



Anna Windischbacher, BSc

Ordinale Regressionsmodelle zur Analyse postoperativer Komplikationen

MASTERARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieurin

Masterstudium Operations Research and Statistics

eingereicht an der

Technischen Universität Graz

Betreuer

Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. H. Friedl

Institut für Statistik

Graz, Mai 2018

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Masterarbeit identisch.

Datum

Unterschrift

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

Date

Signature

ZUSAMMENFASSUNG

Ordinale Regressionsmodelle stellen ein unverzichtbares Konstrukt bei der Analyse von multikategoriellen Daten mit Ordnungsstruktur dar. Diese sind eine Erweiterung der multinomialen Regressionsmodelle, welche aus diesem Grund die Basis dieser Arbeit darstellen. Den theoretischen Schwerpunkt dieser Arbeit bildet die Modellierung ordinaler Responses. Dabei werden im Speziellen die kumulativen und sequentiellen Modelle betrachtet und analysiert. Ein Teil dieser Analyse behandelt den Zusammenhang dieser spezifischen Modelle mit den generalisierten linearen Modellen. Anwendung finden die theoretischen Aspekte in der Modellierung von Patientinnendaten nach einer Mammareduktion. Das Ziel dieser Arbeit ist es, das erworbene Wissen anzuwenden und verschiedene Modelle zur Schätzung der Clavien-Dindo-Klassifikation nach der Mammareduktion zu erstellen.

ABSTRACT

Ordinal regression models are an essential construct in the analysis of multi-categorical data with ordered structure, which are extensions of multinomial regression models. The theoretical focus of this thesis is the modelling of ordinal responses. Therefore, especially cumulative and sequential models are considered and analysed. A part of this analysis deals with the relation between these models and generalized linear models. Furthermore the theoretical aspects are applied for modelling medical data of mamma reductions. The aim of this thesis is to utilize the adopted knowledge and to create models in order to estimate the Clavien-Dindo-classification after mamma reduction.

Inhaltsverzeichnis

1	Einleitung	9
1.1	Daten	9
1.2	Ausblick	11
2	Diskrete Verteilungen	13
2.1	Binomialverteilung	13
2.2	Multinomialverteilung	14
2.3	Poissonverteilung	18
2.4	Exponentialfamilie	20
3	Generalisierte Lineare Modelle	27
3.1	Basisstruktur der Generalisierten Linearen Modelle	27
3.2	Maximum Likelihood Schätzung	29
3.3	Generalisierte Lineare Modelle für diskrete Verteilungen	31
3.3.1	Standardisierte Binomialverteilung	31
3.3.2	Poissonverteilung	32
3.3.3	Multinomialverteilung	33
3.4	Deviance	35
4	Modellierung ordinaler Response	37
4.1	Kumulative Modelle	37
4.1.1	Einfaches kumulatives Modell	38
4.1.2	Generelles kumulatives Modell	39
4.1.3	Testen der Effekte	40
4.2	Linkfunktionen des kumulativen Modells	41
4.2.1	Kumulatives Logit Modell	41
4.2.2	Probit Modell	44
4.2.3	Kumulatives Extremwert Modell	44
4.3	Kumulatives Modell für GLM	45
4.3.1	Maximum Likelihood Schätzung für kumulative Modelle	47
4.4	Sequentielle Modelle	48
4.5	Linkfunktionen des sequentiellen Modells	49
4.5.1	Sequentielles Logit Modell	50

4.5.2	Sequentielles Extremwert Modell	50
4.6	Sequentielles Modell für GLM	51
4.6.1	Maximum Likelihood Schätzung für sequentielle Modelle	51
5	Modellierung in R	53
5.1	Package <code>VGAM</code>	53
5.2	Package <code>MASS</code> und <code>nnet</code>	56
5.3	Package <code>EffectStars2</code>	56
6	Modelle zur Analyse der Clavien-Dindo-Klassifikation	59
6.1	Beschreibung des Datensatzes	59
6.2	Binomiale Modelle	62
6.3	Multinomiale Modelle	67
6.3.1	Analyse mit <code>vglm</code>	67
6.3.2	Analyse mit <code>multinom</code>	72
6.4	Ordinale Modelle	75
6.4.1	Kumulative Modelle	75
6.4.2	Sequentielle Modelle	84
7	Resümee	89
A	Anhang	91
A.1	Iteratively Reweighted Least Squares (IRLS) Methode	91
A.2	Odds und Odds Ratio	92
A.3	Maximum Likelihood Schätzer im saturierten Modell	93
A.4	Deviance der diskreten Verteilungen	94
A.5	Linkfunktionen	96
A.6	Hauck-Donner Effekt	99
	Literatur	103

1 Einleitung

Jede Operation ist mit einem Risiko verbunden, wobei nicht nur während, sondern auch nach dem operativen Eingriff Komplikationen, sogenannte postoperative Komplikationen, auftreten können. Um weltweit eine Verallgemeinerung des Schwierigkeitsgrades dieser Komplikationen zu erhalten, wurde 1992 eine Klassifikation der Komplikationen, genannt Clavien-Dindo-Klassifikation, eingeführt. Diese Klassifikation ermöglicht es, einfach, reproduzierbar und flexibel anwendbar, unabhängig von der Herkunft, die postoperative Befindlichkeit einzuteilen und basiert auf den Therapien, welche für den Heilungsprozess benötigt werden. In dieser Arbeit werden nun mittels statistischen Methoden Faktoren untersucht, welche einen Einfluss auf die Clavien-Dindo-Klassifikation haben können. Die Grundlage dieser Arbeit bildet ein Datensatz, der die Daten von Patientinnen enthält, welche sich an der klinischen Abteilung für plastische, ästhetische und rekonstruktive Chirurgie der Universitätsklinik Graz einer Mammareduktion unterzogen haben und von Dr.med.univ. Raimund Winter für die Analyse zur Verfügung gestellt wurden.

1.1 Daten

Der Datensatz beinhalten die Informationen von 486 Patienten, welche sich in einem Zeitraum von zehn Jahren an der Universitätsklinik Graz aus gesundheitlichen oder ästhetischen Gründen einer Mammareduktion unterzogen haben. Beim vorliegenden Datenmaterial handelt es sich um ausschließlich weibliche Patienten, wobei deren Informationen vor, während und nach der Operation vom medizinischen Personal dokumentiert wurden. Auf Grund der Vollständigkeit dieser Daten ist die Zuverlässigkeit der statistischen Auswertungen garantiert. Der Datensatz wurde bereits für ein Statistikprojekt zur Verfügung gestellt, in welchem acht verschiedene Komplikationsarten modelliert und die Einflussgrößen dieser analysiert wurden. Dabei wurden die präklinischen und klinischen Größen zuerst separat modelliert und dann ein allgemeines Modell für die binären Responses erstellt, für weitere Details siehe Windischbacher (2017). In dieser Arbeit wird nun der Schwerpunkt auf die Clavien-Dindo-Klassifikation (CD-Klassifikation) und die Modellschätzung für diese Response gelegt. Im Unterschied

zu den binären Komplikationen im Projekt umfasst die CD-Klassifikation mehrstufige, ordinale Daten. Ordinale Daten unterliegen im Gegensatz zu nominalen Daten einer Ordnungsstruktur und können in eine natürliche Reihenfolge gebracht werden. Allerdings sind die Abstände bei ordinalen Daten zwischen den verschiedenen Ausprägungen nicht quantifizierbar und dadurch nicht vergleichbar. Ein Beispiel dafür wären Schulnoten. Die Note “sehr gut“ ist zwar besser als “gut“, jedoch ist “genügend“ nicht doppelt so schlecht wie “gut“.

Die CD-Klassifikation besitzt eine Ordnungsstruktur und besteht aus fünf Graden, wobei Grad 3 und 4 zusätzlich in die zwei Stufen *A* und *B* unterteilt werden können. Eine genaue Auflistung und Erklärung dieser Stufen liefert nachfolgende Tabelle 1.1.

Clavien-Dindo-Klassifikation	Beschreibung
Grad 1	jede Abweichung des postoperativen Verlaufs ohne den Bedarf einer pharmakologischen oder operativen, radiologischen oder endoskopischen Behandlung ausgenommen sind Medikamente wie Antiemetika, Antipyretika, Analgetika, Diuretika, Elektrolyt- sowie Physiotherapie, dieser Grad beinhaltet auch Wundinfektionen geöffnet an der Bettseite
Grad 2	beinhaltet pharmakologische Behandlungen mittels Medikamenten (nicht von Grad 1) auch Bluttransfusionen sowie totale parenterale Ernährung
Grad 3 - 3A - 3B	umfasst operativen, radiologischen oder endoskopischen Eingriff - Eingriff nicht unter genereller Anästhesie - Eingriff unter genereller Anästhesie
Grad 4 - 4A - 4B	lebensbedrohliche Komplikationen (inkludiert CNS Komplikationen), IC bzw. ICU Management erforderlich - single Organfehlfunktion (inkludiert Dialyse) - multi Organfehlfunktion
Grad 5	Tod des/der Patienten/in

Tabelle 1.1: Auflistung und Beschreibung der Grade der Clavien-Dindo-Klassifikation

Der Grad 0 beschreibt dabei eine komplikationsfreie Operation und keine postoperativen Beschwerden. Der vorliegende Datensatz der Patientinnen der Universitätsklinik Graz beinhaltet Klassifikationen von Patientinnen mit Grad 0, 1, 2, 3A und 3B. Die Häufigkeiten in absoluten sowie prozentuellen Zahlen sind der Tabelle 1.2 zu entnehmen. Um diese Klassifikation in Abhängigkeit von beeinflussenden Faktoren modellieren zu können, werden in dieser Arbeit Modelle für ordinale Daten betrachtet.

Klassifikation	Häufigkeiten	Prozente
Grad 0	182 Patientinnen	37.45%
Grad 1	234 Patientinnen	48.15%
Grad 2	46 Patientinnen	9.46%
Grad 3A	2 Patientinnen	0.41%
Grad 3B	22 Patientinnen	4.53%

Tabelle 1.2: Absolute und relative Häufigkeiten der einzelnen Klassifikationen in den vorliegenden Daten

1.2 Ausblick

Im Kapitel 2 werden die Grundlagen dieser Arbeit beschrieben, indem diskrete Verteilungen eingeführt werden. Dabei werden die für die Arbeit relevanten Verteilungen erklärt sowie die allgemeine Klasse der Exponentialfamilie definiert. Das Kapitel 3 befasst sich mit den generalisierten linearen Modellen. Es wird eine allgemeine Erklärung dieser Modelle gegeben und dabei sowohl die Schätzung als auch die Deviance beschrieben. Im Kapitel 4 wird der Fokus auf ordinale Daten gelegt. Dabei werden verschiedene Modelle beschrieben, Schätzungen angeführt und Funktionen sowie Tests gezeigt. Die praktische Anwendung findet sich dann in Kapitel 5. Dieses Kapitel beinhaltet die Auswertungen der modellierten Daten, sowie Darstellungen und Interpretationen der Ergebnisse. Die Modellierung und Schätzungen der Daten wurden mit Hilfe von **R** in der Version 3.4.3 und der Entwicklungsumgebung **RStudio** in der Version 1.1.423 durchgeführt. Die Arbeit basiert auf Inhalte von Tutz (2012), Tutz (2002), McCullagh und Nelder (1989) und Agresti (2010). Falls die Quellen nicht explizit angegeben werden, werden diese verwendet.

2 Diskrete Verteilungen

Diese Arbeit beruht auf der Verwendung diskreter Verteilungen, welche im folgenden Abschnitt beschrieben werden. Dabei beinhaltet die Einführung alle in der Arbeit auftretenden Verteilungen und Familien. In diesem Kapitel werden diverse Eigenschaften und Herleitungen dieser Verteilungen gezeigt. Zu Beginn wird die Binomialverteilung erklärt, welche die Grundlage für die weitere Abhandlung bildet. Mit Hilfe der Binomialverteilung kann die Multinomialverteilung formuliert werden, welche eine Erweiterung ist. Die Multinomialverteilung kann auch in analytischer Weise durch die Poissonverteilung hergeleitet werden. Somit stellt die Poissonverteilung eine weitere relevante Verteilung in dieser Arbeit dar. Im Anschluss wird mit der Exponentialfamilie eine Familie von Verteilungen betrachtet, die besonders bei den Generalisierten Linearen Modellen von Relevanz ist. Dabei wird die Zugehörigkeit der Binomial-, der Poisson- sowie der Multinomialverteilung zur Exponentialfamilie gezeigt.

2.1 Binomialverteilung

Bei der Binomialverteilung werden die Wiederholungen eines Experiments untersucht, bei welchem ein bestimmtes Ereignis eintritt oder nicht. Die Binomialverteilung beschreibt dabei die Verteilung der Ergebnisse dieser Wiederholungen. Ein Ergebnis bezeichnet den beobachtbaren Wert, den das Experiment nach einmaliger Durchführung annimmt. Für die i -te Wiederholung dieses Experiments kann eine binäre Zufallsvariable \tilde{Y}_i definiert werden durch

$$\tilde{Y}_i = \begin{cases} 1 & \text{Ereignis tritt ein} \\ 0 & \text{Ereignis tritt nicht ein.} \end{cases} \quad (2.1)$$

Das Ergebnis wird auch als Realisierung der Zufallsvariable bezeichnet. Mittels der Zufallsvariable können die Ergebnisse in zwei Kategorien unterschieden werden. Hat die Zufallsvariable \tilde{Y}_i den Wert 1, so wird das Ergebnis der ersten Kategorie zugeordnet, das Ergebnis mit dem Wert der Zufallsvariable 0 wird der zweiten Kategorie zugeordnet. Bezeichnet π_1 die Wahrscheinlichkeit

$$\pi_1 = P(\tilde{Y}_i = 1), \quad (2.2)$$

so ist $\pi_1 \in [0, 1]$ und beschreibt die Wahrscheinlichkeit, mit der ein Ergebnis der ersten Kategorie angehört. Mit der Definition von π_1 ergibt sich die Wahrscheinlichkeit π_2 durch

$$\pi_2 = P(\tilde{Y}_i = 0) = 1 - \pi_1. \quad (2.3)$$

Diese bezeichnet die Wahrscheinlichkeit, mit der ein Ergebnis der zweiten Kategorie zugeordnet wird.

Bei m -facher, unabhängiger Wiederholung desselben Experiments kann eine Wahrscheinlichkeitsfunktion $f : \mathbb{N} \rightarrow \mathbb{R}$ formuliert werden, welche die Wahrscheinlichkeit angibt, dass das Ereignis y mal eintritt. Dieses ist eine binomiale Wahrscheinlichkeit mit

$$P(Y = y) = f(y|m, \pi_1) = \begin{cases} \frac{m!}{y!(m-y)!} \pi_1^y (1 - \pi_1)^{m-y} & y \in \{0, \dots, m\} \\ 0 & \text{sonst.} \end{cases} \quad (2.4)$$

Für diese m Wiederholungen wird die Anzahl an Ereignissen mit Y bezeichnet, es gilt also $Y = \sum_{i=1}^m \tilde{Y}_i$. Die Abkürzung $Y \sim B(m, \pi_1)$ beschreibt, dass die Zufallsvariable Y mit der Wahrscheinlichkeitsfunktion einer Binomialverteilung in der Gestalt (2.4) in Abhängigkeit der Parameter n und π_1 verteilt ist. Die Binomialverteilung $B(m, \pi_1)$ hat die folgenden Momente

$$\begin{aligned} E(Y) &= m\pi_1 \\ \text{Var}(Y) &= m\pi_1(1 - \pi_1). \end{aligned}$$

Beispiel 1. *Es wird eine Gruppe von 30 Studierenden betrachtet mit 17 Männern und 13 Frauen. Somit gilt für diese Population die Wahrscheinlichkeit $\pi_1 = \frac{13}{30}$, dass ein Studierender weiblich ist. Die Wahrscheinlichkeit, dass in einer Übungsgruppe von $m = 10$ Personen genau $y = 6$ Frauen enthalten sind, beträgt*

$$f\left(6|10, \frac{13}{30}\right) = \frac{10!}{6!(10-6)!} \left(\frac{13}{30}\right)^6 \left(\frac{17}{30}\right)^4 = 0.143.$$

Mit einer Wahrscheinlichkeit von 14.3% sind in der Gruppe 6 Frauen. Die Wahrscheinlichkeitsfunktion zu dieser Situation ist für alle $y \in \{0, \dots, 10\}$ in der Abbildung 2.1 dargestellt.

2.2 Multinomialverteilung

Die Multinomialverteilung ist eine Erweiterung der Binomialverteilung. Wie in Abschnitt 2.1 der Binomialverteilung werden auch bei dieser Verteilung die Wiederholungen eines Experiments untersucht. Im Unterschied zur Binomialverteilung, bei welcher

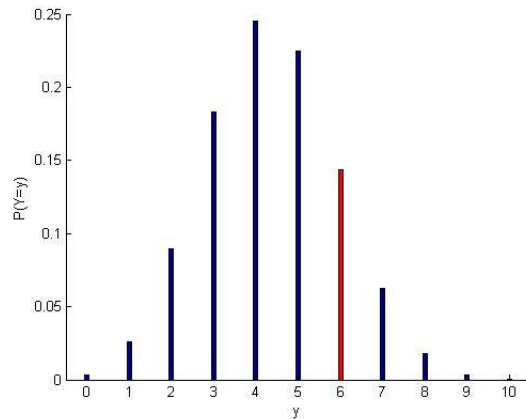


Abbildung 2.1: Darstellung der Wahrscheinlichkeitsfunktion zu Beispiel 1

ein bestimmtes Ereignis eintritt oder nicht, existieren bei der Multinomialverteilung mehrere, sich ausschließende Ereignisse. Bei der Durchführung der i -ten Wiederholung eines Experiments tritt dabei genau eines dieser Ereignisse auf. Existieren für ein Experiment k verschiedene Ereignisse, so können Zufallsvariablen $\tilde{Y}_{i1}, \dots, \tilde{Y}_{ik}$ definiert werden durch

$$\tilde{Y}_{ij} = \begin{cases} 1 & \text{Ereignis } j \text{ tritt im } i\text{-ten Versuch ein} \\ 0 & \text{Ereignis } j \text{ tritt im } i\text{-ten Versuch nicht ein.} \end{cases} \quad (2.5)$$

Analog zur Binomialverteilung können die Ergebnisse mit Hilfe der Zufallsvariablen in Kategorien unterteilt werden. Bei der Multinomialverteilung existieren, in Abhängigkeit zur Anzahl der unterschiedlichen Ereignisse, k verschiedene Kategorien. Weist die Zufallsvariable \tilde{Y}_{ij} den Wert 1 auf, so wird das Ergebnis der Kategorie j zugeordnet. Mit der Bezeichnung

$$\pi_j = P(\tilde{Y}_{ij} = 1) \quad (2.6)$$

ist $\pi_j \in [0, 1]$ und beschreibt die Wahrscheinlichkeit, mit welcher das Ergebnis der j -ten Kategorie zugeordnet wird. Da bei einem Experiment genau eines der k Ereignisse eintritt, erfüllen die Wahrscheinlichkeiten die Bedingung

$$\sum_{j=1}^k \pi_j = 1. \quad (2.7)$$

Bei m -facher, unabhängiger Wiederholung dieses Experiments kann eine Wahrscheinlichkeitsfunktion $f : \mathbb{N}^k \rightarrow \mathbb{R}$ formuliert werden, welche für den Vektor $\mathbf{y} = (y_1, \dots, y_k)$ die Wahrscheinlichkeit angibt, dass das j -te Ereignis genau y_j mal auftritt. Somit gilt $y_j = \sum_{i=1}^m \tilde{Y}_{ij}$ für alle $j = 1, \dots, k$ bei gegebener Wahrscheinlichkeit π_j mit

$$\boldsymbol{\pi} = (\pi_1, \dots, \pi_k)$$

$$P(\mathbf{Y} = \mathbf{y}) = f(\mathbf{y}|m, \boldsymbol{\pi}) = \begin{cases} \frac{m!}{y_1! \cdots y_k!} \pi_1^{y_1} \cdots \pi_k^{y_k} & y_j \in \{0, \dots, m\}, \sum_{j=1}^k y_j = m \\ 0 & \text{sonst.} \end{cases} \quad (2.8)$$

Die Dimension dieses Problems kann jedoch reduziert werden. Auf Grund der Forderung $y_1 + \cdots + y_k = m$ genügt die Kenntnis von $k - 1$ Komponenten des Vektors (y_1, \dots, y_k) . Werden y_1, \dots, y_{k-1} eindeutig festgelegt, so hat y_k die Darstellung

$$y_k = m - y_1 - \cdots - y_{k-1}.$$

Analog dazu hat die Wahrscheinlichkeit π_k die Darstellung

$$\pi_k = 1 - \pi_1 - \cdots - \pi_{k-1}.$$

Es ergibt sich die reduzierte Form der Wahrscheinlichkeitsfunktion $f : \mathbb{N}^{k-1} \rightarrow \mathbb{R}$ mit $\mathbf{y}' = (y_1, \dots, y_{k-1})$ und $\boldsymbol{\pi}' = (\pi_1, \dots, \pi_{k-1})$

$$f(\mathbf{y}'|m, \boldsymbol{\pi}') = \begin{cases} \frac{m!}{y_1! \cdots y_{k-1}! (m - \sum_{j=1}^{k-1} y_j)!} \pi_1^{y_1} \cdots \pi_{k-1}^{y_{k-1}} (1 - \sum_{j=1}^{k-1} \pi_j)^{m - \sum_{j=1}^{k-1} y_j} \\ 0 & \text{sonst.} \end{cases} \quad (2.9)$$

Der Vektor $\mathbf{Y} = (Y_1, \dots, Y_k)$ bezeichnet die Summe der Vektoren der Zufallsvariablen $(\tilde{Y}_{i1}, \dots, \tilde{Y}_{ik})$ einer jeden Wiederholung. Die Abkürzung $\mathbf{Y} \sim M(m, \boldsymbol{\pi})$ beschreibt, dass der Vektor \mathbf{Y} mit der Wahrscheinlichkeitsfunktion einer Multinomialverteilung in der Gestalt (2.8) verteilt ist.

Die Momente der Multinomialverteilung können mittels der Eigenschaften der Binomialverteilung hergeleitet werden. Da jede Zufallsvariable Y_j als Summe der Zufallsvariablen \tilde{Y}_{ij} binomialverteilt ist, gilt

$$\begin{aligned} \mathbb{E}(Y_j) &= m\pi_j \\ \text{Var}(Y_j) &= m\pi_j(1 - \pi_j). \end{aligned}$$

Die Kovarianz $\text{Cov}(Y_{j_1}, Y_{j_2})$ mit $j_1, j_2 \in \{1, \dots, k\}$ ergibt sich durch

$$\begin{aligned} \text{Cov}(Y_{j_1}, Y_{j_2}) &= \mathbb{E}[(Y_{j_1} - \mathbb{E}(Y_{j_1}))(Y_{j_2} - \mathbb{E}(Y_{j_2}))] \\ &= \mathbb{E}[(Y_{j_1} - m\pi_{j_1})(Y_{j_2} - m\pi_{j_2})] \\ &= \mathbb{E}[Y_{j_1}Y_{j_2} - m(\pi_{j_1}Y_{j_2} + \pi_{j_2}Y_{j_1}) + m^2\pi_{j_1}\pi_{j_2}] \\ &= \mathbb{E}(Y_{j_1}Y_{j_2}) - m\pi_{j_1}\mathbb{E}(Y_{j_2}) - m\pi_{j_2}\mathbb{E}(Y_{j_1}) + m^2\pi_{j_1}\pi_{j_2} \\ &= \mathbb{E}(Y_{j_1}Y_{j_2}) - m^2\pi_{j_1}\pi_{j_2}. \end{aligned}$$

Der Term $E(Y_{j_1}Y_{j_2})$ kann umgeformt werden, indem die Zufallsvariablen Y_{j_1} und Y_{j_2} als Summe der Zufallsvariablen einer jeden Wiederholung dargestellt werden

$$E(Y_{j_1}Y_{j_2}) = E\left(\sum_{i_1=1}^m \sum_{i_2=1}^m \tilde{Y}_{i_1j_1} \tilde{Y}_{i_2j_2}\right) = \sum_{i_1=1}^m \sum_{i_2=1}^m E(\tilde{Y}_{i_1j_1} \tilde{Y}_{i_2j_2}). \quad (2.10)$$

Im Fall $i_1 = i_2$ gilt mit der Definition des Erwartungswerts

$$\begin{aligned} E(\tilde{Y}_{i_1j_1} \tilde{Y}_{i_2j_2}) &= \underbrace{0 \cdot 0 \cdot P(\tilde{Y}_{i_1j_1} = 0, \tilde{Y}_{i_2j_2} = 0)}_{=0} + \underbrace{1 \cdot 0 \cdot P(\tilde{Y}_{i_1j_1} = 1, \tilde{Y}_{i_2j_2} = 0)}_{=0} + \\ &\quad \underbrace{0 \cdot 1 \cdot P(\tilde{Y}_{i_1j_1} = 0, \tilde{Y}_{i_2j_2} = 1)}_{=0} + 1 \cdot 1 \cdot P(\tilde{Y}_{i_1j_1} = 1, \tilde{Y}_{i_2j_2} = 1). \end{aligned}$$

Da sich die Ereignisse bei der Multinomialverteilung laut Definition gegenseitig ausschließen, können bei der einmaligen Durchführung eines Experiments ($i_1 = i_2$) nicht beide Ereignisse j_1 und j_2 gleichzeitig eintreten. Somit ergibt sich die Wahrscheinlichkeit $P(\tilde{Y}_{i_1j_1} = 1, \tilde{Y}_{i_1j_2} = 1) = 0$ und daher ist in diesem Fall $E(\tilde{Y}_{i_1j_1} \tilde{Y}_{i_1j_2}) = 0$.

Für $i_1 \neq i_2$ gilt für die Ereignisse j_1 und j_2 $E(\tilde{Y}_{i_1j_1} \tilde{Y}_{i_2j_2}) = \pi_{j_1} \pi_{j_2}$, daher folgt

$$\sum_{\substack{i_1=1 \\ i_1 \neq i_2}}^m \sum_{i_2=1}^m E(\tilde{Y}_{i_1j_1} \tilde{Y}_{i_2j_2}) = m(m-1)\pi_{j_1}\pi_{j_2}.$$

Werden die beiden Fallunterscheidungen zusammengefasst, folgt die Darstellung

$$\sum_{i_1=1}^m \sum_{i_2=1}^m E(\tilde{Y}_{i_1j_1} \tilde{Y}_{i_2j_2}) = m(m-1)\pi_{j_1}\pi_{j_2}. \quad (2.11)$$

Mit (2.11) ergibt sich die Kovarianz

$$\text{Cov}(Y_{j_1}, Y_{j_2}) = m(m-1)\pi_{j_1}\pi_{j_2} - m^2\pi_{j_1}\pi_{j_2} = -m\pi_{j_1}\pi_{j_2}. \quad (2.12)$$

Beispiel 2. Erneut werden dieselben Daten wie in Beispiel 1 betrachtet. Diese Daten werden nun zusätzlich unterschieden in den Alterskategorien < 20 Jahre, $20 - 25$ Jahre und > 25 Jahre. Die erste Kategorie beinhalte 5 Personen, die zweite Kategorie 21 Personen und in der letzten Kategorie seien 4 Personen enthalten. Die Wahrscheinlichkeit, dass in einer Übungsgruppe von $m = 10$ Personen $y_1 = 4$ Personen aus der ersten Kategorie, nur $y_2 = 2$ Personen zwischen $20 - 25$ Jahren und weitere $y_3 = 4$ Personen > 25 Jahren sind, beträgt

$$f\left((4, 2, 4) | 10, \frac{5}{30}, \frac{21}{30}, \frac{4}{30}\right) = \frac{10!}{4!2!4!} \left(\frac{5}{30}\right)^4 \left(\frac{21}{30}\right)^2 \left(\frac{4}{30}\right)^4 = 0.00038.$$

Die Wahrscheinlichkeit für diese Gruppenkonstellation ist 0.038%. Dargestellt ist die Wahrscheinlichkeitsfunktion dieses Beispiels in der Abbildung 2.2.

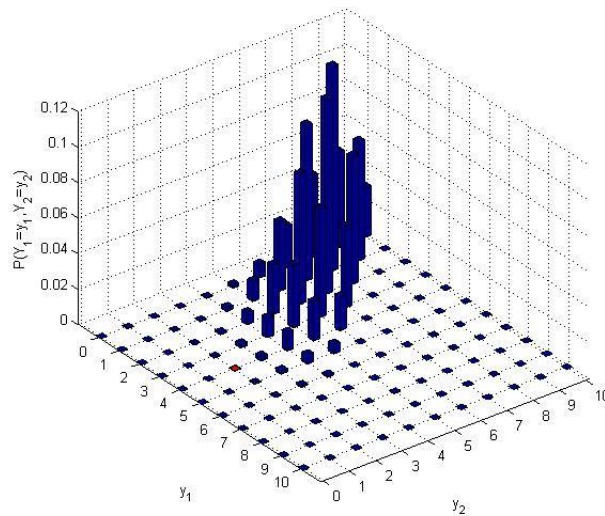


Abbildung 2.2: Darstellung der Wahrscheinlichkeitsfunktion zu Beispiel 2

2.3 Poissonverteilung

Eine weitere, wichtige diskrete Verteilung, welche in einem engen Bezug zur Binomial- und der Multinomialverteilung steht, ist die Poissonverteilung. Die Wahrscheinlichkeitsfunktion der Poissonverteilung ist gegeben durch

$$P(Y = y) = f(y|\lambda) = \frac{\lambda^y}{y!} \exp(-\lambda), \quad y = 0, 1, 2, \dots, \quad \lambda > 0, \quad (2.13)$$

wobei die Funktion von einem fest gewählten Parameter λ abhängig ist. Die Poissonverteilung hat die Momente

$$\begin{aligned} E(Y) &= \lambda \\ \text{Var}(Y) &= \lambda. \end{aligned}$$

Mit Hilfe der Poissonverteilung kann die Multinomialverteilung hergeleitet werden. Seien dazu Y_1, \dots, Y_k unabhängige Zufallsvariablen aus k verschiedenen Poissonverteilungen mit Parametern $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)$. Die gemeinsame Wahrscheinlichkeitsfunktion der Beobachtungen ergibt sich als Produkt der einzelnen Wahrscheinlichkeitsfunktionen mit

$$f(\mathbf{y}|\boldsymbol{\lambda}) = \prod_{j=1}^k \frac{\lambda_j^{y_j} \exp(-\lambda_j)}{y_j!} = \frac{\lambda_1^{y_1} \dots \lambda_k^{y_k} \exp(-\sum_{j=1}^k \lambda_j)}{y_1! \dots y_k!} \quad (2.14)$$

für $\mathbf{y} = (y_1, \dots, y_k)$ und $y_j = 0, 1, 2, \dots$. Dabei ist die gemeinsame Verteilung (2.14) eine Verteilung für nicht negative Beobachtungen y_j . Durch späteres Konditionieren auf $y^+ = \sum_{j=1}^k y_j = n$, wobei $n \in \mathbb{N}$ fest ist und die Anzahl der Wiederholungen eines Experiments beschreibt, muss noch die Verteilung der Summe von unabhängigen Poissonvariablen $Y^+ = \sum_{j=1}^k Y_j$ mit Hilfe der momenterzeugenden Funktion hergeleitet werden. Diese ist gegeben durch

$$M_{Y^+}(t) = M_{Y_1}(t) \cdots M_{Y_k}(t) = \prod_{j=1}^k \exp(\lambda_j(e^t - 1)) = \exp\left(\sum_{j=1}^k \lambda_j(e^t - 1)\right),$$

was der Momentenfunktion einer Poisson $\left(\sum_{j=1}^k \lambda_j\right)$ entspricht. Die bedingte Wahrscheinlichkeitsfunktion $f(\mathbf{y}|y^+; \boldsymbol{\lambda})$ kann berechnet werden mit

$$\begin{aligned} f(\mathbf{y}|y^+; \boldsymbol{\lambda}) &= \frac{f(\mathbf{y}|\boldsymbol{\lambda})}{f(y^+|\boldsymbol{\lambda})} \\ &= \frac{\exp(-\sum_{j=1}^k \lambda_j) \lambda_1^{y_1} \cdots \lambda_k^{y_k}}{y_1! \cdots y_k!} \left(\frac{\exp\left(-\sum_{j=1}^k \lambda_j\right) \left(\sum_{j=1}^k \lambda_j\right)^{y^+}}{y^+!} \right)^{-1} \\ &= \frac{y^+!}{y_1! \cdots y_k!} \frac{\lambda_1^{y_1} \cdots \lambda_k^{y_k}}{\left(\sum_{j=1}^k \lambda_j\right)^{y_1 + \cdots + y_k}} \\ &= \frac{y^+!}{y_1! \cdots y_k!} \left(\frac{\lambda_1}{\sum_{j=1}^k \lambda_j}\right)^{y_1} \cdots \left(\frac{\lambda_k}{\sum_{j=1}^k \lambda_j}\right)^{y_k}. \end{aligned}$$

Mit der Wahl von $\pi_i = \frac{\lambda_i}{\sum_{j=1}^k \lambda_j}$, $i = 1, \dots, k$ entspricht dieses der Wahrscheinlichkeitsfunktion einer Multinomialverteilung (2.8). Daher unterliegt der Vektor (Y_1, \dots, Y_k) von k unabhängig Poisson-verteilten Zufallsvariablen gegeben der Summe Y^+ einer Multinomialverteilung $(Y_1, \dots, Y_k|Y^+) \sim M(y^+, \boldsymbol{\pi})$. Bei den Beispielen zuvor war die Population gegeben, im folgenden Beispiel wird aus der Stichprobe geschätzt.

