



Amanda Schalk BSc

Regressionsmodelle für Prozentwerte

MASTERARBEIT

zur Erlangung des akademischen Grades eines Diplom-Ingenieurs

Operations Research und Statistik

eingereicht an der

Technischen Universität Graz

Betreuer:

Ao. Univ.-Prof. Dipl.-Ing. Dr.techn. Herwig FRIEDL

Institut für Statistik

Graz, Mai 2016

EIDESSTATTLICHE ERKLÄRUNG

AFFIDAVIT

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig 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.

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.

Datum/Date

Unterschrift/Signature

ZUSAMMENFASSUNG

Diese Arbeit behandelt die Modellierung von Prozentwerten und vergleicht die dazu geeigneten Regressionsmodelle. Basierend auf den Kenntnissen über die Lineare Regression wird zunächst das Generalisierte Lineare Modell eingeführt, das im Gegensatz zu den Linearen Modellen nicht länger von normalverteilten Response-Werten ausgeht, sondern Verteilungen aus der linearen einparametrischen Exponentialfamilie zulässt. Auch die Quasi-Likelihoodfunktion mit ihrer Möglichkeit zum Vergleich der Varianzfunktionen und die Extended-Quasi-Likelihoodfunktion werden besprochen und auf Vor- und Nachteile im Umgang mit Prozentwerten untersucht. Als zweite große Klasse an Regressionsmodellen wird die Beta-Regression eingeführt. Durch die Flexibilität der Beta-Verteilung scheinen diese Modelle besonders geeignet zur Modellierung von Werten aus dem offenen Einheitsintervall $(0, 1)$. Bezüglich der Beta-Regression wird ergänzend zum Grundmodell auch eine Erweiterung zur gemeinsamen Modellierung von Erwartungswert und Dispersion behandelt. Die eben genannten Regressionsmodelle werden sowohl theoretisch eingeführt und besprochen, als auch anhand eines Beispiels diskutiert und auf ihre Anwendungsmöglichkeiten in der Praxis getestet.

ABSTRACT

This thesis deals with the modelling of percentage values and compares the appropriate regression models. Based on the knowledge about linear regression, the generalized linear model is introduced first. Unlike the linear models, the generalized linear model does not assume a normal distributed response, but permits distributions from the exponential family. Also the quasi likelihood function with the possibility of comparing variance functions and the extended quasi likelihood function are discussed and tested for advantages and disadvantages, when dealing with percentages. As a second major class of regression models, the beta regression model is introduced. The flexibility of the beta distribution makes these models particularly suitable for modelling responses from the open unit interval $(0, 1)$. According to the beta regression, in addition to the basic model, also the variable dispersion beta regression model is considered. The mentioned regression models are introduced, contemplated theoretically and evaluated based on an example. Moreover, they are analysed for possible practical applications.

1	Einleitung	11
2	Generalisierte Lineare Modelle	13
2.1	Motivation	13
2.2	Die Exponentialfamilie	14
2.2.1	Mitglieder der einparametrischen linearen Exponentialfamilie	16
2.3	Modell und Linkfunktion	17
2.3.1	Beispiele für kanonische Linkfunktionen	18
2.4	Maximum Likelihood Schätzung	19
2.4.1	Iterative Weighted Least Squares	20
2.4.2	IWLS mit kanonischer Linkfunktion und Fisher-Scoring	21
2.5	Anpassungsgüte	23
2.5.1	Pearson Statistik	23
2.5.2	Deviance	24
2.5.3	Analysis of Deviance	26
2.6	Generalisierte Hebelwirkung	26
3	Quasi-Likelihood	31
3.1	Motivation	31
3.2	Die Quasi-Likelihoodfunktion	31
3.3	Beispiele für Quasi-Likelihoodmodelle	34
3.4	Quasi-Dichte	35
3.5	Maximum Quasi-Likelihood Schätzung	36
3.6	Quasi-Deviance	38
3.6.1	Beispiele zur Quasi-Deviance	39
3.7	Extended Quasi-Likelihood	40
3.7.1	Schätzung der Parameter	41
3.7.2	Eigenschaften der EQL-Funktion	42
3.7.3	Parametrisierte Varianzfunktion	42

4	Beta-Regression	45
4.1	Motivation	45
4.2	Grundmodell	46
4.2.1	Maximum Likelihood Schätzung	49
4.2.2	Fisher-Information	51
4.2.3	Newton-Verfahren und Fisher-Scoring	53
4.2.4	Modelldiagnostik	54
4.3	Gemeinsame Modellierung von Erwartungswert und Dispersion	57
4.3.1	Maximum Likelihood Schätzung	58
4.3.2	Fisher-Information	61
4.3.3	Modelldiagnostik	64
4.4	Korrektur und Reduktion des Bias	65
4.4.1	Algorithmus	65
4.5	Ausblick	67
4.5.1	Heterogenität bei Beta-Regressionsmodellen	67
4.5.2	Beta-Regression mit erweitertem Definitionsbereich	69
5	Praktische Anwendung	73
5.1	Motivation	73
5.2	Datensatz	73
5.2.1	Hausarbeit	74
5.2.2	Unterstützung	74
5.2.3	Kinder	75
5.2.4	Anstellung	76
5.2.5	Beschäftigung	77
5.2.6	Hausarbeit der Mutter	78
5.2.7	Modell	80
5.3	GLM mit Binomialverteilung	80
5.4	Quasi-Likelihoodmodelle	82
5.4.1	Varianzfunktion proportional zur Bernoulli-Varianz	82
5.4.2	Quadratische Varianzfunktion	84
5.5	EQL-Modell	87
5.6	Beta-Regressionsmodelle	91
5.6.1	Grundmodell	92
5.6.2	Gemeinsame Modellierung von Erwartungswert und Dispersion	95
5.6.3	Korrektur und Reduktion des Bias	98
5.7	Interpretation der Ergebnisse	99
6	Fazit	101
	Literaturverzeichnis	103

Abbildungsverzeichnis

4.1	Beta-Dichten bei Variation der Parameter α und β	46
4.2	Dichtefunktionen $f(y \mu, \phi)$ der Beta-Verteilung bei Variation der Parameter μ und ϕ	48
5.1	Histogramm der Haus- und Familienarbeit.	74
5.2	Boxplot: Hausarbeit gegen Unterstützung.	75
5.3	Boxplot: Hausarbeit gegen Kinder.	76
5.4	Boxplot: Hausarbeit gegen Art der Anstellung.	77
5.5	Boxplot: Hausarbeit gegen Beschäftigung.	78
5.6	Histogramm: Hausarbeit der Mutter.	79
5.7	Boxplot: Hausarbeit der befragten Frauen gegen Hausarbeit der Mütter.	79
5.8	Residuengrafiken des Modells mQB.	84
5.9	Residuengrafiken des Modells mQW.	86
5.10	EQL-Konturplot für das Modell mEQL1.	88
5.11	EQL-Konturplot für das Modell mEQL2.	89
5.12	Residuengrafiken des Modells mEQL2.	91
5.13	Residuengrafik und generalisierte Hebelpunkte des Modells mBeta.	94
5.14	Generalisierte Hebelpunkte von mBeta gegen die Variable Hausarbeit.	95
5.15	Residuengrafik und generalisierte Hebelpunkte des Modells mBetaD1.	98

KAPITEL 1

Einleitung

Ob bei Umfragen, Meinungsforschungen oder Leistungsbeurteilungen, Prozentwerte sind eine praktische und oft verwendete Darstellungsform von Response-Werten. Die Analyse der Daten könnte sich bei der Verwendung von Prozentsätzen jedoch schwierig gestalten. Die Lineare Regression stellt mit ihrer einfachen Anwendung und den Möglichkeiten zur Interpretation eines der bevorzugten Werkzeuge zum Auffinden von Zusammenhängen zwischen der beobachteten Response und den erklärenden Variablen dar. Da das Lineare Modell (LM) allerdings auf der Annahme reeller Beobachtungen basiert und die Prozentwerte lediglich auf dem Intervall $[0, 1]$ definiert sind, müsste eine Transformation der Response durchgeführt werden. Doch selbst dann würde die vom LM getroffene Annahme der Normalverteilung nicht mit den vorliegenden Datenpunkten übereinstimmen. Aus diesem Grund werden in dieser Arbeit verschiedenen Regressionsmodelle eingeführt und diskutiert, die sich für die Modellierung von Prozentwerten eignen.

Es wird mit dem Übergang vom LM zum Generalisierten Linearen Modell begonnen. Dabei werden alle relevanten Schätzer hergeleitet, sowie Methoden zur Untersuchung der Anpassungsgüte vorgestellt. Das dritte Kapitel behandelt die Verwendung der auf Wedderburn (1974) zurückgehenden Quasi-Likelihoodfunktion und die daraus resultierende Möglichkeit zum Vergleich unterschiedlicher Varianzfunktionen. Zusätzlich wird die von Nelder und Pregibon (1987) definierte Extended-Quasi-Likelihoodfunktion (EQL-Funktion) eingeführt und die parametrisierte Varianzstruktur zusammen mit der optimalen Schätzung der Parameter diskutiert. Kapitel vier erläutert das von Ferrari und Cribari-Neto (2004) stammende Beta-Regressionsmodell mit seinen zahlreichen Anwendungsmöglichkeiten. Beginnend beim Grundmodell werden die Annahmen des Modells beschrieben, die Herleitung der Schätzer angeführt und die Methoden der Modelldiagnostik besprochen. Als Erweiterung wird das Beta-Regressionsmodell zur gemeinsamen Modellierung von Erwartungswert und Dispersion eingeführt und anschließend die Korrektur bzw. Reduktion einer möglichen, bei der Schätzung entstandenen Verzerrung

erklärt. Einen Ausblick und Motivation zur Weiterführung der Beta-Regressionsmodelle dieser Arbeit stellen der Umgang mit der Heterogenität und die Ausdehnung des Definitionsbereichs $(0, 1)$ der Beta-Verteilung dar.

Als praktischer Teil dieser Arbeit wird in Kapitel fünf der von Cernko (2015) erhobene Datensatz analysiert. Dabei wird untersucht, wie stark die Ergebnisse der beschriebenen Regressionsmodelle unter Verwendung realer Daten variieren. Zusätzlich wird versucht die Vor- und Nachteile der Modelle zu beschreiben, um so auch in Zukunft die Wahl des passenden Regressionsmodells bei vorliegenden Daten in Form von Prozentwerten zu vereinfachen.

2.1 Motivation

Bei der statistischen Analyse von Daten ist das Lineare Modell (LM) ein gern verwendetes Werkzeug. Zur Modellanpassung wird angenommen, dass die Response-Variablen y_i für $i = 1, \dots, n$ unabhängig und normalverteilt sind, mit konstanter Varianz σ^2 . Die Modellierung des Erwartungswertes μ_i erfolgt unter Verwendung der erklärenden Variablen $\mathbf{x}_i \in \mathbb{R}^p$ und des Parametervektors $\boldsymbol{\beta} \in \mathbb{R}^p$

$$\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}, \quad i = 1, \dots, n.$$

In der Praxis sind diese Annahmen allerdings nicht immer vertretbar. So werden zum Beispiel beobachtete Häufigkeiten oder Proportionen nicht aus der Normalverteilung stammen. Auch eine mögliche Abhängigkeit der Varianz vom Erwartungswert und die daraus resultierende Variabilität sollten in Betracht gezogen werden.

Aus diesen Gründen wird nun das auf Nelder und Wedderburn (1972) zurückgehende Generalisierte Lineare Modell (GLM) vorgestellt. Es umgeht die Einschränkungen und die daraus entstehenden Probleme des Linearen Modells, indem es die Annahmen verallgemeinert. So müssen zum Beispiel die Response-Variablen nicht mehr normalverteilt sein, sondern werden nun einem Mitglied der einparametrischen linearen Exponentialfamilie zugeordnet. Die Annahme der konstanten Varianz wird verworfen, eine vom Erwartungswert abhängige Varianzfunktion wird eingeführt. Und schließlich wird auch die lineare Modellierung des Erwartungswertes erweitert. Beim GLM entspricht der lineare Prädiktor nicht mehr dem Erwartungswert, sondern einer von μ_i abhängigen bekannten Funktion $g(\mu_i)$, genannt Linkfunktion. Die folgenden Abschnitte orientieren sich an McCullagh und Nelder (1989).

2.2 Die Exponentialfamilie

Definition 2.1. (Allgemeine Exponentialfamilie)

Eine Zufallsvariable y stammt aus der allgemeinen Exponentialfamilie, falls die Dichte- oder Wahrscheinlichkeitsfunktion der Form

$$f(y|\boldsymbol{\theta}) = h(y)p(\boldsymbol{\theta}) \exp \left\{ \sum_{j=1}^k t_j(y)w_j(\boldsymbol{\theta}) \right\} \quad (2.1)$$

entspricht, wobei $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^T$ den Parametervektor darstellt. Dabei gilt $h(y) \geq 0$ und $p(\boldsymbol{\theta}) \geq 0$, während $t_1(y), \dots, t_k(y)$ und $w_1(\boldsymbol{\theta}), \dots, w_k(\boldsymbol{\theta})$ reellwertige Funktionen sind.

Bemerkung:

Eine Exponentialfamilie, für die $d < k$ gilt, wird gekrümmte Exponentialfamilie genannt. Ist jedoch $d = k$, handelt es sich um eine volle Exponentialfamilie. Der Parameter $\boldsymbol{\theta}$ muss daher nicht zwingend vektorwertig sein. Genaueres bezüglich der Exponentialfamilie findet sich in Casella und Berger (2002).

Bemerkung:

Gilt in (2.1) $w_j(\boldsymbol{\theta}) = \boldsymbol{\theta}$ für alle $j = 1, \dots, k$, so wird $\boldsymbol{\theta}$ als kanonischer Parameter und die in Definition 2.1 eingeführte Familie als Exponentialfamilie in kanonischer Form bezeichnet.

Die Ableitung einer logarithmierten Dichtefunktion nach dem Parameter wird Scorefunktion bzw. Scorevektor genannt. Für Scorevektoren gelten unter Einhaltung gewisser Regularitätsbedingungen, welche für die Exponentialfamilie erfüllt sind, die Eigenschaften

$$\mathbb{E} \left(\frac{\partial \log f(y|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) = \mathbf{0} \quad (2.2)$$

$$\text{Var} \left(\frac{\partial \log f(y|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) = \mathbb{E} \left(\frac{\partial \log f(y|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \log f(y|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right) = \mathbb{E} \left(-\frac{\partial^2 \log f(y|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right), \quad (2.3)$$

wobei die Varianz des Scorevektors in (2.3) der Fisher-Informationsmatrix entspricht. Details finden sich in Casella und Berger (2002).

Zur Betrachtung der Generalisierten Linearen Modelle ist jedoch nicht die allgemeine Exponentialfamilie von Interesse, sondern die von McCullagh und Nelder (1989) vorgestellte Darstellung der einparametrischen linearen Exponentialfamilie.

Definition 2.2. (Einparametrische lineare Exponentialfamilie)

Sei y eine Zufallsvariable aus einer Verteilung mit Dichtefunktion

$$f(y|\theta) = \exp \left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right) \quad (2.4)$$

für bekannte Funktionen $a(\cdot)$, $b(\cdot)$ und $c(\cdot, \cdot)$ mit $a(\phi) > 0$. Bei bekanntem ϕ bezeichnet die Menge der Verteilungen mit Dichtefunktion $f(y|\theta)$ die einparametrische lineare Exponentialfamilie in kanonischer Form.

Satz 2.3. Sei y eine Zufallsvariable mit Dichtefunktion aus der einparametrischen linearen Exponentialfamilie. Dann gilt für Erwartungswert und Varianz von y

$$\begin{aligned} \mathbb{E}(y) &= b'(\theta) \\ \text{Var}(y) &= a(\phi)b''(\theta). \end{aligned}$$

Beweis:

Unter Verwendung von (2.2) gilt für den Erwartungswert der Scorefunktion

$$0 = \mathbb{E} \left(\frac{\partial \log f(y|\theta)}{\partial \theta} \right) = \mathbb{E} \left(\frac{y - b'(\theta)}{a(\phi)} \right)$$

und da $a(\phi)$ und $b'(\theta)$ aus dem Erwartungswert herausgezogen werden können, folgt die Behauptung

$$\mathbb{E}(y) = b'(\theta).$$

Zur Herleitung der Varianz müssen zuerst die für (2.3) relevanten Ausdrücke betrachtet werden.

$$\begin{aligned} \mathbb{E} \left(\frac{\partial \log f(y|\theta)}{\partial \theta} \right)^2 &= \mathbb{E} \left(\frac{y - b'(\theta)}{a(\phi)} \right)^2 = \frac{\mathbb{E}(y - b'(\theta))^2}{a^2(\phi)} = \frac{\text{Var}(y)}{a^2(\phi)} \\ \mathbb{E} \left(\frac{\partial^2 \log f(y|\theta)}{\partial \theta^2} \right) &= -\frac{b''(\theta)}{a(\phi)}. \end{aligned}$$

Da diese zwei Identitäten laut (2.3) zusammen 0 ergeben, resultiert

$$\frac{\text{Var}(y)}{a^2(\phi)} - \frac{b''(\theta)}{a(\phi)} = 0.$$

Mit Hilfe einfacher Umformungen folgt schließlich das gewünschte Ergebnis

$$\text{Var}(y) = a(\phi)b''(\theta). \quad \square$$

Bei der linearen einparametrischen Exponentialfamilie entspricht die Varianz also dem Produkt zweier Funktionen. Der erste Teil der Varianz besteht aus der von θ unabhängigen Funktion $a(\phi)$, hierbei wird ϕ als Dispersionsparameter bezeichnet. Die Funktion $b''(\theta)$ hängt nur von dem kanonischen Parameter θ ab. Da $\mu = b'(\theta)$ laut Satz 2.3 gilt, kann $b''(\theta)$ demnach auch als eine vom Erwartungswert abhängige Funktion betrachtet werden. Aus diesem Grund wird die von μ abhängige Varianzfunktion $V(\mu) = b''(\theta)$ eingeführt. Die Varianz von y kann also geschrieben werden als

$$\text{Var}(y) = a(\phi)V(\mu). \quad (2.5)$$

2.2.1 Mitglieder der einparametrischen linearen Exponentialfamilie

Bei den beiden Beispielen, die nun betrachtet werden, kann für die Dispersion das multiplikative Modell $a(\phi) = a\phi$ verwendet werden, wobei das Gewicht a als bekannt angenommen wird.

- **Normalverteilung**

Sei $y \sim \mathcal{N}(\mu, \sigma^2)$ und $y \in \mathbb{R}$, dann kann die Dichtefunktion geschrieben werden als

$$\begin{aligned} f(y|\mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \\ &= \exp\left(\frac{y\mu - \mu^2/2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{\log(2\pi\sigma^2)}{2}\right). \end{aligned}$$

Wird als Parametrisierung $\theta = \mu$ und $\phi = \sigma^2$ gewählt, resultiert daraus die lineare einparametrische Exponentialfamilie mit

$$\begin{aligned} a &= 1 \\ b(\theta) &= \frac{\theta^2}{2} = \frac{\mu^2}{2} \\ c(y, \phi) &= -\frac{y^2}{2\sigma^2} - \frac{\log(2\pi\sigma^2)}{2}. \end{aligned}$$

Unter Anwendung von Satz 2.3 folgt für die ersten beiden Momente von y

$$\begin{aligned} \mathbb{E}(y) &= b'(\theta) = \theta = \mu \\ \text{Var}(y) &= a(\phi)b''(\theta) = \sigma^2, \end{aligned}$$

wobei σ^2 als bekannter Parameter angesehen wird.

