiplikativer Effekt, der nicht vom Startwert x abhängt. Erhöht sich x um 1, so verändert sich die Chance für die Präferenz des Items r im Paarvergleich (r, s) um $\exp(\beta_r - \beta_s)$. Ist der Odds-Ratio kleiner als 1, liegt ein negativer Zusammenhang der untersuchten Merkmale vor, ein Odds-Ratio größer als 1 steht für einen positiver Zusammenhang. Nimmt der Odds-Ratio den Wert 1 an liegt kein Zusammenhang vor [Fahrmeir et al., 2009].

Somit können die Koeffizientenschätzer des Datenbeispiels aus Tabelle 6.1 wie folgt interpretiert werden. Als erstes liegt der Fokus auf der metrischen Kovariable Alter, die standardisiert in die Schätzung mit eingeht. Bereits der Kurvenverlauf in Abbildung 6.4 lässt erkennen, dass Topmodel-Kandidatin Anni einen stark negativen Einfluss besitzt. Betrachtet man den Paarvergleich zwischen Barbara und Anni erhält man für den Odds-Ratio

$$\frac{P(B \succ Ann | (B, Ann), x_{age} + 1)}{P(Ann \succ B | (B, Ann), x_{age} + 1)} \bigg/ \frac{P(B \succ Ann | (B, Ann), x_{age})}{P(Ann \succ B | (B, Ann), x_{age})} = \exp(0.0598 - (-0.4948)) \\ = 1.7412.$$

Erhöht sich das standardisierte Alter um eine Einheit, so steigt die Chance, beim Paarvergleich zwischen Barbara und Anni, Barbara zu präferieren um 74%. Rechnet man diesen Effekt auf das wirkliche Alter der befragten Personen um, entspricht eine Erhöhung des standardisierten Alters um eine Einheit etwa einer Erhöhung des wirklichen Alters um 14.5 Jahre. Somit steigt etwa bei einer knapp 50-jährigen Person die Chance Barbara gegenüber Anni zu präferieren um 74% gegenüber einer 35-jährigen Person. Vergleicht man die Topmodel-Kandidatinnen Anni und Fiona, so ist die Chance Anni gegenüber Fiona zu präferieren etwa 17% geringer, d.h. die Chance, dass eine 35-jährige Person Anni gegenüber Fiona präferiert ist um 17% geringer als bei einer gut 20-jährigen Person. Allgemein steigt mit zunehmenden Alter eine gewisse Abneigung gegenüber Kandidatin Anni, während Barbara an Ansehen gewinnt.

Nun wird der Effekt der binären Kovariable Geschlecht betrachtet. Da die Männer in die Referenzkategorie fallen vergleicht der Odds-Ratio die Präferenzen der Frauen mit den

Präferenzen der Männer. Abbildung 6.4 zeigt eine eindeutige Präferenz der Frauen für Barbara und Anni. Der Odds-Ratio für den Paarvergleich Barbara mit Hana nimmt den Wert 1.4159 an. Somit steigt die Chance bei Frauen Barbara gegenüber Hana zu präferieren um knapp 42%.

Bei den binären Kovariablen q_1 , q_2 und q_3 fallen in die Referenzkategorien jeweils die Personen, die die einzelnen Fragen mit ja beantworteten. Daher ist in den Plots in Abbildung 6.4 immer die Veränderung der Präferenzen der Personen abgetragen, die die Fragen q_1 , q_2 und q_3 mit nein beantworteten. Bei Frage q_2 präferieren Personen, die die TV-Show nicht regelmäßig verfolgen, Kandidatin Anja, die tatsächlich lediglich den 6.Rang in der Sendung erreichte.

In Unterabschnitt 6.1.2 wird das Paarvergleichsmodell nach Bradley, Terry und Luce über Group-LASSO geschätzt. Dabei wird lediglich der Fall mit Kovariablen geschätzt, denn werden keine Kovariablen im Modell berücksichtigt erfolgt keine Penalisierung und man benötigt kein regularisiertes Schätzverfahren.

6.1.2. Anwendung BTL-Modell über Group-LASSO

Das zweite in dieser Arbeit vorgestellte alternative Schätzverfahren für Bradley-Terry-Luce Modelle ist das Group-LASSO Verfahren, siehe Kapitel 5, das im Folgenden auf die Daten des Topmodel-Datensatzes angewendet wird. Wie bereits erwähnt wird hierbei lediglich der Fall mit Kovariablen betrachtet. Der Vorteil gegenüber der Anpassung über Generalisierte Lineare Modelle ist, dass durch Variablenselektion nur diejenigen Kovariablen im finalen Modell berücksichtigt werden, die einen wesentlichen Einfluss auf die Präferenzentscheidung haben.

Die Wahl des besten Modells kann über Modell-Selektionsverfahren wie AIC und BIC erfolgen. Aufgrund der stärkeren Bestrafung bei komplexen Modellen erhält man über das BIC sparsamere Modelle mit weniger Parametern. Da das Ziel bei Group-LASSO die Selektion von informativen Kovariablen ist, wird für dieses Beispiel lediglich das BIC betrachtet. Diese Betrachtung hat bereits bei der Simulationsstudie mit $m = 6$ Items, siehe Abschnitt 5.3, zu einem sparsameren Modell geführt.

Wie ebenfalls aus den Simulationsstudien aus Abschnitt 5.3 hervorgeht, wird bei der Freiheitsgradberechnung unterschieden. Zum Einen kann die Berechnung über das Active Set (AS) erfolgen, indem die Freiheitsgrade mit der Anzahl geschätzter Koeffizienten ungleich Null gleichgesetzt werden, zum anderen kann die Berechnung der Freiheitsgrade über die Approximation nach Yuan and Lin [2006], siehe Gleichung (5.12), für logistische Group-LASSO Modelle (YL) erfolgen.

Zur Überprüfung, welche Anpassung über Group-LASSO bessere Ergebnisse liefert, wurde ein Simulationssetting gemäß den Bedingungen aus dem Topmodel-Datensatz aufgestellt, mit der Annahme, dass eine der fünf Kovariablen keine Information bezüglich der Präferenzentscheidung enthält. Die Ergebnisse sind im Anhang Tabelle A.1 beigefügt. Daraus ist ersichtlich, dass für den Fall mit $m = 6$ Items und fünf Kovariablen eine bessere Vorhersage über die Anpassung eines Group-LASSO Modells mit Freiheitsgradberechnung über die Approximation aus Gleichung (5.12) möglich ist. Somit wird für den Topmodel-Datensatz lediglich die Modellierung über Group-LASSO Modelle mit Freiheitsgradberechnung nach Yuan and Lin [2006] betrachtet.

Wie bei der Modellierung über GLMs unter Berücksichtigung von Kovariablen werden an die Grundstruktur der Designmatrix \mathbf{X} die fünf standardisierten Kovariablen `gender`, `q1`, `q2`, `q3` und `age` mit angehängt.

$$\begin{array}{c}
 \\
 1 \\
 2 \\
 \vdots \\
 2880
 \end{array}
 \left(
 \begin{array}{c|c|c|c|c|c}
 & \text{Intercept} & & \text{gender} & & \text{q1} & & \text{q2} & & \text{q3} & & \text{age} \\
 \hline
 & 1 & -1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & \dots \\
 & 1 & -1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & \dots \\
 & & & \ddots & & & & & \ddots & & & \ddots \\
 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & \dots
 \end{array}
 \right)$$

Da die standardisierten Kovariablen mit dem Effekt der einzelnen Paarvergleiche (Intercept) multipliziert werden, belegt jede einzelne $(m - 1)$ Spalten in der Designmatrix. Da in der Studie exakt gleich viele Männer wie Frauen bezüglich ihrer Präferenz befragt werden, nehmen die gerundeten standardisierten Werte der Kovariable -1 für Männer und 1 für Frauen an. Bei der ersten Person, die den 1. Paarvergleich durchführt, handelt es sich um einen Mann, da sich die Vorzeichen des Effekts aus dem ersten Paarvergleich umdrehen, da mit -1 multipliziert wird. Für den 15. Paarvergleich der 192. Person ändert sich der Effekt nicht, so dass es sich hierbei um eine Frau handelt. Dieses Vorgehen ist für die weiteren Kovariablen analog.

Über den Index werden für den Intercept und die einzelnen Kovariablen die Gruppengrößen festgelegt, die in diesem Beispiel für alle Gruppen $(m - 1) = 5$ beträgt. Da der Intercept nicht penalisiert wird, wird hierfür der Index NA angegeben. Für den Bestrafungsparameter λ wird, wie bei den Simulationen in Abschnitt 5.3, eine Sequenz von 40 Werten angenommen, beginnend mit dem hier maximalen Wert von 50.019 bis zum Wert 0.001. Bei der Schätzung der einzelnen Modelle über die R-Funktion `grplasso` wird über das BIC das Beste selektiert. Die Wahl fällt beim Topmodel-Datensatz auf das 28. Modell, bei dem λ den Wert 15.39 annimmt. Die Koeffizienten der Kovariable `q3` werden bei

diesem Modell auf Null geschätzt, so dass die Kovariable als uninformativ eingestuft wird und aus dem Modell fällt. Die Koeffizientenschätzer des 28. Modells sind in Tabelle 6.2 dargestellt. Wie bei der Schätzung über Generalisierte Lineare Modelle wird aus Identifikationsgründen der Koeffizient des letzten Items, Kandidatin Anja, mit dem Wert 0 festgelegt.

	Barbara	Anni	Hana	Fiona	Mandy	Anja
Intercept	0.6900	0.2520	0.7300	0.5300	-0.1100	0
gender	0.1757	0.1517	-0.0582	-0.0136	-0.0645	0
q1	-0.0450	0.0150	-0.0856	0.0051	-0.0295	0
q2	-0.0356	-0.1024	-0.2140	0.0734	-0.0997	0
q3	0	0	0	0	0	0
age	0.0734	-0.2550	-0.1105	-0.1431	0.0582	0

Tabelle 6.2.: Koeffizientenschätzer über Group-LASSO für Topmodel-Datensatz.

Für den Effekt aus den Paarvergleichen erhält man annähernd die selben Schätzer wie bei der direkten Schätzung über BTL-Modelle bzw. bei der Schätzung über GLMs. Somit bleibt die Rangordnung der Kandidatinnen nach ihrer Attraktivität die selbe. Zum Vergleich der weiteren Schätzergebnisse sind in Abbildung 6.5 sowohl die Schätzer der GLMs (schwarz) als auch die Schätzer über Group-LASSO (blau) eingezeichnet.

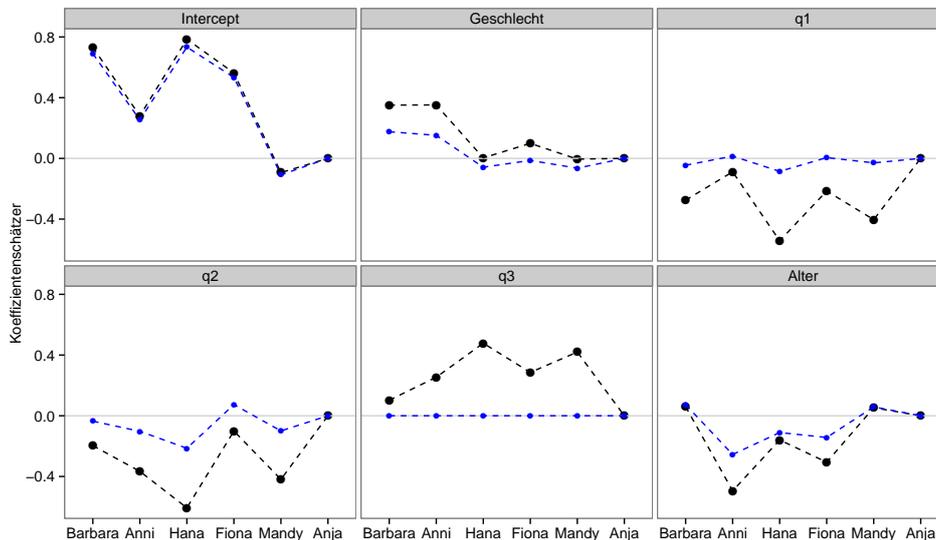


Abbildung 6.5.: Vergleich der Koeffizientenschätzer zwischen Group-LASSO (blau) und GLM (schwarz) für Topmodel-Datensatz.

Es ist gut zu erkennen, dass die blauen Kurven, die sich über die Schätzung von Group-

LASSO ergeben, allgemein etwas flacher verlaufen als die schwarzen. Dieser Effekt kann auf den Shrinkage-Effekt zurückgeführt werden, da über das BIC das 28. Modell mit $\lambda = 15.39$ gewählt wird. Würde man das 40. Modell mit $\lambda = 0$ einzeichnen, würden die blaue und die schwarze Kurve übereinander liegen.

Die Interpretation der Parameter kann aus Unterabschnitt 6.1.1 übernommen werden. Durch den Shrinkage-Effekt sind die Unterschiede bei den Präferenzen geringer, allerdings behalten die allgemeinen Aussagen ihre Gültigkeit, da sich die Anordnungen bei den Topmodel-Kandidatinnen über die einzelnen Kovariablen nicht ändern. Die Fragestellung, ob die befragten Personen das Finale gesehen haben bzw. ob sie wissen wer die Sendung gewonnen hat, wird bei der Schätzung über Group-LASSO als irrelevant bei den Präferenzentscheidungen eingestuft.