Beispiel 3. Bei einer Prüfung an der Universität gibt es eine Prüfungsstatistik von 500 Studierenden, dabei mussten 350 der Studierenden die Prüfung nicht wiederholen, 120 Studierende hatten eine einmalige Wiederholung, 25 Studierende zweimal. Weitere 4 Studierende mussten dreimal zur Prüfung antreten, ein Studierender benötigte vier Versuche bevor er die Prüfung abschließen konnte.

Wahrscheinlichkeit y	0	1	2	3	4
Häufigkeit	350	120	25	4	1

Auf $n = 500$ Studierende kommen somit 186 Prüfungswiederholungen. Für einen beliebigen Studenten ergibt sich eine mittlere (erwartete) Anzahl von Wiederholungen von $\hat{\lambda} = 186/500 = 0.372$. Mit Hilfe dieser Daten können nun Wahrscheinlichkeiten mittels Poissonverteilung geschätzt werden. Es ergibt sich

$$P(Y = 2 | \hat{\lambda} = 0.372) = \frac{\hat{\lambda}^2}{2!} \exp(-\hat{\lambda}) = \frac{0.372^2}{2!} \exp(-0.372) = 0.0477,$$

$$P(Y = 3 | \hat{\lambda} = 0.372) = \frac{\hat{\lambda}^3}{3!} \exp(-\hat{\lambda}) = \frac{0.372^3}{3!} \exp(-0.372) = 0.00591.$$

Die geschätzte Wahrscheinlichkeit, dass ein Studierender die Prüfung zweimal wiederholen muss beträgt 4.77% und für eine dreifache Wiederholung 0.591%. Das Beispiel wird in der Abbildung 2.3 dargestellt.

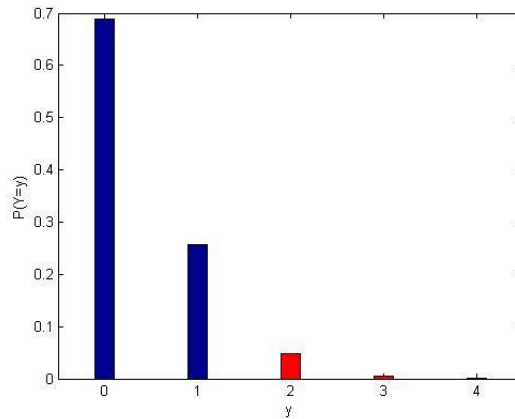


Abbildung 2.3: Darstellung der Wahrscheinlichkeitsfunktion zu Beispiel 3

2.4 Exponentialfamilie

In diesem Abschnitt wird gezeigt, dass die Binomial-, die Multinomial- und die Poissonverteilung aus einer allgemeinen, gemeinsamen Klasse von Verteilungsfunktionen abgeleitet werden können. Mittels Umformung können dabei die Wahrscheinlichkeitsfunktionen auf eine allgemeine Form zurückgeführt werden, welche auf Grund ihrer Darstellung *Exponentialfamilie* genannt wird. Diese Familie stellt eine sehr vorteilhafte Klasse von Verteilungen dar und führt durch ihre verallgemeinerte Form zu vielen Vorteilen und Vereinfachungen.

Definition 2.1. (*allgemeine Exponentialfamilie*) Sei y die Realisierung einer Zufallsvariable aus einer Verteilung mit Dichte- oder Wahrscheinlichkeitsfunktion $f(y|\boldsymbol{\theta})$. Eine allgemeine, k -parametrische Exponentialfamilie ist definiert durch

$$f(y|\boldsymbol{\theta}) = h(y)p(\boldsymbol{\theta}) \exp(\mathbf{t}(y)^\top \mathbf{w}(\boldsymbol{\theta})) \quad (2.15)$$

mit $\mathbf{t}(y) = (t_1(y), \dots, t_k(y))^\top$ und $\mathbf{w}(\boldsymbol{\theta}) = (w_1(\boldsymbol{\theta}), \dots, w_k(\boldsymbol{\theta}))^\top$ reellwertige Funktionen. Weiters sind $h(y) \geq 0$ und $p(\boldsymbol{\theta}) \geq 0$.

Diese allgemeine Exponentialfamilie hat die äquivalente Darstellung

$$f(y|\boldsymbol{\theta}) = \exp[\mathbf{t}(y)^\top \mathbf{w}(\boldsymbol{\theta}) + \log(h(y)) + \log(p(\boldsymbol{\theta}))].$$

Im einparametrischen Fall $k = 1$ erhält man mit der Wahl von $t(y) = y$ (linear), $w(\theta) = \theta$ (kanonisch), $\log(p(\theta)) = -b(\theta)$ und $\log(h(y)) = c(y)$ die einparametrische, lineare Exponentialfamilie in kanonischer Form.

Definition 2.2. (*einparametrische Exponentialfamilie*) Sei y die Realisierung einer Zufallsvariable aus einer Verteilung mit Dichte- oder Wahrscheinlichkeitsfunktion

$$f(y|\theta) = \exp[y\theta - b(\theta) + c(y)] \quad (2.16)$$

für spezielle, bekannte Funktionen $b(\cdot)$ und $c(\cdot)$. $f(y|\theta)$ stellt die einparametrische, lineare Exponentialfamilie mit dem kanonischen Parameter θ dar.

Für die allgemeine Exponentialfamilie (2.15) gelten die folgenden Identitäten

$$\mathbb{E} \left(\frac{\partial \log(f(y|\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} \right) = \mathbf{0} \quad (2.17)$$

$$\text{Var} \left(\frac{\partial \log(f(y|\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} \right) = \mathbb{E} \left(-\frac{\partial^2 \log(f(y|\boldsymbol{\theta}))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right). \quad (2.18)$$

Nun werden die drei in den Abschnitten zuvor erklärten Verteilungen umgeformt, sodass gezeigt werden kann, dass diese der Klasse der Exponentialfamilie zugeordnet werden können.

Standardisierte Binomialverteilung

Die Binomialverteilung $B(m, \pi)$ kann in eine standardisierte Form überführt werden. Anstelle der absoluten Häufigkeiten $y \in \{0, 1, \dots, m\}$ werden die relativen Häufigkeiten $\bar{y} \in \{0, \frac{1}{m}, \frac{2}{m}, \dots, 1\}$ betrachtet. Für die Zufallsvariable der relativen Häufigkeiten gilt $m\bar{Y} \sim B(m, \pi)$. Die Wahrscheinlichkeitsfunktion ergibt sich wie folgend

$$f(y|m, \pi) = P(Y = y) = P(m\bar{Y} = m\bar{y}) = \frac{m!}{m\bar{y}!(m - m\bar{y})!} \pi^{m\bar{y}} (1 - \pi)^{m - m\bar{y}}$$

$$\begin{aligned}
&= \frac{m!}{m\bar{y}!(m-m\bar{y})!} \pi^{m\bar{y}} (1-\pi)^m (1-\pi)^{-m\bar{y}} \\
&= \exp \left[m\bar{y} \log(\pi) + m \log(1-\pi) - m\bar{y} \log(1-\pi) + \log \left(\frac{m!}{m\bar{y}!(m-m\bar{y})!} \right) \right] \\
&= \exp \left[m\bar{y} \log \left(\frac{\pi}{1-\pi} \right) + m \log(1-\pi) + \log \left(\frac{m!}{m\bar{y}!(m-m\bar{y})!} \right) \right] \\
&= \exp \left[m\bar{y} \log \left(\frac{\pi}{1-\pi} \right) - m \log \left(\frac{1}{1-\pi} \right) + \log \left(\frac{m!}{m\bar{y}!(m-m\bar{y})!} \right) \right] \\
&= \exp \left[\underbrace{m\bar{y} \log \left(\frac{\pi}{1-\pi} \right)}_{=: \theta} - m \log \left(1 + \frac{\pi}{1-\pi} \right) + \log \left(\frac{m!}{m\bar{y}!(m-m\bar{y})!} \right) \right] \\
&= \exp \left[m\bar{y}\theta - \underbrace{m \log(1 + \exp(\theta))}_{=: b(\theta)} + \underbrace{\log \left(\frac{m!}{m\bar{y}!(m-m\bar{y})!} \right)}_{=: c(\bar{y})} \right] \\
&= \exp \left[\frac{\bar{y}\theta - b(\theta)}{1/m} + c(\bar{y}) \right]. \tag{2.19}
\end{aligned}$$

Der Term $1/m$ ist konstant und verändert die Eigenschaften der Verteilung nicht, weshalb dieser Term keinen relevanten Einfluss hat. Mit dieser Äquivalenz kann die Binomialverteilung der Klasse der einparametrischen, linearen Exponentialfamilie zugeordnet werden.

In weiterer Folge können die Eigenschaften der behandelten Verteilungen angegeben werden. Mit Hilfe der Gleichungen (2.17) und (2.18) gilt für die Binomialverteilung

$$\begin{aligned}
\mathbb{E} \left(\frac{\partial \log(f(y|\theta))}{\partial \theta} \right) &= \mathbb{E} \left(\frac{\partial}{\partial \theta} \left(m\bar{y}\theta - m \log(1 + \exp(\theta)) + \log \left(\frac{m!}{m\bar{y}!(m-m\bar{y})!} \right) \right) \right) \\
&= \mathbb{E} \left(m\bar{y} - m \frac{\exp(\theta)}{1 + \exp(\theta)} \right) = \mathbb{E}(m\bar{y}) - m\pi = 0 \\
&\implies \mathbb{E}(m\bar{y}) = m\pi \\
\text{Var} \left(\frac{\partial \log(f(y|\theta))}{\partial \theta} \right) &= \mathbb{E} \left(- \frac{\partial^2 \log(f(y|\theta))}{\partial \theta^2} \right) = \mathbb{E} \left(m \frac{\exp(\theta)}{(1 + \exp(\theta))^2} \right) = m\pi(1-\pi) \\
&\implies \text{Var}(m\bar{y}) = m\pi(1-\pi)
\end{aligned}$$

Poissonverteilung

Analog ist auch die Poissonverteilung ein Mitglied der Exponentialfamilie. Durch Umformung der Wahrscheinlichkeitsfunktion der Poissonverteilung ergibt sich

$$\begin{aligned}
 f(y|\lambda) &= P(Y = y) = \frac{\lambda^y}{y!} \exp(-\lambda) \\
 &= \exp\left[y \underbrace{\log(\lambda)}_{=: \theta} - \lambda - \log(y!)\right] = \exp\left[y\theta - \underbrace{\exp(\theta)}_{=: b(\theta)} - \underbrace{\log(y!)}_{=: c(y)}\right] \\
 &= \exp[y\theta - b(\theta) + c(y)].
 \end{aligned} \tag{2.20}$$

Mit diesen Umformungen hat die Wahrscheinlichkeitsfunktion die Form (2.16) der Exponentialfamilie. Die Momente der Poissonverteilung können in analoger Weise mit Hilfe von (2.17) und (2.18) hergeleitet werden

$$\begin{aligned}
 \mathbb{E}\left(\frac{\partial \log(f(y|\theta))}{\partial \theta}\right) &= \mathbb{E}(y - \exp(\theta)) = \mathbb{E}(y) - \lambda = 0 \\
 &\implies \mathbb{E}(y) = \lambda \\
 \text{Var}\left(\frac{\partial \log(f(y|\theta))}{\partial \theta}\right) &= \mathbb{E}\left(-\frac{\partial^2 \log(f(y|\theta))}{\partial \theta^2}\right) = \mathbb{E}(\exp(\theta)) = \lambda \\
 &\implies \text{Var}(y) = \lambda
 \end{aligned}$$

Wesentliche Eigenschaften der einparametrischen Exponentialfamilie (2.16) können mit Hilfe der Eigenschaften für die k -parametrische Exponentialfamilie hergeleitet werden, diese ergeben sich folgend

$$\begin{aligned}
 \mathbb{E}\left(\frac{\partial \log(f(y|\theta))}{\partial \theta}\right) &= \mathbb{E}\left(\frac{\partial \log(\exp[y\theta - b(\theta) + c(y)])}{\partial \theta}\right) \\
 &= b'(\theta) \\
 &\implies \mathbb{E}(y) = b'(\theta) \\
 \text{Var}\left(\frac{\partial \log(f(y|\theta))}{\partial \theta}\right) &= \mathbb{E}\left(-\frac{\partial^2 \log(f(y|\theta))}{\partial \theta^2}\right) = \mathbb{E}\left(-\frac{\partial^2 \log(\exp[y\theta - b(\theta) + c(y)])}{\partial \theta^2}\right) \\
 &= b''(\theta) \\
 &\implies \text{Var}(y) = b''(\theta)
 \end{aligned}$$

Multinomialverteilung

Ebenso wie die Wahrscheinlichkeitsfunktionen der Binomial- und der Poissonverteilung kann die Wahrscheinlichkeitsfunktion der Multinomialverteilung umgeschrieben

werden mittels

$$\begin{aligned}
f(\mathbf{y}|m, \boldsymbol{\pi}) &= \frac{m!}{y_1! \cdots y_k!} \pi_1^{y_1} \cdots \pi_k^{y_k} \\
&= \frac{m!}{y_1! \cdots y_k!} \exp[y_1 \log(\pi_1) + \cdots + y_k \log(\pi_k)] \\
&= \frac{m!}{\underbrace{y_1! \cdots y_k!}_{=:h(\mathbf{y}_1, \dots, \mathbf{y}_k)}} \exp \left[\sum_{j=1}^k y_j \log(\pi_j) \right] \\
&= h(\mathbf{y}_1, \dots, \mathbf{y}_k) \exp \left[\mathbf{t}(\mathbf{y}_1, \dots, \mathbf{y}_k)^\top \boldsymbol{\omega}(\theta_1, \dots, \theta_k) \right], \tag{2.21}
\end{aligned}$$

wobei $\mathbf{t}(\mathbf{y}_1, \dots, \mathbf{y}_k) = (y_1, \dots, y_k)^\top$ und $\boldsymbol{\omega}(\theta_1, \dots, \theta_k) = (\theta_1, \dots, \theta_k)^\top$ mit den Parametern $(\theta_1, \dots, \theta_k)^\top = (\log(\pi_1), \dots, \log(\pi_k))^\top$. Auf Grund dieser Darstellung gehört die Multinomialverteilung der allgemeinen, k -parametrischen Exponentialfamilie an. Als letztes werden Identitäten der Multinomialverteilung betrachtet, wobei diese mittels der Eigenschaften der allgemeinen k -parametrischen Exponentialfamilie (2.17) und (2.18) hergeleitet werden. Als Wahrscheinlichkeitsfunktion wird die reduzierte Form (2.9) verwendet und es ergibt sich

$$\begin{aligned}
\mathbb{E} \left(\frac{\partial \log(f(\mathbf{y}|\boldsymbol{\theta}))}{\partial \theta_s} \right) &= \mathbb{E} \left(\frac{\partial}{\partial \theta_s} \left(\sum_{j=1}^{k-1} y_j \theta_j + y_k \log \left(1 - \sum_{j=1}^{k-1} \exp(\theta_j) \right) \right) \right) \\
&= \mathbb{E} \left(y_s - y_k \frac{\pi_s}{\pi_k} \right) = \mathbb{E}(y_s) - \mathbb{E} \left(y_k \frac{\pi_s}{\pi_k} \right) = 0, \quad s = 1, \dots, k-1.
\end{aligned}$$

Also

$$\mathbb{E}(y_s) = \frac{\mathbb{E}(y_k)}{\pi_k} \pi_s.$$

Durch Aufsummieren aller $k-1$ Gleichungen ergibt sich mit der Summe $\sum_{j=1}^k y_j = m$

$$\begin{aligned}
\sum_{s=1}^{k-1} \mathbb{E}(y_s) &= \frac{\mathbb{E}(y_k)}{\pi_k} \sum_{s=1}^{k-1} \pi_s \\
m - \mathbb{E}(y_k) &= \frac{\mathbb{E}(y_k)}{\pi_k} (1 - \pi_k) \\
m\pi_k &= \pi_k \mathbb{E}(y_k) + \mathbb{E}(y_k) - \pi_k \mathbb{E}(y_k) = \mathbb{E}(y_k).
\end{aligned}$$

Der Erwartungswert einer anderen Komponente ergibt sich durch Umordnung.

Die Varianz einer Komponente folgt durch die folgenden Äquivalenzen

$$\text{Var} \left(\frac{\partial \log(f(\mathbf{y}|\boldsymbol{\theta}))}{\partial \theta_s} \right) = \mathbb{E} \left(- \frac{\partial^2 \log(f(\mathbf{y}|\boldsymbol{\theta}))}{\partial \theta_s^2} \right)$$

Auf Grund der Gleichungen

$$\begin{aligned}\text{Var}\left(\frac{\partial \log(f(\mathbf{y}|\boldsymbol{\theta}))}{\partial \theta_s}\right) &= \text{Var}\left(y_s - y_k \frac{\pi_s}{\pi_k}\right) \\ \text{E}\left(-\frac{\partial^2 \log(f(\mathbf{y}|\boldsymbol{\theta}))}{\partial \theta_s^2}\right) &= \text{E}\left(y_k \frac{\exp(\theta_s)(1 - \sum_{j=1}^{k-1} \exp(\theta_j)) + \exp(\theta_s) \exp(\theta_s)}{1 - \sum_{j=1}^{k-1} \exp(\theta_j)^2}\right),\end{aligned}$$

gilt die Äquivalenz

$$\text{Var}(y_s) - 2\text{Cov}\left(y_s, y_k \frac{\pi_s}{\pi_k}\right) + \text{Var}\left(y_k \frac{\pi_s}{\pi_k}\right) = m\pi_s + m\frac{\pi_s^2}{\pi_k}.$$

Mit der Kovarianz für die Multinomialverteilung (2.12) ergibt sich

$$\begin{aligned}\text{Var}\left(y_k \frac{\pi_s}{\pi_k}\right) &= -2m\pi_s^2 - \text{Var}(y_s) + m\pi_s + m\frac{\pi_s^2}{\pi_k} \\ \text{Var}\left(\frac{y_k}{\pi_k}\right) &= -2m - \text{Var}\left(\frac{y_s}{\pi_s}\right) + \frac{m}{\pi_s} + \frac{m}{\pi_k} \\ \text{Var}\left(\frac{y_k}{\pi_k}\right) &= m \cdot \left(-2 + \frac{1}{\pi_s} + \frac{1}{\pi_k}\right) - \text{Var}\left(\frac{y_s}{\pi_s}\right).\end{aligned}\tag{2.22}$$

Ergebnis (2.22) gilt für alle $s = 1, \dots, k-1$ und somit für ein $t \neq s \in \{1, \dots, k-1\}$. Wird die Differenz der Darstellung für ein s und für ein t betrachtet, so folgt

$$0 = m \cdot \left(-2 + \frac{1}{\pi_s} + \frac{1}{\pi_k}\right) - \text{Var}\left(\frac{y_s}{\pi_s}\right) - m \cdot \left(-2 + \frac{1}{\pi_t} + \frac{1}{\pi_k}\right) + \text{Var}\left(\frac{y_t}{\pi_t}\right)$$

und durch Umformungen

$$\text{Var}\left(\frac{y_s}{\pi_s}\right) - \frac{m}{\pi_s} = \text{Var}\left(\frac{y_t}{\pi_t}\right) - \frac{m}{\pi_t}.$$

Da dies für beliebige t gilt, folgt $\text{Var}\left(\frac{y_t}{\pi_t}\right) - \frac{m}{\pi_t} = c$, wobei c eine Konstante ist. Somit gilt

$$\text{Var}\left(\frac{y_s}{\pi_s}\right) = c + \frac{m}{\pi_s}.$$

Rückeinsetzen in (2.22) ergibt

$$\text{Var}\left(\frac{y_k}{\pi_k}\right) = -2m - c + \frac{m}{\pi_k},$$

also

$$\underbrace{\text{Var}\left(\frac{y_k}{\pi_k}\right) - \frac{m}{\pi_k}}_{=c} = -2m - c$$

und da $c = -m$, folgt daraus

$$\text{Var}(y_k) = m\pi_k(1 - \pi_k).$$

3 Generalisierte Lineare Modelle

Das Ergebnis eines Experiments kann in seiner Erwartung von verschiedenen, unabhängigen (physikalischen, biologischen, sozialen, temporären, ...) Größen abhängig sein. In einem statistischen Modell wird versucht, den Einfluss dieser erklärenden Größen auf das Ergebnis zu modellieren.

3.1 Basisstruktur der Generalisierten Linearen Modelle

In diesem Kapitel bezeichnet $n \in \mathbb{N}$ die Anzahl der Response $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ mit dem Erwartungswert $E(\mathbf{Y}) = \boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^\top$. Der Vektor $\mathbf{y} = (y_1, \dots, y_n)^\top$ ist eine Realisierung der Response \mathbf{Y} . Zusätzlich sind zu jedem Y_i die erklärenden Variablen $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ip-1})$, $i = 1, \dots, n$, $p \in \mathbb{N}$ gegeben. Der konstante Term 1 des ersten Eintrags des Vektors \mathbf{x}_i ist dabei der Intercept.

Stammt die Response \mathbf{Y} aus einer Verteilung der Exponentialfamilie, stellen die *Generalisierten Linearen Modelle* (GLMs) ein allgemeines und weit verbreitetes Modell zur Datenbeschreibung dar. In diesen Modellen wird der Erwartungswert der Response als Funktion der erklärenden Variablen dargestellt. Ein GLM ist durch drei Komponenten bestimmt. Die Realisierung \mathbf{y} , die Wahrscheinlichkeitsfunktion $f(y_i|\boldsymbol{\theta}_i)$ und der Erwartungswert $E(\mathbf{Y})$ bilden die *stochastische Komponente* des GLMs. Der Effekt $\boldsymbol{\beta}$ der erklärenden Variablen \mathbf{x}_i wird in der *systematischen Komponente* mit einem linearen Modell dargestellt und mit Hilfe der *Linkfunktion* $g(\cdot)$ an die vorliegende Verteilung angepasst.

Stochastische Komponente: $y_i \stackrel{ind}{\sim} f(y_i|\boldsymbol{\theta}_i)$, $E(Y_i) = \mu_i$

Die stochastische Komponente beschreibt die unabhängig verteilten Y_i , deren Realisierungen y_i mit der Wahrscheinlichkeitsfunktion $f(y_i|\boldsymbol{\theta}_i)$ aus einer Exponentialfamilie

verteilt sind. Der Erwartungswert von Y_i wird mit μ_i bezeichnet, $E(Y_i) = \mu_i$ mit dem Vektor $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^\top$.

Systematische Komponente: $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$

Um den Einfluss der erklärenden Variablen \mathbf{x}_i darstellen zu können, wird in der systematischen Komponente zu jedem Y_i ein Prädiktor η_i definiert. Der Prädiktor wird von unbekanntem Parametern $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{p-1})^\top$, den sogenannten *Effekten*, gebildet, welche anhand der Realisierungen y_i geschätzt werden müssen. Die Darstellung des Prädiktors erfolgt im GLM durch eine Linearkombination der erklärenden Variablen und der Effekte

$$\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta} = \sum_{l=0}^{p-1} x_{il} \beta_l, \quad i = 1, \dots, n,$$

wobei x_{i0} den Intercept 1 bezeichnet. Wird der Vektor der Prädiktoren $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^\top$ betrachtet, so ergibt sich das Gleichungssystem

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

Die Matrix \mathbf{X} , welche Designmatrix genannt wird, ist eine $(n \times p)$ -Matrix mit

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np-1} \end{pmatrix}. \quad (3.1)$$

Linkfunktion: $g(\mu_i) = \eta_i$

Die Linkfunktion $g: \mathbb{R} \rightarrow \mathbb{R}$ ist eine monotone und differenzierbare Funktion, welche den Erwartungswert μ_i auf den linearen Prädiktor η_i mit der Vorschrift

$$g(\mu_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta} \quad (3.2)$$

abbildet. Die inverse Linkfunktion $g^{-1}(\cdot) = h(\cdot)$ wird Responsefunktion genannt und bildet die Transformation des linearen Prädiktors auf den Erwartungswert

$$\mu_i = h(\eta_i) = h(\mathbf{x}_i^\top \boldsymbol{\beta}). \quad (3.3)$$

Für jede Verteilung der einparametrischen, linearen Exponentialfamilie (2.16) kann eine einfache Linkfunktion, die sogenannte kanonische Linkfunktion angegeben werden. Diese bildet die direkte Verbindung des Prädiktors η_i mit dem Parameter θ_i der Wahrscheinlichkeitsfunktion $f(y_i|\theta_i)$ in der Form

$$\theta_i = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta} = g(\mu_i). \quad (3.4)$$

Das Ergebnis (3.4) zeigt, dass der Parameter θ_i eindeutig durch den Erwartungswert μ_i festgelegt ist. Auf Grund der Gleichung (3.2) folgt, dass der Parameter θ_i somit vom Vektor der Effekte $\boldsymbol{\beta}$ abhängig ist. Im speziellen Fall der Binomial- und Multinomialverteilung gilt komponentenweise der Zusammenhang

$$\mu_i = \pi_i \quad \text{bzw.} \quad \mu_{ij} = \pi_{ij},$$

für die Poissonverteilung gilt der Zusammenhang

$$\mu_i = \lambda_i.$$

3.2 Maximum Likelihood Schätzung

Die Wahrscheinlichkeitsfunktion $f(y_i|\theta_i)$ ist von dem unbekanntem Parameter θ_i mit $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^\top$ abhängig, weshalb dieser Parameter geschätzt werden muss. Eine bekannte und weit verbreitete Methode der Schätzung ist die *Maximum Likelihood Schätzung*. Hierfür wird die gemeinsame Wahrscheinlichkeitsfunktion von \mathbf{y} verwendet, also

$$f(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^n f(y_i|\theta_i).$$

Da $\theta_i = \theta_i(\boldsymbol{\beta})$, ist die Likelihood-Funktion $L(\boldsymbol{\beta}|\mathbf{y})$ mit

$$L(\boldsymbol{\beta}|\mathbf{y}) = f(\mathbf{y}|\boldsymbol{\theta})$$

eine Funktion in $\boldsymbol{\beta}$ abhängig von \mathbf{y} . Auf Grund der einfacheren Darstellung wird die Log-Likelihood-Funktion $l(\boldsymbol{\beta}|\mathbf{y}) = \log(L(\boldsymbol{\beta}|\mathbf{y}))$ betrachtet

$$l(\boldsymbol{\beta}|\mathbf{y}) = \log \left(\prod_{i=1}^n f(y_i|\theta_i) \right).$$

Der Maximum Likelihood Schätzer $\hat{\boldsymbol{\beta}}$ ist als Maximalstelle der Likelihood-Funktion definiert und ist auf Grund der Monotonie des Logarithmus auch die Maximalstelle der Log-Likelihood-Funktion. Die Berechnung von $\hat{\boldsymbol{\beta}}$ erfolgt numerisch über die erste Ableitung der Log-Likelihood-Funktion, der sogenannten Score-Funktion

$$\mathbf{s}(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta}|\mathbf{y})}{\partial \boldsymbol{\beta}}.$$

Durch Nullsetzen ergibt sich die Schätz-Gleichung für $\hat{\boldsymbol{\beta}}$

$$\mathbf{s}(\hat{\boldsymbol{\beta}}) = \mathbf{0}. \tag{3.5}$$

Wird die Hessematrix \mathbf{H} von $l(\boldsymbol{\beta}|\mathbf{y})$ betrachtet,

$$\mathbf{H}(l(\boldsymbol{\beta}|\mathbf{y})) = \frac{\partial^2 l(\boldsymbol{\beta}|\mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top},$$

so zeigt sich, dass diese Matrix zufällige Komponenten beinhaltet und von den Realisierungen abhängt. Der Erwartungswert der negativen Hessematrix wird *Fisher-Matrix* genannt und ist gegeben durch

$$\mathbf{F}(\boldsymbol{\beta}) = \mathbb{E}(-\mathbf{H}(l(\boldsymbol{\beta}|\mathbf{y}))) = \mathbb{E}\left(-\frac{\partial^2 l(\boldsymbol{\beta}|\mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}\right).$$

Die Schätz-Gleichung (3.5) zur Berechnung von $\hat{\boldsymbol{\beta}}$ ist nicht immer analytisch lösbar. In diesem Fall muss die Lösung numerisch approximiert werden. Eine Möglichkeit ist die *iteratively reweighted least-squares* Methode, welche im ψ -ten Schritt die folgende Vorschrift hat

$$\hat{\boldsymbol{\beta}}^{(\psi+1)} = \hat{\boldsymbol{\beta}}^{(\psi)} + \left[\mathbf{F}\left(\hat{\boldsymbol{\beta}}^{(\psi)}\right)\right]^{-1} \mathbf{s}\left(\hat{\boldsymbol{\beta}}^{(\psi)}\right), \quad \psi = 0, 1, \dots, \quad (3.6)$$

wobei $\hat{\boldsymbol{\beta}}^{(0)}$ eine gewählte Startlösung ist. Weiters muss ein entsprechendes Konvergenzkriterium festgelegt werden.