- **Standardisierte Binomialverteilung**

Sei nun $my \sim \mathcal{B}(m, \pi)$ mit $y = 0, \frac{1}{m}, \frac{2}{m}, \dots, 1$, dann lässt sich die Wahrscheinlichkeitsfunktion darstellen als

$$\begin{aligned} f(y|m, \pi) &= \mathbb{P}(Y = y) = \mathbb{P}(mY = my) = \binom{m}{my} \pi^{my} (1-\pi)^{m-my} \\ &= \exp\left(\log\binom{m}{my} + my \log \pi + (m-my) \log(1-\pi)\right) \\ &= \exp\left(\frac{y \log \frac{\pi}{1-\pi} - \log \frac{1}{1-\pi}}{1/m} + \log\binom{m}{my}\right). \end{aligned}$$

Mit der Parametrisierung $\theta = \log \frac{\pi}{1-\pi}$ und $\phi = 1$ führt dies zur einparametrischen linearen Exponentialfamilie mit

$$\begin{aligned} a &= \frac{1}{m} \\ b(\theta) &= \log \frac{1}{1-\pi} = \log(1 + \exp(\theta)) \\ c(y, \phi) &= \log \binom{m}{my}. \end{aligned}$$

Schließlich ergibt sich mit Satz 2.3 als Erwartungswert und Varianz von y

$$\begin{aligned} E(y) &= b'(\theta) = \frac{\exp(\theta)}{1 + \exp(\theta)} = \pi \\ \text{Var}(y) &= a(\phi)b''(\theta) = \frac{1}{m} \frac{\exp(\theta)}{(1 + \exp(\theta))^2} = \frac{1}{m} \pi(1 - \pi). \end{aligned}$$

Es handelt sich hier um die standardisierte Binomialverteilung, da y relative Häufigkeiten beschreibt. Die absolute Häufigkeit my entspricht einer binomialverteilten Zufallsvariable.

2.3 Modell und Linkfunktion

Das Generalisierte Lineare Modell wird unter Annahme der Existenz von $E(y_i)$ und $\text{Var}(y_i)$ durch eine Parametrisierung der Form

$$\begin{aligned} y_i &\stackrel{\text{ind}}{\sim} \text{Exponentialfamilie}(\theta_i) \\ E(y_i) &= \mu_i = \mu(\theta_i) \\ \eta_i &= \mathbf{x}_i^T \boldsymbol{\beta} \\ g(\mu_i) &= \eta_i \end{aligned} \tag{2.6}$$

für $i = 1, \dots, n$ spezifiziert. Hierbei besteht die Response $\mathbf{y} = (y_1, \dots, y_n)^T$ aus n unabhängigen Komponenten mit Dichte- beziehungsweise Wahrscheinlichkeitsfunktion aus der linearen einparametrischen Exponentialfamilie mit Dispersionsparameter ϕ und unbekanntem Parameter θ_i , wobei $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^T$ den Vektor der zu schätzenden Parameter beschreibt. Die erklärenden Variablen sind in den Vektoren $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ enthalten, welche zusammengefasst die Designmatrix $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ ergeben. Weiters bezeichnet $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ den Vektor der zu schätzenden Parameter, $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$ den Vektor der linearen Prädiktoren und $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ den Erwartungswertvektor. Die Linkfunktion $g(\cdot)$ ist eine bekannte monotone und zweimal stetig differenzierbare Funktion, die vom Erwartungswert μ_i abhängt. Aus Satz 2.3 und (2.5) ist bekannt, dass die Varianz von y_i geschrieben werden kann als $\text{Var}(y_i) = a_i(\phi)V(\mu_i)$.

Zusätzlich sei die Dispersion von nun gegeben durch $a_i(\phi) = a_i\phi$.

Hier wird deutlich, was in Abschnitt 2.1 bereits erwähnt wurde. Die wesentlichen Unterschiede zum LM ergeben sich durch die Einführung der Exponentialfamilie und der Linkfunktion. Unter Verwendung der Exponentialfamilie ist die Varianz nicht mehr zwingend konstant, eine Abhängigkeit vom Erwartungswert ist möglich. Auch die Linkfunktion eröffnet neue Möglichkeiten. Nun wird nicht mehr der Erwartungswert, sondern eine Funktion des Erwartungswertes linear modelliert. Dies ist vor allem dann relevant, wenn die Response und folglich auch μ_i nicht aus den reellen Zahlen stammt, sondern, wie zum Beispiel im Falle von Raten und Proportionen, aus dem Einheitsintervall $[0, 1]$. Die Verwendung der Linkfunktion dient demnach dazu, dass der modellierte Erwartungswert aus einem sinnvollen und plausiblen Wertebereich stammt. Eine Übersicht zur Auswahl der passenden Linkfunktion bietet Pregibon (1980). Zur Berechnung des Erwartungswertes $\mu_i = g^{-1}(\eta_i)$ wird die inverse Linkfunktion benötigt, deren Existenz durch Stetigkeit und Monotonie von $g(\cdot)$ gegeben ist.

Jedem Mitglied der linearen einparametrischen Exponentialfamilie kann eine spezielle Linkfunktion, der sogenannte kanonische Link, zugeordnet werden. Unter Verwendung dieser Linkfunktion ist $\mathbf{X}^T \mathbf{y}$ eine suffiziente Statistik für den Parametervektor β . Der kanonische Link tritt auf, wenn

$$\boldsymbol{\theta} = \boldsymbol{\eta}$$

erfüllt ist, das heißt wenn der kanonische Parameter θ_i , für $i = 1, \dots, n$, direkt durch den linearen Prädiktor $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ modelliert wird. Aus Satz 2.3 und der Modellspezifikation in (2.6) ist bekannt, dass

$$\theta_i = b'^{-1}(\mu_i)$$

$$\mu_i = g^{-1}(\eta_i)$$

gilt, woraus schließlich

$$\theta_i = b'^{-1}(g^{-1}(\eta_i)) \tag{2.7}$$

folgt. Daher ist es naheliegend $g(\cdot) = b'^{-1}(\cdot)$ als kanonischen Link zu wählen.

2.3.1 Beispiele für kanonische Linkfunktionen

Zur Berechnung der kanonischen Linkfunktionen werden die Ergebnisse der Beispiele aus Abschnitt 2.2.1 verwendet.

• Normalverteilung

$$\mu_i = b'(\theta_i) = \theta_i.$$

Die Funktion $b'(\cdot)$ ist die identische Abbildung, daher ist auch der kanonische Link die Identität und es gilt

$$\mu_i = g^{-1}(\eta_i) = \eta_i$$

für $i = 1, \dots, n$. Bei normalverteilten Response-Variablen wird der Erwartungswert unter Verwendung der kanonischen Linkfunktion daher linear modelliert. Das stimmt mit den Annahmen des Linearen Modells überein.

• **Standardisierte Binomialverteilung**

$$\mu_i = b'(\theta_i) = \frac{\exp(\theta_i)}{1 + \exp(\theta_i)}.$$

Daraus resultiert als kanonischer Link

$$g(\mu_i) = \log \frac{\mu_i}{1 - \mu_i} = \theta_i,$$

genannt Logit-Link.

2.4 Maximum Likelihood Schätzung

Um den unbekannt Parametervektor $\beta = (\beta_1, \dots, \beta_p)^T$ zu schätzen, wird die Maximum Likelihood Methode verwendet. Dazu wird zunächst die Log-Likelihoodfunktion einer unabhängigen Stichprobe $\mathbf{y} = (y_1, \dots, y_n)^T$ betrachtet, wobei alle y_i demselben Mitglied der Exponentialfamilie zugeordnet werden können, mit konstanter Dispersi-
on ϕ und unbekannt Parameter θ_i . Die Parameter werden wie in (2.6) zum Vektor $\theta = (\theta_1, \dots, \theta_n)^T$ zusammengefasst. Wegen der Unabhängigkeit der Response-Variablen, lässt sich die Log-Likelihood schreiben als

$$l(\theta|\mathbf{y}) = \log f(\mathbf{y}|\theta) = \sum_{i=1}^n \left(\frac{y_i \theta_i - b(\theta_i)}{a_i \phi} + c(y_i, \phi) \right). \quad (2.8)$$

Aus (2.6) und (2.7) geht hervor, dass der Erwartungswert μ_i für $i = 1, \dots, n$ und der Parametervektor θ von β abhängen. Deshalb gilt bei Ableitung der Log-Likelihoodfunktion nach β

$$\frac{\partial l(\theta(\beta)|\mathbf{y})}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial \log f(y_i|\theta_i)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial \log f(y_i|\theta_i)}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j}$$

für $j = 1, \dots, p$. Unter Anwendung der Modellannahmen in (2.6) ergibt sich

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{\partial \mu_i / \partial \theta_i} = \frac{1}{\partial b'(\theta_i) / \partial \theta_i} = \frac{1}{b''(\theta_i)} = \frac{1}{V(\mu_i)},$$

ebenso wie

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial g(\mu_i)} x_{ij} = \frac{x_{ij}}{g'(\mu_i)}.$$

Daraus resultiert schließlich als Scorefunktion

$$\frac{\partial l(\theta(\beta)|\mathbf{y})}{\partial \beta_j} = \sum_{i=1}^n \frac{y_i - \mu_i}{a_i \phi V(\mu_i)} \frac{x_{ij}}{g'(\mu_i)}, \quad j = 1, \dots, p. \quad (2.9)$$

Für die kanonische Linkfunktion $g(\cdot) = b'^{-1}(\cdot)$ kann dieser Ausdruck noch vereinfacht werden. Da $\mu_i = b'(\theta_i)$ laut Satz 2.3 und $g(\mu_i) = \theta_i$ laut (2.7), folgt für die erste Ableitung des kanonischen Links

$$g'(\mu_i) = \frac{\partial g(\mu_i)}{\partial \mu_i} = \frac{\partial \theta_i}{\partial b'(\theta_i)} = \frac{1}{V(\mu_i)}. \quad (2.10)$$

Daher hat die Scorefunktion unter Verwendung einer kanonischen Linkfunktion die Form

$$\frac{\partial l(\boldsymbol{\theta}(\boldsymbol{\beta})|\mathbf{y})}{\partial \beta_j} = \sum_{i=1}^n \frac{y_i - \mu_i}{a_i \phi} x_{ij}, \quad j = 1, \dots, p. \quad (2.11)$$

Der Maximum Likelihood Schätzer (MLE) $\hat{\boldsymbol{\beta}}$ ist im allgemeinen Fall als Nullstelle der Scorefunktion (2.9) und unter Verwendung der kanonischen Linkfunktion als Nullstelle von (2.11) definiert. Nur beim Linearen Modell sind die Gleichungssysteme linear in $\boldsymbol{\beta}$, deshalb kann die Lösung im Allgemeinen nur über numerische Verfahren ermittelt werden. McCullagh und Nelder (1989) beschreiben die Herleitung des Schätzers unter Verwendung der Newton-Raphson-Methode.

2.4.1 Iterative Weighted Least Squares

Um die Nullstelle der Scorefunktion zu berechnen, wird zunächst die Iterationsvorschrift der Newton-Methode betrachtet. Dabei erfolgt die Ermittlung der Iterierten $\boldsymbol{\beta}^{(t+1)}$ durch

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + \left(-\frac{\partial^2 l(\boldsymbol{\theta}(\boldsymbol{\beta})|\mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right)^{-1} \frac{\partial l(\boldsymbol{\theta}(\boldsymbol{\beta})|\mathbf{y})}{\partial \boldsymbol{\beta}}, \quad t = 0, 1, \dots,$$

wobei die Ableitungen der Log-Likelihood an der Stelle $\boldsymbol{\beta}^{(t)}$ ausgewertet werden. Aus Notationsgründen werden nun die Matrizen $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$ und $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$ mit den Komponenten

$$d_i = g'(\mu_i)$$

$$w_i = \frac{1}{a_i V(\mu_i) (g'(\mu_i))^2}$$

eingeführt. Dadurch kann der Gradient der Log-Likelihoodfunktion dargestellt werden als

$$\frac{\partial l(\boldsymbol{\theta}(\boldsymbol{\beta})|\mathbf{y})}{\partial \boldsymbol{\beta}} = \frac{1}{\phi} \mathbf{X}^T \mathbf{D} \mathbf{W} (\mathbf{y} - \boldsymbol{\mu}),$$

woraus schließlich unter Anwendung der Identität $\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}^T} = \mathbf{D}^{-1}$ die negative Hessematrix

$$\begin{aligned} -\frac{\partial^2 l(\boldsymbol{\theta}(\boldsymbol{\beta})|\mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= -\frac{\partial^2 l(\boldsymbol{\theta}(\boldsymbol{\beta})|\mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\eta}^T} \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}} \\ &= -\frac{1}{\phi} \mathbf{X}^T \left(\frac{\partial \mathbf{D} \mathbf{W}}{\partial \boldsymbol{\eta}^T} (\mathbf{y} - \boldsymbol{\mu}) - \mathbf{D} \mathbf{W} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}^T} \right) \mathbf{X} \\ &= \frac{1}{\phi} \mathbf{X}^T \left(\mathbf{W} - \frac{\partial \mathbf{D} \mathbf{W}}{\partial \boldsymbol{\eta}^T} (\mathbf{y} - \boldsymbol{\mu}) \right) \mathbf{X} \end{aligned}$$

resultiert. Um den Ausdruck in der Klammer zu vereinfachen, werden die Elemente

$$w_i^* = w_i - \frac{\partial d_i w_i}{\partial \eta_i} (y_i - \mu_i), \quad i = 1, \dots, n, \quad (2.12)$$

eingeführt und zur Diagonalmatrix $\mathbf{W}^* = \text{diag}(w_1^*, \dots, w_n^*)$ mit der Eigenschaft $E(\mathbf{W}^*) = \mathbf{W}$ zusammengefasst.

Hierbei erfolgt die Berechnung von w_i^* mit Hilfe von

$$\begin{aligned} \frac{\partial d_i w_i}{\partial \eta_i} &= \frac{\partial}{\partial \eta_i} \left(\frac{1}{a_i V(\mu_i) g'(\mu_i)} \right) \\ &= - \frac{a_i V'(\mu_i) g'(\mu_i) \frac{\partial \mu_i}{\partial \eta_i} + a_i V(\mu_i) g''(\mu_i) \frac{\partial \mu_i}{\partial \eta_i}}{(a_i V(\mu_i) g'(\mu_i))^2} \\ &= - \frac{V'(\mu_i) g'(\mu_i) + V(\mu_i) g''(\mu_i)}{a_i (V(\mu_i))^2 (g'(\mu_i))^3}, \end{aligned} \quad (2.13)$$

wobei die letzte Identität aus

$$\frac{\partial \mu_i}{\partial \eta_i} = \frac{1}{\partial g(\mu_i) / \partial \mu_i} = \frac{1}{g'(\mu_i)}$$

folgt.

Nun sind alle Bausteine für den Newton-Schritt vorhanden. Zur Berechnung der Iterierten $\beta^{(t+1)}$ werden die Ableitungen der Log-Likelihood an der Stelle $\beta^{(t)}$ betrachtet. Um dies zu verdeutlichen, werden auch alle von $\beta^{(t)}$ abhängigen Komponenten mit dem Iterationsindex (t) gekennzeichnet. Zusammengefasst resultiert als Iterationsvorschrift der Newton-Methode

$$\beta^{(t+1)} = \beta^{(t)} + (\mathbf{X}^T \mathbf{W}^{*(t)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{D}^{(t)} \mathbf{W}^{(t)} (\mathbf{y} - \boldsymbol{\mu}^{(t)}), \quad t = 0, 1, \dots \quad (2.14)$$

Hierbei ist anzumerken, dass (2.14) nicht vom Dispersionsparameter ϕ abhängt. Werden nun zusätzlich sogenannte Pseudobeobachtungen

$$\mathbf{z}^{(t)} = \mathbf{X} \beta^{(t)} + (\mathbf{W}^{*(t)})^{-1} \mathbf{D}^{(t)} \mathbf{W}^{(t)} (\mathbf{y} - \boldsymbol{\mu}^{(t)})$$

eingeführt, so kann (2.14) geschrieben werden als

$$\beta^{(t+1)} = (\mathbf{X}^T \mathbf{W}^{*(t)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{*(t)} \mathbf{z}^{(t)}, \quad t = 0, 1, \dots \quad (2.15)$$

Daher kann die Iterationsvorschrift auch als Iterative Weighted Least Squares Prozedur (IWLS) interpretiert werden (Green, 1984).

2.4.2 IWLS mit kanonischer Linkfunktion und Fisher-Scoring

Für kanonische Linkfunktionen lassen sich (2.14) und (2.15) noch weiter vereinfachen. Dazu werden die Identität

$$g'(\mu_i) = \frac{1}{V(\mu_i)}$$

aus (2.10) und die daraus folgende zweite Ableitung

$$g''(\mu_i) = -\frac{V'(\mu_i)}{V^2(\mu_i)}$$

für $i = 1, \dots, n$ betrachtet. Eingesetzt in (2.13) resultiert daraus

$$\begin{aligned} \frac{\partial d_i w_i}{\partial \eta_i} &= -\frac{V'(\mu_i)g'(\mu_i) + V(\mu_i)g''(\mu_i)}{a_i(V(\mu_i))^2(g'(\mu_i))^3} \\ &= -\frac{V'(\mu_i)\frac{1}{V(\mu_i)} - V(\mu_i)\frac{V'(\mu_i)}{V^2(\mu_i)}}{a_i(V(\mu_i))^2\left(\frac{1}{V(\mu_i)}\right)^3} \\ &= -\frac{\frac{V'(\mu_i)}{V(\mu_i)} - \frac{V'(\mu_i)}{V(\mu_i)}}{\frac{a_i}{V(\mu_i)}} = 0. \end{aligned}$$

Folglich gilt für die in (2.12) eingeführten Komponenten $w_i^* = w_i$, was schließlich zur Identität $\mathbf{W}^* = \mathbf{W}$ führt. Für kanonische Linkfunktionen resultiert daher

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + (\mathbf{X}^T \mathbf{W}^{(t)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{D}^{(t)} \mathbf{W}^{(t)} (\mathbf{y} - \boldsymbol{\mu}^{(t)}), \quad t = 0, 1, \dots,$$

als Iterationsvorschrift des Newton-Verfahrens. Zusätzlich können Pseudobeobachtungen, in diesem Fall vereinfacht zu

$$\mathbf{z}^{(t)} = \mathbf{X} \boldsymbol{\beta}^{(t)} + \mathbf{D}^{(t)} (\mathbf{y} - \boldsymbol{\mu}^{(t)}),$$

eingeführt werden, woraus schließlich die IWLS-Notation

$$\boldsymbol{\beta}^{(t+1)} = (\mathbf{X}^T \mathbf{W}^{(t)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(t)} \mathbf{z}^{(t)}, \quad t = 0, 1, \dots, \quad (2.16)$$

folgt.

Um das Schema auch ohne kanonische Linkfunktion zu vereinfachen, schlagen McCullagh und Nelder (1989) die Anwendung der Fisher-Scoring Methode vor. Dabei wird die Hessematrix der Log-Likelihoodfunktion durch ihren Erwartungswert

$$\mathbb{E} \left(\frac{1}{\phi} \mathbf{X}^T \mathbf{W}^{*(t)} \mathbf{X} \right) = \frac{1}{\phi} \mathbf{X}^T \mathbf{W}^{(t)} \mathbf{X}$$

ersetzt. Wie in (2.16) resultiert daraus die Iterationsvorschrift

$$\boldsymbol{\beta}^{(t+1)} = (\mathbf{X}^T \mathbf{W}^{(t)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(t)} \mathbf{z}^{(t)}, \quad t = 0, 1, \dots, \quad (2.17)$$

mit den Pseudobeobachtungen

$$\mathbf{z}^{(t)} = \mathbf{X} \boldsymbol{\beta}^{(t)} + \mathbf{D}^{(t)} (\mathbf{y} - \boldsymbol{\mu}^{(t)}).$$

Bemerkung:

Fahrmeir und Kaufmann (1985) zeigten, dass für den mit Hilfe von (2.16) bzw. (2.17) bestimmten Maximum Likelihood Schätzer $\hat{\beta}$

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} \mathcal{N}_p \left(\mathbf{0}, n\phi(\mathbf{X}^T \mathbf{W}(\beta) \mathbf{X})^{-1} \right)$$

gilt. Hierbei sei anzumerken, dass die Diagonalmatrix \mathbf{W} an der Stelle β , also im wahren Parameter, ausgewertet wird.