Das Anwendungsbeispiel auf den Topmodel-Datensatz hat gezeigt, dass sich sowohl das Generalisierte Lineare Modell als auch das Group-LASSO Verfahren als alternative Schätzverfahren für Bradley-Terry-Luce Modelle eignen. Erfolgt die Schätzung von Paarvergleichsdaten ohne zusätzlichen Kovariablen resultieren über das GLM exakt die selben Schätzer wie über das BTL-Modell. Durch die einfache Einbindung subjekt-spezifischer Informationen sind zusätzlich Aussagen über den Einfluss bestimmter Kovariablen auf die Präferenzentscheidungen möglich. Für den Topmodel-Datensatz konnte etwa eine Antipathie gegenüber Anni bei älteren Personen festgestellt werden.

Im folgenden Unterabschnitt 6.1.3 werden die Ergebnisse der alternativen Schätzverfahren für Bradley-Terry-Luce Modelle mit dem Verfahren des modellbasierten rekursiven Partitionierens von [Strobl et al. \[2009\]](#) verglichen.

6.1.3. Vergleich mit modellbasierten rekursiven Partitionieren

In den vorherigen Unterabschnitten 6.1.1 und 6.1.2 wurden regularisierte Schätzverfahren für Bradley-Terry-Luce Modelle auf den Topmodel-Datensatz angewendet. Der wesentliche Vorteil gegenüber der einfachen Schätzung über BTL-Modelle ist die Einbindung von Kovariablen. Dadurch können bei den Präferenzentscheidungen neben den Eigenschaften der zu beurteilenden Items auch Charaktereigenschaften der zu beurteilenden Personen selbst bei der Schätzung berücksichtigt werden.

Ein neuerer Ansatz zur Einbindung von subjekt-spezifischen Kovariablen in BTL-Modellen ist das Verfahren des modellbasierten rekursiven Partitionierens von [Strobl et al. \[2009\]](#), das nach [Breiman \[2001\]](#) zu den Algorithmischen Modelle zählt. Es wird untersucht, ob Gruppen von Subjekten identifiziert werden können, die sich in ihren Präferenzentscheidungen unterscheiden. Zum Beispiel können Präferenzentscheidungen bei Personen ab

einem bestimmten Alter abweichen oder Frauen beurteilen die Stimuli gegenüber Männern differenziert. Zur Illustration ihres Verfahrens führten [Strobl et al. \[2009\]](#) ebenfalls eine Anwendung auf den Topmodel-Datensatz aus.

Aufgrund der selben Datenbasis wird ein Vergleich der Verfahren durchgeführt, wobei ein direkter Vergleich der Ergebnisse nicht möglich ist, da die Verfahren unterschiedliche statistische Hintergründe besitzen. Denn bei den regularisierten Schätzverfahren werden zwar Kovariablen bei der Schätzung berücksichtigt, diese werden allerdings nicht auf Unterschiede innerhalb der einzelnen Kovariablen untersucht, wie es beim Verfahren des modellbasierten rekursiven Partitionierens der Fall ist. Somit sind beim modellbasierten rekursiven Partitionieren Aussagen für bestimmte Gruppen möglich, die sich in ihren Eigenschaften entsprechen, wie etwa eine Gruppe von Personen, die alle unter 45 Jahre alt sind.

Zunächst wird die Methode des modellbasierten rekursiven Partitionierens für Bradley-Terry-Luce Modelle kurz vorgestellt und auf den Topmodel-Datensatz angewendet. Die Ergebnisse werden vorgestellt und mit den Ergebnissen der regularisierten Schätzverfahren verglichen.

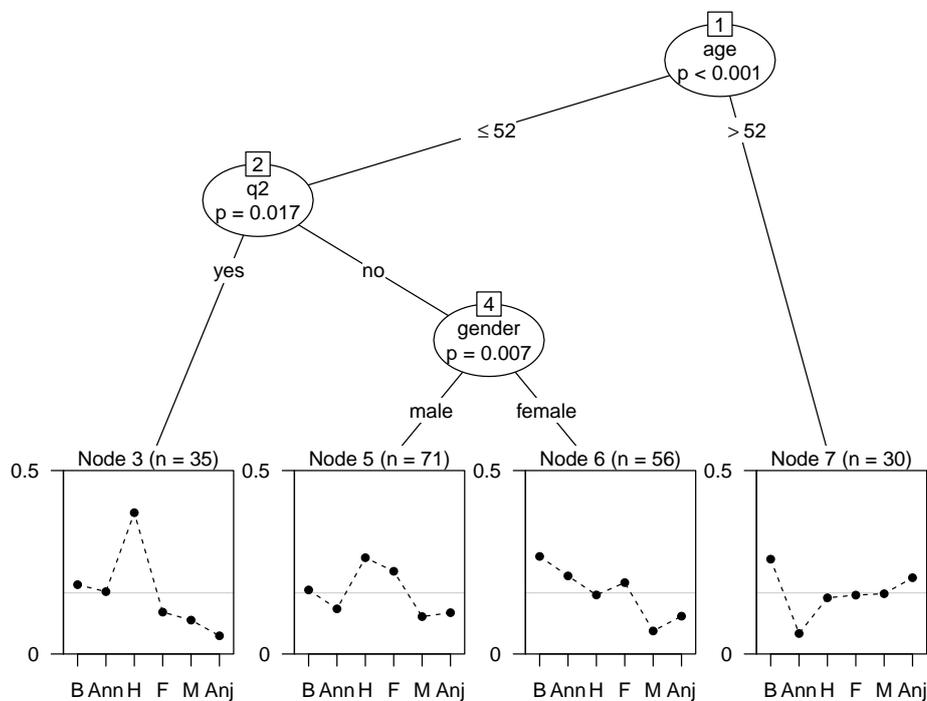
Die Idee der modellbasierten rekursiven Partitionierung für BTL-Modelle ist das rekursive Aufteilen der gesamten Stichprobe in Submodelle, die sich durch bestimmte Eigenschaften voneinander unterscheiden. Die Schnittstelle, an der die Aufteilung in Submodelle stattfindet, wird als „Knoten“ oder auch als „Cutpoint“ bezeichnet und aus den Daten ermittelt. Dabei sei explizit darauf hingewiesen, dass nach Parameterwerten gesplittet wird und nicht nach Variablenausprägungen. Diese Parameter, auch „Worth-Parameter“ genannt, geben Aufschluss über eine Strukturänderung, also einen Strukturbruch, im Präferenzverhalten der Personen. Somit erhält man Submodelle, die jeweils ein homogenes Präferenzverhalten beschreiben, aber untereinander heterogen sind [[Strobl et al., 2009](#)].

Das resultierende rekursiv partitionierte BTL-Modell, wie in [Abbildung 6.6](#), wird anhand eines einfachen Algorithmus generiert, der die folgenden Schritte durchläuft:

1. Schätzung eines Bradley-Terry-Luce Modells für die Paarvergleiche aller Personen der aktuellen (Teil-)Stichprobe, beginnend mit der gesamten Stichprobe.
2. Bestimmung der Parameterstabilität bezüglich jeder verfügbaren Kovariable.
3. Aufteilung/Split der Daten bei signifikanter Instabilitätsstelle („Cutpoint“).
4. Wiederhole Schritte 1-3 bis keine signifikanten Instabilitäten mehr auftreten oder Subgruppen die Mindestgröße unterschreiten.

Für eine detaillierte Beschreibung der einzelnen Schritte sei auf [Strobl et al. \[2009\]](#) verwiesen.

Zur Anwendung des Verfahrens der modellbasierten rekursiven Partitionierung auf den Topmodel-Datensatz wird die R-Funktion `bttree` aus dem R-Paket `psychotree` [[Zeileis et al., 2012b](#)] verwendet. Der BTL-basierte Baum ist in [Abbildung 6.6](#) dargestellt. Die Präferenzentscheidungen, abgeleitet aus dem BTL-Modell, variieren zwischen Gruppen von Personen entsprechend der folgenden Knoten: Personen bis zu einem Alter von 52 Jahren, die Frage `q2` mit ja beantworteten ([Abbildung Node 3 ganz links](#)) präferieren eindeutig Topmodel-Kandidatin Hana gegenüber allen anderen, während, zum Beispiel, Personen über 52 Jahren ([Abbildung Node 7 ganz rechts](#)) alle weiteren Kandidatinnen gegenüber Kandidatin Anni präferieren.



[Abbildung 6.6.](#): Partitioniertes Paarvergleichsmodell für Topmodel-Datensatz, aus dem sich ergibt, dass beurteilte Attraktivität über Gruppen, resultierend aus den Kovariablen Alter, `q2` und Geschlecht, variiert.

Zusätzlich zur graphischen Darstellung des partitionierten Modells in [Abbildung 6.6](#) können die Ergebnisse über die Schätzer der Worth-Parameter in jeder Teilstichprobe, wie in [Tabelle 6.3](#) zusammengefasst werden. Diese zeigen, dass das Rating der Personen bis 52 Jahren, die die TV-Show regelmäßig sehen ([Node 3](#)), im Wesentlichen mit der Entscheidung der Jury übereinstimmt - außer das Rating für Kandidatin Hana, die von den Zuschauern der Sendung als doppelt so attraktiv eingestuft wurde wie Barbara, die ei-

gentliche Gewinnerin. Diese starke Präferenz für Hana kann in keiner anderen Gruppe, die die Show nicht regelmäßig gesehen haben, nachgewiesen werden. Bei den Personen unter 52 Jahren, die die Show nicht regelmäßig verfolgen, zeigen Männer (Node 5) Präferenzen für Hana und Fiona, während Frauen (Node 6) Barbara am meisten präferieren, gefolgt von Anni und Fiona.

	Barbara	Anni	Hana	Fiona	Mandy	Anja
Node 3	0.19	0.17	0.39	0.11	0.09	0.05
Node 5	0.17	0.12	0.26	0.23	0.10	0.11
Node 6	0.27	0.21	0.16	0.19	0.06	0.10
Node 7	0.26	0.06	0.15	0.16	0.16	0.21

Tabelle 6.3.: Schätzer der Worth-Parameter der jeweiligen Teilstichproben des partitionierten Paarvergleichsmodell für den Topmodel-Datensatz.

Interessant ist, dass sich die Präferenzen älterer Teilnehmer (Node 7) von allen anderen Gruppen unterscheiden. Personen über 52 Jahre stufen Anja beinahe so attraktiv wie Barbara ein, während sie gegenüber Anni eine Abneigung haben (ihr Attraktivitätsrating ist nur etwa 20% von Barbara's). Zwischen Hana, Fiona und Mandy zeigt diese Gruppe kaum Unterschiede.

Der Effekt, dass ältere Personen unterschiedliche Präferenzen haben als jüngere, kann darauf zurückgeführt werden, dass zwischen Generationen unterschiedliche Vorstellungen/Wahrnehmungen von Schönheit vorliegen. Hingegen kann die starke Präferenz für eine Kandidatin, wie bei den Personen, die die TV-Show regelmäßig verfolgen, darauf hinweisen, dass die Persönlichkeit der Kandidatin gegenüber ihrer physischen Erscheinung entscheidend bei der Präferenzwahl der Zuschauer beiträgt.

Ein direkter Vergleich der Ergebnisse der partitionierten Modell mit den Ergebnissen aus den regularisierten Schätzungen ist nicht möglich. Beim modellbasierten rekursiven Partitionieren wird nach Unterschieden bei den Präferenzentscheidungen innerhalb der Kovariablen gesucht, sodass differenzierte Aussagen bezüglich Gruppen mit verschiedenen Charaktereigenschaften gemacht werden können. Die Schätzung über regularisierte Verfahren überprüft den Einfluss der einzelnen Kovariablen auf die Präferenzentscheidung und selektiert Kovariablen ohne Information.

Alle betrachteten Modelle zeigen einen Einfluss der Kovariablen Alter, q2 und Geschlecht bei der Präferenzwahl. Die Variablenselektion über Group-LASSO verursacht lediglich den Ausschluss von Kovariable q3 aus dem Modell. Am besten kann der Effekt des Alters über die verschiedenen Modellansätze verglichen werden. Bei der Schätzung über Generalisier-

te Lineare Modelle bzw. über Group-LASSO wurde festgestellt, dass bei zunehmenden Alter eine gewisse Antipathie gegenüber Kandidatin Anni vorliegt. Eine ähnliche Aussage resultiert bei der Modellierung über modellbasiertes rekursives Partitionieren, nur dass hierbei ein konkretes Alter, nämlich 52 Jahre, angegeben wird, ab wann die Abneigung gegenüber Anni auftritt. Bei den regularisierten Schätzverfahren ergab sich zudem eine Präferenz der Frauen für Barbara und Anni. Über das Verfahren von [Strobl et al. \[2009\]](#) erhält man dieselbe Aussage über die Präferenzentscheidungen der Frauen, jedoch nur für diejenigen, die jünger als 52 Jahre sind und die TV-Sendung nicht regelmäßig verfolgen.

Man sieht, dass ein Vergleich der unterschiedlichen Verfahren aufgrund ihrer Motivationen sehr schwierig ist. Dennoch können ähnliche Effekte bei den Kovariablen bezüglich der Präferenzentscheidungen festgestellt werden. Ein weiteres Anwendungsbeispiel, siehe Abschnitt [6.2](#), der regularisierten Schätzverfahren auf Paarvergleichsdaten wird auf den `CEMSChoice`-Datensatz durchgeführt. Hierbei erfolgt ein Vergleich der Ergebnisse mit einem weiteren möglichen Schätzverfahren über loglineare Modelle.