Die Score-Funktion hat die Darstellung

$$\mathbf{s}(\boldsymbol{\beta}) = \mathbf{X}^\top \mathbf{D}(\boldsymbol{\beta}) \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) = \mathbf{X}^\top \mathbf{W}(\boldsymbol{\beta}) [\mathbf{D}(\boldsymbol{\beta})]^{-1} (\mathbf{y} - \boldsymbol{\mu})$$

mit der Designmatrix \mathbf{X} (3.1) und

$$\begin{aligned} \mathbf{D}(\boldsymbol{\beta}) &= \text{diag}\left(\frac{\partial h(\eta_1)}{\partial \eta}, \dots, \frac{\partial h(\eta_n)}{\partial \eta}\right) \\ \boldsymbol{\Sigma} &= \text{diag}(\text{Var}(y_1), \dots, \text{Var}(y_n)) \\ \mathbf{W}(\boldsymbol{\beta}) &= \mathbf{D}(\boldsymbol{\beta}) \boldsymbol{\Sigma}^{-1} [\mathbf{D}(\boldsymbol{\beta})]. \end{aligned}$$

Für die Fisher-Matrix $\mathbf{F}(\boldsymbol{\beta})$ gilt

$$\mathbf{F}(\boldsymbol{\beta}) = \mathbf{X}^\top \mathbf{D}(\boldsymbol{\beta}) \boldsymbol{\Sigma}^{-1} \mathbf{D}(\boldsymbol{\beta}) \mathbf{X} = \mathbf{X}^\top \mathbf{W}(\boldsymbol{\beta}) \mathbf{X}.$$

Unter gewissen Regularitätsbedingungen, wie in Fahrmeir und Kaufmann (1985) nachzulesen, gilt für den Maximum Likelihood Schätzer $\hat{\boldsymbol{\beta}}$

- i) $\hat{\boldsymbol{\beta}}$ existiert asymptotisch für $n \rightarrow \infty$
- ii) $\hat{\boldsymbol{\beta}}$ ist konsistent

iii) $\hat{\boldsymbol{\beta}}$ ist asymptotisch normalverteilt mit

$$\hat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, [\mathbf{F}(\hat{\boldsymbol{\beta}})]^{-1}\right).$$

Die inverse Fisher-Matrix $[\mathbf{F}(\hat{\boldsymbol{\beta}})]^{-1}$ ergibt den Standardfehler des Schätzers $\hat{\boldsymbol{\beta}}$.

3.3 Generalisierte Lineare Modelle für diskrete Verteilungen

In Abschnitt 2.4 wurde gezeigt, dass die standardisierte Binomial-, die Poisson- und die Multinomialverteilung Mitglieder der Exponentialfamilie sind. Im folgenden Abschnitt werden daher die GLMs für diese diskreten Verteilungen formuliert.

3.3.1 Standardisierte Binomialverteilung

Anhand der Äquivalenzen in (2.19) wurde gezeigt, dass die standardisierte Binomialverteilung mit $m_i \bar{y}_i \sim B(m_i, \pi_i)$ Mitglied der einparametrischen, linearen Exponentialfamilie mit $\theta = \log\left(\frac{\pi}{1-\pi}\right)$, $\pi \in (0, 1)$ ist. Für den Erwartungswert der standardisierten Binomialverteilung gilt $\mu_i = E(\bar{Y}_i) = \pi_i$, die Varianz ist gegeben durch $\text{Var}(\bar{Y}_i) = \frac{1}{m_i} \pi_i(1-\pi_i)$. Die kanonische Linkfunktion kann, wie in Abschnitt 3.2 erklärt, angegeben werden durch

$$\theta_i = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta} = g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right). \quad (3.7)$$

Die inverse Linkfunktion ergibt sich dann zu

$$\pi_i = g^{-1}(\theta_i) = h(\theta_i) = \frac{\exp(\theta_i)}{1 + \exp(\theta_i)}.$$

Die Log-Likelihood-Funktion der standardisierten Binomialverteilung hat die Darstellung

$$l(\boldsymbol{\beta} | \bar{y}_1, \dots, \bar{y}_n) = \sum_{i=1}^n m_i (\bar{y}_i \log(\pi_i(\boldsymbol{\beta})) + (1 - \bar{y}_i) \log(1 - \pi_i(\boldsymbol{\beta})) + c(\bar{y}_i)).$$

Die Score-Funktion ergibt sich zu

$$\mathbf{s}(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{x}_i \frac{\partial h(\eta_i)}{\partial \eta_i} \frac{(\bar{y}_i - \pi_i(\boldsymbol{\beta}))}{\pi_i(\boldsymbol{\beta})(1 - \pi_i(\boldsymbol{\beta}))},$$

wobei die Score-Funktion mit der Wahl der kanonischen Linkfunktion (3.7) die Form

$$\mathbf{s}(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{x}_i (\bar{y}_i - \pi_i(\boldsymbol{\beta})).$$

hat. Die Schätz-Gleichung für $\hat{\boldsymbol{\beta}}$ ist gegeben durch

$$\sum_{i=1}^n \mathbf{x}_i \bar{y}_i = \sum_{i=1}^n \mathbf{x}_i \pi_i(\boldsymbol{\beta}).$$

Analog zur Score-Funktion vereinfacht sich mit der Wahl der kanonischen Linkfunktion die Fisher-Matrix zu

$$\mathbf{F}(\boldsymbol{\beta}) = \mathbf{X}^\top \mathbf{W}(\boldsymbol{\beta}) \mathbf{X}$$

mit

$$\begin{aligned} \mathbf{W}(\boldsymbol{\beta}) &= \text{diag}(\text{Var}(\bar{Y}_1), \dots, \text{Var}(\bar{Y}_n)) \\ &= \text{diag}\left(\frac{1}{m_1} \pi_1(\boldsymbol{\beta})(1 - \pi_1(\boldsymbol{\beta})), \dots, \frac{1}{m_n} \pi_n(\boldsymbol{\beta})(1 - \pi_n(\boldsymbol{\beta}))\right). \end{aligned}$$

3.3.2 Poissonverteilung

Wie in (2.20) gezeigt ist die Wahrscheinlichkeitsfunktion der Poissonverteilung ein Mitglied der einparametrischen, linearen Exponentialfamilie mit $\theta = \log(\lambda)$. Für den Erwartungswert und die Varianz der Poissonverteilung gilt $\mu_i = \text{E}(Y_i) = \text{Var}(Y_i) = \lambda_i$. Die kanonische Linkfunktion dieser Verteilung ist gegeben durch

$$\theta_i = \mathbf{x}_i^\top \boldsymbol{\beta} = g(\lambda_i) = \log(\lambda_i). \quad (3.8)$$

Die inverse Linkfunktion ergibt die Responsefunktion mit

$$\lambda_i = g^{-1}(\theta_i) = h(\theta_i) = \exp(\theta_i).$$

Die Log-Likelihood-Funktion der Poissonverteilung hat die folgende Gestalt

$$l(\boldsymbol{\beta}|\mathbf{y}) = \sum_{i=1}^n (y_i \log(\lambda_i(\boldsymbol{\beta})) - \lambda_i(\boldsymbol{\beta}) - \log(y_i!)).$$

Mit der Wahl der kanonischen Linkfunktion (3.8) ergeben sich die folgenden Ergebnisse

$$\begin{aligned} l(\boldsymbol{\beta}|\mathbf{y}) &= \sum_{i=1}^n (y_i \mathbf{x}_i^\top \boldsymbol{\beta} - \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) - \log(y_i!)), \\ \mathbf{s}(\boldsymbol{\beta}) &= \sum_{i=1}^n \mathbf{x}_i (y_i - \exp(\mathbf{x}_i^\top \boldsymbol{\beta})), \\ \mathbf{F}(\boldsymbol{\beta}) &= \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top \exp(\mathbf{x}_i^\top \boldsymbol{\beta}). \end{aligned}$$

3.3.3 Multinomialverteilung

Auf Grund der Gleichungen (2.21) ist die Multinomialverteilung ein Mitglied der k -parametrischen Exponentialfamilie mit $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)^\top = (\log(\pi_1), \dots, \log(\pi_k))^\top$. Da die Multinomialverteilung kein Mitglied der einparametrischen Exponentialfamilie ist, kann keine kanonische Linkfunktion entsprechend den Linkfunktionen der Binomial- oder Poissonverteilung angegeben werden. Im Gegensatz zu den Mitgliedern der einparametrischen, linearen Exponentialfamilie ist der Prädiktor $\boldsymbol{\eta}_i$ ein Vektor mit k Einträgen $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{ik})^\top$. Daher werden in der systematischen Komponente diese Prädiktoren in der folgenden Matrixschreibweise zusammengefasst

$$\begin{pmatrix} \eta_{i1} \\ \vdots \\ \eta_{ik} \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{x}_i^\top & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{x}_i^\top & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \mathbf{x}_i^\top & 0 \\ 0 & \cdots & \cdots & 0 & \mathbf{x}_i^\top \end{pmatrix}}_{=: \mathbf{X}_i} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_k \end{pmatrix}. \quad (3.9)$$

Auf Grund des vektorwertigen Prädiktors gilt für die Linkfunktion $g: \mathbb{R}^k \rightarrow \mathbb{R}^k$. Mit der Abbildung $\mathbf{g}(m_i \boldsymbol{\pi}_i) = \boldsymbol{\eta}_i$ folgt aus dem Gleichungssystem (3.9) das allgemeine Gleichungssystem für das GLM der Multinomialverteilung

$$\begin{pmatrix} g(m_i \pi_{i1}) \\ \vdots \\ g(m_i \pi_{ik}) \end{pmatrix} = \begin{pmatrix} \mathbf{x}_i^\top & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{x}_i^\top & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \mathbf{x}_i^\top & 0 \\ 0 & \cdots & \cdots & 0 & \mathbf{x}_i^\top \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_k \end{pmatrix}, \quad i = 1, \dots, n. \quad (3.10)$$

Eine sehr oft verwendete Linkfunktion ist die sogenannte Logit-Funktion. Hierbei wird eine Referenzkategorie gewählt und das Verhältnis der Wahrscheinlichkeit π_i zur Wahrscheinlichkeit dieser Kategorie betrachtet. Ohne Beschränkung der Allgemeinheit wird die Kategorie k als Referenz gewählt, wodurch die Linkfunktion die folgende Gestalt hat

$$g_j(\pi_{i1}, \dots, \pi_{ik}) = \log\left(\frac{\pi_{ij}}{\pi_{ik}}\right), \quad j = 1, \dots, k-1.$$

Die Responsefunktion hat die Darstellung

$$h_j(\eta_{i1}, \dots, \eta_{ik}) = \frac{\exp(\eta_{ij})}{1 + \sum_{s=1}^{k-1} \exp(\eta_{is})}, \quad j = 1, \dots, k-1.$$

Als Likelihood-Funktion ergibt sich die Form

$$L(\boldsymbol{\beta}|\mathbf{y}) = \prod_{i=1}^n h(\mathbf{y}_i) \pi_{i1}^{y_{i1}} \cdots \pi_{ik-1}^{y_{ik-1}} \left(1 - \sum_{j=1}^{k-1} \pi_{ij}\right)^{m_i - \sum_{j=1}^{k-1} y_{ij}}. \quad (3.11)$$

Somit resultiert als Log-Likelihood $l(\boldsymbol{\beta}) = \sum_{i=1}^n l_i(\boldsymbol{\pi}_i|\mathbf{y}_i)$ mit

$$l_i(\boldsymbol{\pi}_i|\mathbf{y}_i) = m_i \left\{ \sum_{j=1}^{k-1} \frac{y_{ij}}{m_i} \log\left(\frac{\pi_{ij}}{1 - \sum_{s=1}^{k-1} \pi_{is}}\right) + \log\left(1 - \sum_{s=1}^{k-1} \pi_{is}\right) + \log(h(\mathbf{y}_i)) \right\}. \quad (3.12)$$

Weiters ergibt sich die Score-Funktion mit

$$s(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta}|\mathbf{y})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \mathbf{X}_i^T \mathbf{D}_i(\boldsymbol{\beta}) \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}) \left(\frac{\mathbf{y}_i}{m_i} - \boldsymbol{\pi}_i\right)$$

mit

$$\mathbf{D}_i(\boldsymbol{\beta}) = \begin{pmatrix} \frac{\partial g_1(\boldsymbol{\pi}_i)}{\partial \pi_1} & \cdots & \frac{\partial g_1(\boldsymbol{\pi}_i)}{\partial \pi_{k-1}} \\ & \ddots & \\ \frac{\partial g_{k-1}(\boldsymbol{\pi}_i)}{\partial \pi_1} & \cdots & \frac{\partial g_{k-1}(\boldsymbol{\pi}_i)}{\partial \pi_{k-1}} \end{pmatrix}^{-1}$$

und der Kovarianz $\boldsymbol{\Sigma}_i(\boldsymbol{\beta})$

$$\begin{aligned} \boldsymbol{\Sigma}_i(\boldsymbol{\beta}) &= \frac{1}{m_i} \begin{pmatrix} \pi_{i1}(1 - \pi_{i1}) & -\pi_{i1}\pi_{i2} & \cdots & -\pi_{i1}\pi_{ik-1} \\ -\pi_{i1}\pi_{i2} & \pi_{i2}(1 - \pi_{i2}) & & \\ & & \ddots & \\ -\pi_{ik-1}\pi_{i1} & & & \pi_{ik-1}(1 - \pi_{ik-1}) \end{pmatrix} \\ &= [\text{Diag}(\boldsymbol{\pi}_i) - \boldsymbol{\pi}_i \boldsymbol{\pi}_i^T] / m_i. \end{aligned}$$

Die Fisher Matrix kann dargestellt werden durch

$$\mathbf{F}(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{X}_i^T \mathbf{W}_i(\boldsymbol{\beta}) \mathbf{X}_i. \quad (3.13)$$

3.4 Deviance

Nach der Erstellung eines GLMs wird versucht, die Güte der Modellanpassung zu prüfen. Die Deviance stellt eine Möglichkeit der Überprüfung dar und bewertet diese Güte mit Hilfe des Likelihood Quotienten. Dieser stellt das Verhältnis des Supremums der Likelihoodfunktion des erstellten Modells zum Supremum der Likelihoodfunktion des uneingeschränkten Modells dar. Aufgrund der Abhängigkeit $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\beta})$ kann die Likelihoodfunktion $L(\boldsymbol{\beta}|\mathbf{y})$ geschrieben werden als $L(\boldsymbol{\mu}|\mathbf{y})$. Dadurch ergibt sich

$$\Lambda(\mathbf{y}) = \frac{\sup_{g^{-1}(\boldsymbol{\eta})=\boldsymbol{\mu}} L(\boldsymbol{\mu}|\mathbf{y})}{\sup_{\boldsymbol{\mu}} L(\boldsymbol{\mu}|\mathbf{y})}.$$

Der Schätzer $\hat{\boldsymbol{\mu}}$ maximiert dabei das betrachtete Modell. Das uneingeschränkte Modell (*saturierte Modell*) enthält einen freien Parameter μ_i für jede Beobachtung y_i und liefert dadurch eine perfekte Anpassung. Der Maximum Likelihood Schätzer unter dem saturierten Modells ergibt $\hat{\mu}_i = y_i$ (Beweis im Anhang A.3) und somit folgt

$$\Lambda(\mathbf{y}) = \frac{L(\hat{\boldsymbol{\mu}}|\mathbf{y})}{L(\mathbf{y}|\mathbf{y})}.$$

Durch Logarithmieren erhält man die Deviance $D(\mathbf{y}, \hat{\boldsymbol{\mu}})$ mit

$$D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = -2(l(\hat{\boldsymbol{\mu}}|\mathbf{y}) - l(\mathbf{y}|\mathbf{y})).$$

Die Deviance des saturierten Modells ist somit Null. Im Allgemeinen gilt, dass $D(\mathbf{y}, \hat{\boldsymbol{\mu}})$ immer positiv ist, da $l(\mathbf{y}|\mathbf{y}) \geq l(\hat{\boldsymbol{\mu}}|\mathbf{y})$. Eine größere Deviance bedeutet somit eine größere Abweichung des erstellten Modells vom saturierten Modell. Dadurch wird versucht, das Minimum der Deviance zu erhalten, welches der Maximierung des Log-Likelihood des erstellten Modells entspricht. Die Deviance für die standardisierte Binomialverteilung, die Poissonverteilung und die Multinomialverteilung ist im Anhang A.4 angeführt.

4 Modellierung ordinaler Response

Ordinale Daten treten in Kategorien auf, welche eine Ordnung aufweisen, jedoch kein metrisches Verhältnis zwischen den einzelnen Kategorien haben. Diese Daten können diskrete Werte einer Skala oder die kategorielle Einteilung einer kontinuierlichen Größe sein. Dabei werden den Kategorien die Werte $1, \dots, k$ der Ordnung entsprechend zugeordnet. Y bezeichnet die ordinale Response mit $Y \in \{1, \dots, k\}$. Die Modellierung der ordinalen Response Y wird mit Hilfe des Vektors der erklärenden Variablen $\mathbf{x} = (1, x_1, \dots, x_{p-1})^\top$ und dem Vektor der Effekte $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{p-1})^\top$ durchgeführt. Mit Hilfe von vorliegenden erklärenden Variablen kann eine Wahrscheinlichkeit angegeben werden, mit der die Response einer Kategorie zugeordnet werden kann, welche durch

$$P(Y = j|\mathbf{x}) = \pi_j, \quad j = 1, \dots, k$$

dargestellt ist. In dieser Arbeit werden zwei verschiedene Modellklassen vorgestellt, zum einen die kumulativen Modelle, welche für ordinale Daten sehr verbreitet sind und zum anderen sequentielle Modelle, welche sich bei Response Variablen mit sukzessiv steigender Ordnung am besten eignen.

4.1 Kumulative Modelle

Eine sehr verbreitete Variante zur Modellierung ordinaler Daten ist das kumulative Modell. Die Idee der kumulativen Modelle ist die Betrachtung von kumulativen Wahrscheinlichkeiten. Für die Response Y gilt aufgrund der Ordnung der Daten

$$P(Y \leq j|\mathbf{x}) = P(Y = 1|\mathbf{x}) + \dots + P(Y = j|\mathbf{x}) = \pi_1 + \dots + \pi_j, \quad j = 1, \dots, k.$$

Daraus folgt die Abschätzung

$$P(Y \leq j_1|\mathbf{x}) \leq P(Y \leq j_2|\mathbf{x}), \quad j_1, j_2 \in \{1, \dots, k\}, \quad j_1 < j_2.$$

Werden die Kategorien $\{1, \dots, k\}$ auf die zwei Mengen $\{1, \dots, j\}$ und $\{j + 1, \dots, k\}$ für $j \in \{1, \dots, k - 1\}$ aufgeteilt, so reduziert sich das Modell auf ein binäres Problem.

Aus diesem Grund kann eine binäre Zufallsvariable Y_j definiert werden mit

$$Y_j := \begin{cases} 1 & Y \leq j \\ 0 & Y > j, \end{cases} \quad (4.1)$$

für welche der Zusammenhang

$$P(Y_j = 1|\mathbf{x}) = P(Y \leq j|\mathbf{x})$$

gilt. Auf Grund der binären Form von Y_j kann ein Modell mit Hilfe der erklärenden Variable \mathbf{x} angegeben werden, welches die Wahrscheinlichkeit $Y_j = 1$ modelliert

$$P(Y_j = 1|\mathbf{x}) = P(Y \leq j|\mathbf{x}) = F(\mathbf{x}^\top \boldsymbol{\beta}_j), \quad j = 1, \dots, k-1.$$

Die Funktion $F(\cdot)$ ist eine Verteilungsfunktion, der Vektor $\boldsymbol{\beta}_j$ bezeichnet den Vektor der Effekte des Modells in Abhängigkeit der gewählten Stelle der Teilung bei j . Abhängig von den Kategorien $\{1, \dots, k\}$ werden zwei Fälle unterschieden

- i) $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \dots = \boldsymbol{\beta}_{k-1}$
- ii) $\exists j_1, j_2 \in \{1, \dots, k-1\}, j_1 \neq j_2$ mit $\boldsymbol{\beta}_{j_1} \neq \boldsymbol{\beta}_{j_2}$.

Der erste Fall wird im *einfachen kumulativen Modell* behandelt, der zweite Fall wird mit Hilfe des *generellen kumulativen Modells* beschrieben.

4.1.1 Einfaches kumulatives Modell

Das einfache kumulative Modell beruht auf der Annahme, dass der diskreten Response Y eine kontinuierliche, eventuell nicht beobachtbare (*latente*) Zufallsvariable Y' mit der Darstellung

$$Y' = -\mathbf{x}^\top \boldsymbol{\beta} + \varepsilon, \quad (4.2)$$

zugrunde liegt, wobei der Term ε einen Zufallsfehler mit Verteilungsfunktion $F(\cdot)$ darstellt. Die Einteilung der Zufallsvariable Y' in die Kategorien erfolgt in der Form

$$Y = j \iff b_{j-1} \leq Y' \leq b_j, \quad j = 1, \dots, k, \quad (4.3)$$

wobei $-\infty = b_0 < b_1 < \dots < b_k = \infty$ latente Werte, sogenannte Schwellenwerte sind. Mit der Annahme i) gilt die Darstellung von Y' für alle Kategorien $1, \dots, k$. Wird die kumulative Wahrscheinlichkeit $P(Y \leq j|\mathbf{x})$ betrachtet, so gilt

$$P(Y \leq j|\mathbf{x}) = P(Y = 1|\mathbf{x}) + \dots + P(Y = j|\mathbf{x})$$

$$\begin{aligned}
&= P(Y' \leq b_1|\mathbf{x}) + (P(Y' \leq b_2|\mathbf{x}) - P(Y' \leq b_1|\mathbf{x})) \\
&\quad + \cdots + (P(Y' \leq b_j|\mathbf{x}) - P(Y' \leq b_{j-1}|\mathbf{x})) \\
&= P(Y' \leq b_j|\mathbf{x}).
\end{aligned}$$

Daher folgt mit der Unterteilung (4.3)

$$\begin{aligned}
P(Y \leq j|\mathbf{x}) &= P(Y' \leq b_j) \\
&= P(-\mathbf{x}^\top \boldsymbol{\beta} + \varepsilon \leq b_j) \\
&= P(\varepsilon \leq b_j + \mathbf{x}^\top \boldsymbol{\beta}) \\
&= F(b_j + \mathbf{x}^\top \boldsymbol{\beta}), \quad j = 1, \dots, k-1.
\end{aligned} \tag{4.4}$$

Aus dieser Gleichung folgt die Darstellung für $P(Y = j|\mathbf{x})$

$$\begin{aligned}
P(Y = j|\mathbf{x}) &= P(Y \leq j|\mathbf{x}) - P(Y \leq j-1|\mathbf{x}) \\
&= F(b_j + \mathbf{x}^\top \boldsymbol{\beta}) - F(b_{j-1} + \mathbf{x}^\top \boldsymbol{\beta}), \quad j = 1, \dots, k-1.
\end{aligned} \tag{4.5}$$

Die Wahrscheinlichkeit $P(Y \leq j|\mathbf{x})$ ist vom Vektor der erklärenden Variablen \mathbf{x} abhängig. Um einen Vergleich der Wahrscheinlichkeiten für zwei verschiedene Vektoren \mathbf{x} und $\tilde{\mathbf{x}}$ erklärender Variablen zu erhalten, müssen transformierte Wahrscheinlichkeiten betrachtet werden. Für diese gilt die Gleichung

$$F^{-1}(P(Y \leq j|\mathbf{x})) - F^{-1}(P(Y \leq j|\tilde{\mathbf{x}})) = (\mathbf{x} - \tilde{\mathbf{x}})^\top \boldsymbol{\beta}. \tag{4.6}$$

Die Differenz $(\mathbf{x} - \tilde{\mathbf{x}})^\top \boldsymbol{\beta}$ ist unabhängig von j und daher für alle Kategorien gültig.

4.1.2 Generelles kumulatives Modell

Dieses Modell behandelt den Fall ii), bei welchem der Vektor $\boldsymbol{\beta}$ von der Stelle j der Teilung abhängig ist. Die Gleichung (4.4) hat nun auf Grund der unterschiedlichen Effekte die Darstellung

$$P(Y \leq j|\mathbf{x}) = F(b_j + \mathbf{x}^\top \boldsymbol{\beta}_j), \quad j = 1, \dots, k-1. \tag{4.7}$$

Dabei können sich in der allgemeinsten Form des generellen kumulativen Modells bis zu $k-1$ verschiedene Parametervektoren $\boldsymbol{\beta}_j$ ergeben. Da die Wahrscheinlichkeiten $P(Y \leq j|\mathbf{x})$ kumulativ sind, gilt die Abschätzung

$$P(Y \leq j-1|\mathbf{x}) \leq P(Y \leq j|\mathbf{x})$$

und mit der Gleichung (4.7) und der Monotonie der Verteilungsfunktion folgt

$$b_{j-1} + \mathbf{x}^\top \boldsymbol{\beta}_{j-1} \leq b_j + \mathbf{x}^\top \boldsymbol{\beta}_j, \quad j = 2, \dots, k-1.$$

Werden das einfache und das generelle kumulative Modell miteinander verglichen, so besteht der einzige Unterschied in der Abhängigkeit der Effekte $\boldsymbol{\beta}$ von den Kategorien. Im einfachen kumulativen Modell hat der lineare Prädiktor η_j die Darstellung

$$\eta_j = b_j + \mathbf{x}^\top \boldsymbol{\beta}, \quad j = 1, \dots, k-1,$$

während im generellen kumulativen Modell

$$\eta_j = b_j + \mathbf{x}^\top \boldsymbol{\beta}_j, \quad j = 1, \dots, k-1$$

gilt. Durch die Kombination dieser beiden Formen ergibt sich eine weitere Darstellung des linearen Prädiktors η_j mit

$$\eta_j = b_j + \mathbf{x}^\top \boldsymbol{\beta}_{j'} + \mathbf{x}^\top \bar{\boldsymbol{\beta}}_j. \quad (4.8)$$

Die Kategorie j' wird als Referenz gewählt und der Parametervektor $\bar{\boldsymbol{\beta}}_j = \boldsymbol{\beta}_j - \boldsymbol{\beta}_{j'}$, $j = 1, \dots, k-1$, stellt die Differenz zur Referenzkategorie dar. Die Darstellung (4.8) ist nützlich, wenn sich nicht alle Vektoren $\boldsymbol{\beta}_j$ für $j = 1, \dots, k-1$ unterscheiden.

4.1.3 Testen der Effekte

Im einfachen kumulativen Modell wird für die Effekte angenommen, dass

$$\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \dots = \boldsymbol{\beta}_{k-1}.$$

Diese Annahme muss mit Hilfe eines Tests überprüft werden. Die Nullhypothese zu dieser Annahme lautet

$$H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \dots = \boldsymbol{\beta}_{k-1}. \quad (4.9)$$

Da für einzelne erklärende Variablen der Effekt unabhängig von der Kategorie sein kann, kann die Nullhypothese (4.9) komponentenweise betrachtet werden mit

$$H_0 : \beta_{1l} = \dots = \beta_{k-1l}, \quad l = 0, \dots, p-1. \quad (4.10)$$

Um diese Hypothesen zu testen, kann der Likelihood-Quotienten-Test oder auch der Wald-Test verwendet werden.

4.2 Linkfunktionen des kumulativen Modells

In den Abschnitten des einfachen und generellen kumulativen Modells 4.1.1 und 4.1.2 wurde eine allgemeine Verteilungsfunktion F als Linkfunktion betrachtet, welche auf Grund der Annahme (4.2) aus der Verteilung des Zufallsfehler ε resultiert. Im Folgenden werden einige Verteilungsfunktionen explizit angegeben. Die Definitionen der Verteilungsfunktionen sowie deren Momente befinden sich im Anhang A.5.

4.2.1 Kumulatives Logit Modell

Das weit verbreitetste Modell ist das *kumulative Logit Modell*. Dieses beruht auf der Annahme, dass F eine logistische Verteilungsfunktion ist.

Definition 4.1. (*Logit Funktion und logistische Verteilungsfunktion*) Sei $\pi \in (0, 1)$, dann ist der Logit von π definiert durch

$$\text{logit}(\pi) = \frac{\pi}{1 - \pi}.$$

Werden Wahrscheinlichkeiten betrachtet, so kann der Logit als der Logarithmus von dem Verhältnis der Wahrscheinlichkeit zur Gegenwahrscheinlichkeit, den sogenannten Odds, betrachtet werden. Die Umkehrfunktion des Logit hat die Gestalt

$$\text{logit}^{-1}(\pi) = \frac{1}{1 + \exp(-\pi)} = \frac{\exp(\pi)}{1 + \exp(\pi)}$$

und entspricht der logistischen Verteilungsfunktion

$$F(x) = P(Y \leq x) = \frac{1}{1 + \exp(-x)} = \frac{\exp(x)}{1 + \exp(x)}. \quad (4.11)$$

Weiters kann nun der *kumulative Logit* definiert werden.

Definition 4.2. (*kumulativer Logit*) Sei $Y \in \{1, \dots, k\}$ eine ordinale Response mit $P(Y = j|\mathbf{x}) = \pi_j$. Dann ist der kumulative Logit definiert als

$$\begin{aligned} \text{logit}(P(Y \leq j|\mathbf{x})) &= \log\left(\frac{P(Y \leq j|\mathbf{x})}{1 - P(Y \leq j|\mathbf{x})}\right) \\ &= \log\left(\frac{P(Y \leq j|\mathbf{x})}{P(Y > j|\mathbf{x})}\right) \\ &= \log\left(\frac{\pi_1 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_k}\right) \quad j = 1, \dots, k - 1. \end{aligned}$$

Einfaches kumulatives Logit Modell

Wird das einfache kumulative Modell mit der speziellen Wahl der logistischen Verteilungsfunktion (4.11) betrachtet, so ergibt sich mit (4.4) die Darstellung

$$\text{logit}(P(Y \leq j|\mathbf{x})) = \log\left(\frac{P(Y \leq j|\mathbf{x})}{P(Y > j|\mathbf{x})}\right) = b_j + \mathbf{x}^\top \boldsymbol{\beta}. \quad (4.12)$$

Für jedes $j = 1, \dots, k-1$ existiert in diesem Modell ein eigener Intercept b_j , der Vektor der Effekte $\boldsymbol{\beta}$ ändert sich jedoch nicht. Durch die Umkehrfunktion kann die Gleichung (4.12) umgeformt werden zu

$$P(Y \leq j|\mathbf{x}) = \frac{\exp(b_j + \mathbf{x}^\top \boldsymbol{\beta})}{1 + \exp(b_j + \mathbf{x}^\top \boldsymbol{\beta})}, \quad j = 1, \dots, k-1.$$

Auf Grund der Abschätzung der kumulativen Wahrscheinlichkeiten

$$P(Y \leq j-1|\mathbf{x}) \leq P(Y \leq j|\mathbf{x})$$

und der strikten Monotonie der Logit Funktion gilt für den Intercept $b_1 < \dots < b_{k-1}$.