Bemerkung:

Um die Varianz-Kovarianzmatrix von $\hat{\beta}$ zu schätzen, wird das asymptotische Ergebnis der Varianz-Kovarianzmatrix an der Stelle $\hat{\beta}$ ausgewertet, das heißt

$$\widehat{\text{Var}}(\hat{\beta}) = \phi(\mathbf{X}^T \mathbf{W}(\hat{\beta}) \mathbf{X})^{-1}.$$

2.5 Anpassungsgüte

Nach der Schätzung der Modellparameter, ist vor allem die Güte der Modellanpassung von Interesse. Dabei wird gemessen, wie weit der geschätzte Erwartungswert $\hat{\mu}$ von der beobachteten Response \mathbf{y} abweicht. Um einen ordentlichen Vergleich aufzustellen, wird jedoch ein geeignetes Maß benötigt. In diesem Zusammenhang empfehlen McCullagh und Nelder (1989) die Betrachtung der Pearson Statistik und der Deviance.

2.5.1 Pearson Statistik

Definition 2.4. (Fehlerquadratsumme)

Seien y_1, \dots, y_n unabhängige Beobachtungen und $\hat{\mu}_i$ die korrespondierenden, von $\hat{\beta}$ abhängigen, geschätzten Erwartungswerte. Dann ist die Fehlerquadratsumme $\text{SSE}(\hat{\beta})$ definiert als

$$\text{SSE}(\hat{\beta}) = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2.$$

Definition 2.5. (Pearson Statistik)

Für n unabhängige Beobachtungen y_1, \dots, y_n mit Erwartungswerten μ_i und Varianzen $\text{Var}(y_i) = a_i \phi V(\mu_i)$ ist die generalisierte Pearson Statistik definiert als

$$X^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{a_i V(\hat{\mu}_i)}.$$

Für normalverteilte Response-Variablen y_1, \dots, y_n mit $a_i = 1$ und $\phi = \sigma^2$ gilt laut Abschnitt 2.2.1 $V(\mu_i) = 1$, woraus für die Pearson Statistik

$$X^2 = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 = \text{SSE}(\hat{\beta})$$

folgt. Für normalverteilte Beobachtungen ist X^2/σ^2 demnach exakt χ_{n-p}^2 -verteilt, für alle anderen Verteilungen gilt dies eher nur approximativ.

Die Pearson Statistik kann nicht nur als Maß der Anpassungsgüte verwendet werden, sondern auch zur Schätzung des Dispersionsparameters ϕ . Bei bekanntem Parameter β ist

$$\frac{1}{n} \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{a_i V(\mu_i)}$$

erwartungstreu für ϕ . Da β jedoch im Allgemeinen nicht gegeben ist, wird

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{a_i V(\hat{\mu}_i)} = \frac{1}{n-p} X^2$$

als Schätzer für den Dispersionsparameter verwendet.

Bemerkung:

Im Zusammenhang mit der Pearson Statistik können auch Residuen definiert werden. Die Wurzel der Summanden von X^2

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{a_i V(\hat{\mu}_i)}} \tag{2.18}$$

werden Pearson Residuen genannt.

2.5.2 Deviance

Ein weiteres Maß zur Messung der Anpassungsgüte stellt die Deviance dar. Dabei wird ein Likelihood Quotienten Test bezüglich der Nullhypothese $H_0 : \boldsymbol{\mu} = g^{-1}(\boldsymbol{\eta})$ und der Alternativhypothese $H_1 : \boldsymbol{\mu} \neq g^{-1}(\boldsymbol{\eta})$ betrachtet. Als Likelihood Quotienten Teststatistik ergibt sich

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

wobei die Likelihoodfunktion $L(\boldsymbol{\mu}|\mathbf{y})$ hierbei nicht in Abhängigkeit von $\boldsymbol{\theta}$, sondern von $\boldsymbol{\mu}$ dargestellt wird. Unter H_0 erreicht die Likelihoodfunktion an der Stelle $\hat{\boldsymbol{\mu}} = g^{-1}(\mathbf{X}\hat{\boldsymbol{\beta}})$ ihr Maximum. Im uneingeschränkten Fall wird die Likelihoodfunktion an der Stelle $\hat{\boldsymbol{\mu}} = \mathbf{y}$ maximal. Das heißt es werden gleich viele Parameter β_i wie Beobachtungen y_i in das Modell aufgenommen. Ein solches Modell wird volles oder saturiertes Modell genannt. In der Praxis sind diese Modelle jedoch uninteressant, da sie die Daten nicht zusammenfassen, sondern lediglich reproduzieren. Nach der Bestimmung der für die Teststatistik notwendigen Maximalstellen, kann der Quotient vereinfacht werden zu

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

Unter manchmal kritischen Regularitätsbedingungen ergibt sich daraus die asymptotisch χ_d^2 -verteilte Größe

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

Der Freiheitsgrad d ergibt sich hierbei als Differenz der Dimensionen der Parameterräume bezüglich der beiden Hypothesen, für gewöhnlich gilt $d = n - p$.

Definition 2.6. (Skalierte Deviance)

Sei $\mathbf{y} = (y_1, \dots, y_n)^T$ ein Vektor mit n unabhängigen Beobachtungen und $\hat{\boldsymbol{\mu}}$ der Maximum Likelihood Schätzer bezüglich der Log-Likelihoodfunktion $l(\boldsymbol{\mu}|\mathbf{y})$. Dann ist die skalierte Deviance definiert als

$$\frac{1}{\phi} D(\mathbf{y}|\hat{\boldsymbol{\mu}}) = -2(l(\hat{\boldsymbol{\mu}}|\mathbf{y}) - l(\mathbf{y}|\mathbf{y})),$$

wobei ϕ den Dispersionsparameter darstellt.

Bemerkung:

Der Ausdruck $l(\mathbf{y}|\mathbf{y})$ in Definition 2.6 ist bei gegebener Response ein fester Wert. Daraus folgt, dass die Maximalstelle $\hat{\boldsymbol{\mu}}$ der Log-Likelihoodfunktion gleichzeitig die Minimalstelle der Deviance darstellt.

Bemerkung:

Sei $\tilde{\boldsymbol{\theta}}$ der unter dem vollen Modell geschätzte Wert für $\boldsymbol{\theta}$. Dann gilt für die lineare einparametrische Exponentialfamilie

$$\begin{aligned} l(\hat{\boldsymbol{\mu}}|\mathbf{y}) - l(\mathbf{y}|\mathbf{y}) &= \sum_{i=1}^n \left(\frac{y_i \hat{\theta}_i - b(\hat{\theta}_i)}{a_i \phi} + c(y_i, \phi) \right) - \sum_{i=1}^n \left(\frac{y_i \tilde{\theta}_i - b(\tilde{\theta}_i)}{a_i \phi} + c(y_i, \phi) \right) \\ &= \frac{1}{\phi} \sum_{i=1}^n \frac{y_i \hat{\theta}_i - y_i \tilde{\theta}_i - b(\hat{\theta}_i) + b(\tilde{\theta}_i)}{a_i}. \end{aligned}$$

Als Darstellung der Deviance folgt daher

$$\begin{aligned} D(\mathbf{y}|\hat{\boldsymbol{\mu}}) &= -2\phi(l(\hat{\boldsymbol{\mu}}|\mathbf{y}) - l(\mathbf{y}|\mathbf{y})) \\ &= \sum_{i=1}^n \frac{-2(y_i \hat{\theta}_i - y_i \tilde{\theta}_i - b(\hat{\theta}_i) + b(\tilde{\theta}_i))}{a_i} \\ &= \sum_{i=1}^n d_i, \end{aligned} \tag{2.19}$$

mit passender Wahl der Summanden d_i als

$$d_i = \frac{-2(y_i \hat{\theta}_i - y_i \tilde{\theta}_i - b(\hat{\theta}_i) + b(\tilde{\theta}_i))}{a_i}, \quad i = 1, \dots, n.$$

Bemerkung:

Die neue Schreibweise der Deviance in (2.19) motiviert zu einer weiteren Form von Residuen, genannt Deviance Residuen. Diese sind definiert als

$$r_i^D = \text{sign}(y_i - \hat{\mu}_i) \sqrt{d_i}.$$

2.5.3 Analysis of Deviance

Ähnlich zur ANOVA bei Linearen Modellen, kann bei den Generalisierten Linearen Modellen die Analysis of Deviance durchgeführt werden. Dabei wird sukzessive jenes Modell bestimmt, welches die beste Anpassungsgüte aufweist. In jedem Schritt werden zwei Submodelle zu Hypothesen der Form

$$H_0 : \eta_i = \beta_1 x_{i1} + \cdots + \beta_q x_{iq}$$

$$H_1 : \eta_i = \beta_1 x_{i1} + \cdots + \beta_q x_{iq} + \cdots + \beta_p x_{ip}$$

für $i = 1, \dots, n$ betrachtet. Sei M_1 das mittels H_1 spezifizierte Modell und M_0 das aus H_0 resultierende Submodell. Hierbei sei anzumerken, dass M_0 deshalb ein Submodell von M_1 darstellt, da es als Spezialfall aus M_1 resultiert, wenn die Parameter $\beta_{q+1}, \dots, \beta_p$ Null gesetzt werden.

Bei den sukzessive erzeugten Submodellen wird dann eine Parameterschätzung durchgeführt. Seien $\hat{\boldsymbol{\mu}}_0$ und $\hat{\boldsymbol{\mu}}_1$ die beiden geschätzten Erwartungswerte unter M_0 und M_1 , dann folgt mit Definition 2.6

$$D(\mathbf{y}|\hat{\boldsymbol{\mu}}_0) = -2\phi(l(\hat{\boldsymbol{\mu}}_0|\mathbf{y}) - l(\mathbf{y}|\mathbf{y}))$$

$$D(\mathbf{y}|\hat{\boldsymbol{\mu}}_1) = -2\phi(l(\hat{\boldsymbol{\mu}}_1|\mathbf{y}) - l(\mathbf{y}|\mathbf{y})).$$

Die Differenz

$$D(\mathbf{y}|\hat{\boldsymbol{\mu}}_0) - D(\mathbf{y}|\hat{\boldsymbol{\mu}}_1) = -2\phi(l(\hat{\boldsymbol{\mu}}_0|\mathbf{y}) - l(\hat{\boldsymbol{\mu}}_1|\mathbf{y}))$$

beschreibt den Unterschied in der Güte der Modellanpassung und entspricht nach Division durch ϕ der Likelihood Quotienten Teststatistik der gegebenen Hypothesen. Für bekanntes ϕ gilt deshalb asymptotisch

$$\frac{D(\mathbf{y}|\hat{\boldsymbol{\mu}}_0) - D(\mathbf{y}|\hat{\boldsymbol{\mu}}_1)}{\phi} \sim \chi_{p-q}^2.$$

Ist der Dispersionsparameter ϕ jedoch unbekannt, wird stattdessen

$$\frac{(D(\mathbf{y}|\hat{\boldsymbol{\mu}}_0) - D(\mathbf{y}|\hat{\boldsymbol{\mu}}_1))/(p - q)}{D(\mathbf{y}|\hat{\boldsymbol{\mu}}_1)/(n - p)} \sim F_{p-q, n-p}$$

zur Messung der Anpassungsgüte verwendet.

2.6 Generalisierte Hebelwirkung

Ergänzend zur Anpassungsgüte des Modells sind möglicherweise auch Informationen über jene Datenpunkte interessant, welche einen großen Einfluss auf die Schätzung der Parameter ausüben. Um ebendiese Response-Werte zu finden, könnte die von Wei, Hu und Fung (1998) stammende Idee der generalisierten Hebelwirkung (generalized leverage) von Vorteil sein.

Ausgehend von den Linearen Modellen versuchen Wei et al. (1998) das Prinzip der Hebelpunkte auch auf allgemeinere Modelle und Schätzer auszuweiten. Bei den Linearen Modellen kann der Eintrag h_{ij} der Hutmatrix \mathbf{H} dargestellt werden als

$$h_{ij} = \frac{\partial \hat{\mu}_i}{\partial y_j}$$

und spiegelt somit den Einfluss der Beobachtung y_j auf den Maximum Likelihood Schätzer $\hat{\mu}_i$ des Erwartungswertes wieder. Diese Idee soll nun verallgemeinert werden.

Sei dazu $\mathbf{y} = (y_1, \dots, y_n)^T$ ein Response-Vektor bestehend aus n unabhängigen Beobachtungen mit den Erwartungswerten $E(y_i) = \mu_i$, welche zu dem Vektor $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ zusammengefasst werden. Weiters seien $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ die Vektoren der erklärenden Variablen, $\boldsymbol{\alpha} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ bezeichne den zu schätzende Parametervektor und $\tilde{\boldsymbol{\alpha}} = \tilde{\boldsymbol{\alpha}}(\mathbf{y})$ dessen M-Schätzer, der eine Verallgemeinerung des Maximum Likelihood Schätzers darstellt (vgl. Staudte und Sheather, 1990). Hierbei sei anzumerken, dass $\boldsymbol{\beta}$ jenen Parametervektor beschreibt, welcher zur Modellierung von $\boldsymbol{\mu}$ benötigt wird und $\boldsymbol{\gamma}$ die Störparameter zusammenfasst. Die generalisierte Hebelwirkung $\mathbf{GL}(\tilde{\boldsymbol{\alpha}})$ ist dann definiert als

$$\mathbf{GL}(\tilde{\boldsymbol{\alpha}}) = \frac{\partial \tilde{\boldsymbol{\mu}}}{\partial \mathbf{y}^T}, \quad (2.20)$$

wobei der Erwartungswertvektor $\boldsymbol{\mu}$ bei den in dieser Arbeit diskutierten Regressionsmodellen eine Funktion in $\boldsymbol{\beta}$ darstellt und der Schätzwert $\tilde{\boldsymbol{\mu}}$ je nach betrachtetem Regressionsmodell entweder vom M-Schätzer $\tilde{\boldsymbol{\alpha}}$ oder lediglich vom Schätzwert $\tilde{\boldsymbol{\beta}}$ abhängt.

Die Stelle (i, j) der Matrix $\mathbf{GL}(\tilde{\boldsymbol{\alpha}})$ bewertet die Abhängigkeit des i -ten Vorhersagewertes von der j -ten Response. Beobachtungen mit großem $\mathbf{GL}_{ij}(\tilde{\boldsymbol{\alpha}})$ weisen darauf hin, dass die Beobachtung y_j einen großen Einfluss auf den Schätzwert $\tilde{\mu}_i$ ausübt und y_j somit einen Hebelpunkt darstellt.

Um die generalisierte Hebelwirkung für M-Schätzer und somit auch für den Maximum Likelihood Schätzer schnell und effizient berechnen zu können, entwickelten Wei et al. (1998) folgendes Lemma.

Lemma 2.7. (Wei et al., 1998)

Sei $\tilde{\boldsymbol{\alpha}}(\mathbf{y})$ der eindeutige M-Schätzer, der die Zielfunktion

$$Q(\boldsymbol{\alpha}|\mathbf{y}) = \sum_{i=1}^n q_i(f(\mathbf{x}_i, \boldsymbol{\beta})|\boldsymbol{\gamma}) \quad (2.21)$$

minimiert, mit $\boldsymbol{\alpha} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$, $\boldsymbol{\gamma}$ als Vektor der Störparameter und $\mu_i = f(\mathbf{x}_i, \boldsymbol{\beta})$. Falls (2.21) zweimal stetig differenzierbar ist bezüglich $\boldsymbol{\alpha}$ und \mathbf{y} , so kann die Matrix der generalisierten Hebelwirkung geschrieben werden als

$$\mathbf{GL}(\tilde{\boldsymbol{\alpha}}) = \left\{ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\alpha}^T} \left(-\frac{\partial^2 Q(\boldsymbol{\alpha}|\mathbf{y})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T} \right)^{-1} \frac{\partial^2 Q(\boldsymbol{\alpha}|\mathbf{y})}{\partial \boldsymbol{\alpha} \partial \mathbf{y}^T} \right\}_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}(\mathbf{y})}. \quad (2.22)$$

Beweis:

Ausgehend von (2.20) folgt für die generalisierte Hebelwirkung die Darstellung

$$GL(\tilde{\alpha}) = \frac{\partial \tilde{\mu}}{\partial \mathbf{y}^T} = \left\{ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\alpha}^T} \frac{\partial \tilde{\boldsymbol{\alpha}}}{\partial \mathbf{y}^T} \right\}_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}(\mathbf{y})}. \quad (2.23)$$

Da der M-Schätzer $\tilde{\boldsymbol{\alpha}}(\mathbf{y})$ die Zielfunktion (2.21) optimiert, gilt für den Scorevektor

$$\left\{ \frac{\partial Q(\boldsymbol{\alpha}|\mathbf{y})}{\partial \boldsymbol{\alpha}} \right\}_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}(\mathbf{y})} = \mathbf{0}. \quad (2.24)$$

Um zur gewünschten Formel zu gelangen, müssen beide Seiten von (2.24) nach \mathbf{y} abgeleitet werden. Dabei kann die linke Seite der Gleichung als Funktion in $\boldsymbol{\alpha}$ und \mathbf{y} gesehen werden, sei also

$$F(\boldsymbol{\alpha}, \mathbf{y}) = \frac{\partial Q(\boldsymbol{\alpha}|\mathbf{y})}{\partial \boldsymbol{\alpha}}.$$

Unter Verwendung von (2.24) folgt dann für den Gradienten von $F(\boldsymbol{\alpha}, \mathbf{y})$ ausgewertet an der Stelle $\tilde{\boldsymbol{\alpha}}(\mathbf{y})$

$$\left\{ \frac{\partial F(\boldsymbol{\alpha}, \mathbf{y})}{\partial \boldsymbol{\alpha}^T} \frac{\partial \tilde{\boldsymbol{\alpha}}(\mathbf{y})}{\partial \mathbf{y}^T} + \frac{\partial F(\boldsymbol{\alpha}, \mathbf{y})}{\partial \mathbf{y}^T} \frac{\partial \mathbf{y}}{\partial \mathbf{y}^T} \right\}_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}(\mathbf{y})} = \mathbf{0}$$

\Leftrightarrow

$$\left\{ \frac{\partial^2 Q(\boldsymbol{\alpha}|\mathbf{y})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T} \frac{\partial \tilde{\boldsymbol{\alpha}}(\mathbf{y})}{\partial \mathbf{y}^T} + \frac{\partial^2 Q(\boldsymbol{\alpha}|\mathbf{y})}{\partial \boldsymbol{\alpha} \partial \mathbf{y}^T} \right\}_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}(\mathbf{y})} = \mathbf{0}$$

\Leftrightarrow

$$\frac{\partial \tilde{\boldsymbol{\alpha}}(\mathbf{y})}{\partial \mathbf{y}^T} = \left\{ \left(-\frac{\partial^2 Q(\boldsymbol{\alpha}|\mathbf{y})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T} \right)^{-1} \frac{\partial^2 Q(\boldsymbol{\alpha}|\mathbf{y})}{\partial \boldsymbol{\alpha} \partial \mathbf{y}^T} \right\}_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}(\mathbf{y})}.$$

Wird dieses Ergebnis in (2.23) eingesetzt, so resultiert für die generalisierte Hebelwirkung der gewünschte Ausdruck

$$GL(\tilde{\alpha}) = \left\{ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\alpha}^T} \left(-\frac{\partial^2 Q(\boldsymbol{\alpha}|\mathbf{y})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T} \right)^{-1} \frac{\partial^2 Q(\boldsymbol{\alpha}|\mathbf{y})}{\partial \boldsymbol{\alpha} \partial \mathbf{y}^T} \right\}_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}(\mathbf{y})}.$$

□

Bemerkung:

Die Maximum Likelihood Schätzung ist ein Spezialfall der M-Schätzung, da die zu maximierende Log-Likelihoodfunktion als negative Zielfunktion (2.21) interpretiert werden kann. Daraus folgt, dass Lemma 2.7 auch für Maximum Likelihood Schätzer hält. Die generalisierte Hebelwirkung des MLE $\hat{\boldsymbol{\alpha}}$ kann somit unter Verwendung von (2.22) über den Ausdruck

$$GL(\hat{\alpha}) = \left\{ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\alpha}^T} \left(-\frac{\partial^2 l(\boldsymbol{\alpha}|\mathbf{y})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T} \right)^{-1} \frac{\partial^2 l(\boldsymbol{\alpha}|\mathbf{y})}{\partial \boldsymbol{\alpha} \partial \mathbf{y}^T} \right\}_{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}(\mathbf{y})}. \quad (2.25)$$

ermittelt werden.