6.2. CEMS-Universitäten

Das zweite Anwendungsbeispiel behandelt die Untersuchung des `CEMSChoice`-Datensatzes, der ebenfalls im R-Paket `psychotree` [[Zeileis et al., 2012b](#)] verfügbar ist. Studenten der Wirtschaftsuniversität Wien haben die Möglichkeit einen Teil ihres Studiums im Ausland an einer der derzeit 17 CEMS Universitäten (Community of European Management Schools and International Companies) zu absolvieren. Da beobachtet wurde, dass einige Universitäten äußerst beliebt bei den Studenten sind und andere weniger, wurde im Jahr 1995 von [Dittrich et al. \[1998\]](#) eine Studie durchgeführt, um mögliche Gründe für die Präferenzentscheidungen der Studenten zu untersuchen.

Bei der Studie wurden $m = 6$ europäische CEMS-Universitäten ausgewählt und über Paarvergleiche systematisch miteinander verglichen. Dadurch ergeben sich, wie im ersten Anwendungsbeispiel, $I = 15$ Paarvergleiche zwischen den Universitäten in London (London School of Economics and Political Science), Paris (Haut Etudes Commerciales (HEC)), Mailand (Università Luigi Bocconi), Barcelona (Escuela Superior de Administración y Dirección de Empresas (ESADE)), St. Gallen (Hochschule St. Gallen) und Stockholm (Stockholm School of Economics).

Um mögliche Gründe für die Präferenzen der Studenten zu finden, wurden zusätzlich subjekt-spezifische Kovariablen erhoben. Weiter sind die Universitäten nach ihren Spezialisierungen - Volkswirtschaftslehre VWL, Management MA und Finanzwesen FW - sowie

nach ihrem Standort charakterisiert. Unter Standort wird zwischen Ländern, in denen romanische Sprachen gesprochen werden, wie zum Beispiel in Italien, Frankreich oder Spanien, und anderen europäischen Ländern unterschieden. Alle Charakteristiken sind binär (0 = nein, 1 = ja) und können wie in Tabelle 6.4 dargestellt werden.

Universität	VWL	MA	FW	LAT
London	1	0	0	0
Paris	0	1	0	1
Mailand	1	0	0	1
St. Gallen	0	0	1	0
Barcelona	0	1	0	1
Stockholm	0	0	1	0

Tabelle 6.4.: Eigenschaften der CEMS-Universitäten.

Die beobachteten Präferenzen der Studenten bei den $I = 15$ Paarvergleichen sind in Abbildung 6.7 aufgelistet.

	>	=	<	NA's
London : Paris	186	26	91	0
London : Milano	221	26	56	0
Paris : Milano	121	32	59	91
London : St. Gallen	208	22	73	0
Paris : St. Gallen	165	19	119	0
Milano : St. Gallen	135	28	140	0
London : Barcelona	217	19	67	0
Paris : Barcelona	157	37	109	0
Milano : Barcelona	104	67	132	0
St. Gallen : Barcelona	144	25	134	0
London : Stockholm	250	19	34	0
Paris : Stockholm	203	30	70	0
Milano : Stockholm	157	46	100	0
St. Gallen : Stockholm	155	50	98	0
Barcelona : Stockholm	172	41	90	0

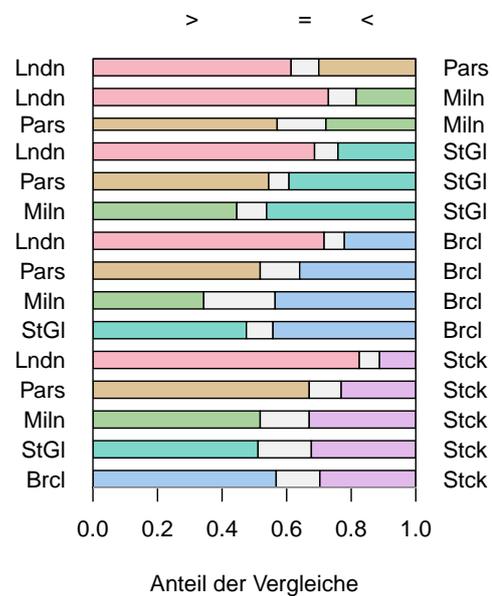


Abbildung 6.7.: Beobachtete Häufigkeiten der Paarvergleiche für CEMSChoice-Datensatz.

Im Gegensatz zum Topmodel-Datensatz sind hier Unentschieden zwischen den Stimuli erlaubt, falls der befragte Student keine eindeutige Entscheidung für eine der beiden Universitäten treffen kann. Bei 91 Studenten wurde der Vergleich Paris : Milan absichtlich ausgelassen, um eine unbalancierte Stichprobe zu erhalten [Dittrich et al., 1998; Strobl et al., 2009].

Die in der Analyse berücksichtigten Kovariablen sind Geschlecht, Studienschwerpunkt, und Indikatoren für die Fragen, ob der Student neben dem Studium Vollzeit arbeitet, ob ein internationaler Abschluss angestrebt wird und ob man gute Sprachkenntnisse in Englisch, Französisch, Spanisch und Italienisch hat. Zusammenfassend enthält der Datensatz `CEMSChoice` 303 Beobachtungen und 10 Variablen:

- `preference`: Präferenz aus den 15 Paarvergleichen der 6 Universitäten: London, Paris, Mailand, St. Gallen, Barcelona, Stockholm
- `study`: Studienschwerpunkt der befragten Studenten (Faktorvariable: commerce (2)/other(economics, business administration, business education) (1))
- `english`: Sprachkenntnisse Englisch (Faktorvariable: good(1)/poor(2))
- `french`: Sprachkenntnisse Französisch (Faktorvariable: good(1)/poor(2))
- `spanish`: Sprachkenntnisse Spanisch (Faktorvariable: good(1)/poor(2))
- `italian`: Sprachkenntnisse Italienisch (Faktorvariable: good(1)/poor(2))
- `work`: Vollzeit-Job während des Studiums (Faktorvariable: no(1)/yes(2))
- `gender`: Geschlecht der befragten Studenten (Faktorvariable: female(1)/male(2))
- `intdegree`: Ziel internationaler Abschluss (Faktorvariable: no(1)/yes(2))
- `preference1998`: Präferenz aus den 15 Paarvergleichen siehe `preference`; Präferenzen des Vergleichs Barcelona:Stockholm waren fälschlicherweise vertauscht

Insgesamt wurden 152 weibliche und 151 männliche Studenten befragt.

Da im Jahr 1998 fälschlicherweise die Präferenzen des 15. Paarvergleichs Barcelona:Stockholm vertauscht wurden, d.h. es wurden 172 Präferenzen für Stockholm anstelle der 90 tatsächlichen Präferenzen für Stockholm notiert, sind die Ergebnisse in Dittrich et al. [1998] falsch bzw. zeigen die Ergebnisse aus Dittrich et al. [1998] die Schätzung mit den vertauschten Präferenzen beim Paarvergleich Barcelona:Stockholm. Die richtigen Schätzergebnisse sind in Dittrich et al. [2001] gegeben.

6.2.1. Anwendung BTL-Modell über GLM

Wendet man für diesen Beispiel-Datensatz die Schätzung über Generalisierte Lineare Modelle an, ist zunächst der Umgang mit den Unentschieden festzulegen. Das hier vorgestellte Bradley-Terry-Luce Modell, siehe Gleichung (3.1), das als Basis zur Schätzung von Paarvergleichsdaten über GLMs dient, lässt beim direkten Vergleich zweier Items kein Unentschieden zu. Somit werden diese auf die Präferenzentscheidungen ja/nein der entsprechenden Paarvergleiche aufgeteilt. Ist die Anzahl an Unentschieden beim Vergleich der Items (i, j) gerade, so wird die Hälfte als Präferenz für Item i und die Hälfte als Präferenz für Item j umkodiert. Für den ersten Paarvergleich im CEMSChoice-Datensatz zwischen den Universitäten in London und Paris mit 26 unentschiedenen Studenten erhöhen sich sowohl die Präferenz für London als auch für Paris um jeweils 13 Stimmen auf $199 > 104$. Bei einer ungeraden Anzahl an Unentschieden wird die Mehrstimme immer dem ersten Item im Vergleich zugewiesen.

Weiter ist zu beachten, dass die 91 Paarvergleiche zwischen Paris und Mailand bei der Schätzung unberücksichtigt bleiben, und somit beim Responsevektor als auch in der Designmatrix ausgeschlossen werden müssen. Aufgrund dessen ergibt sich für die Grundstruktur der Designmatrix die Dimension (4454×5) , die sich bei der Einbindung von, der in Abschnitt 6.2 genannten, acht Kovariablen auf (4454×45) erhöht. Ansonsten ist das Vorgehen, wie beim Topmodel-Datensatz, identisch, d.h. es wird auch hier zwischen der Berücksichtigung von Kovariablen und keiner Berücksichtigung von Kovariablen unterschieden.

Keine Berücksichtigung von Kovariablen

Werden die Präferenzen bezüglich der CEMS-Universitäten über das Generalisierte Lineare Modell ohne Kovariablen geschätzt, kann ein eindeutiges Ranking für die Universitäten nach ihrer Beliebtheit für eine Auslandssemester aufgestellt werden. Die Ränge können anhand der Schätzer abgeleitet werden, die in folgender Tabelle dargestellt sind:

	London	Paris	Mailand	St. Gallen	Barcelona	Stockholm
Schätzer:	1.6035	0.9118	0.3915	0.5278	0.5371	0
Rangzuordnung:	1	2	5	4	3	6

Auf der Beliebtheitsskala der Universitäten für ein Auslandssemester rangiert eindeutig London vor Paris, gefolgt von Barcelona, St. Gallen, Mailand und Stockholm. Graphisch sind die Schätzergebnisse in Abbildung 6.8 dargestellt. Die Ergebnisse lassen vermuten, dass die Sprachkenntnisse der Studenten einen starken Einfluss bei der Wahl der Univer-

sitäten spielen. Die starke Präferenz für London zeigt, dass die englische Sprache in der heutigen Gesellschaft einen sehr hohen Stellenwert besitzt. Nicht zu vergessen ist auch der Fakt, dass Englisch als erste Fremdsprache und Französisch als zweite Fremdsprache in Österreich jahrelang in der Schule gelehrt werden, sodass sich Paris nach London auf Platz 2 einreicht.

Wenn man davon ausgeht, dass die Sprachkenntnisse einen entscheidenden Einfluss bei der Wahl einer Universität haben, ist der letzte Platz für die Universität in Stockholm interessant, da dort ebenfalls auf Englisch unterrichtet wird. Möglicherweise befürchten die Studenten, dass zusätzlich Kurse, die auf Schwedisch gehalten werden, besucht werden müssen.

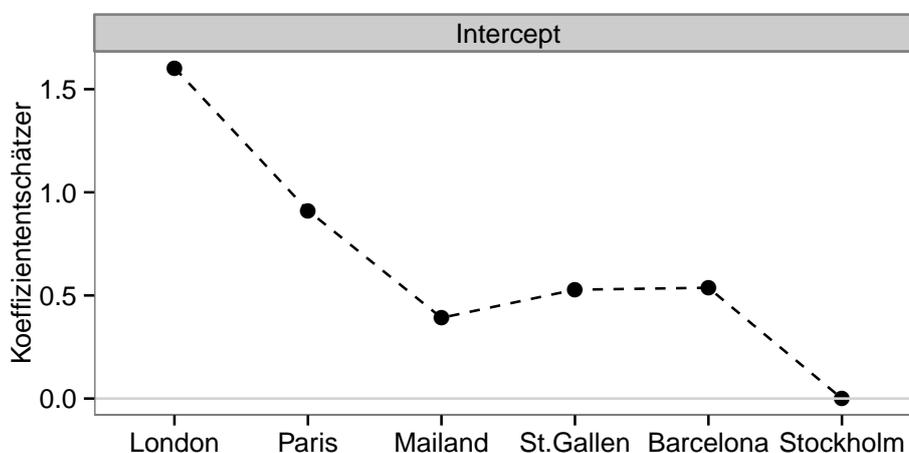


Abbildung 6.8.: Koeffizientenschätzer des Generalisierten Linearen Modells ohne Kovariablen für CEMSChoice-Datensatz.

Um zu überprüfen, ob und inwieweit Sprachkenntnisse einen Einfluss bei der Präferenzentscheidung haben, wird im Folgenden das Generalisierte Lineare Modell mit zusätzlichen subjekt-spezifischen Kovariablen betrachtet. In Unterabschnitt 6.2.2 wird daraufhin untersucht, welche Kovariablen einen entscheidenden Einfluss bei der Präferenzwahl besitzen bzw. ob es Kovariablen gibt, die sich bei der Präferenzentscheidung als unwichtig herausstellen.

Berücksichtigung von Kovariablen

Die Schätzung von Paarvergleichsdaten ohne Berücksichtigung von Kovariablen lässt bereits vermuten, dass bei den Präferenzentscheidungen der Studenten der Wirtschaftsuniversität Wien subjekt-spezifische Kovariablen einen starken Einfluss haben. Bei der Schätzung des CEMSChoice-Datensatzes werden somit als zusätzliche Kovariablen binäre

Größen bezüglich Sprachkenntnissen (`english`, `french`, `spanish`, `italian`), Studienrichtung (`study`), Arbeitssituation (`work`), internationaler Abschluss (`intdegree`) und Geschlecht (`gender`) eingebunden. Dabei ist die Standardisierung der Kovariablen nicht zu vergessen.

Die Koeffizientenschätzer, die man bei der Schätzung von Paarvergleichsdaten über Generalisierte Lineare Modelle mit Kovariablen erhält, sind in Tabelle 6.5 dargestellt. Auch hier wird aus Gründen der Identifizierbarkeit das letzte Item (Universität in Stockholm) auf Null gesetzt. Bei den binären Größen fallen Personen mit guten Sprachkenntnissen, Personen mit anderem Studienschwerpunkt als Handel, Personen, die keine Vollzeitjob neben dem Studium haben, Frauen und Personen, die keinen internationalen Abschluss erlangen möchten in die jeweiligen Referenzkategorien.