Für die Wahrscheinlichkeit $P(Y = j|\mathbf{x}) = P(Y \leq j|\mathbf{x}) - P(Y \leq j-1|\mathbf{x})$ ergibt sich

$$P(Y = j|\mathbf{x}) = \begin{cases} \frac{\exp(b_j + \mathbf{x}^\top \boldsymbol{\beta})}{1 + \exp(b_j + \mathbf{x}^\top \boldsymbol{\beta})}, & j = 1 \\ \frac{\exp(b_j + \mathbf{x}^\top \boldsymbol{\beta})}{1 + \exp(b_j + \mathbf{x}^\top \boldsymbol{\beta})} - \frac{\exp(b_{j-1} + \mathbf{x}^\top \boldsymbol{\beta})}{1 + \exp(b_{j-1} + \mathbf{x}^\top \boldsymbol{\beta})}, & j = 2, \dots, k-1 \\ 1 - \frac{\exp(b_{j-1} + \mathbf{x}^\top \boldsymbol{\beta})}{1 + \exp(b_{j-1} + \mathbf{x}^\top \boldsymbol{\beta})}, & j = k. \end{cases} \quad (4.13)$$

Wird die Differenz von zwei Vektoren erklärender Variablen \mathbf{x} und $\tilde{\mathbf{x}}$ betrachtet, so ergibt sich die Gleichung (4.6) für das kumulative Logit Modell zu

$$\text{logit}(P(Y \leq j|\mathbf{x})) - \text{logit}(P(Y \leq j|\tilde{\mathbf{x}})) = (b_j + \mathbf{x}^\top \boldsymbol{\beta}) - (b_j + \tilde{\mathbf{x}}^\top \boldsymbol{\beta}) = (\mathbf{x} - \tilde{\mathbf{x}})^\top \boldsymbol{\beta}.$$

Für die Differenz der zwei Logits gilt die folgende Äquivalenz

$$\text{logit}(P(Y \leq j|\mathbf{x})) - \text{logit}(P(Y \leq j|\tilde{\mathbf{x}})) = \log\left(\frac{P(Y \leq j|\mathbf{x})/[1 - P(Y \leq j|\mathbf{x})]}{P(Y \leq j|\tilde{\mathbf{x}})/[1 - P(Y \leq j|\tilde{\mathbf{x}})]}\right).$$

Für den Quotienten der beiden Odds ergibt sich daher

$$\frac{P(Y \leq j|\mathbf{x})/[1 - P(Y \leq j|\mathbf{x})]}{P(Y \leq j|\tilde{\mathbf{x}})/[1 - P(Y \leq j|\tilde{\mathbf{x}})]} = \frac{P(Y \leq j|\mathbf{x})/P(Y > j|\mathbf{x})}{P(Y \leq j|\tilde{\mathbf{x}})/P(Y > j|\tilde{\mathbf{x}})} = \exp((\mathbf{x} - \tilde{\mathbf{x}})^\top \boldsymbol{\beta}).$$

Dieses Modell beschreibt das proportionale Verhältnis der Odds für zwei verschiedene erklärende Vektoren. Für das Logit Modell mit $Y \leq j$ gilt, dass die Chance des Vektors \mathbf{x} das $\exp((\mathbf{x} - \tilde{\mathbf{x}})^\top \boldsymbol{\beta})$ -fache der Chance des Vektors $\tilde{\mathbf{x}}$ beträgt. Dieses Modell ist weiters von j unabhängig, wodurch die Interpretation nicht kategoriespezifisch ist.

Generelles kumulatives Logit Modell

Für das generelle kumulative Modell hat das Logit Modell die Gestalt

$$\begin{aligned} \text{logit}(P(Y \leq j|\mathbf{x})) &= \log\left(\frac{P(Y \leq j|\mathbf{x})}{1 - P(Y \leq j|\mathbf{x})}\right) \\ &= \log\left(\frac{P(Y \leq j|\mathbf{x})}{P(Y > j|\mathbf{x})}\right) \\ &= b_j + \mathbf{x}^\top \boldsymbol{\beta}_j. \end{aligned} \tag{4.14}$$

Mittels Umformungen ergibt sich

$$P(Y \leq j|\mathbf{x}) = \frac{\exp(b_j + \mathbf{x}^\top \boldsymbol{\beta}_j)}{1 + \exp(b_j + \mathbf{x}^\top \boldsymbol{\beta}_j)}.$$

Für gemischte Modelle werden die Vektoren \mathbf{x} und $\boldsymbol{\beta}$ in $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)^\top$ und $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)^\top$ partitioniert, wobei $\boldsymbol{\beta}_1$ die Effekte aus dem Fall i) mit den dazugehörigen erklärenden Variablen \mathbf{x}_1 und $\boldsymbol{\beta}_2$ jene Effekte aus dem Fall ii) mit den Variablen \mathbf{x}_2 enthält. Das kumulative Modell (4.14) wird mit dieser Partitionierung partielles kumulatives Logit Modell genannt und hat die Form

$$\log\left(\frac{P(Y \leq j|\mathbf{x})}{P(Y > j|\mathbf{x})}\right) = b_j + \mathbf{x}_1^\top \boldsymbol{\beta}_1 + \mathbf{x}_2^\top \boldsymbol{\beta}_2, \tag{4.15}$$

wobei die Unabhängigkeit von j nur für die Variable \mathbf{x} gilt. Werden nun zwei Populationen $(\mathbf{x}_1, \mathbf{x}_2)$ und $(\tilde{\mathbf{x}}_1, \mathbf{x}_2)$ betrachtet, welche sich nur in der Variable \mathbf{x}_1 und $\tilde{\mathbf{x}}_1$ unterscheiden, ergibt sich das proportionale Verhältnis der Odds

$$\begin{aligned} \frac{P(Y \leq j|\mathbf{x}_1, \mathbf{x}_2)/[1 - P(Y \leq j|\mathbf{x}_1, \mathbf{x}_2)]}{P(Y \leq j|\tilde{\mathbf{x}}_1, \mathbf{x}_2)/[1 - P(Y \leq j|\tilde{\mathbf{x}}_1, \mathbf{x}_2)]} &= \frac{P(Y \leq j|\mathbf{x}_1, \mathbf{x}_2)/P(Y > j|\mathbf{x}_1, \mathbf{x}_2)}{P(Y \leq j|\tilde{\mathbf{x}}_1, \mathbf{x}_2)/P(Y > j|\tilde{\mathbf{x}}_1, \mathbf{x}_2)} \\ &= \exp((\mathbf{x}_1 - \tilde{\mathbf{x}}_1)^\top \boldsymbol{\beta}_1). \end{aligned}$$

Dieses Verhältnis ist für alle $j = 1, \dots, k - 1$ gültig.

4.2.2 Probit Modell

Das Probit Modell basiert auf der Annahme, dass F die Verteilungsfunktion der Standardnormalverteilung $N(0, 1)$ ist. Das kumulative Modell hat dadurch die Gestalt

$$P(Y \leq j|\mathbf{x}) = \Phi(b_j + \mathbf{x}^\top \boldsymbol{\beta})$$

mit Φ als Verteilungsfunktion der Standardnormalverteilung. Das kumulative Probit Modell ergibt ähnliche Ergebnisse wie das kumulative Logit Modell. Feine Unterschiede der beiden Linkfunktionen gibt es, da die Standardabweichung der Standardnormalverteilung geringer ist als die der Standard-logistischen Verteilung, was in der Abbildung 4.1 verdeutlicht wird.

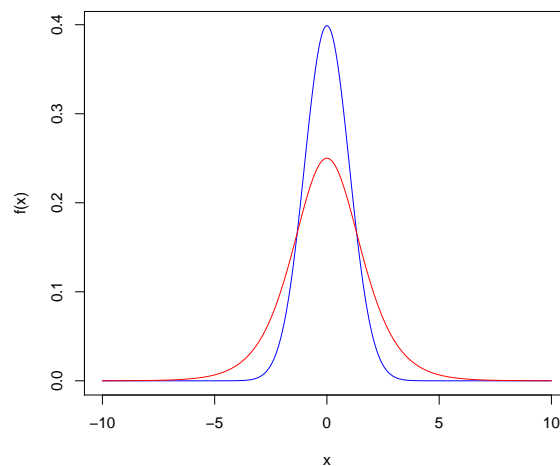


Abbildung 4.1: Vergleich der Dichtefunktionen der Standardnormalverteilung (blau) mit $Var(x) = 1$ und der Standard logistischen Verteilung (rot) mit $Var(x) = \pi^2/3$

4.2.3 Kumulatives Extremwert Modell

Wird als Verteilungsfunktion F die Minimum-Extremwertverteilung verwendet, alternativ auch *Gompertz-Verteilung* genannt, so ergibt sich das sogenannte *kumulative Minimum-Extremwert-Modell* in der Form

$$P(Y \leq j|\mathbf{x}) = 1 - \exp(-\exp(b_j + \mathbf{x}^\top \boldsymbol{\beta})).$$

Umformungen ergeben somit

$$\log(-\log(P(Y > j|\mathbf{x}))) = b_j + \mathbf{x}^\top \boldsymbol{\beta}.$$

Wird der Quotient der Wahrscheinlichkeiten in Abhängigkeit von zwei Populationen \mathbf{x} und $\tilde{\mathbf{x}}$ betrachtet, so gilt

$$\begin{aligned} \frac{\log(1 - P(Y \leq j|\mathbf{x}))}{\log(1 - P(Y \leq j|\tilde{\mathbf{x}}))} &= \frac{\log(P(Y > j|\mathbf{x}))}{\log(P(Y > j|\tilde{\mathbf{x}}))} \\ &= \frac{\exp(b_j + \mathbf{x}^\top \boldsymbol{\beta})}{\exp(b_j + \tilde{\mathbf{x}}^\top \boldsymbol{\beta})} \\ &= \exp((\mathbf{x} - \tilde{\mathbf{x}})^\top \boldsymbol{\beta}). \end{aligned}$$

Durch Verwendung der *Gumbel*-Verteilungsfunktion $F(x) = \exp(-\exp(x))$ folgt das *kumulative Maximum-Extremwert-Modell* mit

$$P(Y \leq j|\mathbf{x}) = \exp(-\exp(b_j + \mathbf{x}^\top \boldsymbol{\beta}))$$

und weiters

$$\log(-\log(P(Y \leq j|\mathbf{x}))) = b_j + \mathbf{x}^\top \boldsymbol{\beta}.$$

Analog zum kumulativen Minimum-Extremwert-Modell ergibt sich

$$\begin{aligned} \frac{\log(P(Y \leq j|\mathbf{x}))}{\log(P(Y \leq j|\tilde{\mathbf{x}}))} &= \frac{\exp(b_j + \mathbf{x}^\top \boldsymbol{\beta})}{\exp(b_j + \tilde{\mathbf{x}}^\top \boldsymbol{\beta})} \\ &= \exp((\mathbf{x} - \tilde{\mathbf{x}})^\top \boldsymbol{\beta}). \end{aligned}$$

4.3 Kumulatives Modell für GLM

Da ordinale Modelle ein Spezialfall der multinomialen GLM sind, werden diese im Zusammenhang mit GLMs im Folgenden genauer beschrieben. Die Notation wird dabei dem Kapitel 3 folgend verwendet.

Einfaches kumulatives Modell

Das einfache kumulative Modell mit der Basis eines multinomialen generalisierten linearen Modells hat die Form

$$P(Y_i \leq j|\mathbf{x}_i) = F(b_j + \mathbf{x}_i^\top \boldsymbol{\beta}),$$

dabei bezeichnet F die Verteilungsfunktion. Für die Wahrscheinlichkeiten $\pi_{ij} = P(Y_i = j|\mathbf{x}_i)$ ergibt sich damit

$$\pi_{ij} = P(Y_i = j|\mathbf{x}_i) = P(Y_i \leq j|\mathbf{x}_i) - P(Y_i \leq j-1|\mathbf{x}_i) = F(b_j + \mathbf{x}_i^\top \boldsymbol{\beta}) - F(b_{j-1} + \mathbf{x}_i^\top \boldsymbol{\beta})$$

mit $-\infty = b_0 < b_1 < \dots < b_k = \infty$. Wird mit

$$\eta_{ij} = b_j + \mathbf{x}_i^\top \boldsymbol{\beta}$$

der Prädiktor an der Stelle der Teilung j bezeichnet, so gilt daher

$$\pi_{ij} = F(\eta_{ij}) - F(\eta_{ij-1}), \quad j = 1, \dots, k-1.$$

Damit ergibt sich das Gleichungssystem

$$\begin{pmatrix} \pi_{i1} \\ \vdots \\ \pi_{ik-1} \end{pmatrix} = h \left\{ \begin{pmatrix} 1 & & & \mathbf{x}_i^\top \\ & 1 & & \vdots \\ & & \ddots & \vdots \\ & & & \ddots & \vdots \\ & & & & 1 & \mathbf{x}_i^\top \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_{k-1} \\ \boldsymbol{\beta} \end{pmatrix} \right\}. \quad (4.16)$$

Aus diesem Gleichungssystem folgt die Darstellung der Responsefunktion $\mathbf{h} = (h_1, \dots, h_{k-1}) : \mathbb{R}^{k-1} \rightarrow \mathbb{R}^{k-1}$ mit

$$h_j(\eta_{i1}, \dots, \eta_{ik-1}) = F(\eta_{ij}) - F(\eta_{ij-1}).$$

Weiters kann die Linkfunktion $\mathbf{g} = (g_1, \dots, g_{k-1}) : \mathbb{R}^{k-1} \rightarrow \mathbb{R}^{k-1}$ mit Hilfe der Äquivalenzen

$$\begin{aligned} F^{-1}(P(Y_i \leq j|\mathbf{x}_i)) &= b_j + \mathbf{x}_i^\top \boldsymbol{\beta} \\ F^{-1}(\pi_{i1} + \dots + \pi_{ij}|\mathbf{x}_i) &= \eta_{ij} \end{aligned}$$

hergeleitet werden durch

$$g_j(\pi_{i1}, \dots, \pi_{ik-1}) = F^{-1}(\pi_{i1} + \dots + \pi_{ij}).$$

Generelles kumulatives Modell

Das generelle kumulative Modell beinhaltet im Unterschied zum einfachen kumulativen Modell kategorie-spezifische Parameter und kann somit in der folgenden Form angegeben werden

$$P(Y_i \leq j|\mathbf{x}_i) = F(b_j + \mathbf{x}_i^\top \boldsymbol{\beta}_j).$$

Aufgrund der Abhängigkeit der Effekte $\boldsymbol{\beta}$ von der Stelle der Teilung j hat der Prädiktor die Darstellung

$$\eta_{ij} = b_j + \mathbf{x}_i^\top \boldsymbol{\beta}_j = (1, \mathbf{x}_i^\top) \begin{pmatrix} b_j \\ \boldsymbol{\beta}_j \end{pmatrix},$$

daraus resultiert das Gleichungssystem

$$\begin{pmatrix} \pi_{i1} \\ \vdots \\ \pi_{ik-1} \end{pmatrix} = h \left\{ \begin{pmatrix} 1 & & & & & & & & & & \mathbf{x}_i^\top \\ & 1 & & & & & & & & & \mathbf{x}_i^\top \\ & & 1 & & & & & & & & \mathbf{x}_i^\top \\ & & & \ddots & & & & & & & \mathbf{x}_i^\top \\ & & & & 1 & & & & & & \mathbf{x}_i^\top \\ & & & & & \ddots & & & & & \mathbf{x}_i^\top \\ & & & & & & 1 & & & & \mathbf{x}_i^\top \end{pmatrix} \begin{pmatrix} b_0 \\ \vdots \\ b_{k-1} \\ \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_{k-1} \end{pmatrix} \right\}. \quad (4.17)$$

4.3.1 Maximum Likelihood Schätzung für kumulative Modelle

Mit Hilfe der Maximum Likelihood Methode können Schätzer für b und $\boldsymbol{\beta}$ berechnet werden. Wie zuvor erklärt hat die Wahrscheinlichkeit π_{ij} die Gestalt

$$\pi_{ij} = P(Y_i = j | \mathbf{x}_i) = F(b_j + \mathbf{x}_i^\top \boldsymbol{\beta}) - F(b_{j-1} + \mathbf{x}_i^\top \boldsymbol{\beta}), \quad j = 1, \dots, k.$$

Die Formulierung einer Indikatorvariablen, welche die Response der Kategorien beschreibt, kann angegeben werden durch

$$\alpha_{ij} = \begin{cases} 1 & Y_i = j \\ 0 & \text{sonst.} \end{cases} \quad (4.18)$$

Mit Hilfe der Indikatorvariable α_{ij} ergibt sich für die Wahrscheinlichkeit π_{ij} die Produktschreibweise

$$\pi_{ij} = P(Y_i = j | \mathbf{x}_i) = \prod_{j=1}^k P(Y_i = j | \mathbf{x}_i)^{\alpha_{ij}} = \prod_{j=1}^k [F(b_j + \mathbf{x}_i^\top \boldsymbol{\beta}) - F(b_{j-1} + \mathbf{x}_i^\top \boldsymbol{\beta})]^{\alpha_{ij}}.$$

Da die Likelihood Funktion die gemeinsame Wahrscheinlichkeitsfunktion der Realisierung \mathbf{y} ist, kann diese dargestellt werden als

$$\begin{aligned} L(\mathbf{b}, \boldsymbol{\beta} | \mathbf{y}) &= \prod_{i=1}^n \left\{ \prod_{j=1}^k P(Y_i = j | \mathbf{x}_i)^{\alpha_{ij}} \right\} \\ &= \prod_{i=1}^n \left\{ \prod_{j=1}^k [F(b_j + \mathbf{x}_i^\top \boldsymbol{\beta}) - F(b_{j-1} + \mathbf{x}_i^\top \boldsymbol{\beta})]^{\alpha_{ij}} \right\}. \end{aligned}$$

Die Maximum Likelihood Schätzer $\hat{\mathbf{b}}$ und $\hat{\boldsymbol{\beta}}$ werden durch die partiellen Ableitungen der Log-Likelihood Funktion berechnet. Diese haben in diesem Fall die Gestalt

$$l(\mathbf{b}, \boldsymbol{\beta} | \mathbf{y}) = \log(L(\mathbf{b}, \boldsymbol{\beta} | \mathbf{y})) = \sum_{i=1}^n \left\{ \sum_{j=1}^k \alpha_{ij} \log [F(b_j + \mathbf{x}_i^\top \boldsymbol{\beta}) - F(b_{j-1} + \mathbf{x}_i^\top \boldsymbol{\beta})] \right\}.$$

Um die Schätzer zu erhalten, muss die Schätzgleichung betrachtet werden. Dabei wird mit f die Dichtefunktion von F bezeichnet, für welche gilt $F' = f$. Weiters wird mit δ_{il} das Kronecker Delta bezeichnet. Die Ableitungen nach b_j und β_l haben die Form

$$\frac{\partial l(\mathbf{b}, \boldsymbol{\beta} | \mathbf{y})}{\partial \beta_l} = \sum_{i=1}^n \sum_{j=1}^k \alpha_{ij} x_{il} \frac{f(b_j + \mathbf{x}_i^\top \boldsymbol{\beta}) - f(b_{j-1} + \mathbf{x}_i^\top \boldsymbol{\beta})}{F(b_j + \mathbf{x}_i^\top \boldsymbol{\beta}) - F(b_{j-1} + \mathbf{x}_i^\top \boldsymbol{\beta})} \quad (4.19)$$

$$\frac{\partial l(\mathbf{b}, \boldsymbol{\beta} | \mathbf{y})}{\partial b_l} = \sum_{i=1}^n \sum_{j=1}^k \alpha_{ij} \frac{\delta_{jl} f(b_j + \mathbf{x}_i^\top \boldsymbol{\beta}) - \delta_{j-1,l} f(b_{j-1} + \mathbf{x}_i^\top \boldsymbol{\beta})}{F(b_j + \mathbf{x}_i^\top \boldsymbol{\beta}) - F(b_{j-1} + \mathbf{x}_i^\top \boldsymbol{\beta})} \quad (4.20)$$

für $j = 1, \dots, k$ und $l = 1, \dots, j$. Da diese Gleichungen oftmals nichtlinear sind, kann keine analytische Lösung gefunden werden. Unter der Annahme der Existenz einer Lösung kann diese mit Hilfe von Iterationsverfahren approximiert werden. Ein Beispiel dafür wäre die *iteratively reweighted least-square* Methode (3.6) siehe Anhang A.1. Von Burrige (1981) wurde gezeigt, dass die Log-Likelihood-Funktion im Logit-, im Probit- und im Extremwert-Modell eine konkave Funktion ist und ein Iterationsverfahren daher schnell konvergiert.

4.4 Sequentielle Modelle

Für ordinale Responses können Fälle auftreten, bei denen die Kategorien sukzessive durchschritten werden müssen. Höhere Kategorien können hierbei nur nach dem Erreichen der vorherigen Kategorien eintreten. Bei der Modellierung dieser Daten finden dann die *sequentuellen Modelle* ihre Anwendung. Mittels der ordinalen Response Y werden schrittweise die Wahrscheinlichkeiten für das Eintreten der Kategorien $1, \dots, k$ geschätzt. Das Modell beginnt mit der Entscheidung zwischen der ersten Kategorie $\{1\}$ und den restlichen Kategorien $\{2, \dots, k\}$ und stellt somit ein binäres Problem dar. Diese Entscheidung wird modelliert durch

$$\pi_1 = P(Y = 1 | \mathbf{x}) = F(b_1 + \mathbf{x}^\top \boldsymbol{\beta})$$

mit den erklärenden Variablen \mathbf{x} und einer Verteilungsfunktion F . Gilt $Y = 1$, dann stoppt das Modell nach der ersten Entscheidung. Andernfalls ist $Y \geq 2$ und es kommt

zur nächsten Entscheidung zwischen der Kategorie $\{2\}$ und den Kategorien $\{3, \dots, k\}$, welche modelliert wird durch

$$P(Y = 2|Y \geq 2, \mathbf{x}) = F(b_2 + \mathbf{x}^\top \boldsymbol{\beta}).$$

Auf Grund der sukzessive steigenden Kategorien müssen die Kategorien von zuvor nicht überprüft werden. Im Schritt j ist die Entscheidung zwischen Kategorie $\{j\}$ und den Kategorien $\{j+1, \dots, k\}$ modelliert mit Hilfe des binären sequentiellen Modells

$$P(Y = j|Y \geq j, \mathbf{x}) = F(b_j + \mathbf{x}^\top \boldsymbol{\beta}), \quad j = 1, \dots, k-1. \quad (4.21)$$

Bei der Entscheidung zwischen der Kategorie $\{j\}$ und den Kategorien $\{j+1, \dots, k\}$ wird die Annahme getroffen, dass für alle Schritte j das gleiche binäre Modell mit der Verteilungsfunktion F zugrunde liegt.

In weiterer Folge kann auch die Wahrscheinlichkeit $P(Y = j|\mathbf{x})$ angegeben werden. Angenommen $Y = j$, dann wurde in dem sequentiellen Prozess für alle $s < j$ mit $s = 1, \dots, j-1$ eine Entscheidung für die Kategorie $\{s+1, \dots, k\}$ getroffen. Dieses kann durch die Wahrscheinlichkeit

$$\prod_{s=1}^{j-1} P(Y > s|Y \geq s, \mathbf{x})$$

angegeben werden. Im Schritt j beträgt die Wahrscheinlichkeit der Entscheidung für die Kategorie j

$$P(Y = j|Y \geq j, \mathbf{x}).$$

Somit kann die Wahrscheinlichkeit $P(Y = j|\mathbf{x})$ als Produkt der beiden Wahrscheinlichkeiten dargestellt werden

$$\begin{aligned} \pi_j = P(Y = j|\mathbf{x}) &= P(Y = j|Y \geq j, \mathbf{x}) \prod_{s=1}^{j-1} P(Y > s|Y \geq s, \mathbf{x}) \\ &= P(Y = j|Y \geq j, \mathbf{x}) \prod_{s=1}^{j-1} (1 - P(Y = s|Y \geq s, \mathbf{x})) \\ &= F(b_j + \mathbf{x}^\top \boldsymbol{\beta}) \prod_{s=1}^{j-1} (1 - F(b_s + \mathbf{x}^\top \boldsymbol{\beta})). \end{aligned}$$

4.5 Linkfunktionen des sequentiellen Modells

Das sequentielle Modell (4.21) wird mit Hilfe der Verteilungsfunktion F beschrieben. Im Folgenden werden sequentielle Modelle mit verschiedenen Linkfunktionen angegeben.

4.5.1 Sequentielles Logit Modell

Bei den sequentiellen Modellen findet das *sequentielle Logit Modell* die häufigste Anwendung. Analog zum kumulativen Logit Modell (4.4) ist F dabei die logistische Verteilungsfunktion $F(x) = \exp(x)/(1 + \exp(x))$. Unter anderem wird die Definition des sequentiellen Logit benötigt.

Definition 4.3. (*sequentieller Logit*) Es seien $1, \dots, k$ Kategorien mit Wahrscheinlichkeiten π_1, \dots, π_k , dann ist der *sequentielle Logit* definiert durch

$$\text{logit}(P(Y = j|\mathbf{x})) = \log \left(\frac{P(Y = j|\mathbf{x})}{P(Y > j|\mathbf{x})} \right) = \log \left(\frac{\pi_j}{\pi_{j+1} + \dots + \pi_k} \right) \quad j = 1, \dots, k - 1$$

für jede Kategorie in Relation zu den höheren Kategorien.

Mit dieser Definition kann nun das sequentielle Logit Modell dargestellt werden als

$$\begin{aligned} \text{logit}(P(Y = j|\mathbf{x})) &= \log \left(\frac{P(Y = j|\mathbf{x})}{P(Y > j|\mathbf{x})} \right) \\ &= b_j + \mathbf{x}^\top \boldsymbol{\beta} \end{aligned}$$

$$P(Y = j|Y \geq j, \mathbf{x}) = \frac{\exp(b_j + \mathbf{x}^\top \boldsymbol{\beta})}{1 + \exp(b_j + \mathbf{x}^\top \boldsymbol{\beta})}.$$

Das proportionale Verhältnis für zwei Vektoren erklärender Variablen \mathbf{x} und $\tilde{\mathbf{x}}$ ist gegeben durch

$$\frac{P(Y = j|\mathbf{x})/(1 - Y \leq j|\mathbf{x})}{P(Y = j|\tilde{\mathbf{x}})/(1 - Y \leq j|\tilde{\mathbf{x}})} = \frac{P(Y = j|\mathbf{x})/(Y > j|\mathbf{x})}{P(Y = j|\tilde{\mathbf{x}})/(Y > j|\tilde{\mathbf{x}})} = \exp((\mathbf{x} - \tilde{\mathbf{x}})^\top \boldsymbol{\beta}).$$

4.5.2 Sequentielles Extremwert Modell

Wird als Verteilungsfunktion F die Extremwertverteilung $F(x) = 1 - \exp(-\exp(x))$ verwendet, so ergibt sich das *sequentielle Minimum-Extremwert-Modell* mit

$$P(Y = j|Y \geq j, \mathbf{x}) = 1 - \exp(-\exp(b_j + \mathbf{x}^\top \boldsymbol{\beta})).$$

Dieses Modell ist äquivalent zum kumulativen Minimum-Extremwert Modell 4.2.3.

4.6 Sequentielles Modell für GLM

Mit Hilfe des sequentiellen Modells können GLMs für ordinale Response modelliert werden. Für multinomiale Daten hat das sequentielle Modell die allgemeine Darstellung

$$P(Y_i = j | Y_i \geq j, \mathbf{x}_i) = F(b_j + \mathbf{x}_i^\top \boldsymbol{\beta}).$$

Wie in der allgemeinen Beschreibung des Modells gezeigt, gilt für π_{ij} die äquivalente Darstellung

$$\pi_{ij} = P(Y_i = j | \mathbf{x}_i) = F(b_j + \mathbf{x}_i^\top \boldsymbol{\beta}) \prod_{s=1}^{j-1} (1 - F(b_s + \mathbf{x}_i^\top \boldsymbol{\beta})).$$

Der lineare Prädiktor an der Stelle der Teilung j hat die Gestalt $\eta_{ij} = b_j + \mathbf{x}_i^\top \boldsymbol{\beta}$, daraus kann die multinomiale Responsefunktion $\mathbf{h}(\boldsymbol{\eta}) = (h_1(\boldsymbol{\eta}), \dots, h_{k-1}(\boldsymbol{\eta})) : \mathbb{R}^{k-1} \rightarrow \mathbb{R}^{k-1}$ angegeben werden mit

$$h_j(\eta_{i1}, \dots, \eta_{ik-1}) = F(\eta_{ij}) \prod_{s=1}^{j-1} (1 - F(\eta_{is})).$$

Die Linkfunktion $\mathbf{g}(\cdot) = \mathbf{h}^{-1}(\cdot)$ hat die Gestalt

$$g_j(\pi_{i1}, \dots, \pi_{ik-1}) = F^{-1} \left(\frac{\pi_{ij}}{1 - \pi_{i1} - \dots - \pi_{ij-1}} \right).$$

Die Designmatrix sowie die Form der Response Funktion entsprechen der Darstellung im kumulativen Modell.

4.6.1 Maximum Likelihood Schätzung für sequentielle Modelle

Für sequentielle Modelle ergibt sich die Maximum Likelihood Schätzung durch die Betrachtung binärer Modelle. Es bezeichnet im Folgenden

$$\nu(j) = P(Y = j | Y \geq j), \quad j = 1, \dots, k-1$$

die konditionalen Übergangswahrscheinlichkeiten, für welche die Darstellung

$$P(Y \geq j) = \prod_{s=1}^{j-1} (1 - \nu(s))$$

gilt. Mit den obigen Ergebnissen folgt

$$P(Y = j) = P(Y = j | Y \geq j)P(Y \geq j) = \nu(j) \prod_{s=1}^{j-1} (1 - \nu(s)).$$

Die Form $\nu(j|\mathbf{x}_i)$ bezeichnet die Übergangswahrscheinlichkeit für die gegebenen erklärenden Variablen \mathbf{x}_i . Die Likelihood Funktion hat die Gestalt

$$L_i(\mathbf{b}, \boldsymbol{\beta} | \mathbf{y}) = P(Y_i = j_i | \mathbf{x}_i) = \nu(j_i | \mathbf{x}_i) \prod_{s=1}^{j_i-1} (1 - \nu(s | \mathbf{x}_i)).$$

Die Definition eines Vektors $(y_{i1}, \dots, y_{ij_i}) = (0, \dots, 0, 1)$ führt zur Umformulierung der Likelihood Funktion in

$$L_i(\mathbf{b}, \boldsymbol{\beta} | \mathbf{y}) = \prod_{s=1}^{j_i} \nu(s | \mathbf{x}_i)^{y_{is}} (1 - \nu(s | \mathbf{x}_i))^{1-y_{is}}$$

und dadurch ergibt sich die Äquivalenz von L_i und einem binären Response Modell mit y_{i1}, \dots, y_{ij_i} . Es folgt dadurch, dass $\nu(s | \mathbf{x}_i) = F(\eta_{is})$. Wird dieses nun in L_i eingesetzt, so ist

$$L_i(\mathbf{b}, \boldsymbol{\beta} | \mathbf{y}) = \prod_{s=1}^{j_i} F(\eta_{is})^{y_{is}} (1 - F(\eta_{is}))^{1-y_{is}}.$$

Zur Berechnung der Schätzer wird die Log-Likelihood-Funktion l betrachtet. Diese hat die Form

$$l(\mathbf{b}, \boldsymbol{\beta} | \mathbf{y}) = \log(L_i(\mathbf{b}, \boldsymbol{\beta} | \mathbf{y})) = \sum_{i=1}^n \sum_{s=1}^{j_i} (y_{is} \log(F(\eta_{is})) + (1 - y_{is}) \log(1 - F(\eta_{is}))).$$

5 Modellierung in R

Die zwei nachfolgenden Kapitel beinhalten die praktische Anwendung der zuvor besprochenen theoretischen Aspekte mit Hilfe von den gegebenen Daten. Es wird versucht mittels der in 1.1 erwähnten Daten geeignete Modelle für die Clavier-Dindo-Klassifikation (CD-Klassifikation) zu erstellen und diese zu analysieren. Die Analyse und Modellschätzung wird mit der Programmiersprache R durchgeführt. Dazu werden in diesem Kapitel die in der Analyse verwendeten Packages der Software und deren Funktionen erklärt.