Bemerkung:

Bei der Linearen Regression besteht α lediglich aus dem Parametervektor β , da keine Störparameter vorkommen. Für die generalisierte Hebelwirkung heißt das $\mathbf{GL}(\hat{\alpha}) = \mathbf{GL}(\hat{\beta})$. Unter den Annahmen eines LM sind bei beobachtetem Response-Vektor \mathbf{y} und gegebener Designmatrix \mathbf{X} die Hutmatrix \mathbf{H} und die Schätzung $\hat{\beta}$ gegeben durch

$$\mathbf{H} = \mathbf{X} \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T,$$

$$\hat{\beta} = \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{y},$$

wobei natürlich die Beziehung

$$\mathbb{E}(\mathbf{y}) = \boldsymbol{\mu} = \mathbf{X}\beta$$

gilt. Mit (2.23) folgt sofort

$$\mathbf{GL}(\hat{\beta}) = \left\{ \frac{\partial \boldsymbol{\mu}}{\partial \beta^T} \frac{\partial \hat{\beta}}{\partial \mathbf{y}^T} \right\}_{\beta=\hat{\beta}(\mathbf{y})} = \mathbf{X} \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T = \mathbf{H}.$$

Unter Verwendung der Normalverteilungsannahme eines LM resultiert auch für Gleichung (2.25)

$$\begin{aligned} \mathbf{GL}(\hat{\beta}) &= \left\{ \frac{\partial \boldsymbol{\mu}}{\partial \beta^T} \left(-\frac{\partial^2 l(\beta|\mathbf{y})}{\partial \beta \partial \beta^T} \right)^{-1} \frac{\partial^2 l(\beta|\mathbf{y})}{\partial \beta \partial \mathbf{y}^T} \right\}_{\beta=\hat{\beta}(\mathbf{y})} \\ &= \mathbf{X} \left(\frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X} \right)^{-1} \left(\frac{1}{\sigma^2} \mathbf{X}^T \right) = \mathbf{H}. \end{aligned}$$

Somit wurde gezeigt, dass für LM die generalisierte Hebelwirkung mit der Hutmatrix übereinstimmt. Dieses Ergebnis war vorhersehbar, da zu Beginn dieses Abschnitts bereits erwähnt wurde, dass die Grundidee der generalisierten Hebelpunkte von dem LM und dessen Einträgen der Hutmatrix stammt.

Bemerkung:

Auch bei den Generalisierten Linearen Modellen stimmt die generalisierte Hebelwirkung mit der Hutmatrix überein. Zusätzlich gilt erneut $\mathbf{GL}(\hat{\alpha}) = \mathbf{GL}(\hat{\beta})$. Hier kommt zwar die Dispersion ϕ als Störparameter ins Spiel, da aber weder der Erwartungswert, noch der Schätzwert $\hat{\beta}$ und somit auch nicht der geschätzte Erwartungswert $\hat{\boldsymbol{\mu}}$ von ϕ abhängt, spielt die Dispersion bei der Berechnung der generalisierten Hebelwirkung keine Rolle.

3.1 Motivation

Wie in Abschnitt 2.1 erläutert, führten einige zu strenge Annahmen bezüglich Verteilung und Modellierung zur Erweiterung vom Linearen Modell zum Generalisierten Linearen Modell. Die wichtigsten Veränderungen waren die Einführung der linearen einparametrischen Exponentialfamilie und der Linkfunktion. Im nächsten Schritt soll nun betrachtet werden was geschieht, wenn nicht von einer konkreten Verteilung ausgegangen werden kann, sondern lediglich von einer gegebenen Beziehung zwischen Varianz und Erwartungswert. Im Speziellen kann diese Beziehung einem Mitglied der Exponentialfamilie zugeordnet werden, im Allgemeinen gilt dies jedoch nicht. Für den allgemeinen Fall wird nun die von Wedderburn (1974) eingeführte Quasi-Likelihoodfunktion betrachtet. Es sei anzumerken, dass in den folgenden Abschnitten die Annahme $a_i(\phi) = \phi$ für alle $i = 1, \dots, n$ getroffen wird.

3.2 Die Quasi-Likelihoodfunktion

Definition 3.1. (Quasi-Likelihoodfunktion, Wedderburn, 1974)

Sei y eine Zufallsvariable mit Erwartungswert $E(y) = \mu$ und Varianz $\text{Var}(y) = \phi V(\mu)$, mit bekannter Varianzfunktion $V(\cdot)$. Die Quasi-Likelihoodfunktion $q(\mu|y)$ ist dann definiert durch die Beziehung

$$\frac{\partial q(\mu|y)}{\partial \mu} = \frac{y - \mu}{\phi V(\mu)}, \quad (3.1)$$

wobei die Ableitung (3.1) als Quasi-Scorefunktion bezeichnet wird. Äquivalent kann die Quasi-Likelihoodfunktion als

$$q(\mu|y) = \int^{\mu} \frac{y - t}{\phi V(t)} dt + \text{Funktion in } y \text{ und } \phi \quad (3.2)$$

3.2. Die Quasi-Likelihoodfunktion

definiert werden.

Bemerkung:

Streng genommen müsste die in Definition 3.1 eingeführte Funktion $q(\mu|y)$ Log-Quasi-Likelihoodfunktion genannt werden. Da sie von Wedderburn (1974) jedoch als Quasi-Likelihoodfunktion definiert wurde, wird diese Namensgebung in dieser Arbeit übernommen.

Sei nun $\mathbf{y} = (y_1, \dots, y_n)$ ein Vektor mit n unabhängigen Response-Variablen und $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ jener Vektor, welcher die Erwartungswerte $\mu_i = E(y_i)$ zusammenfasst. Zusätzlich gelte $\text{Var}(y_i) = \phi V(\mu_i)$ für $i = 1, \dots, n$. Die Quasi-Likelihoodfunktion für den Vektor \mathbf{y} kann dann geschrieben werden als

$$q(\boldsymbol{\mu}|\mathbf{y}) = \sum_{i=1}^n q(\mu_i|y_i)$$

was schließlich bei Ableitung nach μ_j für $j = 1, \dots, n$ zur Quasi-Scorefunktion

$$\frac{\partial q(\boldsymbol{\mu}|\mathbf{y})}{\partial \mu_j} = \frac{y_j - \mu_j}{\phi V(\mu_j)} \quad (3.3)$$

führt. Aus (2.8) ist bekannt, dass die Log-Likelihoodfunktion der linearen einparametrischen Exponentialfamilie unter der Annahme $a_i(\phi) = \phi$ die Form

$$l(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^n \left(\frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi) \right)$$

aufweist. Daraus ergibt sich

$$\begin{aligned} \frac{\partial l(\boldsymbol{\mu}|\mathbf{y})}{\partial \mu_j} &= \frac{\partial \theta_j}{\partial \mu_j} \frac{\partial}{\partial \theta_j} \sum_{i=1}^n \left(\frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi) \right) \\ &= \frac{\partial \theta_j}{\partial \mu_j} \frac{y_j - b'(\theta_j)}{\phi}. \end{aligned}$$

Unter Anwendung von $\mu_j = b'(\theta_j)$ und

$$\frac{\partial \theta_j}{\partial \mu_j} = \frac{1}{\partial b'(\theta_j)/\partial \theta_j} = \frac{1}{b''(\theta_j)} = \frac{1}{V(\mu_j)}$$

folgt somit als Scorefunktion

$$\frac{\partial l(\boldsymbol{\mu}|\mathbf{y})}{\partial \mu_j} = \frac{y_j - \mu_j}{\phi V(\mu_j)}, \quad (3.4)$$

für $j = 1, \dots, n$. Für unabhängige Beobachtungen y_1, \dots, y_n ist die Quasi-Scorefunktion in (3.3) daher äquivalent zur Scorefunktion der linearen einparametrischen Exponentialfamilie in (3.4).

Die Äquivalenz von (3.3) und (3.4) führt zu der Frage ob eventuell auch die Quasi-Likelihoodfunktion mit der Log-Likelihoodfunktion der einparametrischen linearen Exponentialfamilie übereinstimmt. Der nachfolgende Satz von Wedderburn (1974) zeigt, wann dies zutrifft.

Satz 3.2. (Wedderburn, 1974)

Sei y eine Beobachtung mit $E(y) = \mu$ und $\text{Var}(y) = \phi V(\mu)$, mit bekannter Varianzfunktion $V(\cdot)$. Die Log-Likelihoodfunktion $l(\mu|y)$ hat genau dann die Eigenschaft

$$\frac{\partial l(\mu|y)}{\partial \mu} = \frac{y - \mu}{\phi V(\mu)}, \quad (3.5)$$

wenn die Verteilung von y aus der linearen einparametrischen Exponentialfamilie stammt.

Beweis:

„ \Rightarrow “

Sei zunächst (3.5) erfüllt. Unter Anwendung des Hauptsatzes der Differential- und Integralrechnung ergibt sich

$$\begin{aligned} l(\mu|y) &= \int \frac{\partial l(\mu|y)}{\partial \mu} d\mu = \int \frac{y - \mu}{\phi V(\mu)} d\mu \\ &= \frac{1}{\phi} \left(y \int \frac{1}{V(\mu)} d\mu - \int \frac{\mu}{V(\mu)} d\mu \right). \end{aligned}$$

Werden zusätzlich die Bezeichnungen

$$\begin{aligned} \theta &:= \int \frac{1}{V(\mu)} d\mu \\ b(\theta) &:= \int \frac{\mu}{V(\mu)} d\mu \end{aligned}$$

eingeführt, folgt schließlich für die Log-Likelihoodfunktion

$$l(\mu|y) = \frac{y\theta - b(\theta)}{\phi} + c(y, \phi).$$

Diese Funktion entspricht der Dichte- oder Wahrscheinlichkeitsfunktion aus Definition 2.2, wodurch gezeigt wurde, dass die Verteilung von y ein Mitglied der linearen einparametrischen Exponentialfamilie ist.

„ \Leftarrow “

Sei nun die Verteilung von y aus der linearen einparametrischen Exponentialfamilie. Aus (3.4) ist bekannt, dass

$$\frac{\partial l(\mu|y)}{\partial \mu} = \frac{y - \mu}{\phi V(\mu)}$$

für die Scorefunktion bei Ableitung nach μ gilt.

□

Bemerkung:

Für gegebene Varianzfunktion $V(\cdot)$ und festen Dispersionsparameter ϕ liefert Satz 3.2 den Ansatz zur Herleitung des Parameters θ und der Quasi-Likelihoodfunktion $q(\mu|y)$. Zusätzlich geht aus dem Satz hervor, dass die Quasi-Likelihoodfunktion und die Log-Likelihoodfunktion äquivalent sind, falls die Varianzstruktur einem Mitglied der linearen einparametrischen Exponentialfamilie zugeordnet werden kann.

3.3 Beispiele für Quasi-Likelihoodmodelle

In diesem Abschnitt werden nun zwei Varianzfunktionen betrachtet, die für die Modellierung von Prozentwerten eine wichtige Rolle spielen. Mit Hilfe von Satz 3.2 und der Integraldarstellung in (3.2) wird sowohl der korrespondierende Parameter θ , als auch die Quasi-Likelihoodfunktion hergeleitet.

• **Varianzfunktion proportional zur Bernoulli-Varianz**

Falls ein beobachteter Wert y einem Prozentwert aus dem offenen Einheitsintervall $(0, 1)$ entspricht, legt dies eine Varianzannahme proportional zur Bernoulli-Varianz nahe. Betrachtet wird demnach die Varianz

$$\text{Var}(y) = \phi V(\mu) = \phi\mu(1 - \mu),$$

mit $0 < \mu < 1$ und unbekannter Dispersion ϕ . Für den Parameter θ und die Quasi-Likelihoodfunktion $q(\mu|y)$ ergibt sich

$$\theta = \int \frac{1}{V(\mu)} d\mu = \int \frac{1}{\mu(1 - \mu)} d\mu = \log \frac{\mu}{1 - \mu},$$

$$q(\mu|y) = \int^\mu \frac{y - t}{\phi t(1 - t)} dt = \frac{1}{\phi} \left(y \log \frac{\mu}{1 - \mu} + \log(1 - \mu) \right).$$

• **Quadratische Varianzfunktion**

Da die Funktion $V(\mu) = \mu(1 - \mu)$ manchmal zur Überschätzung der Variabilität neigt, schlagen Wedderburn (1974) und McCullagh und Nelder (1989)

$$V(\mu) = \mu^2(1 - \mu)^2$$

als Varianzfunktion und als Basis zur Berechnung des Quasi-Likelihoodmodells für Prozentwerte vor. Gilt zusätzlich $\mu \in (0, 1)$, so folgt für θ und $q(\mu|y)$

$$\theta = \int \frac{1}{\mu^2(1 - \mu)^2} d\mu = 2 \log \frac{\mu}{1 - \mu} + \frac{1}{1 - \mu} - \frac{1}{\mu},$$

$$q(\mu|y) = \int^\mu \frac{y - t}{\phi t^2(1 - t)^2} dt = \frac{1}{\phi} \left((2y - 1) \log \frac{\mu}{1 - \mu} - \frac{y}{\mu} - \frac{1 - y}{1 - \mu} \right).$$

Es sei anzumerken, dass diese Varianzfunktion keinem Mitglied der linearen einparametrischen Exponentialfamilie zugeordnet werden kann.

3.4 Quasi-Dichte

Aus Satz 3.2 ist nun bekannt, dass Quasi-Likelihood und Log-Likelihood übereinstimmen, wenn die Varianzfunktion einem Mitglied der linearen einparametrischen Exponentialfamilie zugeordnet werden kann. Ist dies nicht der Fall, kann zwar die Quasi-Likelihoodfunktion berechnet werden, jedoch keine Likelihoodfunktion und somit auch keine Dichtefunktion. Nelder und Lee (1992) lösten dieses Problem, indem sie die Quasi-Dichte einführten.

Definition 3.3. (Quasi-Dichte, Nelder und Lee, 1992)

Sei y eine Zufallsvariable mit Erwartungswert $E(y) = \mu$ und Varianz $\text{Var}(y) = \phi V(\mu)$. Weiters sei $q(\mu|y)$ die korrespondierende Quasi-Likelihoodfunktion. Dann ist die Quasi-Dichte von y definiert als

$$f_q(y|\mu) = \frac{\exp(q(\mu|y))}{\omega(\mu)}, \quad (3.6)$$

wobei $\omega(\cdot)$ eine Normalisierungsfunktion darstellt, gegeben durch

$$\omega(\mu) = \int_{\mathbb{R}} \exp(q(\mu|y)) dy.$$

Zu der in (3.6) definierten Quasi-Dichte $f_q(y|\mu)$ kann nun wieder die zugehörige Log-Likelihoodfunktion

$$l_q(y|\mu) = \log(f_q(y|\mu)) = q(\mu|y) - \log \omega(\mu)$$

berechnet werden. Aus dieser Log-Likelihoodfunktion ergibt sich die Scorefunktion

$$\frac{\partial l_q(\mu|y)}{\partial \mu} = \frac{\partial q(\mu|y)}{\partial \mu} - \frac{\partial \log \omega(\mu)}{\partial \mu}.$$

Der Unterschied zwischen dieser Scorefunktion und der Quasi-Scorefunktion ist daher gegeben durch

$$\begin{aligned} \frac{\partial q(\mu|y)}{\partial \mu} - \frac{\partial l_q(\mu|y)}{\partial \mu} &= \frac{\partial \log \omega(\mu)}{\partial \mu} = \frac{1}{\omega(\mu)} \frac{\partial \omega(\mu)}{\partial \mu} = \frac{1}{\omega(\mu)} \int_{\mathbb{R}} \frac{\partial \exp(q(\mu|y))}{\partial \mu} dy \\ &= \frac{1}{\omega(\mu)} \int_{\mathbb{R}} \exp(q(\mu|y)) \frac{\partial q(\mu|y)}{\partial \mu} dy = \int_{\mathbb{R}} f_q(y|\mu) \frac{y - \mu}{\phi V(\mu)} dy \\ &= E_q \left(\frac{y - \mu}{\phi V(\mu)} \right) = \frac{\mu_q - \mu}{\phi V(\mu)}. \end{aligned}$$

Der Ausdruck

$$\mu_q = E_q(y) = \int_{\mathbb{R}} y f_q(y|\mu) dy$$

wird hierbei als Quasi-Mean bezeichnet.

Bemerkung:

Die Differenz zwischen Erwartungswert μ und Quasi-Mean μ_q im Vergleich zur Differenz

$y - \mu$ ist ein Maß für den Unterschied zwischen Maximum Quasi-Likelihood Schätzer und Maximum Likelihood Schätzer.

Bemerkung:

Laut Satz 3.2 gilt $\omega(\mu) = 1$, falls die Varianzstruktur einem Mitglied der linearen einparametrischen Exponentialfamilie zugeordnet werden kann, da die Quasi-Likelihoodfunktion dann einer Log-Likelihoodfunktion entspricht.

3.5 Maximum Quasi-Likelihood Schätzung

Aus Abschnitt 3.2 ist bekannt, dass die Quasi-Scorefunktion in (3.3) mit der Scorefunktion der linearen einparametrischen Exponentialfamilie in (3.4) übereinstimmt. Allerdings gilt dies bisher nur bei Ableitung nach dem Erwartungswert μ . Bei der in Abschnitt 2.4.1 vorgestellten Methode zur Berechnung des Maximum Likelihood Schätzers wird jedoch die Ableitung der Log-Likelihoodfunktion nach dem eventuell vektorwertigen Parameter β benötigt.

In diesem Zusammenhang zeigte Wedderburn (1974) mit nachfolgendem Satz, dass die Quasi-Likelihoodfunktion und die Log-Likelihoodfunktion der linearen einparametrischen Exponentialfamilie nicht nur bei Ableitung nach μ gewisse Gemeinsamkeiten aufweisen.

Satz 3.4. (Wedderburn, 1974)

Sei y eine Beobachtung mit $E(y) = \mu$ und $\text{Var}(y) = \phi V(\mu)$. Zusätzlich lasse sich μ als Funktion des Parametervektors $\beta = (\beta_1, \dots, \beta_p)$ darstellen. Für die Quasi-Scorefunktion $q(\mu|y)$ gilt dann für $j, k = 1, \dots, p$:

1. $E\left(\frac{\partial q(\mu|y)}{\partial \mu}\right) = 0$
2. $E\left(\frac{\partial q(\mu|y)}{\partial \beta_j}\right) = 0$
3. $E\left(\frac{\partial q(\mu|y)}{\partial \mu}\right)^2 = -E\left(\frac{\partial^2 q(\mu|y)}{\partial \mu^2}\right) = \frac{1}{\phi V(\mu)}$
4. $E\left(\frac{\partial q(\mu|y)}{\partial \beta_j} \frac{\partial q(\mu|y)}{\partial \beta_k}\right) = -E\left(\frac{\partial^2 q(\mu|y)}{\partial \beta_j \partial \beta_k}\right) = \frac{1}{\phi V(\mu)} \frac{\partial \mu}{\partial \beta_j} \frac{\partial \mu}{\partial \beta_k}$

Beweis:

Siehe Wedderburn (1974). □

Aufgrund der zahlreichen Parallelen zwischen der Log-Likelihoodfunktion und der Quasi-Likelihoodfunktion, schlagen McCullagh und Nelder (1989) auch bei der Maximum Quasi-Likelihood Schätzung die Verwendung des in Abschnitt 2.4.1 vorgestellten Newton-Verfahrens vor.