	London	Paris	Mailand	St. Gallen	Barcelona	Stockholm
Intercept	1.7235	1.0110	0.4579	0.5652	0.5979	0
study	-0.0643	-0.3686	0.0899	0.2159	0.0254	0
english	0.0127	0.0388	0.0942	0.1688	-0.1083	0
french	-0.2180	-0.6647	-0.2104	-0.1476	-0.0260	0
spanish	-0.0482	0.0400	-0.0865	-0.1198	-0.5299	0
italian	-0.2419	0.0313	-0.7729	-0.0788	-0.1048	0
work	0.0456	0.2876	0.1814	-0.0164	0.2087	0
gender	-0.0863	-0.1279	-0.1808	-0.0120	-0.0979	0
intdegree	-0.1446	-0.0375	-0.0925	0.1634	-0.0778	0

Tabelle 6.5.: Koeffizientenschätzer des Generalisierten Linearen Modells mit Kovariablen für CEMSChoice-Datensatz.

Eine übersichtlichere Darstellung der Schätzergebnisse ist in Abbildung 6.9 gegeben. Zur Einschätzung der Verhältnisse ist jeweils die Nulllinie eingezeichnet. Die Effekte aus den einzelnen Paarvergleichen (Intercept) zeigen leichte Veränderungen gegenüber der Schätzung ohne Kovariablen, die daraus resultierende Rangreihung der Universitäten bleibt jedoch unverändert.

Zunächst interessiert der Einfluss der Sprachkenntnisse der Studenten, bei denen jeweils die Personen mit guten Sprachkenntnissen in der jeweiligen Sprache in der Referenzkategorie sind. Die einzelnen Effekte können, wie in Gleichung (6.1), über die Odds-Ratio interpretiert werden. Somit werden die Präferenzentscheidungen der Personen mit schlechten Sprachkenntnissen in der jeweiligen Sprache mit den Präferenzentscheidungen der Personen mit guten Sprachkenntnissen verglichen. Aus Abbildung 6.9 kann bei den Kovariablen

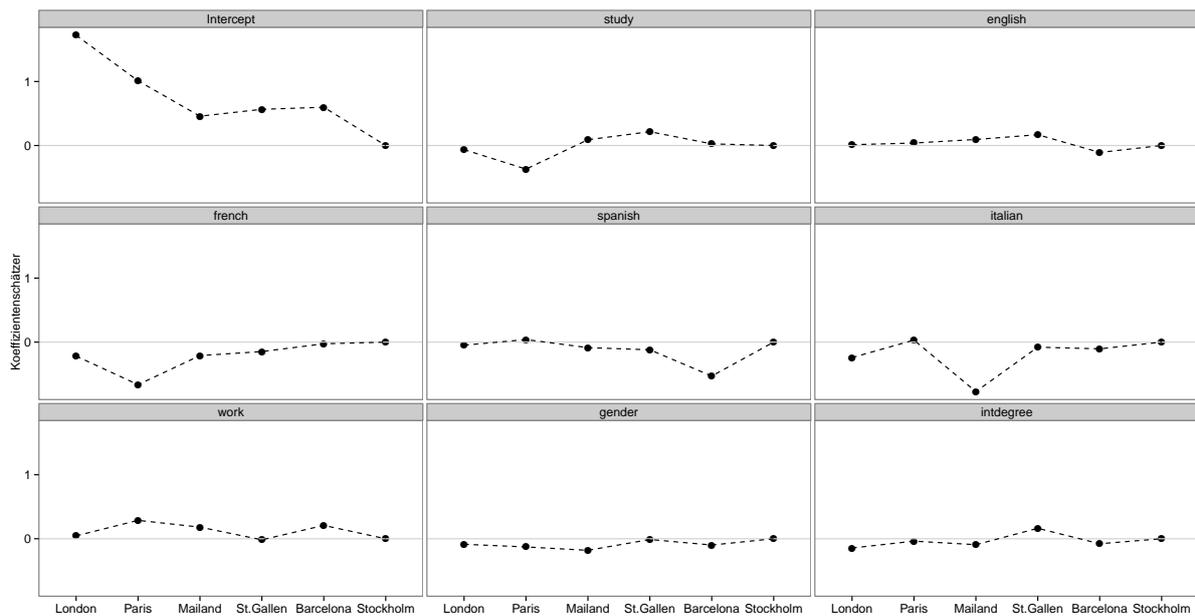


Abbildung 6.9.: Koeffizientenschätzer des Generalisierten Linearen Modells mit Kovariablen für CEMSChoice-Datensatz.

`french`, `spanish` und `italian` eindeutig eine Abneigung gegenüber den Universitäten in den entsprechenden Ländern festgestellt werden. Hat zum Beispiel ein Student schlechte Französischkenntnisse präferiert er alle anderen Universitäten gegenüber der Universität in Paris. Ausnahme dieser Erkenntnis sind schlechte Sprachkenntnisse in Englisch. Hierbei liegt eine Abneigung gegenüber der Universität in Barcelona vor, die beliebteste Universität hingegen ist die in St. Gallen, in der deutschsprachigen Schweiz. Die Chance sich bei schlechten Englischkenntnissen beim Paarvergleich der Universitäten in St. Gallen und London für St. Gallen zu entscheiden, ist um 17% höher.

Betrachtet man den Einfluss der Kovariable `study`, mit Referenzkategorie `other` (economics, business administration, business education), so sind in Abbildung 6.9 die Präferenzentscheidungen der Studenten mit Schwerpunkt in Handel (`commerce`) dargestellt. Ein Student mit Handelsschwerpunkt präferiert die Universität in St. Gallen, gefolgt von der Universität in Mailand. Betrachtet man das Chancenverhältnis beim Paarvergleich der beiden Universitäten, ist die Chance sich für die Universität in St. Gallen zu entscheiden um 13% höher. Die Universität in Paris steht bei Studenten mit Handelsschwerpunkt an letzter Stelle. Anhand dieser Ergebnisse ist es schwierig einen Zusammenhang zwischen den Präferenzentscheidungen der Studenten und der jeweiligen Spezialisierung der Universitäten, siehe Tabelle 6.4, herzustellen.

Die Kovariable `Geschlecht`, mit Referenzkategorie `Frauen`, zeigt in Abbildung 6.9 sehr ge-

ringe Effekte. Der Kurvenverlauf stellt die Präferenzen der männlichen Studenten dar, die die Universität in Stockholm präferieren. Die Variable `intdegree` zeigt, dass Studenten, die die Absicht haben einen internationalen Abschluss zu erreichen, sich am ehesten für die Universität in St. Gallen entscheiden. Die Studenten, die neben dem Studium in Vollzeit arbeiten, präferieren die Universität in Paris.

Die Schätzergebnisse zeigen, dass die Sprachkenntnisse der Studenten die Entscheidung bei den Paarvergleichen wesentlich beeinflussen. Im nächsten Unterabschnitt 6.2.2 wird untersucht, ob diese wirklich einen entscheidenden Einfluss auf die Präferenzwahl haben oder ob es Kovariablen gibt, deren Einfluss als uninformativ geschätzt wird.

6.2.2. Anwendung BTL-Modell über Group-LASSO

Schätzt man die Paarvergleichsdaten des CEMSChoice-Datensatzes über das Group-LASSO Verfahren, ist von wesentlichem Interesse, ob einige Kovariablen bei der Präferenzentscheidung irrelevant sind. Entsprechend den Ergebnissen aus Unterabschnitt 6.2.1 sollten keine Kovariablen, die Informationen über Sprachkenntnisse beinhalten, betroffen sein.

Zur Bestimmung des besten Modells sind die Fälle

- Freiheitsgradberechnung über das Active Set, Selektionskriterium AIC,
- Freiheitsgradberechnung über das Active Set, Selektionskriterium BIC,
- Freiheitsgradberechnung über die Approximation von [Yuan and Lin \[2006\]](#), Selektionskriterium AIC und
- Freiheitsgradberechnung über die Approximation von [Yuan and Lin \[2006\]](#), Selektionskriterium BIC

zu unterscheiden. Die Ergebnisse werden im Folgenden kurz vorgestellt.

Beginnend mit der Freiheitsgradberechnung über das Active Set fällt bei beiden Selektionskriterien AIC und BIC die Wahl auf das 40. Modell, bei dem die Schätzer aufgrund des Penalierungsparameters $\lambda = 0.001$ den ML-Schätzern aus Tabelle 6.5 entsprechen. Somit gelten auch die Interpretationen aus Unterabschnitt 6.2.1.

Die Freiheitsgradberechnung über die Approximation aus Gleichung (5.12) führt beim Selektionskriterium AIC zur Wahl des 39. Modells, bei dem λ den Wert 2.57 annimmt. Bei der Schätzung über Group-LASSO wird demnach eine geringe Penalisierung berücksichtigt, sodass die Koeffizientenschätzer gegenüber den ML-Schätzern aus Tabelle 6.5

einen leichten Shrinkage-Effekt enthalten. Allerdings wird keine der acht Kovariablen als Uninformativ eingestuft und folglich aus dem Modell genommen.

Zum Schluss wird noch die Anpassung über Group-LASSO mit Freiheitsgradberechnung über die Approximation von [Yuan and Lin \[2006\]](#) und Selektionskriterium BIC betrachtet. Gegenüber dem AIC bestraft das BIC komplexere Modelle stärker und man erhält sparsamere Modelle. In diesem Fall wird über das BIC das 36. Modell mit $\lambda = 10.28$ ausgewählt. Somit handelt es sich beim besten Modell über das BIC um ein stärker penalisiertes, dennoch wird keine der Kovariablen aus dem Modell ausgeschlossen. Somit weisen die Koeffizientenschätzer einen stärkeren Shrinkage-Effekt auf und der Einfluss der einzelnen Kovariablen auf die Präferenzwahl lässt sich graphisch, wie in [Abbildung 6.9](#), in leicht abgeflachten Kurven darstellen.

Die Anwendung auf den CEMSChoice-Datensatz hat gezeigt, dass bei der Wahl zwischen Universitäten für ein Auslandssemester insbesondere die Sprachkenntnisse der Studenten eine entscheidende Rolle spielen. Aber auch der Studienschwerpunkt, das Geschlecht, die Absicht einen internationalen Abschluss zu erreichen und die Vollzeitarbeit neben dem Studium haben offensichtlich einen Einfluss, der nicht vernachlässigt werden sollte. Auch bei der Schätzung über Group-LASSO, das durch Modellregularisierung und Variablenselektion möglichst sparsame, interpretierbare Modelle generiert, wurde keine der Kovariablen selektiert.

Für diesen Datensatz wird in [Unterabschnitt 6.2.3](#) ein Vergleich der alternativen Schätzverfahren für Bradley-Terry-Luce Modelle mit der Schätzung über log-lineare Modelle nach [Dittrich et al. \[1998\]](#) durchgeführt. Zum Vergleich der Ergebnisse des CEMSChoice-Datensatzes aus [Unterabschnitt 6.2.1](#) mit der Schätzung über modellbasiertes rekursives Partitionieren, siehe [Unterabschnitt 6.1.3](#), sei auf [Strobl et al. \[2009\]](#) verwiesen.

6.2.3. Vergleich mit Verfahren über Log-Lineare Modelle

Das Anwendungsbeispiel auf den CEMSChoice-Datensatz hat gezeigt, dass subjekt-spezifische Kovariablen bei den Präferenzentscheidungen für eine Universität einen entscheidenden Einfluss haben. [Dittrich et al. \[1998\]](#) stellen eine Methode vor, bei der sowohl subjekt- als auch objekt-spezifische Informationen bei der Schätzung von Bradley-Terry-Luce Modellen berücksichtigt werden können. Hierfür werden BTL-Modelle über loglineare Modelle (siehe, z.B., [Sinclair \[1982\]](#), [Agresti \[2002\]](#)) geschätzt. Diese werden als loglineare Bradley-Terry Modelle (LLBT) bezeichnet. In [Dittrich et al. \[1998\]](#) erfolgt die Anwendung ebenfalls auf den CEMSChoice-Datensatz, sodass auch hier aufgrund der selben Datenbasis ein Vergleich der Methoden durchgeführt wird. Es ist zu beachten, dass bei der Schät-

zung über regularisierte Verfahren keine Unentschieden bei den Präferenzentscheidungen berücksichtigt wurden und somit der Vergleich nur eingeschränkt erfolgen kann.

Über LLBT können fünf unterschiedliche Datensituationen geschätzt werden: (i) einfacher Paarvergleich, (ii) Paarvergleich mit Unentschieden, (iii) Paarvergleich mit subjekt-spezifischen Kovariablen, (iv) Paarvergleich mit objekt-spezifischen Kovariablen und (v) Paarvergleich mit geordneten Effekten. Beim CEMSCochoice-Datensatz sind insbesondere die Fälle (ii), (iii) und (iv) von Interesse.