Im folgenden Abschnitt werden verschiedene Möglichkeiten zur Schätzung der Modelle für ordinale Daten erklärt. Der Schwerpunkt liegt dabei bei dem Package `VGAM`, welches bei ordinalen Daten die häufigste Anwendung findet. Für eine detaillierte Beschreibung der Theorie und der einzelnen Funktionen siehe Yee (2010) und Yee (2018).

5.1 Package VGAM

Das Package `VGAM` (*Vector Generalized Linear and Additive Models*) beinhaltet umfassende Funktionen für die Erstellung von vektorisierten GLMs und vektorisierten generalisierten additiven Modellen (GAMs). Die Grundlage des Packages bildet der *iteratively reweighted least squares* (IRLS) Algorithmus, für nähere Details zum Algorithmus siehe Anhang A.1. Die Schätzung der Modelle für ordinale Response wird in dieser Arbeit mit Hilfe der GLMs durchgeführt. Da die GLMs ein Spezialfall der VGLMs sind, werden in weiterer Folge nur die VGLMs behandelt und deren Modellierung mit dem Package `VGAM` erklärt.

Bei VGLMs wird angenommen, dass die Wahrscheinlichkeitsfunktion der Response (Y_1, \dots, Y_k) in der Gestalt

$$f(y_1, \dots, y_k | \mathbf{x}, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_k) = f((y_1, \dots, y_k), \eta_1, \dots, \eta_k) \quad (5.1)$$

geschrieben werden kann, wobei die Funktion $f(\cdot)$ eine bekannte Funktion darstellt. Für die Funktion $f(\cdot)$ wird vorausgesetzt, dass die Log-Likelihood-Funktion, deren

Ableitung und deren Fisher-Matrix existieren. Die Prädiktoren η_1, \dots, η_k haben in der Funktion die Darstellung

$$\eta_j = \mathbf{x}^\top \boldsymbol{\beta}_j, \quad j = 1, \dots, k. \quad (5.2)$$

Es zeigt sich, dass die VGLMs eine Verallgemeinerung der GLMs für multiple lineare Prädiktoren sind. Dabei umfassen diese Modelle sowohl die univariaten GLMs in der Darstellung (3.2) als auch die multivariaten GLMs mit der Darstellung (3.9). Wird für die Funktion $f(\cdot)$ ein Mitglied der Exponentialfamilie gewählt, so entspricht die Wahrscheinlichkeitsfunktion (5.1) der stochastischen Komponente der Basisstruktur der GLMs in Abschnitt 3.1. Die Struktur der linearen Prädiktoren und der Linkfunktion ist bei den VGLMs analog zu der Struktur der GLMs. Aufgrund der Definition der vektorwertigen GLMs sind in den VGLMs auch Modelle enthalten, deren Verteilung nicht in der Exponentialfamilie ist.

Die Funktion `vglm` aus dem Package `VGAM` bietet umfangreiche Möglichkeiten zur Spezifikation von GLMs. Der Aufruf der Funktion in R erfolgt durch den Befehl `vglm()`. Hierbei können die folgenden Argumente übergeben werden.

```
vglm(formula, family(reverse = TRUE/FALSE, parallel = TRUE/FALSE,
  link = ""), ...)
```

Das Argument `formula` beschreibt die Struktur des GLMs mit Hilfe der erklärenden Variablen. Im Argument `family` wird die Verteilung und Linkfunktion des erstellten und untersuchten Modells spezifiziert. Im Package `VGAM` sind über 150 `family`-Funktionen implementiert. In der folgenden Tabelle sind die wichtigsten `family`-Funktionen für Modelle mit kategorialer Response aufgelistet.

Name	Befehl	Wahrscheinlichkeit
Adjacent-Category	<code>acat(reverse = FALSE)</code>	$P(Y = j + 1 \mathbf{x}) / P(Y = j \mathbf{x})$
Continuous Ratio	<code>cratio(reverse = FALSE)</code>	$P(Y > j Y \geq j, \mathbf{x})$
Kumulatives Modell	<code>cumulative(reverse = FALSE)</code>	$P(Y \leq j \mathbf{x})$
Sequentielles Modell	<code>sratio(reverse = FALSE)</code>	$P(Y = j Y \geq j, \mathbf{x})$
Multinomiales Modell	<code>multinomial()</code>	$\log(P(Y = j \mathbf{x}) / P(Y = k \mathbf{x}))$

Tabelle 5.1: Auflistung der wählbaren Modelle in der Funktion `vglm`

Der Befehl `reverse()` der `family`-Funktionen `acat()`, `cratio()`, `cumulative()` und `sratio()` ist dabei optional und standardmäßig `FALSE`. In der nachfolgenden Tabelle sind die resultierenden Wahrscheinlichkeiten bei einem Aufruf mit dem Argument `reverse = TRUE` aufgelistet. Für das kumulative Modell kann mit Hilfe des Befehls `cumulative(parallel = TRUE)` bestimmt werden, ob eine Gleichheit der Effekte an-

Name	Befehl	Wahrscheinlichkeit
Adjacent-Category	<code>acat(reverse = TRUE)</code>	$P(Y = j \mathbf{x})/P(Y = j + 1 \mathbf{x})$
Continuous Ratio	<code>cratio(reverse = TRUE)</code>	$P(Y < j Y \leq j, \mathbf{x})$
Kumulatives Modell	<code>cumulative(reverse = TRUE)</code>	$P(Y \geq j \mathbf{x})$
Sequentielles Modell	<code>sratio(reverse = TRUE)</code>	$P(Y = j Y \leq j, \mathbf{x})$

Tabelle 5.2: Auflistung der wählbaren Modelle in der Funktion `vglm` mit dem zusätzlichen Attribut `reverse = TRUE`

genommen wird, siehe Abschnitt 4.1.3. Für gemischte Modelle in der Gestalt (4.15) können die Variablen, deren Effekte äquivalent sind, einzeln angegeben werden.

In der Tabelle 5.1 ist zu sehen, dass für multinomiale Modelle mit der Funktion `multinom()` ohne zusätzlicher Angabe die Kategorie k als Referenz verwendet wird. Mit Hilfe des Befehls `refLevel` kann die Referenzkategorie festgelegt werden. Soll die Kategorie 1 als Referenz für das multinomiale Modell gewählt werden, lautet der Befehl `multinomial(refLevel = 1)`.

Für die Erstellung eines GLMs ist die Wahl einer Linkfunktion (3.2) erforderlich. In der Funktion `vglm` stehen hierbei verschiedene Linkfunktionen zur Verfügung, welche mit dem Befehl `link =` festgelegt werden können. Ohne Angabe eines Arguments wird die Logit-Linkfunktion verwendet. In der folgenden Tabelle sind einige Linkfunktionen aufgelistet.

Linkfunktion	$g(\theta)$
<code>cloglog()</code>	$\log(-\log(1 - \theta))$
<code>identity()</code>	θ
<code>logc()</code>	$\log(1 - \theta)$
<code>loge()</code>	$\log(\theta)$
<code>logit()</code>	$\log\left(\frac{\theta}{1-\theta}\right)$
<code>probit()</code>	$\Phi^{-1}(\theta)$
<code>rhobit()</code>	$\log\left(\frac{1+\theta}{1-\theta}\right)$

Tabelle 5.3: Auflistung einiger wählbarer Linkfunktionen in der Funktion `vglm`

Mittels der zuvor beschriebenen Argumente können verschiedene Methoden zur Modellschätzung von ordinalen Daten in R angewandt werden.

5.2 Package MASS und nnet

Als Vorgänger der Funktion `vglm` können die Funktion `multinom` aus dem Package `nnet` und die Funktion `polr` aus dem Package `MASS` angesehen werden. Im Unterschied zu der Funktion `vglm` können mit diesen Funktionen nur spezifische Modelle aufgestellt werden. Mit Hilfe der Funktion `multinom` kann ein multinomiales Logit Modell analysiert werden. Die zwei Aufrufe in R

```
multinom(formula)
```

```
vglm(formula, family = multinomial(refLevel = 1),...)
```

verwenden beide die Kategorie 1 als Referenzkategorie. Die Funktion `polr` (*proportional odds logistic regression*) beschreibt hingegen das einfache kumulative Logit Modell. Für die Funktion `polr` ist die interne Definition gegeben durch

$$\text{logit}(P(Y \leq j|\mathbf{x})) = b_j - \mathbf{x}^\top \boldsymbol{\beta}, \quad (5.3)$$

wohingegen die Funktion `vglm` die Darstellung

$$\text{logit}(P(Y \leq j|\mathbf{x})) = b_j + \mathbf{x}^\top \boldsymbol{\beta}$$

verwendet. Die beiden Definitionen stimmen daher bis auf das Vorzeichen überein, welches den Unterschied im geschätzten Vektor $\hat{\boldsymbol{\beta}}$ liefert. Die Aufrufe der Funktionen in R lauten

```
polr(formula)
```

```
vglm(formula, family = cumulative(parallel = TRUE),...)
```

In dieser Arbeit werden diese beiden Funktionen für den Vergleich mit den Modellen der Funktion `vglm` gewählt. Für weitere Informationen bezüglich der Funktion `polr` siehe Venables und Ripley (2002), bezüglich der Funktion `multinom` siehe Venables und Ripley (2016).

5.3 Package EffectStars2

Für die graphische Darstellung der vektorwertigen GLMs wird die Funktion `effect-stars.vglm` aus dem Package `EffectStars2` verwendet. Diese Funktion ist für multinomiale Logit Modelle, welche mit Hilfe des Attributs `family = multinomial()` erstellt wurden, und für ordinale Modelle, welche mit `family = sratio()`, `family =`

`cratio()`, `family = cumulative()` oder `family = acat()` analysiert wurden, anwendbar. Besitzt ein Faktor für jede Kategorie eine eigene Schätzung, so kann für diesen Faktor ein Effektstern erstellt werden. Der Effektstern stellt die Schätzungen der Effekte sternförmig um einen Mittelpunkt dar und ermöglicht somit einen direkten Vergleich der Effekte mit den einzelnen Kategorien. Da die Schätzung des Effekts als Abstand zum Mittelpunkt dargestellt wird und somit keine negativen Werte verwendet werden können, wird für eine sinnvolle Darstellung der Exponent der Schätzungen gewählt. Jeder Effektstern beinhaltet einen Kreis, welcher Radius $\exp(0) = 1$ hat und somit den Fall von keinem Effekt abbildet. Der Aufruf in R lautet

```
effectstars.vglm(model, plot.parallel = FALSE, symmetric = FALSE,  
                control = star.ctrl(radius=1))
```

wobei `model` für das VGLM steht. Falls das Attribut `plot.parallel = FALSE` gesetzt ist, werden nur die Effekte, welche sich für die einzelnen Kategorien unterscheiden, dargestellt. Für ein multinomiales Modell wird mit `symmetric = FALSE` die Referenzkategorie des Modells als Referenz mit dem Wert $\exp(0) = 1$ dargestellt. Mit dem Zusatz `control = star.ctrl(radius = 1)` kann der Radius des Kreises geändert werden.

6 Modelle zur Analyse der Clavien-Dindo-Klassifikation

In diesem Kapitel folgt die praktische Anwendung der diskutierten Modelle auf die vorliegenden Daten. Die Response CD-Klassifikation kann dabei anhand verschiedener Ausprägungen analysiert werden. Zu Beginn wird eine Übersicht des Datensatzes gegeben. Anschließend wird der binäre Fall mit den Ausprägungen $\{0, 1\}$ betrachtet, wobei einer Patientin mit einer CD-Klassifikation die Ausprägung 1 und einer Patientin ohne CD-Klassifikation die Ausprägung 0 zugeteilt wird. Danach wird die CD-Klassifikation als multinomiale Response betrachtet, welche im Unterschied zu davor alle Klassifikationskategorien beinhaltet. Im nächsten Abschnitt werden die Modelle für ordinale Responses betrachtet. Die CD-Klassifikation wird dabei mit den beiden Modellverfahren der kumulativen und sequentiellen Modelle, welche in Kapitel 4 beschrieben wurden, analysiert.

6.1 Beschreibung des Datensatzes

Die vorliegenden Daten beinhalten die Informationen von 486 ausschließlich weiblichen Patientinnen, welche sich dem operativen Eingriff einer Mammareduktion in einem Zeitraum von 10 Jahren unterzogen. Die Daten umfassen für jede Patientin 9 präklinische und 11 klinische Größen. Während die präklinischen Größen Faktoren und Eigenschaften beinhalten, welche die Patientin vor dem operativen Eingriff aufweist, umfassen die klinischen Größen nur jene, welche bei oder nach der Operation auftreten können. Die vorliegende Arbeit befasst sich mit dem Modell basierend auf sämtliche Größen, welches sich sowohl auf präklinische als auch klinische Größen konzentriert.

Um die relevanten Einflussgrößen zu erhalten, wurde der Einfluss der 9 präklinischen und 11 klinischen Größen auf die CD-Klassifikation überprüft. Zuerst wurden die präklinischen Faktoren separat untersucht und dabei die relevanten Einflussgrößen von diesen für die Klassifikation ermittelt. Anschließend wurden die klinischen Größen analog behandelt. Zuletzt wurden alle Größen in einer gemeinsamen Klasse betrachtet

und die relevanten Faktoren herausgefiltert. Dieses allgemeine Ergebnis wurde mit den einzelnen Ergebnissen für die klinischen und präklinischen Größen verglichen. Traten Faktoren in einer separat getesteten Klasse auf, jedoch nicht im Modell mit allen Faktoren, so wurden diese zusätzlich noch einmal getestet und deren Relevanz für das Modell überprüft. Diese Arbeit beinhaltet die Ergebnisse des allgemeinen Modells, welches nach obigem Verfahren analysiert wurde. Im Nachfolgenden werden die relevanten Faktoren zur Modellierung der CD-Klassifikation beschrieben, welche sich nach Durchführung des zuvor beschriebenen Analysevorgangs herausgefiltert haben. In der Tabelle 6.1 wird eine kurze Übersicht der verwendeten Einflussfaktoren und deren Ausprägungen sowie deren Bezeichnung angegeben.

Namen (Bezeichnung)	Ausprägung	
	Original	Verwendet
Präklinischer Einflussfaktor		
Rauchen (<code>smoke.yn</code>)	Pack year (py)	0,1
Klinische Einflussfaktoren		
Single-Shot-Antibiotikum (<code>single.shot</code>)	0,1,2,...,12	0,1
Resektionsmenge (<code>resection</code>)	[g]	[g]
Revision (<code>revision</code>)	0, Beschreibung	0,1
Operationstechnik (<code>op.tec</code>)	1,2,...,8	1,2,...,5

Tabelle 6.1: Einflussgrößen und deren Ausprägungen

Der präklinische Einflussfaktor Rauchen beschreibt das Rauchverhalten der Patientin. Das Rauchverhalten wurde in Packungsjahren (Pack years) ausgewiesen, welche das Produkt aus der Anzahl an gerauchten Zigaretenschachteln an einem Tag mit der Anzahl an Jahren des Rauchens sind. Für die Modellierung wird eine Unterscheidung in Raucherin und Nichtraucherin verwendet. Dieser Faktor zählt zu den Größen, welche die Patientin bereits vor der Operation aufweist. Die weiteren relevanten Größen sind klinische Faktoren, welche beim Vollzug der Operation eintreten. Der Faktor Single-Shot-Antibiotikum (Single-Shot-AB) bezeichnet eine einmalige Antibiotikungabe vor der Durchführung der Mammareduktion. Diese Antibiotikagabe hat sich in den letzten Jahren verbreitet durchgesetzt. In den Daten wurde protokolliert, welches Antibiotikum verwendet wurde, wobei insgesamt 12 verschiedene Antibiotika verwendet wurden. Im Modell wird zwischen einem Single-Shot-AB und keinem Single-Shot-AB unterschieden. Als Resektionsmenge ist die Menge an Gewebe definiert, welche bei der Operation entfernt wurde. Die Revision bezeichnet eine Wiederholung oder Erweiterung eines zuvor durchgeführten Eingriffes und somit im vorliegenden Datensatz einen weiteren Eingriff bezüglich der Mammareduktion. In den Daten wurde der Eingriff genau ausgewiesen, wobei für das Modell eine Unterscheidung in Revision und keine

Revision getroffen wird. Die Operationstechnik beschreibt die Methode, die bei der Operation verwendet wurde. Hierbei wurden acht verschiedene Techniken angewandt, wobei von Dr. Winter vier Techniken zu einer Gruppe zusammengefasst wurden. Die OP-Techniken weisen die Bezeichnungen `cranial.pedicle`, `free.nipple`, `vertical`, `caudal.pedicle` und `other` auf. Da die Operationstechnik `other` verschiedene Techniken zusammenfasst, kommt es dabei zu einer Inhomogenität. Die Patientinnen in dieser Zuordnung sind nicht miteinander vergleichbar, wodurch auch keine eindeutigen Aussagen bezüglich der Operationstechnik `other` gemacht werden können. Aus diesem Grund wurden die 16 Patientinnen mit dieser Operationstechnik für die Modellierung im Anschluss aus dem Datensatz entfernt.

Die nachfolgenden Tabellen und Abbildungen zeigen die Häufigkeiten der Daten der verbleibenden 470 Patientinnen. Das Durchschnittsalter der Patientinnen liegt bei 39 Jahren. Der Datensatz umfasst einen Zeitraum von 10 Jahren, wobei die Informationen vor, während und nach der Operation vom medizinischen Personal dokumentiert wurden.

CD-Klassifikation	Absolute Häufigkeit	Relative Häufigkeit
Grad 0	178 Patientinnen	37.87%
Grad 1	226 Patientinnen	48.08%
Grad 2	43 Patientinnen	9.15%
Grad 3A	2 Patientinnen	0.43%
Grad 3B	21 Patientinnen	4.47%

Tabelle 6.2: Absolute und relative Häufigkeiten der CD-Klassifikationen in den vorliegenden Daten

Die Tabelle 6.2 stellt die absoluten und relativen Häufigkeiten der einzelnen Klassifikationen der 470 Patientinnen dar. Die meisten Patientinnen wurden nach der Mamma-reduktion dem Grad 0 oder 1 zugeordnet. Nur wenige Patientinnen benötigten einen weiteren Eingriff ohne genereller Anästhesie, weswegen der Grad 3A bei nur 2 Patientinnen zugeteilt wurde. Aufgrund dieser geringen Anzahl werden die zwei Grade 3A und 3B in der Modellierung nicht unterschieden und zum Grad 3 zusammengelegt. Somit ergeben sich 23 Patientinnen für den Grad 3 mit einer relativen Häufigkeit von 4.9%. In den nachfolgenden Auswertungen und der weiteren Arbeit wird immer der Grad 3 behandelt. Die Tabelle 6.3 veranschaulicht die Aufteilung der absoluten Häufigkeiten für die binären Faktoren in die einzelnen Stufen der Klassifikation. Es zeigt sich, dass im vorliegenden Datensatz die Anzahl der Nichtraucherinnen größer ist als die der Raucherinnen. Die Grade 0 und 1 dieses Faktors weisen in beiden Fällen die meisten Daten auf. Die meisten Patientinnen bekamen vor der Operationsdurchführung einen

CD	smoke.yn	smoke.yn	single.shot	single.shot	revision	revision
	= 1	= 0	= 1	= 0	= 1	= 0
0	55	123	155	23	8	170
1	81	145	180	46	22	204
2	19	24	29	14	4	39
3	9	14	19	4	23	0
Gesamt	164	306	383	87	57	413

Tabelle 6.3: Häufigkeitstabelle der Faktoren unterteilt anhand der CD-Klassifikation

Single-Shot-AB. Nur 57 von 470 Patientinnen mussten sich nach der Mammareduktion einer Revisionsoperation unterziehen, wobei der Grad 3 am häufigsten aufgetreten ist. Im Allgemeinen ist der Grad 1 der CD-Klassifikation in allen Fällen am häufigsten. Bei einer revisionslosen Operation wies keine Patientin den Grad 3 auf.

CD	cranial.pedicle	free.nipple	vertical	caudal.pedicle
0	152	2	14	10
1	171	6	34	15
2	33	0	5	5
3	18	0	2	3
Gesamt	374	8	55	33

Tabelle 6.4: Häufigkeitstabelle der Operationstechniken unterteilt anhand der CD-Klassifikation

Die Tabelle 6.4 zeigt die Häufigkeiten der Stufen abhängig von den einzelnen Operationstechniken. Die Technik `free.nipple` wurde nur selten angewandt. Weiters ist ersichtlich, dass bei dieser Durchführung kein Grad 2 und 3 auftreten. Für alle Techniken ist die Häufigkeit beim Grad 1 am größten.

Die Abbildungen 6.1, 6.2 und 6.3 zeigen die Häufigkeiten der Einflussgrößen anhand von Säulendiagrammen. Dabei werden die Resultate aus den Tabellen zuvor veranschaulicht.

6.2 Binomiale Modelle

Zu Beginn wird ein Modell für die Response CD-Klassifikation betrachtet, welche mit den Stufen $\{0, 1\}$ eine binäre Gestalt aufweist. Dabei beinhaltet die binäre Response die zwei Ausprägungen für einen komplikationsfreien Heilungsverlauf (0) und das

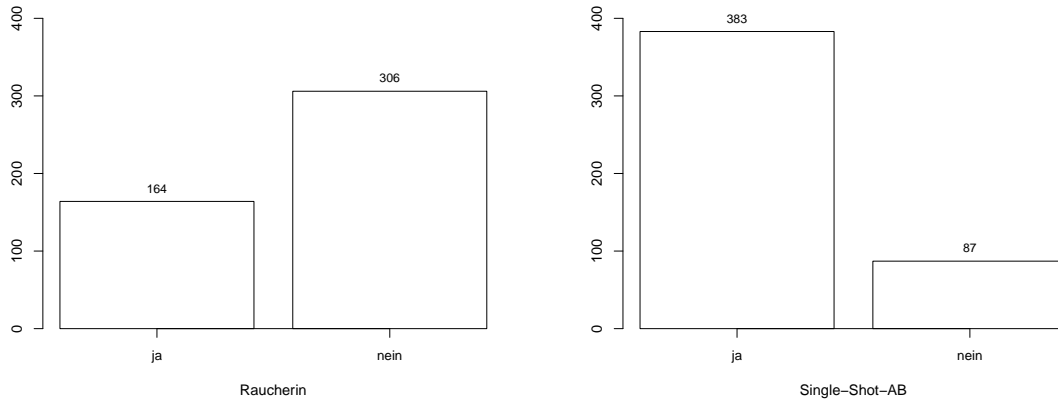


Abbildung 6.1: Häufigkeitsdiagramm Rauchverhalten (links) und Single-Shot-AB (rechts)

Eintreten einer CD-Klassifikation (1). Diese Response Variable wird im Folgenden mit `cd.bin` bezeichnet. Mit Hilfe eines logistischen Modells und den vorliegenden Prädiktoren kann somit eine Schätzung des Modells für die binomiale Response erfolgen. Die relevanten Einflussgrößen umfassen die Größen, welche in Abschnitt 6.1 analysiert und beschrieben wurden.

```
> mod.cd.bin <- glm(cd.bin ~ single.shot + resection + op.tec
+ revision, family = binomial)
> summary(mod.cd.bin)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.6663619	0.3601545	-1.850	0.064283	.
single.shot	-0.7491761	0.2809319	-2.667	0.007659	**
resection	0.0026467	0.0004826	5.484	4.15e-08	***
op.tecfree.nipple	-1.1978538	0.8920942	-1.343	0.179355	
op.tecvertical	1.4203021	0.3624312	3.919	8.90e-05	***
op.teccaudal.pedicle	0.1969918	0.4163789	0.473	0.636137	
revision	1.5638143	0.4104295	3.810	0.000139	***

```
Null deviance: 623.63 on 469 degrees of freedom
Residual deviance: 557.46 on 463 degrees of freedom
```

Die `summary` überprüft zeilenweise die Nullhypothese, dass der Parameter Null und somit auch, ob der dazugehörige Prädiktor, verglichen mit den anderen Prädiktoren, irrelevant ist. Die Signifikanz wird dabei mit Hilfe eines t-Tests ermittelt. Für das Modell mit der binären Response `cd.bin` ergeben sich obige Ergebnisse. Die Resektionsmenge sowie eine Revision der Operation führen zu einer signifikanten Änderung der

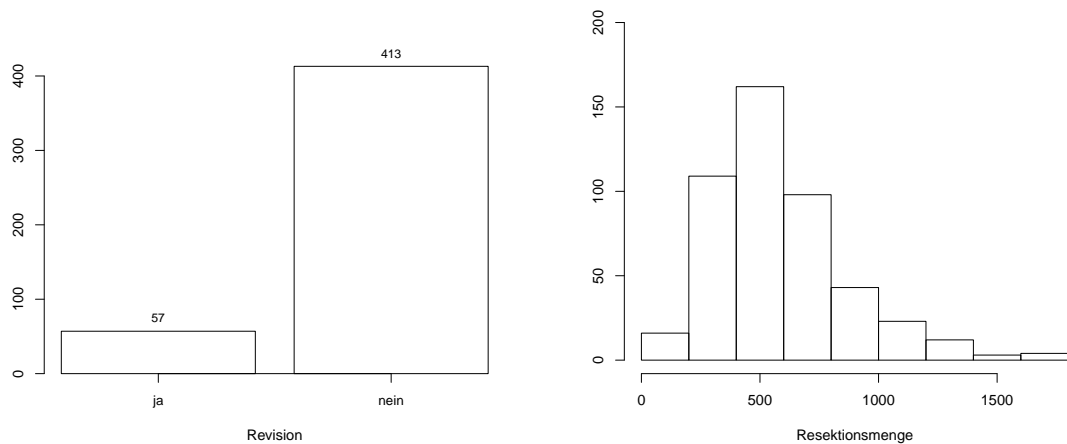


Abbildung 6.2: Häufigkeitsdiagramm Revision (links) und Resektionsmenge (rechts)

Wahrscheinlichkeit für das Eintreten einer CD-Klassifikation. Auch die einmalige Antibiotikungabe Single-Shot-AB beeinflusst die Wahrscheinlichkeit mit einem p-Wert von 0.007659 signifikant. Weiters wurde auch der Faktor `smoke.yn` auf Signifikanz im Modell überprüft. Mit einem p-Wert von 0.0741 wurde aber das Signifikanzniveau von 5% nicht unterschritten und somit Rauchen nicht ins binomiale Modell hinzugefügt.

Um die Relevanz des mehrstufigen Prädiktors `op.tec` zu überprüfen, wird der Befehl `drop1` verwendet. Mit dieser Funktion wird ohne Berücksichtigung der Reihenfolge jeder Prädiktor einzeln aus dem Modell entfernt, das neue Modell angepasst und die Deviance Differenz bewertet.

```
> drop1(mod.cd.bin, test = "Chisq")
```

```

Model:
cd.bin ~ single.shot + resection + op.tec + revision
      Df Deviance   AIC   LRT Pr(>Chi)
<none>      557.46 571.46
single.shot  1   565.01 577.01  7.549 0.0060036 **
resection    1   593.85 605.85 36.395 1.611e-09 ***
op.tec       3   575.59 583.59 18.129 0.0004137 ***
revision     1   576.10 588.10 18.637 1.581e-05 ***

```

Es wird bestätigt, dass die Prädiktoren aus dem zuvor erstellten Modell nicht vernachlässigbar sind und auch der 4-stufige Faktor Operationstechnik von Relevanz für dieses Modell ist.

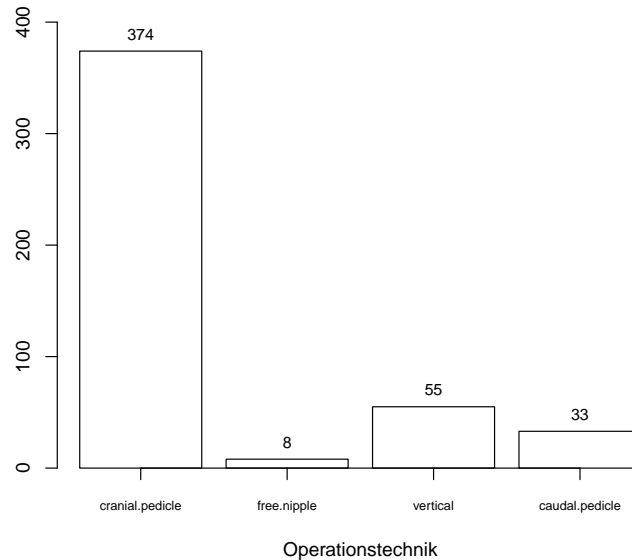


Abbildung 6.3: Häufigkeitsdiagramm Operationstechnik

Weiters werden auch die Odds Ratios angeführt. Die theoretische Grundlage der Odds und des Odds Ratios kann im Anhang A.2 nachgelesen werden.