Seien nun y_1, \dots, y_n unabhängige Response-Variablen mit $E(y_i) = \mu_i$ und $\text{Var}(y_i) = \phi V(\mu_i)$. Zusätzlich gelte $g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$ mit bekannter Linkfunktion $g(\cdot)$. Der Maximum Quasi-Likelihood Schätzer $\hat{\boldsymbol{\beta}}$ resultiert als Nullstelle der Quasi-Scorefunktion

$$\mathbf{U}(\boldsymbol{\beta}|\mathbf{y}) = \frac{\partial q(\boldsymbol{\mu}|\mathbf{y})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \frac{\partial q(\mu_i|y_i)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \frac{y_i - \mu_i}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}}.$$

Die Iterationsvorschrift der Newton-Methode lautet dann

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + \left(-\frac{\partial}{\partial \boldsymbol{\beta}^T} \mathbf{U}(\boldsymbol{\beta}|\mathbf{y}) \right)^{-1} \mathbf{U}(\boldsymbol{\beta}|\mathbf{y}), \quad t = 0, 1, \dots, \quad (3.7)$$

wobei die rechte Seite in $\boldsymbol{\beta}^{(t)}$ ausgewertet wird.

Sowohl McCullagh (1983) als auch McCullagh und Nelder (1989) schlagen eine alternative Darstellung der Quasi-Scorefunktion $\mathbf{U}(\boldsymbol{\beta}|\mathbf{y})$ vor. Dazu werden die Matrizen $\mathbf{V} = \text{diag}(V(\mu_1), \dots, V(\mu_n))$ und

$$\mathbf{P} = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}^T} = \begin{pmatrix} \frac{\partial \mu_1}{\partial \beta_1} & \cdots & \frac{\partial \mu_1}{\partial \beta_p} \\ \vdots & & \vdots \\ \frac{\partial \mu_n}{\partial \beta_1} & \cdots & \frac{\partial \mu_n}{\partial \beta_p} \end{pmatrix}$$

eingeführt. Die Quasi-Scorefunktion kann dann geschrieben werden als

$$\mathbf{U}(\boldsymbol{\beta}|\mathbf{y}) = \frac{1}{\phi} \mathbf{P}^T \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\mu}).$$

Aus den Punkten 2 und 4 von Satz 3.4 ergibt sich für den Erwartungswert der Quasi-Scorefunktion

$$E(\mathbf{U}(\boldsymbol{\beta}|\mathbf{y})) = \mathbf{0}$$

und für die Varianz

$$\text{Var}(\mathbf{U}(\boldsymbol{\beta}|\mathbf{y})) = -E \left(\frac{\partial}{\partial \boldsymbol{\beta}^T} \mathbf{U}(\boldsymbol{\beta}|\mathbf{y}) \right) = \frac{1}{\phi} \mathbf{P}^T \mathbf{V}^{-1} \mathbf{P}. \quad (3.8)$$

Wie in Abschnitt 2.4.2 kommt auch hier die Fisher-Scoring Methode zur Anwendung. Daher wird die zweite Ableitung der Quasi-Likelihoodfunktion in (3.7) durch deren Erwartungswert, gegeben in (3.8), ersetzt. Als Iterationsvorschrift ergibt sich schließlich

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + \left(\mathbf{P}^{(t)T} \mathbf{V}^{(t)-1} \mathbf{P}^{(t)} \right)^{-1} \mathbf{P}^{(t)T} \mathbf{V}^{(t)-1} (\mathbf{y} - \boldsymbol{\mu}), \quad t = 0, 1, \dots,$$

woraus bei Konvergenz der Maximum Quasi-Likelihood Schätzer $\hat{\boldsymbol{\beta}}$ resultiert.

Bemerkung:

Die Iterationsvorschrift ist wieder unabhängig vom Dispersionsparameter ϕ . Bei unbekanntem ϕ schlagen McCullagh und Nelder (1989) eine Schätzung mittels

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} = \frac{1}{n-p} X^2$$

vor, wobei X^2 die Pearson-Statistik darstellt.

Bemerkung:

Was die Fisher-Information für die Likelihoodfunktion darstellt, ist die Matrix (3.8) für die Quasi-Likelihoodfunktion. Diese Aussage bedeutet demnach für die Varianz von $\hat{\beta}$, dass

$$\text{Var}(\hat{\beta}) = \phi(\mathbf{P}^T \mathbf{V}^{-1} \mathbf{P})^{-1}$$

asymptotisch hält.

Bemerkung:

Ähnlich zu den asymptotischen Ergebnissen des Schätzers in Abschnitt 2.4.2, zeigte McCullagh (1983)

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} \mathcal{N}_p(\mathbf{0}, n\phi(\mathbf{P}^T \mathbf{V}^{-1} \mathbf{P})^{-1}).$$

Also ist auch hier die Differenz zwischen dem Maximum Quasi-Likelihood Schätzer $\hat{\beta}$ und dem wahren Parameter β asymptotisch normalverteilt mit Erwartungswert $\mathbf{0}$. Das heißt $\hat{\beta}$ ist ein asymptotisch erwartungstreuer Schätzer für β .

3.6 Quasi-Deviance

Analog zur Deviance in Abschnitt 2.5.2 kann nun die Quasi-Deviance definiert werden. Auch sie stellt ein Maß zur Bewertung der Modellanpassungsgüte dar.

Definition 3.5. (Quasi-Deviance)

Sei $\mathbf{y} = (y_1, \dots, y_n)^T$ ein Vektor mit n unabhängigen Beobachtungen und $\hat{\boldsymbol{\mu}}$ der Maximum Quasi-Likelihood Schätzer bezüglich der Quasi-Likelihoodfunktion $q(\boldsymbol{\mu}|\mathbf{y})$. Die Quasi-Deviance ist dann definiert als

$$D^q(\mathbf{y}|\hat{\boldsymbol{\mu}}) = -2\phi(q(\hat{\boldsymbol{\mu}}|\mathbf{y}) - q(\mathbf{y}|\mathbf{y})) = -2 \sum_{i=1}^n \int_{y_i}^{\hat{\mu}_i} \frac{y_i - t}{V(t)} dt, \quad (3.9)$$

wobei ϕ den Dispersionsparameter darstellt.

Bemerkung:

Aus Definition 3.5 ist ersichtlich, dass die Minimierung der Quasi-Deviance äquivalent ist zur Maximierung der Quasi-Likelihoodfunktion.

Bemerkung:

McCullagh (1983) zeigte, dass die Quasi-Deviance unter gewissen Regularitätsbedingungen eine asymptotisch χ^2 -verteilte Größe mit $n - p$ Freiheitsgraden darstellt.

3.6.1 Beispiele zur Quasi-Deviance

Zwei interessante Beispiele zur Berechnung der Quasi-Deviance stellen die aus Abschnitt 3.3 bekannten Varianzmodelle und die daraus resultierenden Quasi-Likelihoodfunktionen dar. Für Beobachtungen aus dem offenen Einheitsintervall $(0, 1)$ kann die Quasi-Deviance einfach unter Verwendung von Definition 3.5 ermittelt werden. Sollten die Response-Werte jedoch auch Werte aus $\{0, 1\}$ enthalten, so ist die Berechnung der Quasi-Deviance nicht sofort ersichtlich. Zur Vereinfachung wird der Ausdruck (3.9) zunächst geschrieben als

$$\begin{aligned} D^q(\mathbf{y}|\boldsymbol{\mu}) &= -2\phi\left(q(\boldsymbol{\mu}|\mathbf{y}) - q(\mathbf{y}|\mathbf{y})\right) \\ &= -2\phi\left(\sum_{i=1}^n q(\mu_i|y_i) - \sum_{i=1}^n q(y_i|y_i)\right) \\ &= 2\phi\sum_{i=1}^n \left(q(y_i|y_i) - q(\mu_i|y_i)\right). \end{aligned} \quad (3.10)$$

- **Varianzfunktion proportional zur Bernoulli-Varianz**

Aus Abschnitt 3.3 ist bekannt, dass die Quasi-Likelihoodfunktion bezüglich der Varianzannahme

$$\text{Var}(y_i) = \phi V(\mu_i) = \phi\mu_i(1 - \mu_i),$$

mit $0 < \mu_i < 1$, gegeben ist durch

$$q(\mu_i|y_i) = \frac{1}{\phi} \left(y_i \log \frac{\mu_i}{1 - \mu_i} + \log(1 - \mu_i) \right).$$

Da $q(y_i|y_i)$ für y_i aus $\{0, 1\}$ auf den ersten Blick nicht bestimmt werden kann, könnte irrtümlicherweise auch darauf geschlossen werden, dass die Deviance nur für Beobachtungen aus $(0, 1)$ berechnet werden kann. Ausgehend von (3.10) kann die Quasi-Deviance für dieses Modell jedoch geschrieben werden als

$$\begin{aligned} D^q(\mathbf{y}|\boldsymbol{\mu}) &= 2\phi \sum_{i=1}^n \left(q(y_i|y_i) - q(\mu_i|y_i) \right) \\ &= 2\phi \sum_{i=1}^n \left(y_i \log \frac{y_i}{1 - y_i} + \log(1 - y_i) - y_i \log \frac{\mu_i}{1 - \mu_i} - \log(1 - \mu_i) \right) \\ &= 2\phi \sum_{i=1}^n \left(y_i \log \frac{y_i}{\mu_i} + (1 - y_i) \log \frac{1 - y_i}{1 - \mu_i} \right). \end{aligned}$$

Unter Verwendung der Regel von de l'Hospital gilt

$$\lim_{y_i \rightarrow 0} y_i \log \frac{y_i}{\mu_i} = 0,$$

genauso wie

$$\lim_{y_i \rightarrow 1} (1 - y_i) \log \frac{1 - y_i}{1 - \mu_i} = 0,$$

womit die Quasi-Deviance auch für Beobachtungen aus $\{0, 1\}$ bestimmt werden kann.

• **Quadratische Varianzfunktion**

Ebenso wurde in Abschnitt 3.3 gezeigt, dass die Verwendung der Varianzfunktion

$$V(\mu_i) = \mu_i^2(1 - \mu_i)^2$$

zur Quasi-Likelihoodfunktion

$$q(\mu_i|y_i) = \frac{1}{\phi} \left((2y_i - 1) \log \frac{\mu_i}{1 - \mu_i} - \frac{y_i}{\mu_i} - \frac{1 - y_i}{1 - \mu_i} \right)$$

führt. Mit (3.10) resultiert somit für die Quasi-Deviance

$$\begin{aligned} D^q(\mathbf{y}|\boldsymbol{\mu}) &= 2\phi \sum_{i=1}^n \left(q(y_i|y_i) - q(\mu_i|y_i) \right) \\ &= 2\phi \sum_{i=1}^n \left((2y_i - 1) \log \frac{y_i}{1 - y_i} - \frac{y_i}{y_i} - \frac{1 - y_i}{1 - y_i} - (2y_i - 1) \log \frac{\mu_i}{1 - \mu_i} + \frac{y_i}{\mu_i} + \frac{1 - y_i}{1 - \mu_i} \right) \\ &= 2\phi \sum_{i=1}^n \left((2y_i - 1) \log \frac{y_i(1 - \mu_i)}{(1 - y_i)\mu_i} + \frac{y_i}{\mu_i} + \frac{1 - y_i}{1 - \mu_i} - 2 \right). \end{aligned}$$

Für y_i aus $\{0, 1\}$ wird dieser Ausdruck unendlich groß. Bei Turner und Firth (2015) findet sich daher die um die Konstante $\epsilon = 0.0005$ korrigierte Version

$$D^q(\mathbf{y}|\boldsymbol{\mu}, \epsilon) = 2\phi \sum_{i=1}^n \left((2y_i - 1) \log \frac{(y_i + \epsilon)(1 - \mu_i)}{(1 - y_i + \epsilon)\mu_i} + \frac{y_i}{\mu_i} + \frac{1 - y_i}{1 - \mu_i} - 2 \right). \quad (3.11)$$

3.7 Extended Quasi-Likelihood

Durch die Einführung der Quasi-Likelihoodfunktion ist eine Analyse mittels des Generalisierten Linearen Modells nicht mehr nur für Beobachtungen mit Verteilung aus der linearen einparametrischen Exponentialfamilie möglich. Die Quasi-Likelihoodfunktion ermöglicht auch die Verwendung von Varianzstrukturen, die keinem Mitglied der eben genannten Familie zugeordnet werden können. Im nächsten Schritt stellt sich die Frage, welche Möglichkeiten bestehen um mehrere Varianzfunktionen miteinander zu vergleichen. Da die bisher kennengelernten Methoden nicht dazu in der Lage sind, wird nun eine auf Nelder und Pregibon (1987) zurückgehende Erweiterung des Quasi-Likelihood Ansatzes betrachtet.

Definition 3.6. (EQL-Funktion, Nelder und Pregibon, 1987)

Sei y eine Zufallsvariable mit Erwartungswert $E(y) = \mu$ und Varianz $\text{Var}(y) = \phi V(\mu)$, mit bekannter Varianzfunktion $V(\cdot)$ und Dispersionsparameter ϕ . Zusätzlich sei $q(\mu|y)$ die korrespondierende Quasi-Likelihoodfunktion und $D^q(y|\mu)$ die daraus resultierende Quasi-Deviance. Die Extended Quasi-Likelihoodfunktion (EQL-Funktion) $q^+(\mu, \phi|y)$ ist dann definiert als

$$q^+(\mu, \phi|y) = -\frac{1}{2} \log(2\pi\phi V(y)) - \frac{1}{2\phi} D^q(y|\mu).$$

Sollten n unabhängige Beobachtungen y_1, \dots, y_n mit $E(y_i) = \mu_i$ und $\text{Var}(y_i) = \phi V(\mu_i)$ vorliegen, wird die EQL-Funktion bezüglich $\mathbf{y} = (y_1, \dots, y_n)^T$ und $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ geschrieben als

$$q^+(\boldsymbol{\mu}, \phi | \mathbf{y}) = \sum_{i=1}^n q^+(\mu_i, \phi | y_i). \quad (3.12)$$

3.7.1 Schätzung der Parameter

Sowohl die Schätzung des Erwartungswertes, als auch die des Dispersionsparameters sind von Interesse. Um diese Schätzer zu erhalten, wird die bei Nelder und Pregibon (1987) und Thaler (2009a) beschriebene Maximierung der EQL-Funktion betrachtet.

Sei nun $\mathbf{y} = (y_1, \dots, y_n)^T$ ein Vektor mit n unabhängigen Beobachtungen und $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ der Vektor der Erwartungswerte $E(y_i) = \mu_i$. Zusätzlich gelte $\text{Var}(y_i) = \phi V(\mu_i)$ für $i = 1, \dots, n$. Die EQL-Funktion weist folglich die aus (3.12) bekannte Form auf.

Wird der Erwartungswert $\boldsymbol{\mu}$ wie in den vorherigen Abschnitten als Funktion im Parameter $\boldsymbol{\beta}$ dargestellt, kann der Vektor $\boldsymbol{\beta}$ geschätzt werden, indem $q^+(\boldsymbol{\mu}, \phi | \mathbf{y})$ maximiert wird. Aus Definition 3.6 ist ersichtlich, dass die Maximierung von $q^+(\boldsymbol{\mu}, \phi | \mathbf{y})$ äquivalent ist zur Minimierung der Quasi-Deviance $D^q(\mathbf{y} | \boldsymbol{\mu})$. Diese Minimierung ist jedoch gleichzusetzen mit der Maximierung der Quasi-Likelihoodfunktion $q(\boldsymbol{\mu} | \mathbf{y})$. Das heißt jener Schätzer $\hat{\boldsymbol{\beta}}$, welcher aus der Maximierung von $q^+(\boldsymbol{\mu}, \phi | \mathbf{y})$ hervorgeht, stellt auch den besten Schätzer von $\boldsymbol{\beta}$ bei Maximierung von $q(\boldsymbol{\mu} | \mathbf{y})$ dar.

Zur Berechnung der Schätzer $\hat{\boldsymbol{\beta}}$ und $\hat{\phi}$ wird die EQL-Funktion simultan nach $\boldsymbol{\beta}$ und ϕ abgeleitet und Null gesetzt, das heißt

$$\frac{\partial q^+(\boldsymbol{\mu}, \phi | \mathbf{y})}{\partial (\boldsymbol{\beta}^T, \phi)^T} \stackrel{!}{=} \mathbf{0}. \quad (3.13)$$

Um dieses Gleichungssystem zu lösen erfolgt zunächst die Berechnung von $\hat{\boldsymbol{\beta}}$ beziehungsweise $\hat{\boldsymbol{\mu}}$ über die Maximierung der Quasi-Likelihoodfunktion $q(\boldsymbol{\mu} | \mathbf{y})$. Für die Schätzung der Dispersion ϕ ergibt sich aus (3.13) das Gleichungssystem

$$\begin{aligned} \frac{\partial q^+(\boldsymbol{\mu}, \phi | \mathbf{y})}{\partial \phi} &= \sum_{i=1}^n \frac{\partial q^+(\mu_i, \phi | y_i)}{\partial \phi} \\ &= \sum_{i=1}^n \frac{\partial}{\partial \phi} \left(-\frac{1}{2} \log(2\pi\phi V(y_i)) - \frac{1}{2\phi} D^q(y_i | \mu_i) \right) \\ &= \sum_{i=1}^n \left(-\frac{1}{2\phi} + \frac{1}{2\phi^2} D^q(y_i | \mu_i) \right) = -\frac{n}{2\phi} + \frac{1}{2\phi^2} D^q(\mathbf{y} | \boldsymbol{\mu}) \stackrel{!}{=} 0. \end{aligned}$$

Aus elementaren Umformungen zusammen mit der soeben berechneten Schätzung $\hat{\boldsymbol{\mu}}$ ergibt sich als Schätzer von ϕ die Größe

$$\frac{D^q(\mathbf{y}|\hat{\boldsymbol{\mu}})}{n}.$$

Dieser Ausdruck ist auch unter dem Begriff mittlere Deviance bekannt.

Um die Schätzung des p -dimensionalen Vektors $\boldsymbol{\mu}$ auszugleichen und den dadurch entstandenen Bias zu korrigieren, wird $\hat{\phi}$ definiert als

$$\hat{\phi} = \frac{D^q(\mathbf{y}|\hat{\boldsymbol{\mu}})}{n - p}.$$

3.7.2 Eigenschaften der EQL-Funktion

Zunächst wird die Motivation betrachtet, die Nelder und Pregibon (1987) für die Einführung der aus Definition 3.6 bekannten EQL-Funktion nennen. Wie aus der Quasi-Likelihoodfunktion, kann auch aus der EQL-Funktion eine Dichte abgeleitet werden. Dazu wird $\exp(q^+)$ mit einem passenden Faktor normalisiert, der üblicherweise eine Funktion in $\boldsymbol{\mu}$ und ϕ , sowie in weiteren Parametern, zusammengefasst zu $\boldsymbol{\theta}$, darstellt. Laut Nelder und Pregibon (1987) kann es passieren, dass sich der Normalisierungsfaktor auch bei großen Schwankungen der Parameter nur geringfügig ändert. Aus diesem Grund kann die Normalisierung ignoriert werden. Es reicht aus, die Maximierung von $\exp(q^+)$ zu betrachten.