Die Methodik der loglinearen Schätzung von Bradley-Terry-Luce Modellen wird im Folgenden kurz vorgestellt, für detaillierte Ausführungen sei auf [Dittrich et al. \[1998\]](#) sowie auf [Dittrich and Hatzinger \[2009\]](#) verwiesen. [Fienberg and Larntz \[1976\]](#) und [Fienberg \[1979\]](#) geben eine detaillierte Beschreibung der logistischen Formulierung von BT-Modellen (basierend auf der Binomialverteilung) und ihrer loglinearen Darstellung (basierend auf der Poisson-Verteilung). Ein alternativer Ansatz ist in [Sinclair \[1982\]](#) gegeben. Darauf aufbauend ergibt sich für die Datensituation (i) einfacher Paarvergleichsmodelle ohne Kovariablen das loglineare Modell

$$\ln(w_{(jk)j}) = \mu_{(jk)} + \lambda_j^O - \lambda_k^O,$$

$$\ln(w_{(jk)k}) = \mu_{(jk)} - \lambda_j^O + \lambda_k^O.$$

Mit $w_{(jk)j} = n_{(jk)}p_{(jk)j}$ wird dabei die erwartete Anzahl an Vergleichen bezeichnet, bei denen Universität j gegenüber Universität k präferiert wird. Die μ 's können als Interaktionsparameter bezüglich der Universitäten im Vergleich interpretiert werden. Die λ^O 's repräsentieren die Objekt-Parameter und können mit den π 's aus Gleichung (3.1) über $\ln \pi = 2\lambda^O$ in Beziehung gebracht werden. Für die Verteilungen der Zufallsvariablen $Y_{(jk)j}$ und $Y_{(jk)k}$ wird in der loglinearen Formulierung entsprechend die Poisson-Verteilung angenommen.

Für das Anwendungsbeispiel auf den CEMSCochoice-Datensatz ist die Formulierung des loglinearen BTL-Modells mit Unentschieden, Datensituation (ii), von Relevanz. Diese können einfach über die Spezifikation von [Davidson and Beaver \[1977\]](#) in LLBT übertragen werden

$$\ln(w_{(jk)j}) = \mu_{(jk)} + \lambda_j^O - \lambda_k^O,$$

$$\ln(w_{(jk)0}) = \mu_{(jk)} + \delta,$$

$$\ln(w_{(jk)k}) = \mu_{(jk)} - \lambda_j^O + \lambda_k^O,$$

wobei $w_{(jk)j}$ für die erwartete Anzahl an Präferenzen für Universität j , $w_{(jk)k}$ für die erwartete Anzahl an Präferenzen für Universität k und $w_{(jk)0}$ für die erwartete Anzahl keiner Entscheidung beim Vergleich (jk) steht und δ den Unentschieden Effekt darstellt.

In Tabelle 6.6 ist beispielhaft für drei Objekte die Designstruktur bei loglinearen Modellen mit Unentschieden aufgezeigt. Dabei sind in den Spalten die Anteile y , ein Faktor μ , ein Dummy für δ und die Items O_1, O_2 und O_3 für λ_1^O, λ_2^O und λ_3^O dargestellt.

Vergleich	Entscheidung	Bezeichnung	μ	δ	λ_1^O	λ_2^O	λ_3^O
(12)	O_1	$y_{(12)1}$	1	0	1	-1	0
(12)	no	$y_{(12)0}$	1	1	0	0	0
(12)	O_2	$y_{(12)2}$	1	0	-1	1	0
(13)	O_1	$y_{(13)1}$	2	0	1	0	-1
(13)	no	$y_{(13)0}$	2	1	0	0	0
(13)	O_3	$y_{(13)3}$	2	0	-1	0	1
(23)	O_2	$y_{(23)2}$	3	0	0	1	-1
(23)	no	$y_{(23)0}$	3	1	0	0	0
(23)	O_3	$y_{(23)3}$	3	0	0	-1	1

Tabelle 6.6.: Designstruktur eines loglinearen Bradley-Terry Modells mit Unentschieden.

Durch die Berücksichtigung von subjekt-spezifischen Kovariablen können für verschiedene Gruppen unterschiedliche Präferenzskalen entstehen. Ausgangspunkt ist auch hier das LLBT mit Unentschieden. Zur Illustration des Ansatzes mit subjekt-spezifischen Kovariablen wird *eine kategoriale Kovariable* S mit l Levels, $l = 1, 2, \dots, L$ angenommen. Sei $w_{(jk)j|l}$ die erwartete Anzahl an Präferenzen für Item j für Personen in der Kovariablenklasse l . Die loglineare Darstellung dieses erweiterten Bradley-Terry Modells ist in folgenden Gleichungen gegeben:

$$\begin{aligned} \ln(w_{(jk)j|l}) &= \mu_{(jk)|l} + \lambda_j^O - \lambda_k^O + \lambda_l^S + \lambda_{jl}^{OS} - \lambda_{kl}^{OS} \\ \ln(w_{(jk)0|l}) &= \mu_{(jk)|l} + \lambda_l^S + \delta \\ \ln(w_{(jk)k|l}) &= \mu_{(jk)|l} - \lambda_j^O + \lambda_k^O + \lambda_l^S - \lambda_{jl}^{OS} + \lambda_{kl}^{OS} \end{aligned}$$

Der Parameter λ_l^S stellt den Haupteffekt der subjekt-spezifischen Kovariable bezüglich des l -ten Levels dar; λ_{jl}^{OS} und λ_{kl}^{OS} sind die interessierenden subjekt-objekt Interaktionsparameter, die den Effekt der subjekt-spezifischen Kovariable im l -ten Level auf die Präferenz der Objekte j und k beschreibt. δ ist wieder der Unentschieden Effekt. Eine ausführlichere Beschreibung ist in [Dittrich et al. \[1998\]](#) gegeben.

Bei der Einbindung von objekt-spezifischen Kovariablen, Datensituation (iv), sei wiederum als Ausgangssituation der Paarvergleich mit Unentschieden, Datensituation (ii), vorausgesetzt. Die Idee ist, die objekt-spezifischen Kovariablen über eine Linearkombination von P Kovariablen X_1, \dots, X_P darzustellen. Hierfür wird der objekt-bezogene Parameter λ_j^O durch die lineare Formulierung

$$\lambda_j^O = \sum_{p=1}^P x_{jp} \beta_p^X$$

ersetzt, wobei die x_{jp} 's für die Kovariablen mit der p -ten Eigenschaft von Objekt j stehen und die β^X 's unbekannte Regressionsparameter sind. Das LLBT mit einer objekt-spezifischen Kovariable lautet:

$$\begin{aligned} \ln(w_{(jk)j}) &= \mu_{(jk)} + \lambda_j^O - \lambda_k^O \\ &= \mu_{(jk)} + \beta_1^X x_{j1} - \beta_1^X x_{k1} \\ &= \mu_{(jk)} + \beta_1^X (x_{j1} - x_{k1}) \end{aligned}$$

Das LLBT wird wieder über das Generalisierte Lineare Modell mit poissonverteilten Fehler und Logit-Link geschätzt [Dittrich et al., 1998; Dittrich and Hatzinger, 2009].

Nun können für den CEMSChoice-Datensatz verschiedene Modelle über den loglinearen Ansatz angepasst werden, deren Ergebnisse mit den Ergebnissen der regularisierten Schätzungen verglichen werden können. Die Schätzung von LLBT Modellen kann in R mittels den R-Paketen `prefmod` [Hatzinger, 2012] und `gnm` [Turner and Firth, 2012] erfolgen.

Keine Berücksichtigung von Kovariablen

Zunächst wird das Paarvergleichsmodell über LLBT mit Unentschieden aber ohne zusätzliche Kovariablen betrachtet, Datensituation (ii). Man erhält die folgenden Schätzer:

	London	Paris	Mailand	St. Gallen	Barcelona	Stockholm	Unentschieden
Schätzer:	0.911	0.516	0.221	0.299	0.307	0	-1.317
Parameter:	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	δ

Tabelle 6.7.: Koeffizientenschätzer des loglinearen Bradley-Terry Modells ohne Kovariablen für CEMSChoice-Datensatz.

Der negative Wert beim Unentschieden Effekt deutet auf eine starke Tendenz für eine Entscheidung im Paarvergleich hin. Man erkennt eine starke Präferenz für London, mit Paris auf dem zweiten Platz, gefolgt von Barcelona, St. Gallen, Mailand und Stockholm. Anhand eines Goodness-of-Fit Tests zeigt sich, dass die Anpassung des Modells schlecht

ist: Aus der **Residual deviance**: 77.282 mit $df = 24$ Freiheitsgraden resultiert ein sehr kleiner p-Wert von 0, so dass die Anpassung des Modells nicht in Ordnung geht. Ein möglicher Grund könnte der Einfluss von subjekt-spezifischen Kovariablen sein, welche hier nicht berücksichtigt wurden.

Bei der Berechnung der Worth-Parameter ist zu beachten, dass Bradley-Terry Modelle skaleninvariant sind und dass aus Identifikationsgründen die Bedingung $\sum_i \pi_i = 1$ gilt. Die Beziehung zwischen den λ 's und den Worth-Parametern π ist gegeben durch

$$\pi_j = \frac{\exp\{2\lambda_j^O\}}{\sum_i \exp\{2\lambda_i^O\}}, \quad j = 1, 2, \dots, J.$$

Die Schätzer der Worth-Parameter für LLBT Modelle ohne Kovariablen sind in folgender Tabelle gegeben.

	London	Paris	Mailand	St. Gallen	Barcelona	Stockholm
π_i	0.407	0.185	0.102	0.120	0.121	0.066

Graphisch können diese wie in Abbildung 6.10 dargestellt werden.

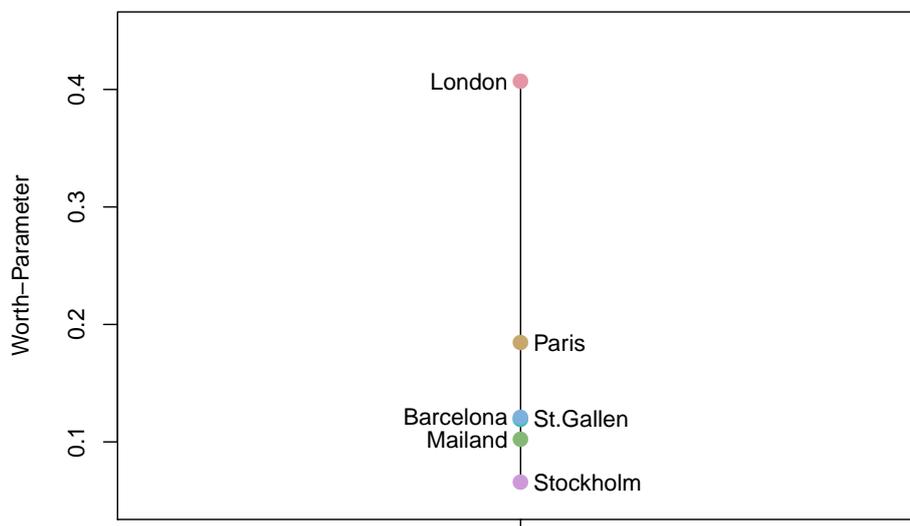


Abbildung 6.10.: Präferenzskala der Universitäten bei Anpassung eines loglinearen Bradley-Terry Modells ohne Kovariablen.

Vergleicht man dieses Ergebnis mit dem Ergebnis aus Unterabschnitt 6.2.1, der Schät-

zung über das Generalisierte Lineare Modell ohne Berücksichtigung von Kovariablen, ist zu erkennen, dass die resultierende Präferenzskala bei den Paarvergleichsmodellen ohne Kovariablen dieselbe ist, obwohl bei der Schätzung über Generalisierte Lineare Modelle keine Unentschieden berücksichtigt wurden.

Berücksichtigung von Kovariablen

Der Vorteil bei der Modellierung über LLBT ist die direkte Einbindung von subjekt- und objekt-spezifischen Kovariablen. Für den CEMSCochoice-Datensatz existieren sowohl subjekt-spezifische Kovariablen, wie etwa das Geschlecht der Studenten, der Studienschwerpunkt und Sprachkenntnisse in ausgewählten Fremdsprachen, als auch objekt-spezifische Kovariablen, siehe Tabelle 6.4, wie zum Beispiel die Studienausrichtung einer Universität. Damit die Ergebnisse aus den Schätzungen über Generalisierte Lineare Modelle und über Group-LASSO mit den Ergebnissen der Schätzung über LLBT verglichen werden können, wird hier nur der Fall betrachtet, bei dem alle subjekt-spezifischen Kovariablen berücksichtigt werden.

Die resultierenden Koeffizientenschätzer sind in Tabelle 6.8 dargestellt:

	London	Paris	Mailand	St. Gallen	Barcelona	Stockholm
Intercept	1.5785	1.1359	1.4330	0.4002	1.3973	0
study	-0.1030	-0.4858	0.0291	0.2138	-0.1060	0
english	-0.0317	0.0330	0.1294	0.1858	-0.1221	0
french	-0.2606	-0.8030	-0.2635	-0.1931	-0.0354	0
spanish	-0.0150	0.0730	-0.1005	-0.1831	-0.8589	0
italian	-0.3871	0.0219	-1.1421	-0.1441	-0.2323	0
work	0.2550	0.8525	0.6302	-0.0008	0.6697	0
gender	-0.0738	-0.1479	-0.1891	0.0179	-0.1118	0
intdegree	-0.1513	-0.0217	-0.0469	0.2227	-0.0756	0

Tabelle 6.8.: Koeffizientenschätzer des loglinearen Bradley-Terry Modells mit Kovariablen.

Zudem wird der Unentschieden Effekt mit einem Wert von -1.2437 geschätzt. Das bedeutet, dass sich die Studenten bevorzugt für eine eindeutige Entscheidung im Paarvergleich entscheiden und selten keine Präferenz angeben.