```
> exp(cbind(OR = coef(mod.cd.bin), confint(mod.cd.bin))[-1])
```

	OR	2.5 %	97.5 %
single.shot	0.4727559	0.26820741	0.8099079
resection	1.0026502	1.00173396	1.0036335
op.tecfree.nipple	0.3018413	0.05853116	2.2813604
op.tecvertical	4.1383706	2.07745686	8.6626215
op.teccaudal.pedicle	1.2177341	0.54909078	2.8548824
revision	4.7770074	2.24909272	11.4532373

Die Chance für das Eintreten einer CD-Klassifikation beträgt bei der Single-Shot-AB Gabe 47.28% der Chance ohne einer Antibiotikumgabe. Bei der Resektionsmenge vergrößert sich die Chance um 0.3% pro 1g Resektionsmenge, für 100g erhöht sich die Chance um 30.35%. Beim Faktor Operationstechnik wird die Standardtechnik `cranial.pedicle` als Referenz für die Vergleiche herangezogen. Die Verwendung der Operationstechnik `vertical` führt somit zur 4fachen Chance im Vergleich zur Chance der Standardtechnik `cranial.pedicle`. Bei einer Revisionsoperation ist die Chance für das Eintreten einer CD-Klassifikation das 4.78fache.

Für einen Vergleich der Auswirkung verschiedener Linkfunktionen wurde das binäre Modell mit der Logit, der Probit und der Cloglog (entspricht der Gumbel-Verteilung) Linkfunktion berechnet. In der Tabelle 6.5 sind die p-Werte der Koeffizienten für die verschiedenen Modelle aufgelistet.

Faktoren	Logit Modell	Probit Modell	Cloglog Modell
<code>single.shot</code>	0.007659	0.005120	0.00202
<code>resection</code>	4.15e-08	1.31e-07	2.52e-06
<code>op.tec</code>	0.0004137	0.0009846	0.004412
<code>revision</code>	0.000139	9.34e-05	7.12e-05

Tabelle 6.5: Vergleich der einzelnen p-Werte der Koeffizienten unter Verwendung verschiedener Linkfunktionen beim binären Modell

Es zeigt sich, dass die p-Werte unter den verschiedenen Linkfunktionen ein ähnliches Verhalten aufweisen. Bei allen drei Linkfunktionen weisen die Prädiktoren ähnliche Signifikanzen auf. Die verschiedenen Operationstechniken wurden durch den Prädiktor `op.tec` zusammengefasst und mittels `drop1` analysiert.

Im Folgenden wird die Wahrscheinlichkeit für das Auftreten einer CD-Klassifikation graphisch dargestellt.

Farben	<code>op.tec</code>	<code>single.shot</code>	<code>revision</code>
blau	<code>cranial.pedicle</code>	1	0
rot	<code>cranial.pedicle</code>	0	1
schwarz	<code>cranial.pedicle</code>	0	0

Tabelle 6.6: Kombinationen der Faktoren und die dazugehörigen Farben

Die Tabelle 6.6 beschreibt die verwendeten Faktorkombinationen und deren Farbe in der Abbildung 6.4. Die Abbildung zeigt die geschätzte Wahrscheinlichkeit für eine CD-Klassifikation in Abhängigkeit von der Resektionsmenge. Als Operationstechnik wurde `cranial.pedicle` gewählt. Die schwarze Linie verdeutlicht die Wahrscheinlichkeit für eine CD-Klassifikation ohne den Einfluss von weiteren Risikofaktoren. Werden nun einzelne Prädiktoren hinzugefügt, so ändert sich die Wahrscheinlichkeit. Dabei zeigt die Graphik, dass die Gabe eines Single-Shot-AB die Wahrscheinlichkeit einer CD-Klassifikation deutlich senken kann, im Gegensatz dazu erhöht die Revision die Wahrscheinlichkeit für das Eintreten einer Klassifikation.

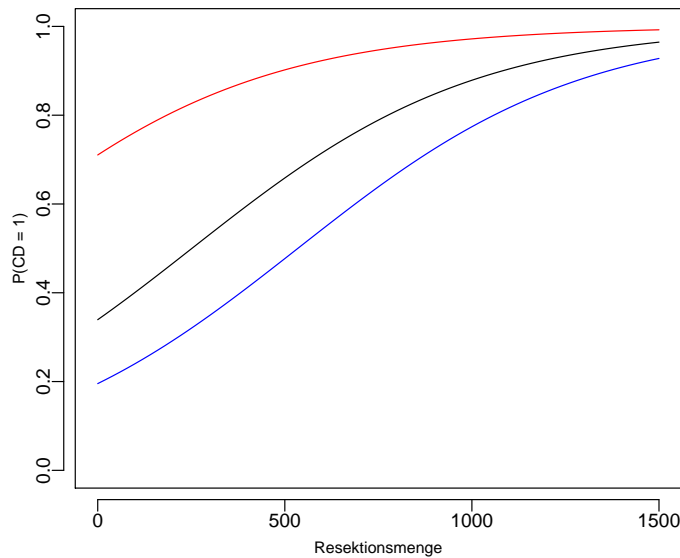


Abbildung 6.4: Geschätzte Wahrscheinlichkeit für das Eintreten der CD-Klassifikation im binomialen Modell

6.3 Multinomiale Modelle

In diesem Abschnitt wird die CD-Klassifikation als multinomiale Response mit den Ausprägungen $\{0, 1, 2, 3\}$ modelliert, wobei die Ausprägung 0 kein Auftreten einer CD-Klassifikation beschreibt. Dabei wird die natürliche Ordnung der Ausprägungen zuerst noch ignoriert und auf Grund dessen die Response Variable nur als nominal skaliert betrachtet. Es werden die beiden Funktionen `vglm` und `multinom` zur Modellschätzung herangezogen. Die Response trägt in der Modellierung den Namen `cd.mult.ges`.

Der Theorie der multikategorialen Modelle entsprechend werden die Klassifikationen $\{0, 1, 2, 3\}$ als Kategorien $\{1, 2, 3, 4\}$ bezeichnet, wodurch die Kategorie 1 die Klassifikation 0 beschreibt. Die Funktion `vglm` verwendet die Bezeichnungen der Kategorien.

6.3.1 Analyse mit `vglm`

Zu Beginn wird die Modellschätzung der Response `cd.mult.ges` mittels der Funktion `vglm` aus dem Package `VGAM` durchgeführt. Es werden zuerst die relevanten Faktoren für

das multinomiale Modell analysiert. Dabei wird das multinomiale Modell `mod.cd.mult` betrachtet, welches alle relevanten Faktoren aus dem binomialen Modell beinhaltet.

```
> mod.cd.mult <- vglm(cd.mult.ges ~ single.shot + resection + op.tec + revision,
                      family = multinomial(refLevel = 1))
```

Für jeden einzelnen Faktor wird nun die Relevanz im multinomialen Modell überprüft, wobei dieses anhand eines Likelihood-Ratio Tests durchgeführt wird. Mit Hilfe des Single-Shot-ABs wird diese Analyse im Nachfolgenden exemplarisch dargestellt.

```
> lrtest(mod.cd.mult, vglm(cd.mult.ges ~ resection + op.tec + revision,
                           family = multinomial(refLevel=1)))
```

Likelihood ratio test

Model 1: `cd.mult.ges ~ single.shot + resection + op.tec + revision`

Model 2: `cd.mult.ges ~ resection + op.tec + revision`

```
  #Df LogLik Df  Chisq Pr(>Chisq)
1 1389 -425.23
2 1392 -430.96  3 11.468  0.009447 **
```

Das Modell `mod.cd.mult` wird mit Hilfe von `lrtest` mit dem Modell ohne den Faktor Single-Shot-AB verglichen. Mit einem p-Wert von 0.009447 kann die Annahme, dass der Faktor Single-Shot-AB im Modell nicht relevant ist, verworfen werden. In der Tabelle 6.7 werden alle Faktoren und deren Ergebnisse beim Vergleich mit dem Modell `mod.cd.mult` dargestellt.

	Df	LogLik	χ^2	p-Wert
<code>mod.cd.mult</code>	-	-425.23	-	-
- <code>single.shot</code>	3	-430.96	11.468	0.009447
- <code>resection</code>	3	-443.97	37.482	3.638e-08
- <code>op.tec</code>	9	-437.53	24.61	0.003434
- <code>revision</code>	3	-481.41	112.37	< 2.2e-16

Tabelle 6.7: Auflistung der Ergebnisse des Likelihood-Ratio Tests auf Relevanz der einzelnen Faktoren

Da alle p-Werte der einzelnen Faktoren kleiner als 5% sind, wird die Relevanz jedes Faktors für das Modell ersichtlich. Somit stellen die Faktoren Single-Shot-AB, Resektionsmenge, Operationstechnik und Revision relevante Größen des multinomialen Modells dar. Da sich der Faktor `smoke.yn` im binomialen Modell nur gering über dem Signifikanzniveau befindet, wird dieser auch in diesem Modell noch einmal separat überprüft.

```
> lrtest(vglm(cd.mult.ges ~ single.shot + resection + op.tec + revision
            + smoke.yn,family = multinomial(refLevel=1)),
        mod.cd.mult)
```

Likelihood ratio test

```
Model 1: cd.mult.ges ~ single.shot + resection + op.tec + revision + smoke.yn
Model 2: cd.mult.ges ~ single.shot + resection + op.tec + revision
```

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	1386	-422.70			
2	1389	-425.23	3	5.0602	0.1674

Es ergibt sich keine Signifikanz für den Faktor Rauchen für das multinomiale Modell mittels `vglm`. Dadurch wird der Faktor in der restlichen Analyse nicht berücksichtigt. Mit Hilfe des Modells `mod.cd.mult` wird nun die Modellschätzung durchgeführt.

Als Referenzkategorie (`level`) wurde die Klassifikation 0 (`level` 1) gewählt, um einen Vergleich der Chance für eine CD-Klassifikation von 1, 2 oder 3 mit einem komplikationsfreien Heilungsverlauf zu erhalten. Wird keine Linkfunktion explizit angegeben, wird die Logit-Linkfunktion verwendet. Für den gegebenen Fall existieren aus diesem Grund drei lineare Prädiktoren

$$\log\left(\frac{\pi_j}{\pi_1}\right), \quad j = 2, 3, 4,$$

wobei die Funktion `vglm` die Bezeichnung μ anstelle von π verwendet. Für jeden linearen Prädiktor existiert ein eigener Effekt, dadurch ergeben sich für jede erklärende Variable drei Schätzungen. Diese werden in der Ausgabe der `summary` aufsteigend nummeriert.

```
> summary(mod.cd.mult)
```

Call:

```
vglm(formula = cd.mult.ges ~ single.shot + resection + op.tec +
      revision, family = multinomial(refLevel = 1))
```

Pearson residuals:

	Min	1Q	Median	3Q	Max
$\log(\mu[,2]/\mu[,1])$	-4.086	-8.561e-01	-3.642e-01	9.386e-01	1.810
$\log(\mu[,3]/\mu[,1])$	-3.405	-3.675e-01	-1.742e-01	-8.976e-02	4.582
$\log(\mu[,4]/\mu[,1])$	-4.033	-4.651e-05	4.445e-05	5.048e-05	1.347

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept):1	-8.567e-01	3.715e-01	-2.306	0.0211 *

(Intercept):2	-2.427e+00	5.445e-01	-4.458	8.29e-06	***
(Intercept):3	-1.999e+01	6.001e+02	NA	NA	
single.shot:1	-6.404e-01	2.909e-01	-2.201	0.0277	*
single.shot:2	-1.350e+00	4.106e-01	-3.289	0.0010	**
single.shot:3	-3.919e-01	7.829e-01	-0.501	0.6167	
resection:1	2.535e-03	4.912e-04	5.160	2.47e-07	***
resection:2	3.212e-03	6.964e-04	4.612	3.98e-06	***
resection:3	2.243e-03	1.191e-03	1.882	0.0598	.
op.tecfree.nipple:1	-7.781e-01	8.935e-01	-0.871	0.3838	
op.tecfree.nipple:2	-1.956e+01	5.536e+03	NA	NA	
op.tecfree.nipple:3	-1.876e+01	4.200e+03	NA	NA	
op.tecvertical:1	1.436e+00	3.685e-01	3.895	9.80e-05	***
op.tecvertical:2	1.346e+00	6.058e-01	2.222	0.0263	*
op.tecvertical:3	9.216e-01	1.074e+00	0.858	0.3910	
op.teccaudal.pedicle:1	7.323e-02	4.393e-01	0.167	0.8676	
op.teccaudal.pedicle:2	5.894e-01	6.033e-01	0.977	0.3286	
op.teccaudal.pedicle:3	8.993e-01	1.024e+00	0.879	0.3797	
revision:1	8.887e-01	4.468e-01	1.989	0.0467	*
revision:2	8.462e-01	6.695e-01	1.264	0.2063	
revision:3	2.010e+01	6.001e+02	0.033	0.9733	

Number of linear predictors: 3

Names of linear predictors:

log(mu[,2]/mu[,1]), log(mu[,3]/mu[,1]), log(mu[,4]/mu[,1])

Residual deviance: 850.4553 on 1389 degrees of freedom

Log-likelihood: -425.2276 on 1389 degrees of freedom

Number of iterations: 18

Warning: Hauck-Donner effect detected in the following estimate(s):

'(Intercept):3', 'op.tecfree.nipple:2', 'op.tecfree.nipple:3'

Reference group is level 1 of the response

Für die Kategorien $\{1, 2, 3, 4\}$ ergeben sich durch die Schätzungen der Effekte die folgenden Darstellungen mit den geschätzten Wahrscheinlichkeiten $\hat{\pi}_j$, $j = 1, 2, 3, 4$,

$$\begin{aligned} \log\left(\frac{\hat{\pi}_2}{\hat{\pi}_1}\right) &= -0.857 - 0.64 \cdot (\text{single.shot} = 1) + 0.003 \cdot \text{resection} \\ &\quad - 0.778 \cdot (\text{free.nipple} = 1) + 1.436 \cdot (\text{vertical} = 1) \\ &\quad + 0.073 \cdot (\text{caudal.pedicle} = 1) + 0.888 \cdot (\text{revision} = 1) \\ \log\left(\frac{\hat{\pi}_3}{\hat{\pi}_1}\right) &= -2.427 - 1.35 \cdot (\text{single.shot} = 1) + 0.003 \cdot \text{resection} \\ &\quad - 19.56 \cdot (\text{free.nipple} = 1) + 1.346 \cdot (\text{vertical} = 1) \\ &\quad + 0.589 \cdot (\text{caudal.pedicle} = 1) + 0.846 \cdot (\text{revision} = 1) \end{aligned}$$

$$\begin{aligned} \log\left(\frac{\hat{\pi}_4}{\hat{\pi}_1}\right) = & -19.99 - 0.392 \cdot (\text{single.shot} = 1) + 0.002 \cdot \text{resection} \\ & - 18.76 \cdot (\text{free.nipple} = 1) + 0.922 \cdot (\text{vertical} = 1) \\ & + 0.899 \cdot (\text{caudal.pedicle} = 1) + 20.1 \cdot (\text{revision} = 1). \end{aligned}$$

Mittels des Arguments `refLevel = 1` wird die Klassifikation 0 als Referenzkategorie des Modells gewählt. Bei den Operationstechniken dient die Technik `cranial.pedicle` als Referenz. Im multinomialen Modell ist ersichtlich, dass die Resektionsmenge bei allen Stufen einen geringen p-Wert und somit eine hohe Signifikanz aufweist. Die Ergebnisse zeigen, dass für den Grad 2 und 3 bei der Operationstechnik `free.nipple` keine Schätzer existieren. Dieses Resultat ergibt sich daraus, dass bei dieser Technik keine Patientin der Stufe 2 oder 3 zugeordnet wurde, welches auch in der Tabelle 6.4 ersichtlich ist. Die `summary` des multinomialen Modells liefert auch eine Warnung, welche auf das Auftreten des *Hauck-Donner Effekts* hinweist. Für Details und Erklärungen zum Hauck-Donner Effekt siehe Anhang A.6.

Das Auftreten des Hauck-Donner Effekts in einem Modell kann im Package `VGAM` mittels des Aufrufes `hdeff.vglm` überprüft werden.

```
> hdeff.vglm(mod.cd.mult)
```

(Intercept):1	(Intercept):2	(Intercept):3
FALSE	FALSE	TRUE
single.shot:1	single.shot:2	single.shot:3
FALSE	FALSE	FALSE
resection:1	resection:2	resection:3
FALSE	FALSE	FALSE
op.tecfree.nipple:1	op.tecfree.nipple:2	op.tecfree.nipple:3
FALSE	TRUE	TRUE
op.tecvertical:1	op.tecvertical:2	op.tecvertical:3
FALSE	FALSE	FALSE
op.teccaudal.pedicle:1	op.teccaudal.pedicle:2	op.teccaudal.pedicle:3
FALSE	FALSE	FALSE
revision:1	revision:2	revision:3
FALSE	FALSE	FALSE

Die Funktion `hdeff.vglm` testet dabei die Ableitung der Wald-Test-Statistik auf das Auftreten von negativen Werten mit zunehmendem Abstand des Parameters von der Nullhypothese. Mit dieser Ausgabe wird ersichtlich, wo der Hauck-Donner Effekt im multinomialen Modell auftritt.

Bis dato können mit Hilfe des Packages `VGAM` keine weiteren statistischen Analysen mittels `anova` oder `drop1` ausgeführt werden, da diese noch nicht implementiert wurden. Dieses ist laut Yee für die Zukunft noch geplant. Die Schätzungen des Modells können jedoch graphisch dargestellt werden. Dabei werden die geschätzten Effekte

durch Effektsterne abgebildet. Dieses erfolgt mittels des nachfolgenden Aufrufs in R.

```
> effectstars.vglm(mod.cd.mult, symmetric = FALSE, p.values = FALSE)
```

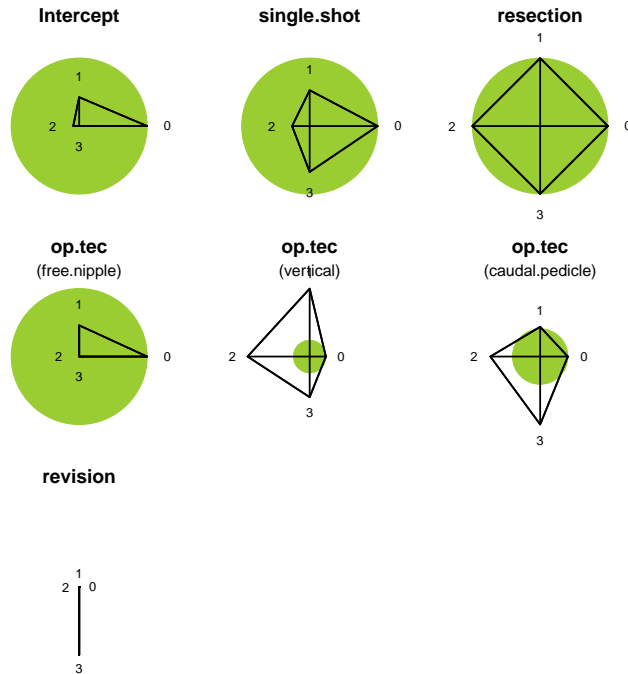


Abbildung 6.5: Effekt-Graphik für das multinomiale Modell der CD-Klassifikation $\{0, 1, 2, 3\}$

Die Abbildung 6.5 zeigt die Effekte der CD-Klassifikation in den vier Stufen $\{0, 1, 2, 3\}$. Es werden dabei die geschätzten Effekte aus dem Modell `mod.cd.mult` für jeden Faktor aufgetragen. Die Referenzklassifikation 0 wird in jedem Effektstern genau auf dem Radius des Kreises mit Wert $\exp(0) = 1$ abgebildet und stellt dadurch die Situation mit keinem Effekt dar. Die ersten vier Effektsterne veranschaulichen, dass sich die Effekte innerhalb des Kreises befinden und weisen somit auf die negativen Effekte der einzelnen Faktoren hin. Die restlichen Darstellungen befinden sich außerhalb des Kreises. Die Revision weist für die CD-Klassifikation 3 große Werte auf, wodurch die Darstellung des Einheitskreises verschwindet.

6.3.2 Analyse mit multinom

Die Response `cd.mult.ges` wird weiters mit Hilfe der Funktion `multinom` des Packages `nnet` modelliert, um einen Vergleich mit den Ergebnissen von zuvor zu erhalten. Das Modell `mod.cd.multinom` beinhaltet dabei die relevanten Faktoren, welche im binomialen Modell und im multinomialen Modell mittels `vglm` verwendet wurden.

```

> library(nnet)
> mod.cd.multinom <- multinom(cd.mult.ges ~ single.shot + resection + op.tec
                               + revision)

# weights: 32 (21 variable)
initial value 651.558350
iter 10 value 439.519701
iter 20 value 425.694101
iter 30 value 425.228351
final value 425.227630
converged

```

Die Anwendung der Funktion `multinom` ergibt ein konvergentes Verhalten und einen `final value` von 425.227630 nach 32 Iterationen mit $7 \cdot 3 = 21$ Parametern. Dieses Ergebnis stimmt im Betrag mit dem maximalen Log-Likelihood Wert der Schätzung des Modells mittels `vglm` überein.

Zuerst wird die Relevanz jedes einzelnen Faktors im Modell anhand einer `anova` überprüft. Exemplarisch wird dieses im Nachfolgenden mit dem Vergleichsmodell ohne Single-Shot-AB dargestellt. Mit Hilfe eines Likelihood-Ratio Tests wird überprüft, welches der beiden Modelle relevanter ist.

```

> anova(mod.cd.multinom, multinom(cd.mult.ges ~ resection + op.tec
                                   + revision), test = "Chisq")

```

Likelihood ratio tests of Multinomial Models

Response: cd.mult.ges

	Model	Resid. df	Resid. Dev	Test
1	resection + op.tec + revision	1392	861.9233	
2	single.shot + resection + op.tec + revision	1389	850.4553	1 vs 2
	Df LR stat.	Pr(Chi)		
1				
2	3	11.468	0.009446603	

Das Modell ohne Single-Shot-AB hat 3 Freiheitsgrade weniger, eine Deviance von 861.9233 und somit eine Differenz der Deviance zum Modell mit Single-Shot-AB von 11.468. Mittels eines Tests ergibt sich ein p-Wert von 0.009446603, welcher kleiner als 5% ist und somit die Relevanz des Faktors Single-Shot-AB im multinomialen Modell bestätigt. In der nachfolgenden Tabelle sind die Ergebnisse der `anova` für jeden einzelnen Faktor dargestellt, wobei das Modell `mod.cd.multinom` immer als Referenz verwendet wird.

Da alle p-Werte der einzelnen Faktoren kleiner als 5% sind, wird die Relevanz jedes Faktors für das Modell ersichtlich. Somit stellen die Faktoren Single-Shot-AB, Re-

	Df	Deviance	Differenz der Deviances	p-Wert
mod.cd.multinom	-	850.4553	-	-
-single.shot	3	861.9233	11.468	0.009446598
-resection	3	887.9373	37.482	3.638259e-08
-op.tec	9	875.0678	24.61254	2.158411e-05
-revision	3	962.8272	112.3719	3.387157e-24

Tabelle 6.8: Ergebnisse des Likelihood-Ratio Tests auf Relevanz der einzelnen Faktoren

sektionsmenge, Operationstechnik und Revision auch in diesem Fall relevante Größen des multinomialen Modells dar. Weiters wird noch der Faktor `smoke.yn` separat überprüft.

```
> anova(mod.cd.multinom, mod.cd.smoke, test="Chisq")
```

```
Likelihood ratio tests of Multinomial Models
```

```
Response: cd.mult.ges
```

	Model	Resid. df	Resid. Dev
1	single.shot + resection + op.tec + revision	1389	850.4553
2	single.shot + resection + op.tec + revision + smoke.yn	1386	845.3950

Test	Df	LR stat.	Pr(Chi)
1			
2 1 vs 2	3	5.060227	0.1674397

Die Anzahl der Freiheitsgrade erhöht sich um 3 bei Hinzufügen des Faktors Rauchen, während sich die Deviance um 5.060227 zum Modell `mod.cd.multinom` verringert. Es ergibt sich ein p-Wert von 0.1674 und somit keine Relevanz des Faktors `smoke.yn` für das multinomiale Modell. Die Modellschätzung wird somit im Folgenden anhand des Modells `mod.cd.multinom` durchgeführt.

```
> summary(mod.cd.multinom)
```

```
Call:
```

```
multinom(formula = cd.mult.ges ~ single.shot + resection + op.tec +
  revision)
```

```
Coefficients:
```

	(Intercept)	single.shot	resection	op.tecfree.nipple	op.tecvertical
1	-0.8566625	-0.6404588	0.002534843	-0.7780992	1.4355159
2	-2.4272029	-1.3505555	0.003212044	-22.1295890	1.3460018
3	-18.6753250	-0.3923885	0.002242216	-19.5929570	0.9209427

	op.teccaudal.pedicle	revision
1	0.07321664	0.8886812
2	0.58941871	0.8462034

3 0.89913673 18.7805645

Std. Errors:

	(Intercept)	single.shot	resection	op.tecfree.nipple	op.tecvertical
1	0.3610554	0.2872087	0.0004866886	8.926567e-01	0.3573824
2	0.5339334	0.4069343	0.0006916056	5.407206e-10	0.5948068
3	0.3433434	0.5998233	0.0010368020	4.956375e-09	0.5726018
	op.teccaudal.pedicle	revision			
1	0.4376629	0.4467879			
2	0.6016376	0.6691212			
3	0.9868799	0.3433436			

Residual Deviance: 850.4553

AIC: 892.4553

Die Funktion `multinom` fitted multinomiale Log-Lineare-Modelle mit Hilfe von neuronalen Netzen. Dabei ergibt sich ein Ergebnis, welches sich mit den Schätzungen von `vglm` vergleichen lässt. Die Ergebnisse der einzelnen Funktionen stimmen weitgehend überein, große Unterschiede sind in den Zeilen erkennbar, in welchen im Modell zuvor der Hauck-Donner Effekt auftrat.

6.4 Ordinale Modelle

Im folgenden Abschnitt wird die Response CD-Klassifikation `ordinal` betrachtet und mit Hilfe der Methoden aus Kapitel 4 analysiert. Die Response muss dabei für die Anwendung dieser Methoden vom Typ `ordered` sein, um Level-Attribute zu haben und somit die Sortierung der Variablen anzunehmen. Diese Response erhält die Bezeichnung `cd.ord` und wird mit Hilfe von kumulativen und sequentiellen Modellen analysiert. Analog zu den multinomialen Modellen werden die Klassifikationen $\{0, 1, 2, 3\}$ mit den Kategorien $\{1, 2, 3, 4\}$ identifiziert.

6.4.1 Kumulative Modelle

Zuerst wird eine Modellschätzung für die ordinale Response `cd.ord` in den Stufen $\{0, 1, 2, 3\}$ mittels des kumulativen Modells durchgeführt.

Analyse mit vglm

Für diese Analyse wird das Argument der Funktion `vglm` als `family = cumulative` gewählt. Zuerst wird die Relevanz der einzelnen Faktoren für das kumulative Modell überprüft. Dabei werden die relevanten Faktoren von den Modellen zuvor durch `mod.kum` zusammengefasst.

```
> cd.ord <- ordered(cd.mult.ges)
> mod.kum<-vglm(cd.ord ~ single.shot + resection + op.tec + revision,
               family = cumulative(parallel=TRUE))

> lrtest(mod.kum, vglm(cd.ord ~ resection + op.tec + revision,
                      family = cumulative(parallel=TRUE)))
```

Likelihood ratio test

```
Model 1: cd.ord ~ single.shot + resection + op.tec + revision
Model 2: cd.ord ~ resection + op.tec + revision
```

```
  #Df  LogLik Df  Chisq Pr(>Chisq)
1 1401 -457.66
2 1402 -461.40  1  7.4754  0.006255 **
```

Mit Hilfe des Likelihood-Ratio Tests kann die Signifikanz des Single-Shot-AB im kumulativen Modell festgestellt werden. Bei einer Deviance Differenz von 7.4754 und einem p-Wert von 0.006255 kann der Faktor Single-Shot-AB nicht aus dem kumulativen Modell entfernt werden. Die Ergebnisse der weiteren relevanten Faktoren sind der nachfolgenden Tabelle zu entnehmen.

	Df	LogLik	χ^2	p-Wert
<code>mod.kum</code>	-	-457.66	-	-
<code>-single.shot</code>	1	-461.40	7.4754	0.006255
<code>-resection</code>	1	-474.82	34.306	4.711e-09
<code>-op.tec</code>	3	-466.55	17.778	0.0004888
<code>-revision</code>	1	-488.77	62.219	3.073e-15

Tabelle 6.9: Auflistung der Ergebnisse des Likelihood-Ratio Tests auf Relevanz der einzelnen Faktoren

Es zeigt sich, dass die p-Werte aller Faktoren eine Signifikanz für das Modell aufweisen. Zusätzlich wird auch der Faktor Rauchen überprüft. Dazu wird das Modell `mod.cd.ord.kum` verwendet, welches zu den restlichen relevanten Faktoren auch das Rauchen hinzufügt.

```
> mod.cd.ord.kum <- vglm(cd.ord ~ single.shot + resection + op.tec
+ revision + smoke.yn,
family = cumulative(parallel = TRUE))
```

```
> lrtest(mod.cd.ord.kum, vglm(cd.ord ~ single.shot + resection + op.tec + revision,
family = cumulative(parallel=TRUE)))
```

Likelihood ratio test

Model 1: cd.ord ~ single.shot + resection + op.tec + revision + smoke.yn
 Model 2: cd.ord ~ single.shot + resection + op.tec + revision

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	1400	-455.06			
2	1401	-457.66	1	5.2079	0.02248 *

Durch den Likelihood-Ratio Test wird die Relevanz des Faktors `smoke.yn` für das kumulative Modell ersichtlich. Für die ordinale Analyse wird auf Grund dieser Ergebnisse das Modell `mod.cd.ord.kum` mit den Prädiktoren Single-Shot-AB, Resektionsmenge, Operationstechnik, Revision und Rauchen verwendet.

Für die kumulative Analyse wird der Theorie entsprechend die Annahme der Gleichheit der Effekte benötigt. Diese Annahme kann auch mit Hilfe eines Likelihood-Ratio Tests überprüft werden. Zu diesem Zweck wird das Modell mit dem zusätzlichen Argument `family = cumulative(parallel = TRUE)` erstellt, welches eine Gleichheit der Effekte annimmt und anschließend mit allen möglichen Modellen ohne Gleichheit der Effekte verglichen, exemplarisch ist der Test für den Effekt `single.shot` gegeben.

```
> lrtest(mod.cd.ord.kum, vglm(cd.ord ~ single.shot + resection + op.tec + revision
+ smoke.yn, family = cumulative(parallel=
FALSE~single.shot)))
```

Likelihood ratio test

Model 1: cd.ord ~ single.shot + resection + op.tec + revision + smoke.yn
 Model 2: cd.ord ~ single.shot + resection + op.tec + revision + smoke.yn

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	1400	-455.06			
2	1398	-453.28	-2	3.5618	0.1685

Da der p-Wert größer als 5% ist, wird die Annahme der Gleichheit der Effekte angenommen. Die Ergebnisse der restlichen Prädiktoren sind in der nachfolgenden Tabelle dargestellt.

	Df	LogLik	χ^2	p-Wert
mod.cd.ord.kum	-	-455.06	-	-
~single.shot	2	-453.28	3.5618	0.1685
~resection	2	-453.15	3.8164	0.1484
~op.tec	6	-451.08	7.9526	0.2416
~smoke.yn	2	-454.82	0.4745	0.7888

Tabelle 6.10: Auflistung der Ergebnisse des Likelihood-Ratio Tests auf parallele Effekte der Faktoren

Die Revisionsoperation kann dabei nicht getestet werden, der Grund dafür ist, dass die Stufe 3 nur Daten von Patientinnen mit einer solchen Operation enthält. Dieses wird auch in der Tabelle 6.3 verdeutlicht. Aufgrund der resultierenden p-Werte der Tests, welche alle größer als 5% sind, wird die Annahme der Gleichheit der Effekte für das kumulative Modell nicht verworfen. Daher wird von nun an das Argument `parallel = TRUE` angegeben. Dadurch wird ein proportionales Odds Modell spezifiziert. Das kumulative Modell beinhaltet somit die linearen Prädiktoren `single.shot`, `resection`, `op.tec`, `revision` und `smoke.yn` und wird als `mod.cd.ord.kum` bezeichnet.

Analog zu den multinomialen Modellen wird die Logit-Linkfunktion verwendet, da keine Linkfunktion explizit angegeben wird. Bei der Verwendung dieser Linkfunktion werden zusätzlich zu den Effekten die Exponenten dieser ausgegeben. Der Theorie der kumulativen Modelle entsprechend existieren für vier Klassifikationen drei lineare Prädiktoren, welche mit der Logit-Linkfunktion die Gestalt

$$\text{logit}(P(Y \leq j|\mathbf{x})), \quad j = 1, 2, 3$$

haben. Aus der Annahme der Gleichheit der Effekte und der Gleichung (4.4) folgt, dass für jeden Prädiktor eine Schätzung des Intercepts berechnet wird, während die restlichen Effekte einen Koeffizienten für alle Prädiktoren aufweisen. Die Intercepts werden in der Ausgabe der Funktion `vglm` entsprechend der Aufzählung nummeriert.