Als weitere Eigenschaft der EQL-Funktion soll nun die mögliche Äquivalenz zur Log-Likelihoodfunktion betrachtet werden. Mit Satz 3.2 wurde gezeigt, dass die Quasi-Likelihoodfunktion und die Log-Likelihoodfunktion nur für Verteilungen aus der einparametrischen linearen Exponentialfamilie übereinstimmen. Nelder und Pregibon (1987) zeigten die Äquivalenz der EQL-Funktion und der Log-Likelihoodfunktion für bestimmte Mitglieder der Exponentialfamilie.

Die Äquivalenz kann für die Normalverteilung und für die inverse Gauß-Verteilung gezeigt werden. Für die Poissonverteilung, die standardisierte Binomialverteilung und die negative Binomialverteilung kann die EQL-Funktion unter Anwendung der Stirling-Approximation aus der Log-Likelihoodfunktion abgeleitet werden. Bei der Gammaverteilung unterscheiden sich EQL-Funktion und Log-Likelihoodfunktion nur um einen von ϕ abhängigen Faktor. Eine ausführliche Diskussion findet sich bei Thaler (2009a).

3.7.3 Parametrisierte Varianzfunktion

Wie bereits erwähnt wurde die EQL-Funktion vor allem wegen der Vergleichsmöglichkeit unterschiedlicher Varianzfunktionen eingeführt. Nelder und Pregibon (1987) folgend wird in diesem Abschnitt besprochen, wie aus einer Familie $\mathcal{F}_{\boldsymbol{\theta}}$ von parametrisierten Varianzfunktionen $V_{\boldsymbol{\theta}}(\boldsymbol{\mu}) := V(\boldsymbol{\mu}|\boldsymbol{\theta})$ der optimale Wert des unbekanntes, möglicherweise

vektorwertigen Parameters $\boldsymbol{\theta}$ gefunden wird.

Sei $\mathbf{y} = (y_1, \dots, y_n)^T$ ein Vektor mit n unabhängigen Beobachtungen, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ der korrespondierende Vektor der vom Parameter $\boldsymbol{\beta}$ abhängigen Erwartungswerte und $V_{\boldsymbol{\theta}}(\cdot)$ die parametrisierte Varianzfunktion. Da die Abhängigkeit der Varianzfunktion von $\boldsymbol{\theta}$ auch zu einer parametrisierten Quasi-Likelihoodfunktion $q_{\boldsymbol{\theta}}(\boldsymbol{\mu}|\mathbf{y})$ und Quasi-Deviance $D_{\boldsymbol{\theta}}^q(\mathbf{y}|\boldsymbol{\mu})$ führt, folgt für die EQL-Funktion

$$q_{\boldsymbol{\theta}}^+(\boldsymbol{\mu}, \phi|\mathbf{y}) = \sum_{i=1}^n q_{\boldsymbol{\theta}}^+(\mu_i, \phi|y_i) = \sum_{i=1}^n \left(-\frac{1}{2} \log(2\pi\phi V_{\boldsymbol{\theta}}(y_i)) - \frac{1}{2\phi} D_{\boldsymbol{\theta}}^q(y_i|\mu_i) \right),$$

mit Dispersionsparameter ϕ . Um jene Schätzer $\hat{\boldsymbol{\beta}}$ und $\hat{\boldsymbol{\theta}}$ zu bestimmen, welche die EQL-Funktion $q_{\boldsymbol{\theta}}^+(\boldsymbol{\mu}, \phi|\mathbf{y})$ maximieren, können sogenannte Profile-Quasi-Likelihood-Intervalle verwendet werden. Eine Beschreibung der Vorgehensweise findet sich sowohl bei Thaler (2009a) als auch bei Nelder und Pregibon (1987).

Binomialähnliche Varianzstruktur

Da für Beobachtungen aus dem offenen Einheitsintervall $(0, 1)$ die Bernoulli-Varianz ein guter Kandidat für die Varianzannahme sein könnte, wird nun die Familie der binomial-ähnlichen Varianzfunktionen betrachtet. In Abschnitt 3.3 wurden bereits die Funktionen $V(\mu) = \mu(1 - \mu)$ und $V(\mu) = \mu^2(1 - \mu)^2$ untersucht. Im nächsten Schritt werden zwei zusätzliche Parameter $k \in \mathbb{R}$ und $l \in \mathbb{R}$ eingeführt und die daraus resultierende Familie von Varianzfunktionen

$$V_{k,l}(\mu) = \mu^k(1 - \mu)^l$$

als Basis zur Berechnung der Quasi-Likelihoodfunktion verwendet. Zur Vereinfachung sei zunächst $k \in \mathbb{N}$ und $l \in \mathbb{N}$. Um die Quasi-Likelihoodfunktion zu bestimmen, muss nach Definition 3.1 das Integral

$$q_{k,l}(\mu|y) = \int^{\mu} \frac{y - t}{\phi V(\mu)} dt = \int^{\mu} \frac{y - t}{\phi t^k(1 - t)^l} dt$$

bestimmt werden. Zur Vereinfachung wird zunächst das Integral

$$\int \frac{1}{\mu^k(1 - \mu)^l} d\mu$$

mittels Partialbruchzerlegung (Thaler, 2009a) umgeformt zu

$$\begin{aligned} S_{k,l}(\mu) &:= \int \frac{1}{\mu^k(1 - \mu)^l} d\mu \\ &= \binom{k+l-2}{k-1} \log \left| \frac{\mu}{1-\mu} \right| - \sum_{i=0}^{k-2} \binom{i+l-1}{l-1} \frac{1}{(k-i-1)\mu^{k-i-1}} \\ &\quad + \sum_{j=0}^{l-2} \binom{j+k-1}{k-1} \frac{1}{(l-j-1)(1-\mu)^{l-j-1}}. \end{aligned}$$

Für $k = 0$ und $l = 0$ werden die Ausdrücke

$$S_{0,l}(\mu) = \begin{cases} \mu & l = 0 \\ -\log |1 - \mu| & l = 1 \\ \frac{1}{(l-1)(1-\mu)^{l-1}} & l \geq 2 \end{cases}$$

und

$$S_{k,0}(\mu) = \begin{cases} \mu & k = 0 \\ \log |\mu| & k = 1 \\ \frac{-1}{(k-1)\mu^{k-1}} & k \geq 2 \end{cases}$$

verwendet. Daraus ergibt sich für die Quasi-Likelihoodfunktion $q_{k,l}(\mu|y)$ die Darstellung

$$\begin{aligned} q_{k,l}(\mu|y) &= \int^{\mu} \frac{y-t}{\phi t^k (1-t)^l} dt = \frac{1}{\phi} \left(\int^{\mu} \frac{y}{t^k (1-t)^l} dt - \int^{\mu} \frac{1}{t^{k-1} (1-t)^l} dt \right) \\ &= \frac{1}{\phi} (y S_{k,l}(\mu) - S_{k-1,l}(\mu)), \end{aligned}$$

mit unbekannter Dispersion ϕ .

Bemerkung:

Für reellwertige Parameter k und l findet sich die Anleitung zur Berechnung der Quasi-Likelihoodfunktion bei Thaler (2009a).

Bemerkung:

In der Praxis kann es vorkommen, dass die beobachteten Response-Werte nicht alle aus dem offenen Einheitsintervall $(0, 1)$ stammen, sondern auch Werte am Rand dieses Intervalls annehmen. Für Beobachtungen aus $\{0, 1\}$ ist die EQL-Funktion bezüglich der binomialähnlichen Varianzstruktur jedoch nicht definiert. Die Parameterschätzer existieren dennoch.

4.1 Motivation

Um die strengen Einschränkungen des Linearen Modells zu umgehen, wurde das Generalisierte Lineare Modell eingeführt. Das GLM ist vor allem für die Modellierung von Häufigkeiten oder binären Größen geeignet, nicht aber zur Modellierung von Prozentwerten. Erst durch die Einführung der Quasi-Likelihoodfunktion konnten die aus dem Intervall $(0, 1)$ stammenden Prozentwerte modelliert werden. Auch die EQL-Funktion scheint für diesen Zweck geeignet zu sein.

Da es in dieser Arbeit um den Vergleich von Regressionsmodellen bezüglich der Modellierung von Prozentwerten geht, wird nun ein weiteres Modell eingeführt. Einen alternativen Ansatz zur Anpassung von Daten aus dem offenen Einheitsintervall $(0, 1)$ stellt das von Ferrari und Cribari-Neto (2004) entwickelte Beta-Regressionsmodell dar. Hierbei wird angenommen, dass die Response-Variable y einer Beta-Verteilung mit den Parametern α und β folgt. Die aus der Variation von α und β resultierende Beta-Verteilungsfamilie scheint vor allem wegen ihrer Flexibilität und der daraus entstehenden Bandbreite an Dichtefunktionen ein geeigneter Kandidat für die Verteilungsannahme eines Regressionsmodells zu sein.

In Abbildung 4.1 lässt sich gut erkennen, welchen Einfluss die Parameter α und β auf Variabilität und Schiefe der Dichtefunktion ausüben. Für $\alpha = \beta$ ist die Dichtefunktion symmetrisch. Der Fall $\alpha < \beta$ führt zu einer rechtsschiefen und $\alpha > \beta$ zu einer linksschiefen Kurve.

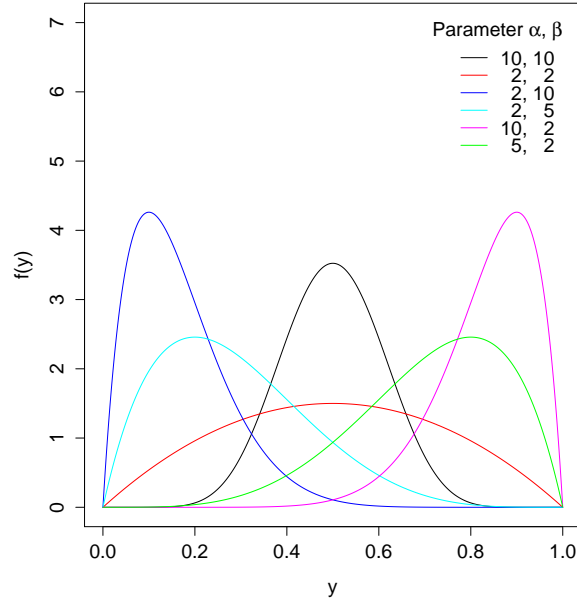


Abbildung 4.1: Beta-Dichten bei Variation der Parameter α und β .

4.2 Grundmodell

Da das Beta-Regressionsmodell von Ferrari und Cribari-Neto (2004) auf der Annahme einer Beta-verteilten Response-Variable y basiert, wird zunächst die Verteilung und deren Dichtefunktion genauer betrachtet.

Definition 4.1. (Beta-Verteilung)

Eine stetige Zufallsvariable y aus dem Intervall $(0, 1)$ ist Beta-verteilt mit den Parametern α und β , d.h. $y \sim \mathcal{B}(\alpha, \beta)$, falls die Dichtefunktion $f(y|\alpha, \beta)$ die Form

$$f(y|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} y^{\alpha-1} (1-y)^{\beta-1} = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1-y)^{\beta-1} \quad (4.1)$$

hat, mit $\alpha > 0$ und $\beta > 0$. Hierbei bezeichnet $B(\cdot, \cdot)$ die Betafunktion, gegeben durch

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt$$

und $\Gamma(\cdot)$ die Gammafunktion, definiert als

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt.$$

Sei nun die Response-Variable y aus dem offenen Einheitsintervall $(0, 1)$, einer Beta-Verteilung mit den Parametern $\alpha > 0$ und $\beta > 0$ folgend. Erwartungswert und Varianz von y sind dann gegeben durch

$$\begin{aligned} E(y) &= \frac{\alpha}{\alpha + \beta} \\ \text{Var}(y) &= \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}. \end{aligned}$$

Wie bereits aus den vorherigen Kapiteln bekannt, geht es bei Regressionsmodellen vorwiegend um die Modellierung des Erwartungswertes. Zusätzlich ist auch das Vorkommen eines Präzisions- bzw. Dispersionsparameters bei den meisten Modellen üblich. Um eine Struktur zu gewährleisten, die sich mit diesen Grundsätzen der Regression vereinbaren lässt, schlagen Ferrari und Cribari-Neto (2004) eine Reparametrisierung der Dichtefunktion (4.1) mittels

$$\begin{aligned} \mu &= \frac{\alpha}{\alpha + \beta} \\ \phi &= \alpha + \beta \end{aligned}$$

vor. Mit dieser alternativen Parametrisierung lassen sich Erwartungswert und Varianz von y schreiben als

$$E(y) = \mu \tag{4.2}$$

$$\text{Var}(y) = \frac{\mu(1 - \mu)}{1 + \phi} = \frac{V(\mu)}{1 + \phi}. \tag{4.3}$$

Hierbei sei anzumerken, dass $0 < \mu < 1$ und $\phi > 0$ durch $\alpha > 0$ und $\beta > 0$ gegeben ist. Die Funktion $V(\mu) = \mu(1 - \mu)$ erinnert stark an die Varianzfunktion von Bernoulli-verteilten Response-Werten. Weiters fällt auf, dass der Dispersionsparameter ϕ als Präzisionsparameter interpretiert werden kann, da die Varianz bei steigendem ϕ abnimmt. Für den Fall $\phi = 1$ stimmt (4.3) mit der Varianz einer Bernoulli-Variablen überein.

Wie Erwartungswert und Varianz, kann auch die in Definition 4.1 eingeführte Dichtefunktion der Beta-Verteilung parametrisiert in μ und ϕ dargestellt werden. Aus (4.1) ergibt sich

$$f(y|\mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1 - \mu)\phi)} y^{\mu\phi - 1} (1 - y)^{(1 - \mu)\phi - 1} \tag{4.4}$$

als Dichtefunktion von y in Abhängigkeit von μ und ϕ .

Wie in Abbildung 4.1 erkennbar, ist die Dichte der Beta-Verteilung genau dann symmetrisch, wenn $\alpha = \beta$ hält. Da aufgrund der Reparametrisierung $\alpha = \mu\phi$ und $\beta = (1 - \mu)\phi$ gilt, ist die Dichtefunktion $f(y|\mu, \phi)$ nur für den Fall $\mu = 1/2$ symmetrisch.

4.2. Grundmodell

Abbildung 4.2 zeigt, wie sich die Änderung der Parameter μ und ϕ auf die Dichtefunktion der Beta-Verteilung auswirkt. Wie bereits erwähnt, ist die Wahl des Erwartungswertes μ für die Symmetrie bzw. Schiefe der Kurve verantwortlich. Für $\mu < 0.5$ ist die Dichte rechtsschief und für $\mu > 0.5$ linksschief. Der Dispersionsparameter ϕ hingegen nimmt Einfluss auf die Varianz der Zufallsvariablen und somit auch auf die Höhe der Dichtefunktion. Je weiter die Dichtefunktion ausschlägt, desto größer ist der Anteil der Daten, die sich in der Nähe des Erwartungswertes befinden. Weitere Vergleichsmöglichkeiten finden sich bei Ferrari und Cribari-Neto (2004).

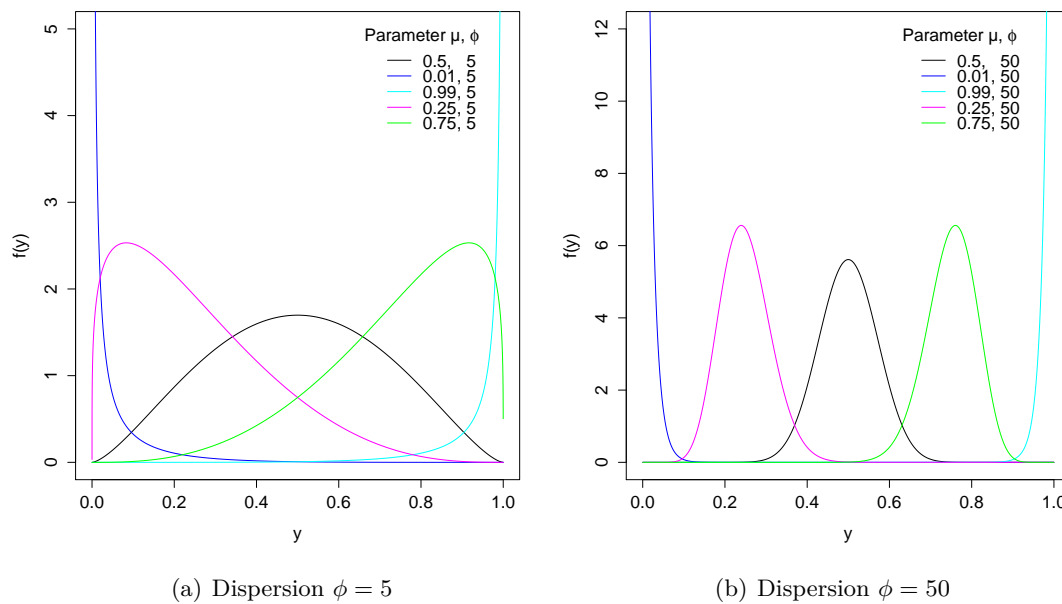


Abbildung 4.2: Dichtefunktionen $f(y|\mu, \phi)$ der Beta-Verteilung bei Variation der Parameter μ und ϕ .

Sei nun $\mathbf{y} = (y_1, \dots, y_n)^T$ ein Vektor bestehend aus n unabhängigen, beta-verteilten Zufallsvariablen mit konstanter Dispersion ϕ und Erwartungswerten $E(y_i) = \mu_i$, zusammengefasst zu $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$. Äquivalent zu den Generalisierten Linearen Modellen wird auch bei der Beta-Regression nicht der Erwartungswert μ_i sondern $g(\mu_i)$ linear modelliert, wobei $g(\cdot)$ wieder eine bekannte Linkfunktion darstellt. Wie auch schon in den vorigen Kapiteln ist $g(\cdot)$ eine monotone, zweimal stetig differenzierbare Funktion und bildet das Intervall $(0, 1)$ auf die reellen Zahlen ab. Die Modellierung erfolgt demnach mittels

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta},$$

mit Parametervektor $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ und linearen Prädiktoren η_1, \dots, η_n . Die Vektoren $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$, $i = 1, \dots, n$, enthalten die erklärenden Variablen und können zur $n \times p$ Designmatrix $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ zusammengefasst werden.

Die Log-Likelihoodfunktion einer einzelnen Variable y_i entspricht

$$\begin{aligned}
 l(\mu_i, \phi | y_i) &= \log f(y_i | \mu_i, \phi) \\
 &= \log \left(\frac{\Gamma(\phi)}{\Gamma(\mu_i \phi) \Gamma((1 - \mu_i) \phi)} y_i^{\mu_i \phi - 1} (1 - y_i)^{(1 - \mu_i) \phi - 1} \right) \\
 &= \log \Gamma(\phi) - \log \Gamma(\mu_i \phi) - \log \Gamma((1 - \mu_i) \phi) + (\mu_i \phi - 1) \log(y_i) \\
 &\quad + [(1 - \mu_i) \phi - 1] \log(1 - y_i).
 \end{aligned} \tag{4.5}$$

Wegen der Unabhängigkeit der Variablen y_1, \dots, y_n folgt für die gemeinsame Log-Likelihoodfunktion die Darstellung

$$l(\boldsymbol{\mu}, \phi | \mathbf{y}) = \sum_{i=1}^n l(\mu_i, \phi | y_i). \tag{4.6}$$

Bemerkung:

So wie das Modell definiert wurde, müssen die Beobachtungen nicht zwingend aus dem offenen Einheitsintervall $(0, 1)$ stammen, sondern können in beliebigen Intervallen (a, b) liegen. Allerdings ist dann lediglich die Modellierung der Transformation $\frac{y_i - a}{b - a}$ der Beobachtungen und nicht die Modellierung der Beobachtungen selbst möglich.