Normalerweise werden die Schätzungen von loglinearen Bradley-Terry Modellen in Worth-Parameter umgerechnet und die resultierenden Präferenzskalen werden wie in Abbildung 6.10 dargestellt. Bei vielen Kovariablen ist diese Darstellung schwierig, sodass hier zum graphischen Vergleich der Schätzergebnisse, dieselbe Darstellung wie bei der Schätzung

über GLM und Group-LASSO gewählt wird. Zum besseren Vergleich werden sowohl die Koeffizientenschätzer aus der Schätzung über LLBT (rot) als auch aus der Schätzung über GLMs (schwarz) in Abbildung 6.11 dargestellt. Es sei noch einmal darauf hingewiesen, dass bei der Schätzung über LLBT die Unentschieden mit berücksichtigt werden können.

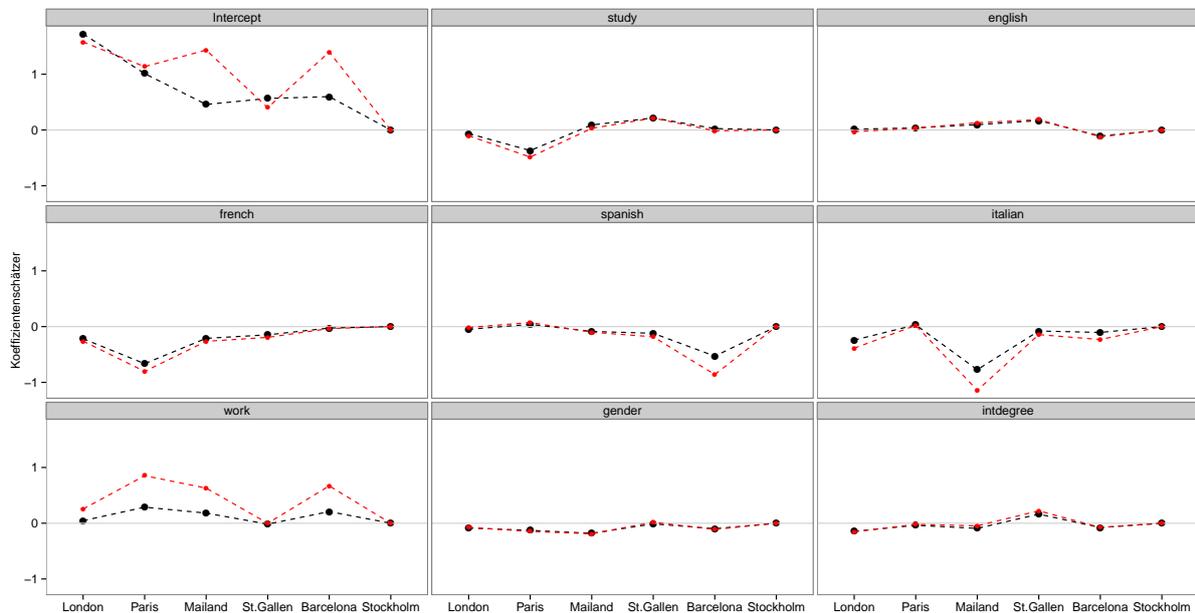


Abbildung 6.11.: Vergleich der Koeffizientenschätzer zwischen loglinearen Bradley-Terry Modell (rot) und Generalisiertem Linearem Modell (schwarz) mit subjekt-spezifischen Kovariablen für CEMSChoice-Datensatz.

Die abgetragenen Präferenzordnungen zeigen den Unterschied gegenüber den Personen in den jeweiligen Referenzkategorien. Demnach sind bei den Kovariablen bezüglich der Sprachkenntnisse jeweils die Präferenzen der Personen dargestellt, welche in der angegebenen Sprache schlechte Kenntnisse besitzen, bei der Kovariable Geschlecht ist die Präferenzskala der Männer dargestellt.

Beim Vergleich der Ergebnisse fällt auf, dass nahezu die selben Präferenzordnungen resultieren. Auch bei der Schätzung über LLBT zeigt sich, dass die Sprachkenntnisse einen großen Einfluss auf die Wahl einer Universität haben. Interessant ist, dass der Intercept nun eine andere Präferenzskala aufzeigt, als bei der Betrachtung ohne Kovariablen und der Schätzung über GLMs. Somit lässt sich zusammenfassen, dass eine Berücksichtigung zusätzlicher Kovariablen bei der Schätzung von Paarvergleichen sinnvoll ist, da für Personengruppen mit unterschiedlichen Eigenschaften sehr wohl differenzierte Präferenzskalen resultieren. Alternative Schätzverfahren für BTL-Modellen, sei es über GLMs, Group-LASSO oder LLBT, stellen eine einfache Möglichkeit dar, personenbezogene Informationen direkt in die Schätzung mit einzubeziehen.

7. Zusammenfassung und Ausblick

Paarvergleiche, die zur Bewertung nicht direkt messbarer, subjektiver Kriterien durchgeführt werden, werden in der angewandten Statistik häufig über das Bradley-Terry-Luce Modell angepasst. Über die Forced-Choice Befragungstechnik resultiert für das zu bewertende subjektive Kriterium eine eindeutige Rangordnung. Oft wird davon ausgegangen, dass die Präferenzentscheidungen der Personen nicht nur von den Eigenschaften der zu beurteilenden Stimuli, sondern auch von den Charaktereigenschaften der Personen selbst abhängen. Diese Informationen können bei der Schätzung über das Bradley-Terry-Luce Modell lediglich durch separate Modellanpassungen, nicht aber direkt im Modell berücksichtigt werden. Abhilfe verschaffen alternative Schätzverfahren wie Generalisierte Lineare Modelle oder Group-LASSO. Die dadurch erhaltenen Ergebnisse werden im Folgenden kurz zusammengefasst und ein Ausblick auf weitere Analysemöglichkeiten gegeben.

Die Schätzung von Paarvergleichsdaten über Generalisierte Lineare Modelle in Kapitel 4 ermöglicht im Gegensatz zur Schätzung über BTL-Modelle eine einfache Einbindung von Kovariablen. Dass sich das GLM als alternatives Schätzverfahren für BTL-Modelle eignet, konnte im Anwendungsbeispiel auf den Topmodel-Datensatz, Abschnitt 6.1, gezeigt werden, da bei der Schätzung ohne Kovariablen, die selben Schätzer resultieren. Ebenfalls wurde deutlich, dass subjekt-spezifische Kovariablen durchaus einen Einfluss auf die Präferenzentscheidungen der Personen haben. Schwierigkeiten ergeben sich bei einer hohen Anzahl an Kovariablen p für einen relativ kleinen Stichprobenumfang n . In entsprechenden Fällen liefert die Anpassung über Generalisierte Lineare Modelle keine zufriedenstellenden Ergebnisse, so dass die Anwendung auf Fälle mit der Bedingung $p < n$ beschränkt ist.

Beim zweiten vorgestellten alternativen Schätzverfahren für Bradley-Terry-Luce Modelle über Group-LASSO, siehe Kapitel 5, handelt es sich um ein Shrinkage-Verfahren. Im Gegensatz zu Verfahren, bei denen Parameter getrennt voneinander penalisiert werden, können über Regularisierungstechniken gesamte Kovariablen selektiert werden. Die Möglichkeit der direkten Variablenselektion führt zu einer Verringerung der Modellkomplexität, während die im Modell verbleibenden Koeffizienten geschrumpft werden. Die Anwendung auf Paarvergleichsdaten wurde anhand von Simulationsstudien überprüft, mit dem Ergebnis, dass Group-LASSO Modelle gute Vorhersagen liefern und somit zu Recht als alter-

natives Schätzverfahren für BTL-Modelle herangezogen werden können. Es konnte jedoch kein allgemein gültiger Modellierungsansatz gefunden werden, da bei einigen Simulationen die Freiheitsgradberechnung über das Active Set und bei anderen die Freiheitsgradberechnung über die Approximation für logistische Group-LASSO Modelle bessere Vorhersagen ermöglichte. Dennoch konnte festgestellt werden, dass eine große Anzahl an Items als auch an Kovariablen bessere Ergebnisse unterstützen. Auch der Fall $p > n$ bereitet bei der Schätzung keine Probleme. Bei der Anwendung auf den Topmodel-Datensatz erfolgte durch den Ausschluss einer Kovariable ein sparsameres Modell gegenüber der Anpassung mittels Generalisierter Linearer Modelle. Dieses Ergebnis veranschaulicht, dass auch hier einzelne subjekt-spezifische Informationen offensichtlich einen Einfluss auf die Präferenzwahl haben.

Zusammenfassend konnte gezeigt werden, dass die Schätzung von Paarvergleichsdaten sowohl über Generalisierte Lineare Modelle als auch über Group-LASSO erfolgen kann, die jeweils adäquate Schätzergebnisse erzielen. Dabei ist lediglich die Paarvergleichsstruktur in den Kontext parametrischer Modelle zu übertragen. Der wesentliche Vorteil gegenüber BTL-Modellen ist die einfache Einbindung von Kovariablen, deren Einfluss auf die Präferenzentscheidungen nachgewiesen werden konnte.

Betrachtet man abschließend die Methodik des paarweisen Vergleichs wird ersichtlich, dass es viele Möglichkeiten zu Erweiterungen gibt. Insbesondere fehlt den Anpassungen über GLM und Group-LASSO die Fähigkeit Paarvergleichsmodelle mit Unentschieden schätzen zu können. Beim Anwendungsbeispiel auf den CEMSChoice-Datensatz in Abschnitt 6.2 mussten deshalb die Unentschieden auf eindeutige Präferenzen aufgeteilt werden. Auch der Vergleich der Methoden mit der Schätzung über loglineare Modelle konnte somit nur beschränkt erfolgen. Hingegen können die Schwierigkeiten beim Vergleich der Datenmodelle mit dem Algorithmischen Modell des modellbasierten rekursiven Partitionierens in Unterabschnitt 6.1.3 nicht durch Erweiterungen genommen werden, da die statistischen Modellierungsansätze unterschiedlich motiviert sind.

Ein interessanter Aspekt bei der Schätzung über Group-LASSO ist die Wahl des optimalen λ 's, die anstatt über das BIC auch über Kreuzvalidierung erfolgen kann, mit dem kleinen Nachteil der etwas längeren Laufzeit. Trotzdem könnte ein Vergleich der beiden Methoden durchgeführt werden. Weiter sind neben der Schätzung über Group-LASSO Anpassungen über das Elastic Net oder über Group-Bridge denkbar. Weitere Möglichkeiten Bradley-Terry-Luce Modelle zu schätzen sind über nichtparametrische Modellanpassung oder Modelle mit ordinalem Response gegeben.

A. Anhang

Zunächst ist in Abschnitt [A.1](#) der mathematische Beweis von Lemma 1 aufgeführt. Im Anschluss folgt in Abschnitt [A.2](#) der R-Code der Simulationsmasken, unterteilt in BTL-Modell über GLM in Abschnitt [A.2.1](#) und BTL-Modell über Group-LASSO in Abschnitt [A.2.2](#). Weiter sind in Abschnitt [A.3](#) zusätzliche Simulationsergebnisse dargestellt. Der hier vorliegenden Masterarbeit ist zudem eine CD mit dem Elektronischen Anhang beigelegt, der in Abschnitt [A.4](#) genauer erläutert wird.

A.1. Beweis von Lemma 1

Zunächst wird der Intercept aus dem Modell genommen. Es wird angenommen, dass β_1, \dots, β_G fest sind. Um den Schätzer für den Intercept zu erhalten, muss eine Funktion der Form

$$g(\beta_0) = - \sum_{i=1}^n [y_i(\beta_0 + c_i) - \log\{1 + \exp(\beta_0 + c_i)\}]$$

minimiert werden, mit der Ableitung

$$g'(\beta_0) = - \sum_{i=1}^n \left\{ y_i - \frac{\exp(\beta_0 + c_i)}{1 + \exp(\beta_0 + c_i)} \right\},$$

wobei $c_i = \sum_{g=1}^G \mathbf{x}_{i,g}^T \beta_g$ eine Konstante ist. Es gilt $\lim_{\beta_0 \rightarrow \inf} g'(\beta_0) = n - \sum_{i=1}^n y_i > 0$ und $\lim_{\beta_0 \rightarrow -\inf} g'(\beta_0) = - \sum_{i=1}^n y_i < 0$. Zudem ist $g'(\cdot)$ stetig und strikt ansteigend. Somit gibt es ein eindeutiges $\beta_0^* \in \mathbb{R}$ für das $g'(\beta_0^*) = 0$ gilt. Beim Satz über implizite Funktionen ist die zugehörige Funktion $\beta_0^*(\beta_1, \dots, \beta_G)$ stetig differenzierbar. Ersetzt man β_0 in $S_\lambda(\beta)$ durch die Funktion $\beta_0^*(\beta_1, \dots, \beta_G)$ und verwendet man das Dualitätstheorem, so kann Gleichung [\(5.6\)](#) als Optimierungsproblem unter der Voraussetzung $\sum_{g=1}^G \|\beta_g\|_2 \leq t$ für ein $t > 0$ umformuliert werden. Dies ist dann ein Optimierungsproblem einer stetigen Funktion über eine kompakte Menge; deshalb wird das Minimum erreicht [\[Meier et al., 2008\]](#).