```
> summary(mod.cd.ord.kum)
```

```
Call:
```

```
vglm(formula = cd.ord ~ single.shot + resection + op.tec + revision +
      smoke.yn, family = cumulative(parallel = TRUE))
```

```
Pearson residuals:
```

```

              Min          1Q      Median          3Q          Max
logit(P[Y<=1]) -1.611 -0.80154 -0.3193  0.9068  8.739
logit(P[Y<=2]) -5.042  0.14873  0.2665  0.4040  2.091
```

```
logit(P[Y<=3]) -3.116  0.07755  0.1005 0.1451 1.042
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept):1	0.6570244	0.3080508	2.133	0.032937	*
(Intercept):2	3.4007580	0.3564710	9.540	< 2e-16	***
(Intercept):3	4.7948622	0.4113924	11.655	< 2e-16	***
single.shot	0.6341913	0.2323534	2.729	0.006344	**
resection	-0.0021713	0.0003661	-5.931	3.01e-09	***
op.tecfree.nipple	1.7085854	0.7340059	2.328	0.019925	*
op.tecvertical	-0.9944524	0.2987976	-3.328	0.000874	***
op.teccaudal.pedicle	-0.5537497	0.3541975	-1.563	0.117960	
revision	-2.5193466	0.2970340	-8.482	< 2e-16	***
smoke.yn	-0.4380727	0.1905721	-2.299	0.021521	*

Number of linear predictors: 3

Names of linear predictors: logit(P[Y<=1]), logit(P[Y<=2]),
logit(P[Y<=3])

Residual deviance: 910.1188 on 1400 degrees of freedom

Log-likelihood: -455.0594 on 1400 degrees of freedom

Number of iterations: 8

No Hauck-Donner effect found in any of the estimates

Exponentiated coefficients:

single.shot	resection	op.tecfree.nipple
1.88549670	0.99783109	5.52114596
revision	smoke.yn	op.tecvertical
0.08051219	0.64527886	0.36992594
op.teccaudal.pedicle		
0.57479045		

Für dieses Modell werden die Intercepts b_1, b_2, b_3 und der Vektor der Effekte β geschätzt. Es ergibt sich die Darstellung

$$\widehat{\text{logit}}(P(Y \leq j|\mathbf{x})) = \hat{b}_j + 0.634 \cdot (\text{single.shot} = 1) - 0.002 \cdot \text{resection} \\ + 1.709 \cdot (\text{free.nipple} = 1) \\ - 0.995 \cdot (\text{vertical} = 1) \\ - 0.554 \cdot (\text{caudal.pedicle} = 1) \\ - 2.519 \cdot (\text{revision} = 1) - 0.438 \cdot (\text{smoke.yn} = 1)$$

mit den geschätzten Intercepts

$$\hat{b}_1 = 0.657 \quad \hat{b}_2 = 3.401 \quad \hat{b}_3 = 4.795.$$

Dem kumulativen Modell folgend gilt die Bedingung $\hat{b}_1 < \hat{b}_2 < \hat{b}_3$. Analog zum multinomialen, ungeordneten Fall zeigt sich, dass die Revision und die Resektionsmenge mit sehr kleinen p-Werten eine große Signifikanz aufweisen. Auch der Single-Shot-AB ist im kumulativen Modell signifikant. Das Rauchverhalten weist zum ersten Mal eine deutliche Signifikanz auf und hat im Modell einen p-Wert von 0.021521.

Mit dem folgenden Funktionsaufruf wird das Modell auf den Hauck-Donner Effekt getestet.

```
> hdeff.vglm(mod.cd.ord.kum)

      (Intercept):1      (Intercept):2      (Intercept):3
      FALSE           FALSE           FALSE
op.tecfree.nipple  op.tecvertical  op.teccaudal.pedicle
      FALSE           FALSE           FALSE
single.shot        resection        revision
      FALSE           FALSE           FALSE
smoke.yn
      FALSE
```

Dieses bestätigt, dass im kumulativen Modell kein Hauck-Donner Effekt auftritt.

In der Tabelle 6.11 sind die p-Werte der Intercepts und der Koeffizienten für das kumulative Modell mit unterschiedlichen Linkfunktionen angegeben.

Faktoren	Logit Modell	Probit Modell	Cloglog Modell
single.shot	0.00634	0.00770	0.03605
resection	3.01e-09	2.49e-08	1.57e-08
op.tec	0.00037	0.00176	0.00237
revision	< 2e-16	< 2e-16	< 2e-16
smoke.yn	0.02152	0.02226	0.01157

Tabelle 6.11: Vergleich der einzelnen p-Werte der Koeffizienten unter Verwendung verschiedener Linkfunktionen beim kumulativen Modell

Die Unterschiede der einzelnen Modelle ergeben sich durch die verschiedenen Linkfunktionen, dabei ist die Ähnlichkeit zwischen dem Logit und dem Probit Modell gut erkennbar.

Die Abbildung 6.6 zeigt links die Graphik des Effektsterns für das kumulative Modell. Da im kumulativen Modell keine Gruppierung vorhanden ist, werden nur die Effekte des Intercepts dargestellt. Der Einheitskreis ist aufgrund der großen Werte des Exponenten der Intercepts nicht erkennbar. Mit dem folgenden Aufruf in R

```
> effectstars.vglm(mod.cd.ord.kum, plot.parallel = TRUE)
```

werden alle Effekte des Modells dargestellt. Aufgrund der Annahme der Gleichheit der Effekte haben die Effektsterne aller Parameter mit Ausnahme des Intercepts den gleichen Abstand für alle Prädiktoren. Dieses ist in der Abbildung 6.6 auf der rechten Seite dargestellt.

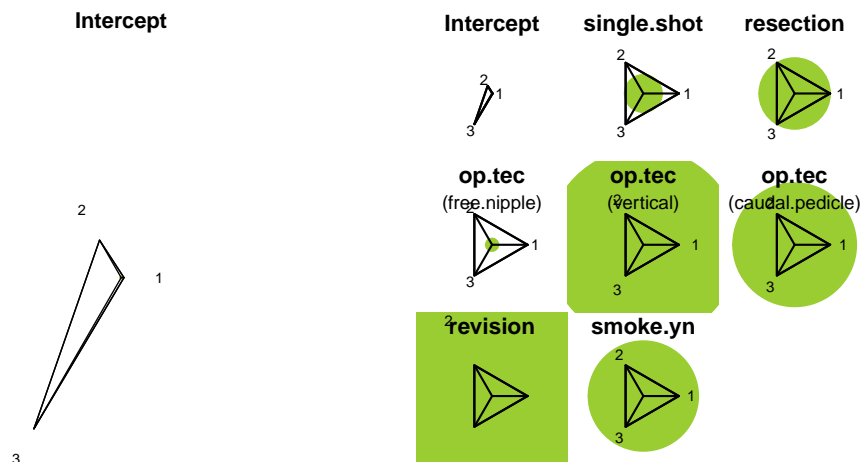


Abbildung 6.6: Effekt-Graphik von \hat{b}_1 , \hat{b}_2 und \hat{b}_3 des kumulativen Modells (links) und Effekt-Graphik für das kumulative Modell mit allen Faktoren (rechts)

Im Nachfolgenden werden Modellabbildungen betrachtet, welche abhängig von der Resektionsmenge die Wahrscheinlichkeit der verschiedenen Risikofaktoren mittels des kumulativen Modells abbilden. In der Tabelle 6.12 sind die verwendeten Kombinationen mit der dazugehörigen Farbe dargestellt.

Farben	op.tec	single.shot	revision	smoke.yn
blau	cranial.pedicle	1	0	0
rot	cranial.pedicle	0	1	0
orange	cranial.pedicle	0	0	1
schwarz	cranial.pedicle	0	0	0

Tabelle 6.12: Kombinationen der Faktoren und die dazugehörigen Farben in den Abbildungen 6.7 und 6.8

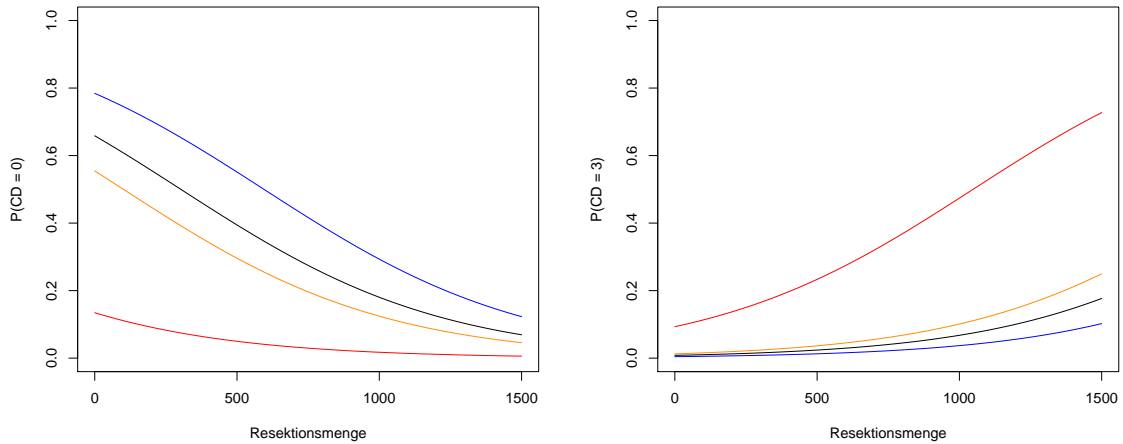


Abbildung 6.7: Die kumulative Wahrscheinlichkeit $P(\text{CD} = 0)$ (links) und $P(\text{CD} = 3)$ (rechts) für verschiedene Faktorkombinationen in Abhängigkeit der Resektionsmenge

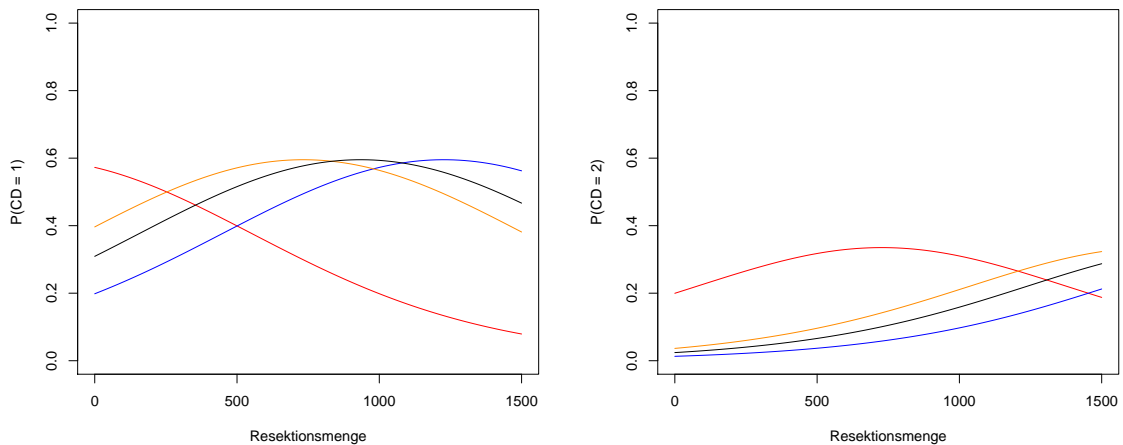


Abbildung 6.8: Die kumulative Wahrscheinlichkeit $P(\text{CD} = 1)$ (links) und $P(\text{CD} = 2)$ (rechts) für verschiedene Faktorkombinationen in Abhängigkeit der Resektionsmenge

Die Abbildung 6.7 (links) zeigt, dass die Wahrscheinlichkeit für die Klassifikation 0 mit steigender Resektionsmenge für alle vier Varianten sinkt. Weiters zeigt sich ein deutlicher Unterschied bei einer Revisionsoperation. Für den Faktor Single-Shot-AB ist die Wahrscheinlichkeit für eine Klassifikation 0 am größten verglichen mit den anderen Faktorkombinationen. Die schwarze Linie verdeutlicht die Wahrscheinlichkeit ohne Einfluss von weiteren Faktoren. Im Vergleich dazu ist in der rechten Abbildung 6.7 die

Wahrscheinlichkeit für die Klassifikation 3 dargestellt. Die Wahrscheinlichkeit nimmt in allen vier Varianten mit steigender Resektionsmenge zu, wobei die Wahrscheinlichkeit bei Vorliegen einer Revision deutlich größer ist und eine weitaus stärkere Steigung aufweist.

In der Abbildung 6.8 (links) ist zu sehen, dass die Wahrscheinlichkeit für eine Klassifikation 1 für die Faktoren Rauchen und Single-Shot-AB mit steigender Resektionsmenge zu Beginn etwas zunimmt, ab ca. 1000g Resektionsmenge jedoch wieder abnimmt. Liegt eine Revisionsoperation vor, so nimmt die Wahrscheinlichkeit dieser Klassifikation mit steigender Resektionsmenge ab. Die Wahrscheinlichkeit für eine Klassifikation 2 ist rechts in der Abbildung 6.8 dargestellt. Die Kurve für den Single-Shot-AB und das Rauchen weisen ein ähnliches Verhalten auf. Die Wahrscheinlichkeit für eine Klassifikation 2 ohne eine Revision nimmt mit steigender Resektionsmenge zu. Bei einer Revision ist die Wahrscheinlichkeit deutlich höher, steigt zu Beginn an, nimmt aber ab einer Resektionsmenge von ca. 1000g wieder ab. Die schwarzen Linien beschreiben die Wahrscheinlichkeiten bei Verwendung der Operationstechnik `cranial.pedicle` ohne den Einfluss weiterer Faktoren.

Analyse mit `polr`

Für einen Vergleich wird das kumulative Logit Modell ebenfalls mit der Funktion `polr` aus dem Package `MASS` analysiert. Dieses gibt die geschätzten Punktwahrscheinlichkeiten für die einzelnen Ergebnisse an. Erneut werden die Levels der CD-Klassifikation benötigt, weshalb auch hier wiederum die Response `cd.ord <- ordered(cd.mult.ges)` verwendet wird.

```
> library(MASS)
> mod.cd.ord.kum2 <- polr(cd.ord ~ single.shot + resection
                          + op.tec + revision + smoke.yn)
> summary(mod.cd.ord.kum2)
```

Call:

```
polr(formula = cd.ord ~ single.shot + resection + op.tec + revision +
      smoke.yn)
```

Coefficients:

	Value	Std. Error	t value
single.shot	-0.634190	0.2306722	-2.749
resection	0.002171	0.0003791	5.728
op.tecfree.nipple	-1.708621	0.6915774	-2.471
op.tecvertical	0.994438	0.2959229	3.360
op.teccaudal.pedicle	0.553750	0.3730575	1.484
revision	2.519372	0.3281250	7.678
smoke.yn	0.438068	0.1923100	2.278

Intercepts:

	Value	Std. Error	t value
0 1	0.6570	0.3127	2.1012
1 2	3.4008	0.3594	9.4629
2 3	4.7949	0.4149	11.5572

Residual Deviance: 910.1188

AIC: 930.1188

Es zeigt sich, dass die Funktionen `vglm` und `polr` mit Ausnahme des Vorzeichens der Effekte die selben Ergebnisse liefern. Die unterschiedlichen Vorzeichen ergeben sich durch die Definitionen der beiden Funktionen (5.3).

Im Folgenden werden die p-Werte dieses Modells betrachtet. Dabei wird mittels eines Koeffiziententests die Relevanz überprüft.

```
> library(AER)
> coeftest(mod.cd.ord.kum2)
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
single.shot	-0.63467331	0.23094782	-2.7481	0.0059937	**
resection	0.00211155	0.00035867	5.8872	3.927e-09	***
op.tecfree.nipple	-1.70158917	0.68888981	-2.4700	0.0135096	*
op.tecvertical	0.97217865	0.29402130	3.3065	0.0009447	***
op.teccaudal.pedicle	0.56625529	0.37341563	1.5164	0.1294129	
op.tecother	0.99674951	0.49140221	2.0284	0.0425217	*
revision	2.56831772	0.32714578	7.8507	4.138e-15	***
smoke.yn	0.49988687	0.18923632	2.6416	0.0082515	**
0 1	0.64194088	0.30424424	2.1100	0.0348625	*
1 2	3.39065943	0.34988940	9.6907	< 2.2e-16	***
2 3	4.84335155	0.40969229	11.8219	< 2.2e-16	***

Auch die p-Werte verdeutlichen die Signifikanzen der einzelnen Prädiktoren und Stufen im ordinalen Modell. Dabei werden zuerst die Prädiktoren ausgegeben und danach die Stufen der Klassifikation in Relation dargestellt.

6.4.2 Sequentielle Modelle

Die Response CD-Klassifikation unterliegt einer sukzessiven Ordnung, daher kann die CD-Klassifikation auch mit Hilfe der sequentiellen Modelle analysiert werden. Für die sequentiellen Modelle wird analog zu den kumulativen Modellen die geordnete,

vierstufige Response `cd.ord` verwendet.

Analyse mit `vglm`

Das Modell wird mit der Funktion `vglm` geschätzt, wobei für das sequentielle Modell das Argument `family = sratio` gewählt sein muss. Wird keine Linkfunktion explizit angegeben, wird die Logit-Linkfunktion verwendet. Äquivalent zu den kumulativen Modellen folgt mit der Gleichung (4.21), dass für die vier Klassifikationen drei lineare Prädiktoren existieren. Mit der Logit-Linkfunktion besitzen die Prädiktoren die Darstellung in der Ausgabe

$$\text{logit}(P(Y = j | Y \geq j | \mathbf{x})), \quad j = 1, 2, 3.$$

```
> mod.cd.ord.seq <- vglm(cd.ord ~ single.shot + resection + op.tec + revision
+ smoke.yn, family = sratio(parallel = TRUE))
> summary(mod.cd.ord.seq)
```

Call:

```
vglm(formula = cd.ord ~ single.shot + resection + op.tec + revision
+ smoke.yn, family = sratio(parallel = T))
```

Pearson residuals:

	Min	1Q	Median	3Q	Max
logit(P[Y=1 Y>=1])	-1.393	-7.840e-01	-4.658e-01	9.576e-01	7.310
logit(P[Y=2 Y>=2])	-5.029	-8.178e-06	2.213e-05	5.118e-01	2.150
logit(P[Y=3 Y>=3])	-2.195	1.655e-06	4.515e-06	1.349e-05	1.869

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept):1	0.5553066	0.2788888	1.991	0.04647 *
(Intercept):2	2.8782107	0.3425661	8.402	< 2e-16 ***
(Intercept):3	2.9112028	0.4565604	6.376	1.81e-10 ***
single.shot	0.4777884	0.2088184	2.288	0.02213 *
resection	-0.0018182	0.0003274	-5.553	2.81e-08 ***
op.tecfree.nipple	1.5767336	0.6624522	2.380	0.01731 *
op.tecvertical	-0.7709245	0.2697743	-2.858	0.00427 **
op.teccaudal.pedicle	-0.4734248	0.3157912	-1.499	0.13383
revision	-2.4065378	0.2683761	-8.967	< 2e-16 ***
smoke.yn	-0.3934961	0.1726289	-2.279	0.02264 *

Number of linear predictors: 3

Names of linear predictors: `logit(P[Y=1|Y>=1])`, `logit(P[Y=2|Y>=2])`,
`logit(P[Y=3|Y>=3])`

Residual deviance: 890.5685 on 1400 degrees of freedom

Log-likelihood: -445.2842 on 1400 degrees of freedom

Number of iterations: 7

No Hauck-Donner effect found in any of the estimates

Analog zum kumulativen Modell werden auch im sequentiellen Modell die Intercepts b_1, b_2, b_3 und der Vektor der Effekte β geschätzt. Damit ergibt sich die Darstellung

$$\begin{aligned} \widehat{\text{logit}}(P(Y = j|Y \geq j, \mathbf{x})) = & \hat{b}_j + 0.478 \cdot (\text{single.shot} = 1) - 0.002 \cdot \text{resection} \\ & + 1.577 \cdot (\text{free.nipple} = 1) \\ & - 0.771 \cdot (\text{vertical} = 1) \\ & - 0.473 \cdot (\text{caudal.pedicle} = 1) \\ & - 2.407 \cdot (\text{revision} = 1) - 0.394 \cdot (\text{smoke.yn} = 1) \end{aligned}$$

mit den geschätzten Intercepts

$$\hat{b}_1 = 0.555 \quad \hat{b}_2 = 2.878 \quad \hat{b}_3 = 2.911.$$

Es zeigt sich, dass die Revision und die Resektionsmenge für das sequentielle Modell von großer Signifikanz sind. Das Rauchverhalten sowie der Single-Shot-AB haben p-Werte kleiner 5% und sind somit signifikant.

In der Tabelle 6.13 sind die p-Werte der Koeffizienten des sequentiellen Modells für verschiedene Linkfunktionen angegeben.

Faktoren	Logit Modell	Probit Modell	Cloglog Modell
single.shot	0.02213	0.01922	0.03604
resection	2.81e-08	7.67e-08	1.58e-08
op.tec	0.00193	0.00407	0.00237
revision	< 2e-16	< 2e-16	< 2e-16
smoke.yn	0.02264	0.02091	0.01158

Tabelle 6.13: Vergleich der einzelnen p-Werte der Koeffizienten unter Verwendung verschiedener Linkfunktionen beim sequentiellen Modell

Diese Tabelle verdeutlicht die Unterschiede der p-Werte bei der Verwendung verschiedener Linkfunktionen. Dabei weisen diese ein ähnliches Verhalten wie im kumulativen Modell auf.

Analog zum kumulativen Modell werden in der Abbildung 6.9 (links) die Intercepts des sequentiellen Modells in einem Effektestern dargestellt. In der Abbildung ist der

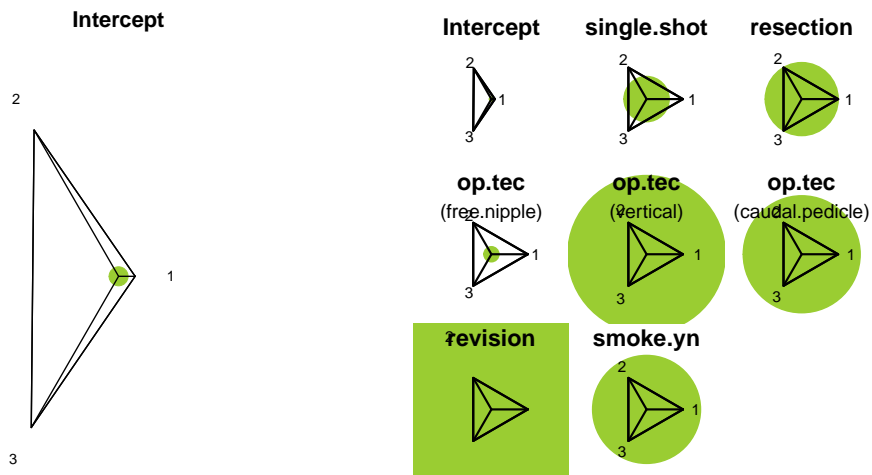


Abbildung 6.9: Effekt-Graphik von \hat{b}_1 , \hat{b}_2 und \hat{b}_3 des sequentiellen Modells (links) und Effekt-Graphik für das sequentielle Modell mit allen Faktoren (rechts)

grüne Kreis mit Radius 1 erkennbar, wobei sich die Effekte des Intercepts deutlich außerhalb befinden. Anschließend können alle Effekte dargestellt werden. Analog zum kumulativen Modell sind alle Effekte mit Ausnahme des Intercepts für alle Prädiktoren gleich, wie in der rechten Abbildung verdeutlicht. Werden die Wahrscheinlichkeiten der einzelnen Stufen im sequentiellen Modell dargestellt, so ergeben sich der Definition des kumulativen und sequentiellen Modells entsprechend die gleichen Abbildungen. Die Wahrscheinlichkeiten für die Klassifikationen entsprechen daher im sequentiellen Modell ebenfalls den Abbildungen 6.7 und 6.8.

7 Resümee

In diesem Kapitel werden die Ergebnisse der statistischen Analysen dieser Arbeit kurz zusammengefasst. Diese Arbeit befasst sich mit einem Datensatz von Patientinnen, welche sich an der Medizinischen Universität Graz einer Mammareduktion unterzogen haben. Dabei wurde in der statistischen Modellierung und den Auswertungen die Clavien-Dindo-Klassifikation betrachtet und der Einfluss verschiedener erklärender Faktoren auf die Clavien-Dindo-Klassifikation (CD-Klassifikation) überprüft. Die vorliegenden Daten beziehen sich auf die Klassifikationsgrade 0, 1, 2, 3A und 3B. Von insgesamt 470 Patientinnen wurde bei 178 die Klassifikationsstufe 0, bei 226 Patientinnen der Grad 1, bei 43 Patientinnen der Grad 2 und bei nur 2 Patientinnen der Grad 3A und bei 21 der Grad 3B nachgewiesen. Durch die geringe Datenanzahl wurden die Stufen 3A und 3B zur Gruppe 3 zusammengefasst. Die CD-Klassifikation entspricht somit mehrstufigen Daten, welche einer Ordnungsstruktur unterliegen und als ordinale Daten bezeichnet werden. In der Arbeit wird die Klassifikation mit Hilfe verschiedener Betrachtungsmöglichkeiten und statistischer Methoden analysiert und die Ergebnisse im Folgenden kurz zusammengefasst.

Binomiales Modell

In dieser Modellanalyse wurden die Patientinnen in zwei Gruppen eingeteilt und zwar in komplikationslose Operation (0) und Operation mit Komplikationen (1). Dabei ergab sich bei der Analyse ein Modell mit relevanten Einflussgrößen, welches die Prädiktoren Single-Shot-AB, Resektionsmenge, Operationstechnik und Revisionsoperation enthält. In diesem Modell verändert sich die Wahrscheinlichkeit für das Eintreten einer CD-Klassifikation bei der Gabe eines Single-Shot-AB mit einem p-Wert von 0.0077. Die Resektionsmenge sowie eine Revisionsoperation beeinflussen die CD-Klassifikation stark, bestätigt wird dieses durch die signifikanten p-Werte von $4.15 \cdot 10^{-8}$ und 0.000139. Auch die Operationstechnik stellt im binären Modell eine signifikante Einflussgröße dar. Weiters wurde diese Analyse graphisch dargestellt, wobei die Wahrscheinlichkeit für das Eintreten einer Klassifikation in Abhängigkeit der Resektionsmenge abgebildet wurde.

Multinomiales Modell

Diese Modellart verwendet alle auftretenden CD-Klassifikationsstufen $\{0, 1, 2, 3\}$, wobei die 0 komplikationslose Operationen beschreibt. Dabei wird die natürliche Ordnung der Ausprägungen zuerst noch ignoriert und auf Grund dessen die Response Variable nur als nominal skaliert betrachtet. Diese Analyse wurde mit Hilfe von zwei verschiedenen Herangehensweisen durchgeführt, wobei beide die selben Ergebnisse liefern. Für das multinomiale Modell ergaben sich die selben relevanten Prädiktoren wie im binomialen Modell mit Single-Shot-AB, Resektionsmenge, Operationstechnik und Revisionsoperation. Dabei wurde für jeden Prädiktor jede Stufe separat ausgewertet und betrachtet. Die graphische Darstellung wurde mit Hilfe von Effektsternen durchgeführt. Diese stellen für jeden Prädiktor die Effekte der einzelnen Stufen relativ zu den Werten dar.

Ordinales Modell

Die ordinalen Modelle verwenden die CD-Klassifikation mit ihrer Ordnungsstruktur. Für die Analyse gibt es zwei zusätzliche Unterscheidungen der Modellklassen, die kumulativen und die sequentiellen Modelle. Die kumulativen Modelle sind für ordinale Daten sehr verbreitet. Die sequentiellen Modelle setzen eine sukzessiv steigende Ordnung bei der Response Variable voraus, wodurch es zu Abhängigkeiten zwischen den Stufen kommt. Für die CD-Klassifikation werden wiederum die Stufen $\{0, 1, 2, 3\}$ betrachtet.

Im kumulativen Modell zeigt sich, dass die Prädiktoren der zuvor untersuchten multinomialen Modelle auch in diesem Modell signifikant sind. Der Prädiktor Single-Shot-AB weist mit einem p-Wert von 0.006344 eine hohe Signifikanz für das ordinale Modell im kumulativen Fall auf. Sehr signifikant sind auch die beiden Größen Resektionsmenge und Revisionsoperation mit sehr kleinen p-Werten. Die Operationstechnik ist ebenso von Bedeutung für das Modell. Weiters ist im kumulativen Modell auch erstmals das Rauchen eine relevante Einflussgröße, welche einen p-Wert von 0.021 besitzt.

Im sequentiellen Fall sind auch die fünf Prädiktoren Single-Shot-AB, Resektionsmenge, Revisionsoperation sowie Operationstechnik und Rauchen die relevanten Einflussgrößen. Durch die unterschiedliche Herangehensweise ergeben sich unterschiedliche p-Werte im Vergleich zum kumulativen Modell. Die Resektionsmenge und die Revisionsoperation weisen auch im sequentiellen Modell sehr geringe p-Werte und somit eine sehr große Signifikanz auf. Der Single-Shot-AB hat einen p-Wert von 0.02213 und das Rauchen weist eine Signifikanz mit 0.02264 auf.

A Anhang

A.1 Iteratively Reweighted Least Squares (IRLS) Methode

Die *iteratively reweighted least squares* (IRLS) Methode ist eine numerische Methode zur Approximation der Lösung der Schätzgleichung

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

Ausgehend von einer gewählten Startlösung $\hat{\boldsymbol{\beta}}^{(0)}$ wird die Scorefunktion $\mathbf{s}(\boldsymbol{\beta})$ mit Hilfe der Taylor-Reihe linearisiert

$$\mathbf{s}(\boldsymbol{\beta}) \approx \mathbf{s}(\hat{\boldsymbol{\beta}}^{(0)}) + \frac{\partial \mathbf{s}(\hat{\boldsymbol{\beta}}^{(0)})}{\partial \boldsymbol{\beta}} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}^{(0)}).$$

Wird nun anstelle der Nullstelle der Schätzgleichung die Nullstelle der linearisierten Schätzgleichung gesucht, so folgt

$$\hat{\boldsymbol{\beta}}^{(1)} = \hat{\boldsymbol{\beta}}^{(0)} - \left(\frac{\mathbf{s}(\hat{\boldsymbol{\beta}}^{(0)})}{\partial \boldsymbol{\beta}} \right)^{-1} \mathbf{s}(\hat{\boldsymbol{\beta}}^{(0)}).$$

Durch Iteration ergibt sich für den ψ -ten Schritt die Vorschrift

$$\hat{\boldsymbol{\beta}}^{(\psi+1)} = \hat{\boldsymbol{\beta}}^{(\psi)} - \left(\frac{\mathbf{s}(\hat{\boldsymbol{\beta}}^{(\psi)})}{\partial \boldsymbol{\beta}} \right)^{-1} \mathbf{s}(\hat{\boldsymbol{\beta}}^{(\psi)}), \quad \psi = 0, 1, \dots \quad (\text{A.1})$$

Dieses Verfahren entspricht in seiner Form dem *Newton-* oder auch *Newton-Raphson-*Verfahren zur Berechnung von Nullstellen. Aufgrund der Definition der Scorefunktion mit $\mathbf{s}(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta}|\mathbf{y})}{\partial \boldsymbol{\beta}}$ gilt für die Ableitung der Scorefunktion

$$\frac{\partial \mathbf{s}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{\partial^2 l(\boldsymbol{\beta}|\mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = \mathbf{H}(l(\boldsymbol{\beta}|\mathbf{y}))$$

und daher hat die Vorschrift (A.1) die äquivalente Darstellung

$$\hat{\beta}^{(\psi+1)} = \hat{\beta}^{(\psi)} - \left[\mathbf{H} \left(l \left(\hat{\beta}^{(\psi)} | \mathbf{y} \right) \right) \right]^{-1} \mathbf{s} \left(\hat{\beta}^{(\psi)} \right), \quad \psi = 0, 1, \dots$$

Wird hierbei anstelle der zuvor angewandten Hessematrix $\mathbf{H} \left(l \left(\hat{\beta} | \mathbf{y} \right) \right)$ die Fisher-Matrix $\mathbf{F} \left(\hat{\beta} \right) = \mathbf{E} \left(-\mathbf{H} \left(l \left(\hat{\beta} | \mathbf{y} \right) \right) \right)$ verwendet, so ergibt sich die Vorschrift

$$\hat{\beta}^{(\psi+1)} = \hat{\beta}^{(\psi)} + \left[\mathbf{F} \left(\hat{\beta}^{(\psi)} \right) \right]^{-1} \mathbf{s} \left(\hat{\beta}^{(\psi)} \right), \quad \psi = 0, 1, \dots \quad (\text{A.2})$$

Diese Vorschrift kann in ihrer Darstellung weiter modifiziert werden. Mit Hilfe der Äquivalenzen

$$\begin{aligned} \mathbf{s}(\beta) &= \mathbf{X}^\top \mathbf{W}(\beta) [\mathbf{D}(\beta)]^{-1} (\mathbf{y} - \boldsymbol{\mu}) \\ \mathbf{F}(\beta) &= \mathbf{X}^\top \mathbf{W}(\beta) \mathbf{X} \end{aligned}$$

folgt die Form

$$\hat{\beta}^{(\psi+1)} = \hat{\beta}^{(\psi)} + \left(\mathbf{X}^\top \mathbf{W} \left(\hat{\beta}^{(\psi)} \right) \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{W} \left(\hat{\beta}^{(\psi)} \right) \left[\mathbf{D} \left(\hat{\beta}^{(\psi)} \right) \right]^{-1} (\mathbf{y} - \hat{\boldsymbol{\mu}}^{(\psi)}).$$

Mit der Definition von sogenannten *Pseudo-Beobachtungen*

$$\tilde{\eta}_i \left(\hat{\beta}^{(\psi)} \right) = \mathbf{x}_i^\top \hat{\beta}^{(\psi)} + \left(\frac{\partial h(\eta_i)}{\partial \eta_i} \right)^{-1} \left(y_i - \mu_i \left(\hat{\beta}^{(\psi)} \right) \right)$$

gilt die Darstellung

$$\hat{\beta}^{(\psi+1)} = \left(\mathbf{X}^\top \mathbf{W} \left(\hat{\beta}^{(\psi)} \right) \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{W} \left(\hat{\beta}^{(\psi)} \right) \tilde{\boldsymbol{\eta}} \left(\hat{\beta}^{(\psi)} \right).$$

Aufgrund dieser Form ist dieses Verfahren ein *iteratively reweighted least squares* Verfahren mit dem Gewicht $\mathbf{W} \left(\hat{\beta}^{(\psi)} \right)$, welches in jeder Iteration neu berechnet wird. Mit der Wahl der kanonischen Linkfunktion reduziert sich die Vorschrift auf die Darstellung

$$\hat{\beta}^{(\psi+1)} = \hat{\beta}^{(\psi)} + \left(\mathbf{X}^\top \mathbf{W} \left(\hat{\beta}^{(\psi)} \right) \mathbf{X} \right)^{-1} \mathbf{X}^\top \left(\mathbf{y} - \hat{\boldsymbol{\mu}}^{(\psi)} \right).$$

Für weitere Ausführungen siehe Green (1984).

A.2 Odds und Odds Ratio

Da sich ordinale Daten nicht einfach miteinander vergleichen lassen, werden Odds und Odds Ratio als Vergleichsmittel verwendet. Auch bei der Interpretation des Logit Modells gibt es durch Odds und Odds Ratio Vorteile.

Definition A.1. (*Odds und Log Odds*) Sei Y eine binäre Zufallsvariable. Die Odds sind definiert als das Verhältnis der Wahrscheinlichkeit von $Y = 1$ zur Wahrscheinlichkeit von $Y = 0$ bei gegebenen Werten \mathbf{x}

$$\text{Odds}(\mathbf{x}) = \frac{P(Y = 1|\mathbf{x})}{P(Y = 0|\mathbf{x})}. \quad (\text{A.3})$$

Die logarithmierten Odds (Log-Odds) werden durch Logarithmieren der Odds gebildet.

Definition A.2. (*Odds Ratio und Log-Odds Ratio*) Das Odds Ratio wird zum Vergleichen zweier Vektoren \mathbf{x} und $\tilde{\mathbf{x}}$ mit unterschiedlichen Werten für die erklärenden Variablen verwendet. Sie ist definiert als das Verhältnis der Odds

$$\frac{\text{Odds}(\mathbf{x})}{\text{Odds}(\tilde{\mathbf{x}})} = \frac{P(Y = 1|\mathbf{x})/P(Y = 0|\mathbf{x})}{P(Y = 1|\tilde{\mathbf{x}})/P(Y = 0|\tilde{\mathbf{x}})}.$$

Das Log-Odds Ratio wird mit Hilfe des Logarithmus gebildet. Es ist dann der Quotient der Exponenten zweier linearer Prädiktoren.

Bemerkung A.1. Das Odds Ratio kann alle nichtnegativen Werte annehmen, Werte > 1 weisen auf eine größere Chance von \mathbf{x} im Vergleich zur Chance von $\tilde{\mathbf{x}}$ hin. Ist das Ergebnis des Odds Ratio ein Wert < 1 so bedeutet dieses, dass die Chance von $\tilde{\mathbf{x}}$ größer der Chance von \mathbf{x} ist, bei Gleichheit sind beide Chancen gleich.

A.3 Maximum Likelihood Schätzer im saturierten Modell

Das saturierte Modell wird in der Definition der Deviance benötigt, wo es als Vergleichsmodell verwendet wird, um die Anpassungsgüte eines betrachteten Modells zu bewerten. Dieses Modell hat zwar eine perfekte Anpassung, ergibt dadurch jedoch auch keine Parameterreduktion.

Satz A.3. Unter dem saturierten Modell und der einparametrischen, linearen Exponentialfamilie gilt für den Maximum Likelihood Schätzer $\hat{\mu}_i = y_i$, $i = 1, \dots, n$.

Beweis. Es seien $Y_i \sim$ Exponentialfamilie (θ_i) unabhängig mit $i = 1, \dots, n$. Im Abschnitt 3.2 wurde aufgrund der Abhängigkeit $\theta_i = \theta_i(\boldsymbol{\beta})$ die dazugehörige Log-Likelihood-Funktion angegeben als

$$l(\boldsymbol{\beta}|\mathbf{y}) = \log \left(\prod_{i=1}^n f(y_i|\theta_i) \right).$$

Der Maximum Likelihood Schätzer ist als Nullstelle der Scorefunktion

$$\mathbf{s}(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta}|\mathbf{y})}{\partial \boldsymbol{\beta}}$$

definiert. Für die Exponentialfamilie kann die Log-Likelihood-Funktion angeführt werden mit

$$l(\boldsymbol{\beta}|\mathbf{y}) = \sum_{i=1}^n (y_i \theta_i - b(\theta_i) + c(y_i)).$$

Für den Parameter θ_i gilt die Abhängigkeit $\theta_i = \theta_i(\mu_i)$. Aufgrund dieser Abhängigkeit und der Kettenregel hat die Score-Funktion für dieses Modell die Gestalt

$$\begin{aligned} \frac{\partial l(\boldsymbol{\beta}|\mathbf{y})}{\partial \mu_i} &= \left(\frac{\partial (y_i \theta_i - b(\theta_i) + c(y_i))}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \right) \\ &= (y_i - b'(\theta_i)) \frac{\partial \theta_i}{\partial \mu_i}, \quad i = 1, \dots, n. \end{aligned}$$

Diese Terme sind Null für

$$b'(\hat{\theta}_i) = \hat{\mu}_i = y_i.$$

□

A.4 Deviance der diskreten Verteilungen

Für die Deviance der verschiedenen Verteilungen können explizit Darstellungen angeführt werden.

Standardisierte Binomialverteilung

Für die Standardisierte Binomialverteilung gilt mit den relativen Häufigkeiten $\bar{y} \in \{0, \frac{1}{m}, \frac{2}{m}, \dots, 1\}$

$$\begin{aligned} l(\hat{\boldsymbol{\mu}}|\mathbf{y}) &= \sum_{i=1}^n m_i \left(\bar{y}_i \log(\hat{\mu}_i) + (1 - \bar{y}_i) \log(1 - \hat{\mu}_i) + c(\bar{y}_i) \right) \\ l(\mathbf{y}|\mathbf{y}) &= \sum_{i=1}^n m_i \left(\bar{y}_i \log(y_i) + (1 - \bar{y}_i) \log(1 - y_i) + c(\bar{y}_i) \right). \end{aligned}$$

Daraus ergibt sich die Deviance

$$D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = -2 \sum_{i=1}^n \left(y_i \log \left(\frac{\hat{\mu}_i}{y_i} \right) - (m_i - y_i) \log \left(\frac{1 - \hat{\mu}_i}{1 - y_i} \right) \right).$$

Poissonverteilung

Für die Poissonverteilung gilt

$$l(\hat{\boldsymbol{\mu}}|\mathbf{y}) = \sum_{i=1}^n \left(y_i \log(\hat{\mu}_i) - \hat{\mu}_i - \log(y_i!) \right)$$

$$l(\mathbf{y}|\mathbf{y}) = \sum_{i=1}^n \left(y_i \log(y_i) - y_i - \log(y_i!) \right).$$

Als Deviance folgt

$$D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = -2 \left[\sum_{i=1}^n (y_i \log(\hat{\mu}_i) - \hat{\mu}_i - \log(y_i!)) - \sum_{i=1}^n (y_i \log(y_i) - y_i - \log(y_i!)) \right]$$

$$= -2 \sum_{i=1}^n \left(y_i \log \left(\frac{\hat{\mu}_i}{y_i} \right) - (y_i - \hat{\mu}_i) \right)$$

$$= -2 \sum_{i=1}^n y_i \log \left(\frac{\hat{\mu}_i}{y_i} \right),$$

da gilt, dass

$$\sum_{i=1}^n (y_i - \hat{\mu}_i) = 0.$$

Multinomialverteilung

Zur Multinomialverteilung gehört die Log-Likelihoodfunktion

$$l_i(\hat{\boldsymbol{\mu}}_i|\mathbf{y}_i) = m_i \sum_{j=1}^k \left\{ \frac{y_{ij}}{m_i} \log \left(\frac{\hat{\mu}_{ij}/m_i}{1 - \sum_{s=1}^{k-1} \hat{\mu}_{is}/m_i} \right) + \log \left(1 - \sum_{s=1}^{k-1} \hat{\mu}_{is}/m_i \right) + \log(h(y_i)) \right\}$$

$$l_i(\mathbf{y}_i|\mathbf{y}_i) = m_i \sum_{j=1}^k \left\{ \frac{y_{ij}}{m_i} \log \left(\frac{y_{ij}/m_i}{1 - \sum_{s=1}^{k-1} y_{is}/m_i} \right) + \log \left(1 - \sum_{s=1}^{k-1} y_{is}/m_i \right) + \log(h(y_i)) \right\}.$$

Die Differenz dieser Funktionen ergibt

$$l_i(\hat{\boldsymbol{\mu}}_i|\mathbf{y}_i) - l_i(\mathbf{y}_i|\mathbf{y}_i) = m_i \sum_{j=1}^k \left(\frac{y_{ij}}{m_i} \log \left(\frac{\hat{\mu}_{ij}/m_i}{1 - \sum_{s=1}^{k-1} \hat{\mu}_{is}/m_i} \right) + \log \left(1 - \sum_{s=1}^{k-1} \hat{\mu}_{is}/m_i \right) \right. \\ \left. + \log(h(y_i)) \right) - m_i \sum_{j=1}^k \left(\frac{y_{ij}}{m_i} \log \left(\frac{y_{ij}/m_i}{1 - \sum_{s=1}^{k-1} y_{is}/m_i} \right) \right.$$

$$\begin{aligned}
& + \log \left(1 - \sum_{s=1}^{k-1} y_{is}/m_i \right) + \log(h(y_i)) \\
& = m_i \sum_{j=1}^k \left\{ \frac{y_{ij}}{m_i} \left(\log \left(\frac{\hat{\mu}_{ij}/m_i}{\hat{\mu}_{ik}/m_i} \right) - \log \left(\frac{y_{ij}/m_i}{y_{ik}/m_i} \right) \right) \right. \\
& \quad \left. + \log \left(\hat{\mu}_{ik}/m_i \right) - \log \left(y_{ik}/m_i \right) \right\} \\
& = m_i \sum_{j=1}^k \left\{ \frac{y_{ij}}{m_i} \log \left(\frac{\hat{\mu}_{ij}}{y_{ij}} \right) + \log \left(\frac{y_{ik}}{\hat{\mu}_{ik}} \right) + \log \left(\frac{\hat{\mu}_{ik}}{y_{ik}} \right) \right\} \\
& = m_i \sum_{j=1}^k \frac{y_{ij}}{m_i} \log \left(\frac{\hat{\mu}_{ij}}{y_{ij}} \right) \\
& = \sum_{j=1}^k y_{ij} \log \left(\frac{\hat{\mu}_{ij}}{y_{ij}} \right).
\end{aligned}$$

Die Deviance für die Multinomialverteilung hat daher die Gestalt

$$D(\mathbf{y}_i | \hat{\boldsymbol{\mu}}_i) = -2 \sum_{i=1}^n \sum_{j=1}^k y_{ij} \log \left(\frac{\hat{\mu}_{ij}}{y_{ij}} \right).$$

A.5 Linkfunktionen

Die im Kapitel 3 verwendeten Linkfunktionen werden im Folgenden definiert sowie wichtige Momente der entsprechenden Verteilungen angeführt.

Logistische Verteilung

Für weitere Informationen siehe Balakrishnan (1991). Sei X eine logistische Zufallsvariable mit $X \sim L(\mu, \sigma)$, wobei $-\infty < \mu < \infty$ und $\sigma > 0$. Die allgemeine logistische Dichtefunktion ist gegeben durch

$$f(x|\mu, \sigma) = \frac{\exp \left(-\frac{x - \mu}{\sigma} \right)}{\sigma \left(1 + \exp \left(-\frac{x - \mu}{\sigma} \right) \right)^2}, \quad -\infty < x < \infty.$$

Die Verteilungsfunktion hat die Gestalt

$$F(x|\mu, \sigma) = \left[1 + \exp\left(-\frac{x - \mu}{\sigma}\right) \right]^{-1}.$$

Die Momente der logistischen Verteilungsfunktion sind gegeben durch

$$\begin{aligned} E(X) &= \mu \\ \text{Var}(X) &= \frac{\sigma^2 \pi^2}{3}. \end{aligned}$$

In dieser Arbeit wird hauptsächlich die standardisierte Form der logistischen Verteilungsfunktion verwendet. Sei $X \sim L(0, 1)$, so ergibt sich

$$\begin{aligned} f(x|0, 1) &= \frac{\exp(-x)}{(1 + \exp(-x))^2} \\ F(x|0, 1) &= \frac{1}{1 + \exp(-x)} \end{aligned}$$

mit

$$\begin{aligned} E(X) &= 0 \\ \text{Var}(X) &= \frac{\pi^2}{3}. \end{aligned}$$

Die logistische Verteilungsfunktion ist äquivalent zur Umkehrfunktion des Logit.

Standard Normalverteilung

Das Probit Modell bei der Schätzung des Modells für ordinale Responses benötigt die standardisierte Normalverteilung. Sei X eine stetige Zufallsvariable mit $X \sim N(0, 1)$. Die Dichtefunktion der Standardnormalverteilung ist gegeben durch

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \quad -\infty < x < \infty$$

mit der Verteilungsfunktion

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{1}{2}t^2\right) dt.$$

Die Momente der Standardnormalverteilung sind

$$\begin{aligned} E(X) &= 0 \\ \text{Var}(X) &= 1. \end{aligned}$$

Für Details siehe Jagdish und Campbell (1996).

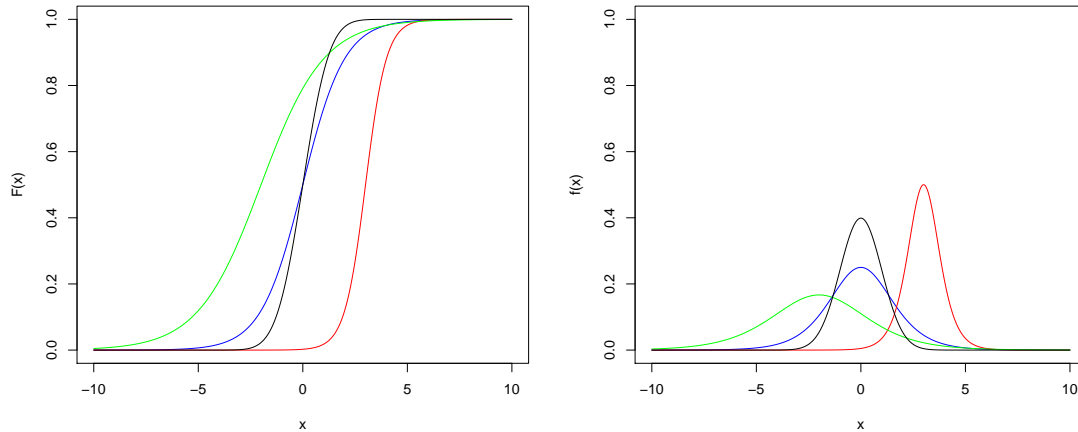


Abbildung A.1: Logistische Verteilungs- (links) und Dichtefunktionen (rechts) mit $\mu = 0, \sigma = 1$ (blau), $\mu = 3, \sigma = 0.5$ (rot) und $\mu = -2, \sigma = 1.5$ (grün) sowie standard-normal (schwarz)

Extremwert Verteilung

Die Gompertz und Gumbel Verteilung gehören zu der Extremwert-Verteilung. Sei X eine Zufallsvariable, dann hat die Gumbel-Verteilungsfunktion die Gestalt

$$F(x|\mu, \sigma) = \exp \left\{ \exp \left\{ -\frac{x - \mu}{\sigma} \right\} \right\}$$

mit $x, \mu \in \mathbb{R}$ und $\sigma > 0$. Die Dichteverteilung der Gumbel Verteilung ist gegeben durch

$$f(x|\mu, \sigma) = \frac{1}{\sigma} \exp \left\{ -\frac{x - \mu}{\sigma} \right\} \exp \left\{ \exp \left\{ -\frac{x - \mu}{\sigma} \right\} \right\}.$$

Mit der Euler-Mascheroni Konstante $\gamma = 0.57722$ können die Momente der Gumbel Verteilung angeführt werden mit

$$\begin{aligned} E(X) &= \mu + \gamma\sigma \\ \text{Var}(X) &= \frac{\sigma^2\pi^2}{6}. \end{aligned}$$

Wird als Zufallsvariable $-X$ verwendet, so hat die Verteilungsfunktion die Darstellung

$$F(x|\mu, \sigma) = 1 - F(-x|\mu, \sigma) = 1 - \exp \left\{ -\exp \left\{ -\frac{-x - \mu}{\sigma} \right\} \right\}.$$

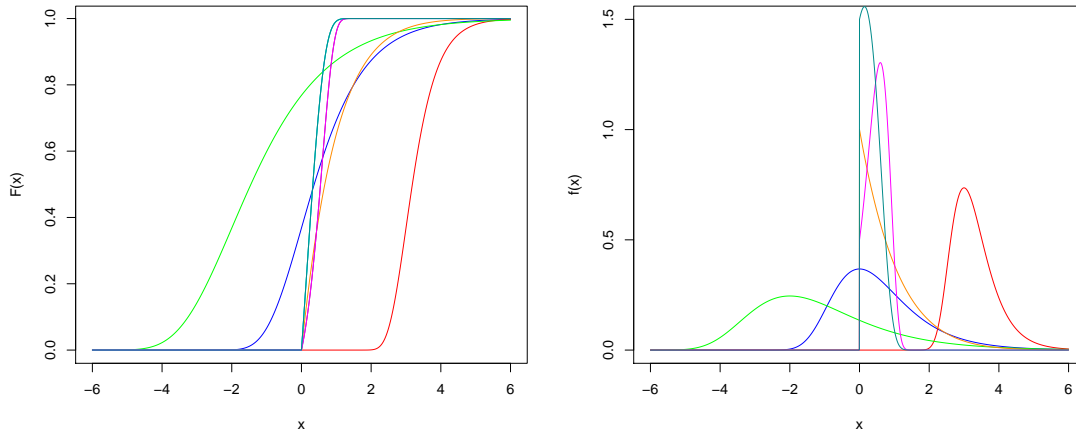


Abbildung A.2: Gumbel Verteilungs- (links) und Dichtefunktionen (rechts) mit $\mu = 0, \sigma = 1$ (blau), $\mu = 3, \sigma = 0.5$ (rot) und $\mu = -2, \sigma = 1.5$ (grün), Gompertz mit $\mu = 0.1, \sigma = 1$ (orange), $\mu = 3, \sigma = 0.5$ (magenta) und $\mu = 2, \sigma = 1.5$ (cyan)

Mit Hilfe dessen kann nun die Gompertz Verteilungsfunktion angegeben werden. Diese ergibt sich in Standardform zu

$$F(x|0, 1) = 1 - F(-x|0, 1) = 1 - \exp \{-\exp \{x\}\}$$

mit der Dichtefunktion

$$f(x|0, 1) = \exp \{x - \exp \{x\}\}.$$

Für weitere Details siehe Hosking, Wallis und Wood (1984).

A.6 Hauck-Donner Effekt

Der Hauck-Donner Effekt beschreibt ein Resultat bei der Schätzung eines binomialen Logit-Modells, bei dem der Wald-Test fehlschlägt und somit keine gesicherte Aussage über einen Schätzer getroffen werden kann. Dieser Effekt wurde von Hauck und Donner (1977) zum ersten Mal wahrgenommen und beschrieben. Die Auswirkungen und Folgen dieses Effekts sind bis heute noch nicht erfasst.

Für das binomiale Logit-Modell seien Y_1, \dots, Y_n unabhängige und binäre Response-Variablen und \mathbf{x}_i der Vektor der erklärenden Variablen. Der Logit der Wahrscheinlich-

keit $P(Y_i = 1|\mathbf{x}_i)$ besitzt die Darstellung

$$\log \frac{P(Y_i = 1|\mathbf{x}_i)}{1 - P(Y_i = 1|\mathbf{x}_i)} = \mathbf{x}_i^\top \boldsymbol{\beta},$$

mit

$$P(Y_i = 1|\mathbf{x}_i) = [1 + \exp(-\mathbf{x}_i^\top \boldsymbol{\beta})]^{-1}.$$

Für $j_1 \in \{0, \dots, p-1\}$ ergibt sich die zu testende Hypothese wie folgend

$$H_0 : \beta_{j_1} = \beta_{j_0} \quad \text{gegen} \quad H_1 : \beta_{j_1} \neq \beta_{j_0}.$$

Der Test auf diese Hypothese ist ein parametrischer Test für den Effekt β_{j_1} . Mit $\hat{\beta}_{j_1}$ als Maximum Likelihood Schätzer von β_{j_1} ergibt sich die Wald-Test-Statistik zu

$$W = \frac{(\hat{\beta}_{j_1} - \beta_{j_0})^2}{[F(\hat{\boldsymbol{\beta}})]^{-1}}. \quad (\text{A.4})$$

Aufgrund dieser Definition der Wald-Test-Statistik ergeben sich verschiedene Problemfälle. Tritt eine Response Kategorie nur selten auf und besitzt ein Faktor für diese Kategorie immer den Wert 1 (oder äquivalent immer den Wert 0), so ist die geschätzte Wahrscheinlichkeit in diesem Fall ≈ 1 (oder äquivalent ≈ 0). Der Schätzer des dazugehörigen Effekts strebt dann gegen ∞ und es gilt die Aussage

$$\lim_{|\beta_{j_1}| \rightarrow \infty} W = 0. \quad (\text{A.5})$$

Diese Aussage führt zu dem Phänomen von Hauck und Donner, da beim Testen eines Parameters die Wald-Test-Statistik zuerst mit steigendem Abstand zum Testparameter größer wird. Bei Annäherung des Parameters an den Rand seiner möglichen Werte nimmt die Wald-Test-Statistik aufgrund der Aussage (A.5) wieder ab. Dieses Verhalten ist im folgenden graphischen Beispiel von Hauck und Donner anhand des Parameters p_2 ersichtlich.

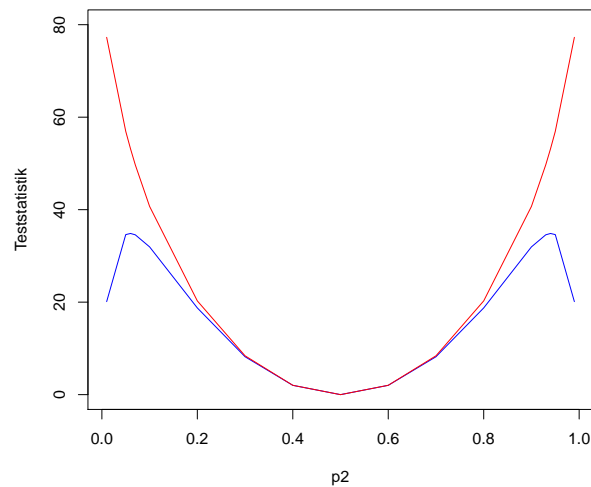


Abbildung A.3: Vergleich der Test-Statistiken von Wald (blau) und Likelihood Ratio (rot)

In der Abbildung A.3 ist die Wurzel der Wald-Test-Statistik für den Parameter $p_2 \in (0, 1)$ dargestellt. Als Vergleich wurde die Test-Statistik des Likelihood-Ratio-Tests abgebildet. Am Rand lässt sich deutlich die Auswirkung des Hauck-Donner Effekts auf die Wald-Test-Statistik beobachten, während die Likelihood-Ratio-Test-Statistik weiter monoton steigt. Für die Daten des zugrundeliegenden Beispiels siehe Hauck und Donner (1977).

Aus der Aussage (A.5) folgt weiters, dass die Varianz des Schätzers im Vergleich zum Schätzer weitaus stärker ansteigt. Aufgrund dieses Verhaltens kann kein sinnhafter Wert der Teststatistik angegeben werden und es kann keine Aussage über die Hypothese H_0 getroffen werden. Die theoretische Grundlage dieses Effekts kann in Hauck und Donner (1977) nachgelesen werden.

Literatur

- Agresti, A. (2010). *Analysis of Ordinal Categorical Data (Second Edition)*. John Wiley and Sons.
- Balakrishnan, N. (1991). *Handbook of the Logistic Distribution*. Marcel Dekker, Inc.
- Burridge, J. (1981). A note on maximum likelihood estimation for regression models using grouped data. *Journal of the Royal Statistical Society, Series B*, 41-45.
- Clavien, P., Barkun, J., Dindo, D. & et al. (2009). The Clavien Dindo classification of surgical complications. *Annals of Surgery*, 250, 187-196.
- Dindo, D., Demartines, N. & Clavien, P. (2004). Classification of surgical complications. *Annals of Surgery*, 240 (2), 205-213.
- Fahrmeir, L. & Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics*, 13, 342-368.
- Green, P. (1984). Iteratively reweighted least squares for maximum likelihood estimation and some robust and resistant alternative. *Journal of the Royal Statistical Society, Series B*, 46, 149-192.
- Hauck, W. & Donner, A. (1977). Wald's test as applied to hypotheses in logit analysis. *Journal of the American Statistical Association*, 72, 851-853.
- Hosking, J., Wallis, J. & Wood, E. (1984). Estimation of the generalized extreme-value distribution by the method of probability weighted moments. *Technical Summary Report, University of Wisconsin-Madison Mathematics Research Center* (2674).
- Jagdish, P. & Campell, K. (1996). *Handbook of the Normal Distribution, Second Edition, Revised and Expanded*. Marcel Dekker, Inc.
- McCullagh, P. & Nelder, J. (1989). *Generalized Linear Models, 2nd Edition*. Chapman and Hall.
- Schauberger, G. (2017). Package 'EffectStars2' [Software-Handbuch].
- Steel, R. (1953). Relation between poisson and multinomial distributions. *Biometrics Unit Technical Report* (BU-39-M).

- Tutz, G. (2000). *Die Analyse kategorialer Daten*. Cambridge University Press.
- Tutz, G. (2012). *Regression for Categorical Data*. Cambridge University Press.
- Venables, W. & Ripley, B. (2002). *Modern Applied Statistics with S, 4th Edition*. Springer.
- Venables, W. & Ripley, B. (2016). Package 'nnet' [Software-Handbuch].
- Windischbacher, A. (2017). *Statistische Analyse potentieller Risikofaktoren für Komplikationen nach einer Mammareduktion* (Masterprojekt Statistik). Institut für Statistik, Technische Universität Graz.
- Yee, T. (2010). The VGAM package for categorical data analysis. *University of Auckland*.
- Yee, T. (2015). *Vector Generalized Linear and Additive Models: With an Implementation in R*. Springer.
- Yee, T. (2018). Package 'VGAM' [Software-Handbuch].