4.2.1 Maximum Likelihood Schätzung

Nach der Definition des Modells, ist der nächste Schritt die Schätzung der Parameter. Ferrari und Cribari-Neto (2004) schlagen dazu die Maximum Likelihood Methode vor. Wie schon von den Generalisierten Linearen Modellen bekannt, handelt es sich bei den zu schätzenden Parametern um die Dispersion ϕ und um den Parametervektor des linearen Prädiktors, $\boldsymbol{\beta}$. Zur Bestimmung der Maximum Likelihood Schätzer (MLE) wird die Scorefunktion berechnet und Null gesetzt, die damit einhergehenden Rechenschritte sind zu vergleichen mit Ferrari und Cribari-Neto (2004) und Ospina, Cribari-Neto und Vasconcellos (2006). Zuerst erfolgt die Ableitung der aus (4.5) und (4.6) bekannten Log-Likelihoodfunktion nach $\boldsymbol{\beta}$. Da $\mu_i = g^{-1}(\eta_i) = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})$ eine Funktion in η_i und somit auch in $\boldsymbol{\beta}$ darstellt, gilt für die Scorefunktion

$$\frac{\partial l(\boldsymbol{\mu}, \phi | \mathbf{y})}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l(\mu_i, \phi | y_i)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l(\mu_i, \phi | y_i)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}, \quad j = 1, \dots, p. \tag{4.7}$$

Zur Vereinfachung werden die drei Teile des Produkts separat betrachtet. Es gilt

$$\begin{aligned}
 \frac{\partial \eta_i}{\partial \beta_j} &= \frac{\partial \mathbf{x}_i^T \boldsymbol{\beta}}{\partial \beta_j} = x_{ij}, \\
 \frac{\partial \mu_i}{\partial \eta_i} &= \frac{1}{\partial g(\mu_i) / \partial \mu_i} = \frac{1}{g'(\mu_i)}.
 \end{aligned}$$

Unter Verwendung der Digamma-Funktion

$$\psi(z) = \frac{d \log \Gamma(z)}{dz}$$

folgt für die Ableitung nach μ_i

$$\frac{\partial l(\mu_i, \phi | y_i)}{\partial \mu_i} = \phi \left(\log \frac{y_i}{1 - y_i} - \psi(\mu_i \phi) + \psi((1 - \mu_i)\phi) \right) = \phi(y_i^* - \mu_i^*), \quad (4.8)$$

mit

$$y_i^* = \log \frac{y_i}{1 - y_i}$$

und

$$\mu_i^* = \psi(\mu_i \phi) - \psi((1 - \mu_i)\phi).$$

Zusammen ergibt sich daher für die Scorefunktion bei Ableitung nach β_j

$$\frac{\partial l(\boldsymbol{\mu}, \phi | \mathbf{y})}{\partial \beta_j} = \phi \sum_{i=1}^n (y_i^* - \mu_i^*) \frac{x_{ij}}{g'(\mu_i)}. \quad (4.9)$$

Werden nun zusätzlich die Matrix

$$\mathbf{D} = \text{diag} \left(\frac{1}{g'(\mu_1)}, \dots, \frac{1}{g'(\mu_n)} \right) \quad (4.10)$$

und die Vektoren $\mathbf{y}^* = (y_1^*, \dots, y_n^*)^T$ und $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_n^*)^T$ eingeführt, so folgt für die Ableitung der Log-Likelihoodfunktion nach $\boldsymbol{\beta}$

$$\mathbf{U}_\beta(\boldsymbol{\beta}, \phi) = \frac{\partial l(\boldsymbol{\mu}, \phi | \mathbf{y})}{\partial \boldsymbol{\beta}} = \phi \mathbf{X}^T \mathbf{D} (\mathbf{y}^* - \boldsymbol{\mu}^*). \quad (4.11)$$

Bemerkung:

Es ist bereits bekannt, dass der Erwartungswert einer Scorefunktion unter Einhaltung gewisser Regularitätsbedingungen stets Null ist. Aus (4.9) folgt daher $E(y_i^*) = \mu_i^*$.

Im nächsten Schritt wird die Scorefunktion bei Ableitung nach der Dispersion ϕ betrachtet. Es gilt

$$\begin{aligned} U_\phi(\boldsymbol{\beta}, \phi) &= \frac{\partial l(\boldsymbol{\mu}, \phi | \mathbf{y})}{\partial \phi} = \sum_{i=1}^n \frac{\partial l(\mu_i, \phi | y_i)}{\partial \phi} \\ &= \sum_{i=1}^n \left(\mu_i \left(\log \frac{y_i}{1 - y_i} - \psi(\mu_i \phi) + \psi((1 - \mu_i)\phi) \right) + \log(1 - y_i) + \psi(\phi) - \psi((1 - \mu_i)\phi) \right) \\ &= \sum_{i=1}^n \left(\mu_i (y_i^* - \mu_i^*) + \log(1 - y_i) + \psi(\phi) - \psi((1 - \mu_i)\phi) \right). \end{aligned}$$

Zur Berechnung der Maximum Likelihood Schätzer $\hat{\boldsymbol{\beta}}$ und $\hat{\phi}$ werden die Scorefunktionen $\mathbf{U}_\beta(\boldsymbol{\beta}, \phi)$ aus (4.11) und $U_\phi(\boldsymbol{\beta}, \phi)$ simultan Null gesetzt. Da beide Gleichungssysteme nicht linear in $\boldsymbol{\beta}$ bzw. ϕ sind, muss ein numerisches Verfahren verwendet werden. Ferrari und Cribari-Neto (2004) schlagen dazu das Newton-Verfahren oder das Quasi-Newton-Verfahren vor.

4.2.2 Fisher-Information

Sowohl für die Verwendung der eben erwähnten Näherungsverfahren, als auch für Verteilungsaussagen bezüglich der geschätzten Parameter ist die Fisher-Information äußerst hilfreich. Aus diesem Grund wird nun, Ferrari und Cribari-Neto (2004) folgend, betrachtet, wie diese hergeleitet werden kann. Die Fisher-Information bezüglich des Parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$ ist gegeben durch

$$\mathbf{I}(\boldsymbol{\theta}) = \mathbb{E} \left(\frac{\partial l(\boldsymbol{\theta}|\mathbf{y})}{\partial \boldsymbol{\theta}} \frac{\partial l(\boldsymbol{\theta}|\mathbf{y})}{\partial \boldsymbol{\theta}^T} \right) = -\mathbb{E} \left(\frac{\partial^2 l(\boldsymbol{\theta}|\mathbf{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right), \quad (4.12)$$

wobei die zweite Identität durch die Einhaltung gewisser Regularitätsbedingungen gegeben ist. Zusätzlich sei anzumerken, dass $l(\boldsymbol{\theta}|\mathbf{y})$ der Log-Likelihoodfunktion $l(\boldsymbol{\mu}, \phi|\mathbf{y})$ entspricht, es wurde nur eine andere Notation verwendet, um die Abhängigkeit vom Parameter $\boldsymbol{\theta}$ zu verdeutlichen. Die Fisher-Informationsmatrix teilt sich demnach in vier Submatrizen auf,

$$\mathbf{I}(\boldsymbol{\beta}, \phi) = \begin{pmatrix} \mathbf{I}_{\beta\beta}(\boldsymbol{\beta}, \phi) & \mathbf{I}_{\beta\phi}(\boldsymbol{\beta}, \phi) \\ \mathbf{I}_{\phi\beta}(\boldsymbol{\beta}, \phi) & \mathbf{I}_{\phi\phi}(\boldsymbol{\beta}, \phi) \end{pmatrix}, \quad (4.13)$$

wobei $\mathbf{I}_{\beta\phi}(\boldsymbol{\beta}, \phi) = \mathbf{I}_{\phi\beta}^T(\boldsymbol{\beta}, \phi)$ gilt.

Der erste Teil, $\mathbf{I}_{\beta\beta}(\boldsymbol{\beta}, \phi)$, ergibt sich bei zweifacher Ableitung der Log-Likelihoodfunktion nach $\boldsymbol{\beta}$, zur Vereinfachung wird zunächst nach β_j und β_k abgeleitet. Unter Verwendung von (4.7) gilt

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\mu}, \phi|\mathbf{y})}{\partial \beta_j \partial \beta_k} &= \sum_{i=1}^n \frac{\partial}{\partial \mu_i} \left(\frac{\partial l(\mu_i, \phi|y_i)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right) \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_k} \\ &= \sum_{i=1}^n \left(\frac{\partial^2 l(\mu_i, \phi|y_i)}{\partial \mu_i^2} \frac{\partial \mu_i}{\partial \eta_i} + \frac{\partial l(\mu_i, \phi|y_i)}{\partial \mu_i} \frac{\partial}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \right) \frac{\partial \mu_i}{\partial \eta_i} x_{ij} x_{ik}. \end{aligned}$$

Die zweite Ableitung der Log-Likelihoodfunktionen nach μ_i lässt sich mit (4.8) darstellen als

$$\frac{\partial^2 l(\mu_i, \phi|y_i)}{\partial \mu_i^2} = \frac{\partial}{\partial \mu_i} \phi(y_i^* - \mu_i^*) = -\phi^2 [\psi'(\mu_i \phi) + \psi'((1 - \mu_i) \phi)], \quad (4.14)$$

zusammen mit der Trigammafunktion $\psi'(z) = \frac{d\psi(z)}{dz}$. Da $\mathbb{E} \left(\frac{\partial l(\mu_i, \phi|y_i)}{\partial \mu_i} \right) = 0$ für die Log-Likelihoodfunktion der Beta-Verteilung erfüllt ist, gilt unter Verwendung von (4.14) für den Erwartungswert der zweiten Ableitung

$$\begin{aligned} \mathbb{E} \left(\frac{\partial^2 l(\boldsymbol{\mu}, \phi|\mathbf{y})}{\partial \beta_j \partial \beta_k} \right) &= -\phi^2 \sum_{i=1}^n [\psi'(\mu_i \phi) + \psi'((1 - \mu_i) \phi)] \frac{x_{ij} x_{ik}}{(g'(\mu_i))^2} \\ &= -\phi \sum_{i=1}^n w_i x_{ij} x_{ik}, \end{aligned}$$

mit

$$w_i = \phi[\psi'(\mu_i\phi) + \psi'((1 - \mu_i)\phi)] \frac{1}{(g'(\mu_i))^2}. \quad (4.15)$$

Zusammen mit der Diagonalmatrix $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$ ergibt sich daher

$$\mathbf{I}_{\beta\beta}(\boldsymbol{\beta}, \phi) = -\mathbf{E} \left(\frac{\partial^2 l(\boldsymbol{\mu}, \phi | \mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right) = \phi \mathbf{X}^T \mathbf{W} \mathbf{X}. \quad (4.16)$$

Um die Submatrix $\mathbf{I}_{\beta\phi}(\boldsymbol{\beta}, \phi)$ zu berechnen, werden zunächst wieder die Ableitungen der Log-Likelihoodfunktion betrachtet. Hierbei wird zuerst nach β_j und anschließend nach ϕ abgeleitet. Unter Verwendung von (4.9) folgt

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\mu}, \phi | \mathbf{y})}{\partial \beta_j \partial \phi} &= \sum_{i=1}^n \frac{\partial}{\partial \phi} \left(\phi(y_i^* - \mu_i^*) \frac{x_{ij}}{g'(\mu_i)} \right) \\ &= \sum_{i=1}^n \left(\frac{\partial}{\partial \phi} \phi y_i^* - \frac{\partial}{\partial \phi} \phi \mu_i^* \right) \frac{x_{ij}}{g'(\mu_i)} \\ &= \sum_{i=1}^n \left((y_i^* - \mu_i^*) - \phi \frac{\partial \mu_i^*}{\partial \phi} \right) \frac{x_{ij}}{g'(\mu_i)}, \end{aligned}$$

wobei für die letzte Identität die Produktregel verwendet wurde, da μ_i^* von ϕ abhängig ist. Die Ableitung von μ_i^* nach ϕ ist gegeben durch

$$\frac{\partial \mu_i^*}{\partial \phi} = \psi'(\mu_i\phi)\mu_i - \psi'((1 - \mu_i)\phi)(1 - \mu_i). \quad (4.17)$$

Somit resultiert für die zweite gemischte Ableitung der Log-Likelihoodfunktion

$$\frac{\partial^2 l(\boldsymbol{\mu}, \phi | \mathbf{y})}{\partial \beta_j \partial \phi} = \sum_{i=1}^n ((y_i^* - \mu_i^*) - c_i) \frac{x_{ij}}{g'(\mu_i)}, \quad (4.18)$$

mit

$$c_i = \phi[\psi'(\mu_i\phi)\mu_i - \psi'((1 - \mu_i)\phi)(1 - \mu_i)]. \quad (4.19)$$

Da $\mathbf{E}(y_i^*) = \mu_i^*$ gilt (vgl. Bemerkung auf Seite 50), folgt für den Erwartungswert von (4.18)

$$\mathbf{E} \left(\frac{\partial^2 l(\boldsymbol{\mu}, \phi | \mathbf{y})}{\partial \beta_j \partial \phi} \right) = - \sum_{i=1}^n c_i \frac{x_{ij}}{g'(\mu_i)}.$$

Unter Verwendung des Vektors $\mathbf{c} = (c_1, \dots, c_n)^T$ und der in (4.10) eingeführten Matrix \mathbf{D} gilt schließlich für die Submatrix $\mathbf{I}_{\beta\phi}(\boldsymbol{\beta}, \phi)$ der Fischer-Informationsmatrix

$$\mathbf{I}_{\beta\phi}(\boldsymbol{\beta}, \phi) = -\mathbf{E} \left(\frac{\partial^2 l(\boldsymbol{\mu}, \phi | \mathbf{y})}{\partial \boldsymbol{\beta} \partial \phi} \right) = \mathbf{X}^T \mathbf{D} \mathbf{c}. \quad (4.20)$$

Auch für die Berechnung der skalaren Größe $\mathbf{I}_{\phi\phi}(\boldsymbol{\beta}, \phi)$ wird zuerst wieder die zweite Ableitung der Log-Likelihoodfunktion betrachtet. Zusammen mit dem Ergebnis von $U_\phi(\boldsymbol{\beta}, \phi)$ und unter Verwendung von (4.17) ergibt sich

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\mu}, \phi | \mathbf{y})}{\partial \phi^2} &= \sum_{i=1}^n \frac{\partial}{\partial \phi} \left(\mu_i (y_i^* - \mu_i^*) + \log(1 - y_i) + \psi(\phi) - \psi((1 - \mu_i)\phi) \right) \\ &= \sum_{i=1}^n \left(-\frac{\partial}{\partial \phi} \mu_i \mu_i^* + \frac{\partial}{\partial \phi} \psi(\phi) - \frac{\partial}{\partial \phi} \psi((1 - \mu_i)\phi) \right) \\ &= -\sum_{i=1}^n \left(\psi'(\mu_i \phi) \mu_i^2 + \psi'((1 - \mu_i)\phi) (1 - \mu_i)^2 - \psi'(\phi) \right), \end{aligned}$$

wobei für den letzten Schritt die binomische Formel benötigt wird. Wird zusätzlich die Diagonalmatrix $\mathbf{V} = \text{diag}(v_1, \dots, v_n)$ mit den Einträgen

$$v_i = \psi'(\mu_i \phi) \mu_i^2 + \psi'((1 - \mu_i)\phi) (1 - \mu_i)^2 - \psi'(\phi) \quad (4.21)$$

eingeführt, so folgt

$$\mathbf{I}_{\phi\phi}(\boldsymbol{\beta}, \phi) = -\mathbf{E} \left(\frac{\partial^2 l(\boldsymbol{\mu}, \phi | \mathbf{y})}{\partial \phi^2} \right) = \text{trace}(\mathbf{V}). \quad (4.22)$$

Die Ergebnisse (4.16), (4.20) und (4.22) ergeben zusammen die Fisher-Informationsmatrix $\mathbf{I}(\boldsymbol{\beta}, \phi)$ aus (4.13).

Bemerkung:

Ferrari und Cribari-Neto (2004) zeigten, dass unter Einhaltung gewisser Regularitätsbedingungen

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\phi} \end{pmatrix} \xrightarrow{D} \mathcal{N}_{p+1} \left(\begin{pmatrix} \boldsymbol{\beta} \\ \phi \end{pmatrix}, n\mathbf{I}^{-1}(\boldsymbol{\beta}, \phi) \right)$$

erfüllt ist.

4.2.3 Newton-Verfahren und Fisher-Scoring

In Abschnitt 4.2.1 wurden die Scorefunktionen $\mathbf{U}_\beta(\boldsymbol{\beta}, \phi)$ und $U_\phi(\boldsymbol{\beta}, \phi)$ hergeleitet. Werden diese Funktionen Null gesetzt, können die Parameterschätzer $\hat{\boldsymbol{\beta}}$ und $\hat{\phi}$ unter Verwendung numerischer Verfahren bestimmt werden. In diesem Abschnitt wird die Anwendung des Newton-Verfahrens zur Schätzung von $\boldsymbol{\beta}$ genauer betrachtet.

Wird die Dispersion ϕ als feste Größe angenommen, kann der Parametervektor $\boldsymbol{\beta}$ durch Nullsetzen der Scorefunktion $\mathbf{U}_\beta(\boldsymbol{\beta}, \phi)$ aus (4.11) geschätzt werden. Bei den Generalisierten Linearen Modellen wurde die Newton-Methode zusammen mit der Fisher-Scoring Technik als numerisches Verfahren gewählt. Für Beta-Regressionsmodelle führt

das Newton-Verfahren zur Iterationsvorschrift

$$\begin{aligned}\boldsymbol{\beta}^{(t+1)} &= \boldsymbol{\beta}^{(t)} + \left(-\frac{\partial^2 l(\boldsymbol{\mu}, \phi | \mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right)^{-1} \frac{\partial l(\boldsymbol{\mu}, \phi | \mathbf{y})}{\partial \boldsymbol{\beta}} \\ &= \boldsymbol{\beta}^{(t)} + \left(-\frac{\partial^2 l(\boldsymbol{\mu}, \phi | \mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right)^{-1} \mathbf{U}_\beta(\boldsymbol{\beta}, \phi), \quad t = 0, 1, \dots,\end{aligned}$$

wobei die Ableitungen der Log-Likelihood an der Stelle $\boldsymbol{\beta}^{(t)}$ ausgewertet werden und $\boldsymbol{\beta}^{(t)}$ den Schätzwert der Iteration t darstellt. Bei der Fisher-Scoring Methode wird die Hessematrix durch ihren Erwartungswert ersetzt, welcher laut (4.12) der Informationsmatrix $I_{\beta\beta}$ entspricht. Zusammen mit (4.11) und (4.16) lautet die Iterationsvorschrift zur Berechnung von $\boldsymbol{\beta}^{(t+1)}$, $t = 1, 2, \dots$, demnach

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + (\mathbf{X}^T \mathbf{W}^{(t)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{D}^{(t)}(\mathbf{y}^* - \boldsymbol{\mu}^{*(t)}).$$

Werden zusätzlich die Pseudobeobachtungen

$$\mathbf{z}^{(t)} = \mathbf{X} \boldsymbol{\beta}^{(t)} + \mathbf{W}^{(t)-1} \mathbf{D}^{(t)}(\mathbf{y}^* - \boldsymbol{\mu}^{*(t)})$$

eingeführt, so kann die Iterationsvorschrift geschrieben werden als

$$\boldsymbol{\beta}^{(t+1)} = (\mathbf{X}^T \mathbf{W}^{(t)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(t)} \mathbf{z}^{(t)}.$$

Der MLE $\hat{\boldsymbol{\beta}}$ resultiert aus der Konvergenz dieses Verfahrens. Die Dispersion ϕ wurde hierbei als feste Größe angenommen, in der Praxis wird diese durch ihren MLE $\hat{\phi}$ ersetzt. Genauer findet sich bei Espinheira, Ferrari und Cribari-Neto (2008b).

4.2.4 Modelldiagnostik

Nach der Aufstellung des Modells und der Schätzung der Parameter, sollten diagnostische Analysen zur Überprüfung der Anpassungsgüte durchgeführt werden. Dazu stehen mehrere Möglichkeiten zur Verfügung, von denen nun einige besprochen werden. Sei dazu M das Modell, das untersucht werden soll und $\hat{\boldsymbol{\mu}}$ der Maximum Likelihood Schätzer des Erwartungswertes unter diesem Modell.