A.2. R-Code

A.2.1. Simulationsmaske: BTL-Modell über GLM

Laden der Funktion, mit der die Designmatrix erstellt wird.

```
> source("Designmatrix.R")
```

Bei der Simulation sind folgende Größen vom Anwender anzugeben:

- *m*: Anzahl zu vergleichender Objekte/Items
- *n*: Anzahl Personen
- *B*: Anzahl binärer Kovariablen im Modell
- *C*: Anzahl metrischer Kovariablen im Modell
- *p*: Anzahl Simulationsdurchläufe

Die Anzahl der Paarvergleiche, die mit *I* bezeichnet wird, wird aus der Anzahl der Objekte/Items über die Formel $I \leftarrow m \cdot (m-1) / 2$ berechnet. Mit diesen Angaben kann auch die Designmatrix an die gewünschte Simulation angepasst werden.

```
> X <- designmatrix(m, n, lastcol=FALSE)
```

Zum Abspeichern der Simulationsergebnisse sind leere Listen zu erstellen.

```
> gamma <- list()
> means_gamma <- list()
```

Über folgende Simulationsmaske können Bradley-Terry-Luce Modelle über Generalisierte Lineare Modelle angepasst werden.

```
> for(i in 1:p){
+   set.seed(i)
+
+   X_neu <- X
+   # Schleife zur Erstellung binärer Kovariablen
+   for(w in 1:B){
+     bin <- scale(sample(c(0,1), n, replace=TRUE))
+     binvec <- rep(bin, I)
```

```
+   binmat <- X * binvec
+   X_kov <- cbind(X_neu, binmat)
+   X_neu <- X_kov
+ }
+
+ # Schleife zur Erstellung metrischer Kovariablen
+ for(z in 1:C){
+   metr <- scale(rnorm(n))
+   metrvec <- rep(metr, I)
+   metrmat <- X * metrvec
+   X_kov <- cbind(X_neu, metrmat)
+   X_neu <- X_kov
+ }
+
+ # Ziehen von normalverteilten Schätzern
+ gamma <- rnorm((m-1)*(B+C+1))
+
+ # Linearer Prädiktor
+ eta <- X_neu %*% gamma
+ pi <- exp(eta)/(1+exp(eta))
+
+ # Dimension Vergleichsvektor
+ K <- 50
+ gammas <- matrix(0,ncol=(B+C+1)*(m-1),nrow=K)
+ for(k in 1:K){
+   # Responsevektor
+   yi <- rbinom(n*I, 1, pi)
+   # Generalisiertes Lineares Modell
+   model <- glm.fit(x = X_kov, y = yi, family = binomial(link="logit"),
+                   intercept = FALSE)
+   gammas[k,] <- model$coefficients
+ }
+ gamma[[i]]
+ means_gamma[[i]] <- colMeans(gammas)
+ }
```

A.2.2. Simulationsmaske: BTL-Modell über Group-LASSO

```
> library(grplasso)
```

Laden der Funktion, mit der die Designmatrix erstellt wird.

```
> source("Designmatrix.R")
```

Bei der Simulation sind folgende Größen vom Anwender anzugeben:

- `m`: Anzahl zu vergleichender Objekte/Items
- `n`: Anzahl Personen
- `B`: Anzahl binärer Kovariablen im Modell
- `b`: Anzahl der `B` binären Kovariablen, welche informativ ins Modell eingehen sollen
- `C`: Anzahl metrischer Kovariablen im Modell
- `c`: Anzahl der `C` metrischen Kovariablen, welche informativ ins Modell eingehen sollen
- `p`: Anzahl Simulationsdurchläufe
- `df.method`: Angabe, welche Art von Freiheitsgradberechnung (`lasso` / `grplasso`) verwendet werden soll

Die Anzahl der Paarvergleiche, die mit `I` bezeichnet wird, wird aus der Anzahl der Objekte/Items über die Formel $I \leftarrow m \cdot (m-1) / 2$ berechnet. Mit diesen Angaben kann auch die Designmatrix an die gewünschte Simulation angepasst werden.

```
> X <- designmatrix(m, n, lastcol=FALSE)
```

Zum Abspeichern der Simulationsergebnisse sind leere Listen bzw. Vektoren zu erstellen.

```
> aic <- list()
```

```
> bic <- list()
```

```
> fit <- list()
```

```
> false.pos <- true.pos <- c()
```

Über folgende Simulationsmaske können Modelle unter Verwendung des Group-LASSO Verfahrens geschätzt werden. Dabei kann die Vorhersagegenauigkeit anhand der Größen `true.pos` und `false.pos` überprüft werden. Im optimalen Fall werden alle Kovariablen mit Information im Modell berücksichtigt und die Kovariablen ohne Information nicht.

```
> for(i in 1:p){
+   set.seed(i)
+
+   X_neu <- X
+   # Schleife zur Erstellung binärer Kovariablen
+   for(w in 1:B){
+     bin <- scale(sample(c(0,1), n, replace=TRUE))
+     binvec <- rep(bin, I)
+     binmat <- X * binvec
+     X_kov <- cbind(X_neu, binmat)
+     X_neu <- X_kov
+   }
+
+   # Schleife zur Erstellung metrischer Kovariablen
+   for(z in 1:C){
+     metr <- scale(rnorm(n))
+     metrvec <- rep(metr, I)
+     metrmat <- X * metrvec
+     X_kov <- cbind(X_neu, metrmat)
+     X_neu <- X_kov
+   }
+
+   # Ziehen von normalverteilten Schätzern
+   gamma_intercept <- rnorm((m-1)*1)
+   gamma_info_bin <- rnorm((m-1)*b)
+   gamma_ohne_info_bin <- rep(0, times = (m-1)*(B-b))
+   gamma_info_metr <- rnorm((m-1)*c)
+   gamma_ohne_info_metr <- rep(0, times = (m-1)*(C-c))
+   gamma <- c(gamma_intercept, gamma_info_bin, gamma_ohne_info_bin,
+             gamma_info_metr, gamma_ohne_info_metr)
+
+   # Linearer Prädiktor
+   eta <- X_neu %*% gamma
+   pi <- exp(eta)/(1+exp(eta))
+
+   # Responsevektor
```

```
+ y <- rbinom(n*I, 1, pi)
+
+ # Index für Gruppen festlegen
+ index <- rep(c(NA,1:(B+C)), each = (m-1))
+
+ # Sequenz von Lambda's bestimmen
+ lambda <- 1.05* lambdamax(x = X_neu, y = y, index = index, model = LogReg(),
+                           center=FALSE, standardize=FALSE)
+ lambda.seq <- seq(from=lambda, to = 0.001, length=40)
+
+ # Group Lasso
+ suppressWarnings(
+ fit[[i]] <- grplasso(x = X_neu, y = y, index=index, lambda=lambda.seq,
+                      model = LogReg(), center=FALSE, standardize=FALSE,
+                      control = grpl.control(trace = 0)))
+
+ # Log-Likelihood des Logit-Modells
+ loglik <- fit[[i]]$nloglik
+
+ # Bestimmung der Freiheitsgrade für jedes Lambda
+ beta <- fit[[i]]$coef[-(1:(m-1)),]
+ norm <- matrix(0,nrow=B+C,ncol=length(lambda.seq))
+ for(u in 1:length(lambda.seq)){
+   beta.help <- matrix(beta[,u],ncol=B+C)
+   norm[,u] <- sqrt(colSums(beta.help^2))
+ }
+ if(df.method=="grplasso"){
+ df.grp <- colSums(norm>0) + colSums(norm/norm[,ncol(norm)])*(m-2)
+ }
+ if(df.method=="lasso"){
+ df.grp <- colSums(beta!=0)
+ }
+ df <- m - 1 + df.grp
+
+ # Berechnung der Selektionskriterien AIC und BIC
+ aic[[i]] <- 2*loglik + 2*df
```

```

+ bic[[i]] <- 2*loglik + log(nrow(X_neu))*df
+
+ trues <- c(1:b,(1:c)+B)
+ falses <- (1:(B+C))[-trues]
+ true.pos[i] <- sum(which(norm[,which.min(bic[[i]])]!=0) %in% trues)
+ false.pos[i] <- sum(which(norm[,which.min(bic[[i]])]!=0) %in% falses)
+
+ cat("Simulation",i,"\n")
+ }

```

A.3. Weitere Simulationsergebnisse

Std.Abw.	Group-LASSO (YL)		Group-LASSO (AS)	
	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)
0.2	83.00	1.00	56.25	7.00
0.25	95.50	1.00	85.25	16.00
0.3	98.75	1.00	94.25	22.00
0.4	99.75	11.00	99.50	33.00
0.5	100.00	24.00	100.00	41.00
1	100.00	74.00	100.00	79.00

Tabelle A.1.: Ergebnisse des Simulationssettings gemäß Topmodel-Datensatz: 6 Items, 5 Kovariablen, davon 1 ohne Information, 200 Personen, 100 Simulationen.

Std.Abw.	Group-LASSO (YL)		Group-LASSO (AS)	
	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)	Anteil korrekt informativ geschätzt (in %)	fälschlicherweise informativ geschätzt (in %)
0.20	100.00	8.33	100.00	0.00
0.25	100.00	15.83	100.00	0.00
0.3	100.00	11.67	100.00	0.00
0.4	100.00	21.67	100.00	0.83
0.5	100.00	10.00	100.00	1.67
1	100.00	33.00	100.00	15.83

Tabelle A.2.: Ergebnisse des Simulationssettings: 20 Items, 10 Kovariablen, davon 6 ohne Information, 500 Personen, 20 Simulationen.

A.4. Elektronischer Anhang

Der auf der beigefügten CD enthaltene elektronische Anhang beinhaltet die Ordner **Bilder**, **Masterthesis**, **R** und **Sweave**.

Im Ordner **Bilder** sind Bild-Dateien enthalten, die nicht mittels der R-Funktion „Sweave“ erstellt wurden, sondern selbst erstellt wurden oder aus einer anderen Quelle entnommen sind.

Der Ordner **Masterthesis** enthält die elektronische Form der hier vorliegenden Masterarbeit im PDF-Format. Außerdem ist die `Masterthesis.tex`-Datei zu finden, welche zur Erzeugung des PDFs im Programm \LaTeX benötigt wird. Die Datei `Literatur.bib` beinhaltet die Liste der verwendeten Literatur/Quellen. Die Dateien `Sweave.sty` und `apalike-url.bst` werden zur Erzeugung des Berichts im PDF-Format benötigt, wobei `Sweave.sty` zur Einbindung von R-Output dient. Die Dateien `Layout.tex`, `Meta.tex` und `Packages.tex` legen den Grundaufbau der Masterarbeit fest, wohingegen die Datei `Deckblatt.tex` das Layout der Titelseite erzeugt. Zudem ist der Ordner **Kapitel** vorhanden, der die einzelnen Kapitel des Berichts in \TeX -Format enthält, die zur besseren Übersicht einzeln in den Bericht eingefügt werden.

Die Auswertungen in R befinden sich im Ordner **R**, der folgende Unterordner enthält:

- **Anwendungsbeispiele:** Hier ist der R-Code der beiden Anwendungsbeispiele enthalten, genauer in den Dateien `Anwendungsbeispiel_CEMSChoice.R` und `Anwendungsbeispiel_Topmodel.R`. Der R-Code der Methoden mit denen die regularisierten

Schätzverfahren verglichen werden ist beim CEMSChoice-Datensatz in der Datei `Vergleich_CEMSChoice.R` (log-lineare Bradley-Terry Modelle) und beim Topmodel-Datensatz in der Datei `Vergleich_Topmodel.R` (modellbasiertes rekursives Partitionieren) abgespeichert.

- **Grafiken:** Dieser Ordner enthält den Code aller in R erzeugten Grafiken, die nicht zu den einzelnen Anwendungsbeispielen zählen. Somit ist der R-Code für die Logit-Funktion, die Grafik über AIC und BIC, die Grafik über die Pfade bei Group-LASSO und die Grafiken mit den Ergebnissen der Simulationen enthalten.
- **Simulationen_Group-Lasso:** In diesem Ordner sind die R-Codes der durchgeführten Simulationen enthalten. Jedes Simulationssetting ist einzeln in einem Unterordner aufgeführt, in dem wiederum ein Ordner mit den Simulationsergebnissen enthalten ist.

Außerdem sind im Ordner **R** die R-Dateien `Designmatrix.R`, `Simulationsmaske_GLM.R` und `Simulationsmaske_GroupLasso.R` enthalten, mit deren Hilfe für beliebige Szenarien Bradley-Terry-Luce Modelle über Generalisierte Lineare Modelle bzw. über Group-LASSO simuliert werden können.

Die Dateien zur Einbindung von R-Outputs und R-Grafiken in \LaTeX mittels der Funktion „Sweave“ befinden sich im Ordner **Sweave**. Dort sind im Unterordner **Grafiken** die PDF-Dateien der erzeugten R-Grafiken enthalten.

Literaturverzeichnis

- Agresti, A. (2002). *Categorical Data Analysis*. Wiley Series in Probability and Statistics, Second edition.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. N. and Csáki, F., editors, *Second International Symposium on Information Theory*, pages 267–281. Akadémiai Kiadó, Budapest.
- Altobelli, C. F. (2007). *Marktforschung: Methoden–Anwendungen–Praxisbeispiele*. Lucius & Lucius Verlagsgesellschaft mbH.
- Antoniadis, A. and Fan, J. (2001). Regularization of wavelet approximations. *Journal of the American Statistical Association*, 96(455):939–955.
- Bakin, S. (1999). *Adaptive Regression and Model Selection in Data Mining Problems*. PhD thesis, Australian National University, Canberra.
- Böckenholt, U. (2001). Thresholds and intransitivities in pairwise judgments: A multilevel analysis. *Journal of Educational and Behavioral Statistics*, 26(3):269–282.
- Bortz, J. and Döring, N. (2006). *Forschungsmethoden und Evaluation – für Human- und Sozialwissenschaftler*. Springer Medizin Verlag. 4. Auflage.
- Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4):324–345.
- Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, 16(3):199–231.
- Cai, T. T. (2001). Discussion of „Regularization of wavelet approximations“ (by Antoniadis A. and Fan J.). *Journal of the American Statistical Association*, 96(455):960–962.
- Choisel, S. and Wickelmaier, F. (2007). Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference. *Journal of the Acoustical Society of America*, 121(1):388–400.

- Cohen, A. L. (2004). Models of choice. Available from: http://people.umass.edu/alc/course_pages/fall_2004/modeling_behavior/lectures/choice.ppt [Online; accessed 14.03.2013].
- Condon, E. M. (2012). *Varying Coefficients in Logistic Regression with Applications to Marketing Research*. PhD thesis, University of Illinois at Urbana-Champaign.
- Critchlow, D. E. and Fligner, M. A. (1991). Paired comparison, triple comparison, and ranking experiments as generalized linear models, and their implementation on GLIM. *Psychometrika*, 56(3):517–533.
- Dahl, D. B. (2012). xtable: Export tables to LaTeX or HTML. R package version 1.7-1. Available from: <http://CRAN.R-project.org/package=xtable> [Online; accessed 14.03.2013].
- David, H. A. (1988). *The Method of Paired Comparisons*. Oxford University Press.
- Davidson, R. R. (1970). On extending the Bradley-Terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association*, 65(329):317–328.
- Davidson, R. R. and Beaver, R. J. (1977). On extending the Bradley-Terry model to incorporate within-pair order effects. *Biometrics*, 33(4):693–702.
- Davidson, R. R. and Farquhar, P. H. (1976). A bibliography on the method of paired comparisons. *Biometrics*, 32(2):241–252.
- Debreu, G. (1960). Review of R. D. Luce, Individual choice behavior: A theoretical analysis. *American Economic Review*, 50(1-2):186–188.
- Dittrich, R., Francis, B., Hatzinger, R., and Katzenbeisser, W. (2006). Modelling dependency in multivariate paired comparisons: A log-linear approach. *Mathematical Social Sciences*, 52:197–209.
- Dittrich, R. and Hatzinger, R. (2009). Fitting loglinear Bradley-Terry models (LLBT) for paired comparisons using the R package pefmod. *Psychology Science Quarterly*, 51(2):216–242.
- Dittrich, R., Hatzinger, R., and Katzenbeisser, W. (1998). Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. *Journal of the Royal Statistical Society, Series C, Applied Statistics*, 47(4):511–525.

- Dittrich, R., Hatzinger, R., and Katzenbeisser, W. (2001). Corrigendum: Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. *Journal of the Royal Statistical Society, Series C, Applied Statistics*, 50(2):247–249.
- Dobson, A. J. (2002). *An Introduction to Generalized Linear Models*. Chapman & Hall/CRC, Second edition.
- Elad, M. (2010). *Sparse and Redundant Representations – From Theory to Applications in Signal and Image Processing*. Springer Verlag.
- Fahrmeir, L., Kneib, T., and Lang, S. (2009). *Regression – Modelle, Methoden und Anwendungen*. Springer Verlag.
- Fienberg, S. E. (1979). Log linear representation for paired comparison models with ties and within-pair order effects. *Biometrics*, 36(2):479–481.
- Fienberg, S. E. and Larntz, K. (1976). Log linear representation for paired and multiple comparisons models. *Biometrika*, 63(2):245–254.
- Frisenfeldt Thuesen, K. (2007). Analysis of Ranked Preference Data. Technical report, Technical University of Denmark, Informatics and Mathematical Modelling. Available from: www.imm.dtu.dk.
- Genkin, A., Lewis, D. D., and Madigan, D. (2007). Large-scale Bayesian logistic regression for text categorization. *Technometrics*, 49(3):291–304.
- Gesellschaft, W. (2012). Rangordnung und Paarvergleich als Frageform. Available from: <http://www.wpgs.de/content/view/462/355/> [Online; accessed 14.03.2013].
- Guttman, L. (1946). An approach for quantifying paired comparisons and rank order. *The Annals of Mathematical Statistics*, 17(2):144–163.
- Hastie, T., Tibshirani, R., and Friedman, J. (2008). *The Elements of Statistical Learning – Data Mining, Inference, and Prediction*. Springer Verlag, Second edition.
- Hatzinger, R. (2012). `prefmod`: Utilities to fit paired comparison models for preferences. R package version 0.8-31. Available from: <http://CRAN.R-project.org/package=prefmod> [Online; accessed 14.03.2013].
- Hatzinger, R. and Dittrich, R. (2012). `prefmod`: An R package for modeling preferences based on paired comparisons, rankings, or ratings. *Journal of Statistical Software*, 48(10).

- Huang, T.-K., Weng, R. C., and Lin, C.-J. (2006). Generalized Bradley-Terry models and multi-class probability estimates. *Journal of Machine Learning Research*, 7:85–115.
- Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3):299–314.
- Kendall, M. G. (1955). Further contributions to the theory of paired comparisons. *Biometrics*, 11(1):43–62.
- Kendall, M. G. and Babington Smith, B. (1940). On the method of paired comparisons. *Biometrika*, 31(3/4):324–345.
- Kim, Y., Kim, J., and Kim, Y. (2006). Blockwise sparse regression. *Statistica Sinica*, 16:375–390.
- Lange, K. (2000). *Numerical Analysis for Statisticians*. Springer Verlag, Second edition.
- Leisch, F. (2002a). Sweave: Dynamic Generation of Statistical Reports Using Literate Data Analysis. In Härdle, W. and Rönz, B., editors, *Compstat 2002 — Proceedings in Computational Statistics*, pages 575–580. Physica Verlag, Heidelberg. ISBN 3-7908-1517-9. Available from: <http://www.stat.uni-muenchen.de/~leisch/Sweave>.
- Leisch, F. (2002b). Sweave, Part I: Mixing R and L^AT_EX. *R News*, 2(3):28–31. Available from: <http://CRAN.R-project.org/doc/Rnews/>.
- Lokhorst, J. (1999). The Lasso and Generalised Linear Models. Technical report, University of Adelaide, Adelaide. Available from: www.adelaide.edu.au.
- Luce, R. D. (1959). *Individual Choice Behavior*. New York: Wiley.
- Mallows, C. L. (1957). Non-null ranking models. *Biometrika*, 44(1/2):114–130.
- Matthews, J. N. S. and Morris, K. P. (1995). An application of Bradley-Terry-Type models to the measurement of pain. *Journal of the Royal Statistical Society, Series C, Applied Statistics*, 44(2):243–255.
- McCullagh, P. J. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall/CRC, Second edition.
- Medialexikon (2012). Paarvergleich (Methode des paarweisen Vergleichs). Available from: <http://www.medialine.de/deutsch/wissen/medialexikon.php?snr=4189> [Online; accessed 14.03.2013].

- Meier, L. (2009). grplasso: Fitting user specified models with group lasso penalty. R package version 0.4-2. Available from: <http://CRAN.R-project.org/package=grplasso> [Online; accessed 14.03.2013].
- Meier, L., van de Geer, S., and Bühlmann, P. (2008). The group lasso for logistic regression. *Journal fo the Royal Statistical Society, Series B, Statistical Methodology*, 70(1):53–71.
- Müller-Herbers, S. (2012). Reader: Methoden zur Beurteilung von Varianten. Institut für Grundlagen der Planung, Universität Stuttgart.
- Mosteller, F. (1951). Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, 16(1):3–9.
- Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal fo the Royal Statistical Society, Series A, General*, 135(3):370–384.
- Novafeel (2013). Schönheit, Schön. Available from: <http://www.novafeel.de/schoenheit/schoenheit.htm> [Online; accessed 14.03.2013].
- Oberfeld, D., Hecht, H., Allendorf, U., and Wickelmaier, F. (2009). Ambient lighting modifies the flavor of wine. *Journal of Sensory Studies*, 24(6):797–832.
- Pfister, R. (WiSe 2007/2008). Mitschrift zur Vorlesung: Methoden der Skalierung. Institut für Psychologie, Julius-Maximilians-Universität Würzburg.
- Psychologielexikon (2012). Präferenz. Available from: <http://www.psychology48.com/deu/d/praeferenz/praeferenz.htm> [Online; accessed 14.03.2013].
- R Core Team and contributors worldwide (2012). stats: The R Stats Package. R package version 2.15.3. Available from: <http://stat.ethz.ch/R-manual/R-patched/library/stats/html/00Index.html> [Online; accessed 14.03.2013].
- R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from: <http://www.R-project.org>.
- Reineking, B. and Schröder, B. (2003). Variablenselektion. In *Habitatmodelle - Methodik, Anwendung, Nutzen*, pages 39–45. UFZ-Umweltforschungszentrum Leipzig-Halle.

- Rodríguez, G. (2012). Generalized Linear Models. Available from: <http://data.princeton.edu/wws509/> [Online; accessed 14.03.2013].
- Roth, V. (2004). The generalized lasso. *IEEE Trans Neural Netw.*, 15(1):16–28.
- Saari, D. G. (2005). The profile structure for Luce’s Choice Axiom. *Journal of Mathematical Psychology*, 49:226–253.
- Schwartz, L. (1965). On Bayes procedures. *Z. Wahrscheinlichkeitstheorie*, 4(1):10–26.
- SDI-Research (2012). Forced-Choice Befragungstechnik. Available from: <http://www.sdi-research.at/lexikon/forced-choice.html> [Online; accessed 14.03.2013].
- Shevade, S. K. and Keerthi, S. S. (2003). A simple and efficient algorithm for gene selection using sparse logistic regression. *Bioinformatics*, 19(17):2246–2253.
- Sinclair, C. D. (1982). GLIM for preference. In Gilrichst, R., editor, *GLIM 82: Proceedings of the International Conference on Generalised Linear Models*. Lecture Notes in Statistics, 14.
- Strobl, C., Wickelmaier, F., and Zeileis, A. (2009). Accounting for Individual Differences in Bradley-Terry Models by Means of Recursive Partitioning. Technical Report 54, University of Munich, Department of Statistics. Available from: <http://www.stat.uni-muenchen.de>.
- Thurstone, L. L. (1927a). A law of comparative judgment. *Psychological Review*, 34(4):273–286.
- Thurstone, L. L. (1927b). Psychophysical analysis. *The American Journal of Psychology*, 38(3):368–389.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal fo the Royal Statistical Society, Series B, Methodological*, 58(1):267–288.
- Tibshirani, R. (2012). The Lasso Page – L1-constrained fitting for statistics and data mining. Available from: <http://www-stat.stanford.edu/~tibs/lasso.html> [Online; accessed 14.03.2013].
- Tseng, P. and Yun, S. (2009). A coordinate gradient descent method for nonsmooth separable minimization. *Math. Programming, Series B*, (117):387–423.

- Tsukida, K. and Gupta, M. R. (2011). How to Analyze Paired Comparison Data. Technical report, University of Washington, Department of Electrical Engineering. Available from: <http://www.ee.washington.edu>.
- Turner, H. and Firth, D. (2012). gnm: Generalized nonlinear models. R package version 1.0-6. Available from: <http://CRAN.R-project.org/package=gnm> [Online; accessed 14.03.2013].
- Tutz, G. (1986). Bradley-Terry-Luce models with an ordered response. *Journal of Mathematical Psychology*, 30(3):306–316.
- Tutz, G. (2000). *Die Analyse Kategorialer Daten*. Oldenbourg Wissenschaftsverlag GmbH.
- Tutz, G. (2012). *Regression for Categorical Data*. Cambridge University Press.
- Tutz, G. (WiSe 2009/2010). Skript zur Vorlesung Generalisierte Regressionsmodelle. Institut für Statistik, LMU München.
- Tutz, G. (WiSe 2012/2013). BTL, unpublished manuscript. Institut für Statistik, LMU München.
- Tutz, G. and Schauburger, G. (2012). A Penalty Approach to Differential Item Functioning in Rasch Models. Technical Report 134, University of Munich, Department of Statistics. Available from: <http://www.stat.uni-muenchen.de>.
- Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer Verlag, Fourth edition.
- Wikipedia (2012). Germany's Next Topmodel - Die freie Enzyklopädie. Available from: http://de.wikipedia.org/wiki/Germany%E2%80%99s_Next_Topmodel [Online; accessed 14.03.2013].
- Wood, S. W. (2006). *Generalized Additive Models – An Introduction with R*. Chapman & Hall/CRC.
- Yang, X. D. (2011). *Statistical Methods for Variable Selection in the Context of High Dimensional Data: Lasso and Extensions*. PhD thesis, Mc Master University, Hamilton, Ontario, Canada.
- Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal fo the Royal Statistical Society, Series B, Statistical Methodology*, 68(1):49–67.

- Zeileis, A. (2011). Schönheit liegt im Auge des Betrachters: Ein statistischer Blick auf Topmodels und Professoren. Available from: <http://eeecon.uibk.ac.at/~zeileis/papers/Ibk-2011.pdf> [Online; accessed 14.03.2013].
- Zeileis, A., Strobl, C., and Wickelmaier, F. (2012a). psychotools: Infrastructure for Psychometric Modeling. R package version 0.1-4. Available from: <http://CRAN.R-project.org/package=psychotools> [Online; accessed 14.03.2013].
- Zeileis, A., Strobl, C., Wickelmaier, F., and Kopf, J. (2012b). psychotree: Recursive Partitioning Based on Psychometric Models. R package version 0.12-2. Available from: <http://CRAN.R-project.org/package=psychotree> [Online; accessed 14.03.2013].
- Zermelo, E. (1929). Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 29:436–460.

Eidesstattliche Erklärung

Hiermit versichere ich, Manuela Bitterlich, dass ich die vorliegende Masterarbeit mit dem Thema

Regularisierte Schätzverfahren für Bradley-Terry-Luce Modelle

selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, dass alle wörtlichen und sinngemäßen Zitate, die aus anderen Quellen übernommen wurden, als solche gekennzeichnet wurden. Die Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Buchloe, den 18. März 2013

MANUELA BITTERLICH