Modelldiskrepanz und Deviance

Wie bei den Generalisierten Linearen Modellen, kann auch bei der Beta-Regression die Deviance eines Modells berechnet werden. Ferrari und Cribari-Neto (2004) führen zunächst den Begriff der Modellierungsdiskrepanz ein. Für bekannte Dispersion ϕ ist die Modelldiskrepanz $D(\mathbf{y} | \hat{\boldsymbol{\mu}}, \phi)$ von M gegeben durch

$$D(\mathbf{y} | \hat{\boldsymbol{\mu}}, \phi) = \sum_{i=1}^n 2(l(\tilde{\mu}_i, \phi | y_i) - l(\hat{\mu}_i, \phi | y_i)) = 2(l(\tilde{\boldsymbol{\mu}}, \phi | \mathbf{y}) - l(\hat{\boldsymbol{\mu}}, \phi | \mathbf{y})),$$

wobei $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \dots, \tilde{\mu}_n)^T$ den Vektor der geschätzten Erwartungswerte des saturierten Modells darstellt. Ferrari und Cribari-Neto (2004) zeigen ebenfalls, dass $\tilde{\boldsymbol{\mu}} \approx \mathbf{y}$ für große Dispersion ϕ gilt. Sollte ϕ eine unbekannte Größe mit MLE $\hat{\phi}$ sein, so stellt $D(\mathbf{y} | \hat{\boldsymbol{\mu}}, \hat{\phi})$ eine Approximation zu $D(\mathbf{y} | \hat{\boldsymbol{\mu}}, \phi)$ dar und kann wie üblich Deviance genannt werden. Ziel ist natürlich wieder die Minimierung der Deviance bzw. der Modelldiskrepanz.

Residuen

Residuen sind ein oft verwendetes Maß der lokalen Anpassungsgüte, da sie gut veranschaulicht und interpretiert werden können. So kann zum Beispiel ein Punktediagramm, das die Residuen in Abhängigkeit der Indizes zeigt, auf mögliche Muster oder Ausschläge untersucht werden. Möglich wäre auch die Untersuchung eines Diagramms, das den Zusammenhang zwischen den Residuen und den geschätzten linearen Prädiktoren darstellt. Ferrari und Cribari-Neto (2004) und Espinheira et al. (2008b) schlagen folgende Residuen vor:

- **Deviance Residuen** (Ferrari und Cribari-Neto, 2004)

Wird r_i^d definiert als

$$r_i^d = \sqrt{2} \operatorname{sign}(y_i - \hat{\mu}_i) \left[l(\tilde{\mu}_i, \hat{\phi} | y_i) - l(\hat{\mu}_i, \hat{\phi} | y_i) \right]^{\frac{1}{2}},$$

so kann die Deviance geschrieben werden als

$$D(\mathbf{y} | \hat{\boldsymbol{\mu}}, \hat{\phi}) = \sum_{i=1}^n (r_i^d)^2.$$

Aus diesem Grund werden die Größen r_i^d , für $i = 1, \dots, n$, Deviance Residuen genannt. Ist das Deviance Residuum einer Beobachtung y_i groß, weist das auf eine schlechte Modellanpassung in diesem Punkt hin.

- **Standardisierte Residuen** (Ferrari und Cribari-Neto, 2004)

Die standardisierten Residuen sind definiert als

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{\operatorname{Var}}(y_i)}} = \frac{y_i - \hat{\mu}_i}{\sqrt{\frac{\hat{\mu}_i(1-\hat{\mu}_i)}{1+\hat{\phi}}}}. \quad (4.23)$$

Hierbei bezeichnet $\widehat{\operatorname{Var}}(y_i)$ die geschätzte Varianz von y_i , also die Varianz von y_i ausgewertet im MLE des Erwartungswertes und der Dispersion. Diese Residuen sind zu vergleichen mit den Pearson Residuen der Generalisierten Linearen Modelle aus Gleichung (2.18).

- **Gewichtete Residuen** (Espinheira et al., 2008b)

Wird die Dispersion ϕ als bekannte Größe angenommen, so ist das gewichtete Residuum definiert als

$$r_i^* = \frac{y_i^* - \hat{\mu}_i^*}{\sqrt{\phi \hat{p}_i}},$$

mit

$$y_i^* = \log\left(\frac{y_i}{1-y_i}\right),$$

$$\hat{\mu}_i^* = \psi(\hat{\mu}_i \phi) - \psi((1-\hat{\mu}_i)\phi),$$

$$\hat{p}_i = \psi'(\hat{\mu}_i \phi) + \psi'((1-\hat{\mu}_i)\phi).$$

- **Standardisierte gewichtete Residuen 1** (Espinheira et al., 2008b)

Espinheira et al. (2008b) zeigten, dass $\text{Var}(y_i^*) = p_i$ gilt. Aus diesem Grund sind die standardisierten gewichteten Residuen 1 definiert als

$$r_i^w = \sqrt{\phi} r_i^* = \frac{y_i^* - \hat{\mu}_i^*}{\sqrt{\hat{p}_i}}.$$

- **Standardisierte gewichtete Residuen 2** (Espinheira et al., 2008b)

Diese Residuen unterscheiden sich von den standardisierten gewichteten Residuen 1 nur im Standardisierungsfaktor. Sie sind gegeben durch

$$r_i^{ww} = \frac{r_i^*}{\sqrt{\phi^{-1}(1 - h_{ii})}} = \frac{r_i^w}{\sqrt{1 - h_{ii}}} = \frac{y_i^* - \hat{\mu}_i^*}{\sqrt{\hat{p}_i(1 - h_{ii})}},$$

wobei h_{ii} die Diagonalelemente der Hutmatrix

$$\mathbf{H} = \widehat{\mathbf{W}}^{\frac{1}{2}} \mathbf{X} (\mathbf{X}^T \widehat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T \widehat{\mathbf{W}}^{\frac{1}{2}} \quad (4.24)$$

darstellen (vgl. Ferrari und Cribari-Neto, 2004). Es sei anzumerken, dass die Varianz von r_i^* gegeben ist durch $\text{Var}(r_i^*) = \phi^{-1}(1 - h_{ii})$.

Bei den letzten drei Residuen von Espinheira et al. (2008b) wurde die Dispersion ϕ als feste, bekannte Größe angenommen. Da dies in der Praxis jedoch meistens nicht der Fall ist, wird ϕ bei der Berechnung der Residuen einfach durch den Maximum Likelihood Schätzer $\hat{\phi}$ ersetzt.

Generalisierte Hebelwirkung

Um jene Datenpunkte zu finden, welche einen großen Einfluss auf die Schätzung der Parameter ausüben, schlagen Ferrari und Cribari-Neto (2004) die Betrachtung der von Wei et al. (1998) eingeführten generalisierten Hebelwirkung vor.

Um die generalisierte Hebelwirkung zu ermitteln, kann die aus Lemma 2.7 bekannte Formel bzw. der auf die Log-Likelihoodfunktion angepasste Ausdruck (2.25) verwendet werden. Im Gegensatz zum Linearen Modell und dem Generalisierten Linearen Modell hängt der geschätzte Erwartungswert des Beta-Regressionsmodells nicht nur vom geschätzten Parametervektor $\hat{\beta}$ sondern auch von der geschätzten Dispersion $\hat{\phi}$ ab. Somit ist auch die generalisierte Hebelwirkung eine von $\hat{\alpha} = (\hat{\beta}^T, \hat{\phi})^T$ abhängige Größe. Ein weiterer Unterschied besteht darin, dass die generalisierte Hebelwirkung des Beta-Regressionsmodells nicht mit der Hutmatrix übereinstimmt, sondern eine etwas komplexere Form annimmt. Das Ergebnis der Matrix $\mathbf{GL}(\hat{\alpha})$ sowie Angaben zur Herleitung finden sich bei Ferrari und Cribari-Neto (2004).

Cook-Distanz

Die Cook-Distanz ist ein Maß für den Einfluss einer einzelnen Beobachtung auf den Parameterschätzer. Ist die Cook-Distanz eines Punktes groß, sollte die Validität dieser

Beobachtung überprüft werden. Die Cook-Distanz der i -ten Beobachtung ist gegeben durch den Ausdruck

$$C_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})^T \mathbf{X}^T \widehat{\mathbf{W}} \mathbf{X} (\hat{\beta} - \hat{\beta}_{(i)})}{p},$$

wobei p die Dimension des Parametervektors β darstellt. Der Ausdruck $\hat{\beta}_{(i)}$ bezeichnet den Maximum Likelihood Schätzer von β nach Elimination der i -ten Beobachtung. Die Cook-Distanz misst demnach die Distanz zwischen dem MLE bezüglich der gesamten Stichprobe und jenem MLE, welcher durch das Weglassen der i -ten Beobachtung generiert wird.

Um zu vermeiden, dass $n + 1$ Modelle angepasst werden müssen, schlagen Ferrari und Cribari-Neto (2004) eine Approximation der Cook-Distanz mittels

$$\tilde{C}_i = \frac{h_{ii} r_i^2}{p(1 - h_{ii})^2}$$

vor. Hierbei bezeichnet h_{ii} die Diagonalelemente der Hutmatrix (4.24) und r_i die standardisierten Residuen aus (4.23). Diagramme, die C_i beziehungsweise \tilde{C}_i gegen den Index i abtragen, liefern eine gute Möglichkeit um Beobachtungen mit großer Cook-Distanz leicht und schnell zu erkennen.

Bemerkung:

Ergänzend zu den bereits vorgestellten Methoden der Modelldiagnostik, diskutieren Espinheira, Ferrari und Cribari-Neto (2008a) Cook-ähnliche Distanzen sowie lokale Einflussfaktoren.

4.3 Gemeinsame Modellierung von Erwartungswert und Dispersion

Das Grundmodell von Ferrari und Cribari-Neto (2004) liefert eine gute Basis zur Modellierung von Response-Werten aus dem offenen Einheitsintervall $(0, 1)$. Um die Flexibilität des Modells zu erhöhen, wird nun eine Erweiterung betrachtet, die von Smithson und Verkuilen (2006) angewendet und später von Simas, Barreto-Souza und Rocha (2010) formell eingeführt wurde. Dabei wird die Dispersion ϕ nicht länger als konstant angenommen, sondern jeder Beobachtung eine eigene Dispersion zugeordnet. Zusätzlich werden die Dispersionsparameter mittels Linkfunktion und linearen Prädiktoren modelliert. Das heißt es werden nicht länger nur die Erwartungswerte modelliert, von nun an erfolgt eine gemeinsame Modellierung von Erwartungswert und Dispersion. Dieses Modell ist bekannt unter den Namen „variable dispersion beta regression model“ (vgl. Cribari-Neto und Zeileis, 2010) und „double index beta regression model“ (vgl. Grün, Kosmidis und Zeileis, 2012).

Smithson und Verkuilen (2006) beschränken sich bei der Wahl der Linkfunktion zur Modellierung der Dispersion auf den Logarithmus. Simas et al. (2010) hingegen schlagen

ergänzend auch die Quadratwurzel und die identische Abbildung vor. In dieser Arbeit werden die Ergebnisse von Simas et al. (2010) zur gemeinsamen Modellierung von Erwartungswert und Dispersion adaptiert.

Sei nun $\mathbf{y} = (y_1, \dots, y_n)^T$ ein Vektor bestehend aus n unabhängigen Zufallsvariablen, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ der Vektor der Erwartungswerte und $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)^T$ der Dispersionsvektor. Das Beta-Regressionsmodell zur gemeinsamen Modellierung von Erwartungswert und Dispersion ist durch die Annahmen

$$\begin{aligned} y_i &\stackrel{ind}{\sim} \mathcal{B}(\mu_i, \phi_i), \\ g_1(\mu_i) &= \eta_{1i} = \mathbf{x}_i^T \boldsymbol{\beta}, \\ g_2(\phi_i) &= \eta_{2i} = \mathbf{z}_i^T \boldsymbol{\delta}, \end{aligned} \tag{4.25}$$

für $i = 1, \dots, n$ spezifiziert. Hierbei stellen $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ und $\boldsymbol{\delta} = (\delta_1, \dots, \delta_q)^T$ die beiden zu schätzenden Vektoren der Regressionsparameter und $\boldsymbol{\eta}_1 = (\eta_{11}, \dots, \eta_{1n})^T$ und $\boldsymbol{\eta}_2 = (\eta_{21}, \dots, \eta_{2n})^T$ die linearen Prädiktoren dar. Die bekannten, erklärenden Variablen $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ und $\mathbf{z}_i = (z_{i1}, \dots, z_{iq})^T$ können zu den Designmatrizen $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ und $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)^T$ zusammengefasst werden. Die Funktionen $g_1(\cdot)$ und $g_2(\cdot)$ bezeichnen die beiden bekannten Linkfunktionen und müssen demnach monoton und zweimal stetig differenzierbar sein.

Die Log-Likelihoodfunktion einer $\mathcal{B}(\mu_i, \phi_i)$ -verteilten Größe y_i resultiert aus (4.5) und ist gegeben durch

$$\begin{aligned} l(\mu_i, \phi_i | y_i) &= \log \Gamma(\phi_i) - \log \Gamma(\mu_i \phi_i) - \log \Gamma((1 - \mu_i) \phi_i) + (\mu_i \phi_i - 1) \log(y_i) \\ &\quad + [(1 - \mu_i) \phi_i - 1] \log(1 - y_i). \end{aligned} \tag{4.26}$$

Da y_1, \dots, y_n unter den Modellannahmen von (4.25) unabhängig sind, folgt für die gemeinsame Log-Likelihoodfunktion

$$l(\boldsymbol{\mu}, \boldsymbol{\phi} | \mathbf{y}) = \sum_{i=1}^n l(\mu_i, \phi_i | y_i).$$

4.3.1 Maximum Likelihood Schätzung

Nachdem das Modell definiert wurde, folgt wieder die Schätzung der Parameter. Es ist bereits bekannt, dass die Erwartungswerte $\mu_i = g_1^{-1}(\eta_{1i}) = g_1^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})$ Funktionen in $\boldsymbol{\beta}$ darstellen. Äquivalent dazu stellen die Dispersionen $\phi_i = g_2^{-1}(\eta_{2i}) = g_2^{-1}(\mathbf{z}_i^T \boldsymbol{\delta})$ Funktionen in $\boldsymbol{\delta}$ dar. Bei den zu schätzenden Parametern handelt es sich demnach um die Vektoren $\boldsymbol{\beta}$ und $\boldsymbol{\delta}$. Wie auch schon beim Grundmodell in Abschnitt 4.2 werden zur Berechnung der Maximum Likelihood Schätzer zunächst die Scorevektoren bezüglich der Ableitungen nach den beiden Parametervektoren hergeleitet und anschließend simultan Null gesetzt.

Zuerst wird der Scorevektor bezüglich β hergeleitet. Vergleichbar mit Abschnitt 4.2.1 ergibt sich für die Ableitung der Log-Likelihoodfunktion nach einem beliebigen β_j

$$\frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\phi} | \mathbf{y})}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l(\mu_i, \phi_i | y_i)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l(\mu_i, \phi_i | y_i)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_j}, \quad (4.27)$$

wobei die ersten beiden Teile des Produkts gegeben sind durch

$$\begin{aligned} \frac{\partial \eta_{1i}}{\partial \beta_j} &= \frac{\partial \mathbf{x}_i^T \boldsymbol{\beta}}{\partial \beta_j} = x_{ij}, \\ \frac{\partial \mu_i}{\partial \eta_{1i}} &= \frac{1}{\partial g_1(\mu_i) / \partial \mu_i} = \frac{1}{g_1'(\mu_i)}. \end{aligned}$$

Ähnlich zum Ergebnis (4.8) resultiert für die Ableitung der Log-Likelihoodfunktion nach μ_i

$$\frac{\partial l(\mu_i, \phi_i | y_i)}{\partial \mu_i} = \phi_i \left(\log \frac{y_i}{1 - y_i} - \psi(\mu_i \phi_i) + \psi((1 - \mu_i) \phi_i) \right) = \phi_i (y_i^* - \ddot{\mu}_i^*), \quad (4.28)$$

mit dem bereits bekannten Ausdruck

$$y_i^* = \log \frac{y_i}{1 - y_i}$$

und der ergänzenden Bezeichnung

$$\ddot{\mu}_i^* = \psi(\mu_i \phi_i) - \psi((1 - \mu_i) \phi_i).$$

Es sei anzumerken, dass die zwei Punkte über den Buchstaben verwendet werden, um die vorkommenden Vektoren und Matrizen von den bereits eingeführten Ausdrücken des Grundmodells unterscheiden zu können.

Zusammengefasst folgt für die Ableitung der Log-Likelihoodfunktion nach β_j

$$\frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\phi} | \mathbf{y})}{\partial \beta_j} = \sum_{i=1}^n \phi_i (y_i^* - \ddot{\mu}_i^*) \frac{x_{ij}}{g_1'(\mu_i)}. \quad (4.29)$$

Durch Definition der Diagonalmatrizen

$$\begin{aligned} \ddot{\mathbf{D}}_1 &= \text{diag} \left(\frac{1}{g_1'(\mu_1)}, \dots, \frac{1}{g_1'(\mu_n)} \right), \\ \boldsymbol{\Phi} &= \text{diag}(\phi_1, \dots, \phi_n) \end{aligned}$$

und der Vektoren $\mathbf{y}^* = (y_1^*, \dots, y_n^*)^T$ und $\ddot{\boldsymbol{\mu}}^* = (\ddot{\mu}_1^*, \dots, \ddot{\mu}_n^*)^T$ resultiert der Scorevektor

$$\dot{\mathbf{U}}_{\beta}(\boldsymbol{\beta}, \boldsymbol{\delta}) = \frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\phi} | \mathbf{y})}{\partial \boldsymbol{\beta}} = \mathbf{X}^T \boldsymbol{\Phi} \ddot{\mathbf{D}}_1 (\mathbf{y}^* - \ddot{\boldsymbol{\mu}}^*). \quad (4.30)$$

4.3. Gemeinsame Modellierung von Erwartungswert und Dispersion

Im nächsten Schritt soll der Scorevektor bezüglich des Parametervektors $\boldsymbol{\delta}$ betrachtet werden. Für die Ableitung nach einem beliebigen δ_j des Dispersionsmodells ergibt sich

$$\frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\phi} | \mathbf{y})}{\partial \delta_j} = \sum_{i=1}^n \frac{\partial l(\mu_i, \phi_i | y_i)}{\partial \delta_j} = \sum_{i=1}^n \frac{\partial l(\mu_i, \phi_i | y_i)}{\partial \phi_i} \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \delta_j}. \quad (4.31)$$

Ähnlich wie zuvor können die einzelnen Teile geschrieben werden als

$$\begin{aligned} \frac{\partial \eta_{2i}}{\partial \delta_j} &= \frac{\partial \mathbf{z}_i^T \boldsymbol{\delta}}{\partial \delta_j} = z_{ij}, \\ \frac{\partial \phi_i}{\partial \eta_{2i}} &= \frac{1}{\partial g_2(\phi_i) / \partial \phi_i} = \frac{1}{g_2'(\phi_i)}, \end{aligned}$$

$$\frac{\partial l(\mu_i, \phi_i | y_i)}{\partial \phi_i} = \mu_i(y_i^* - \check{\mu}_i^*) + \log(1 - y_i) + \psi(\phi_i) - \psi((1 - \mu_i)\phi_i) = \check{a}_i, \quad (4.32)$$

wobei das Ergebnis (4.32) aus Notationsgründen von nun an mit \check{a}_i bezeichnet wird. Zusammengesetzt ergibt sich für die Scorefunktion

$$\begin{aligned} \frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\phi} | \mathbf{y})}{\partial \delta_j} &= \sum_{i=1}^n \left(\mu_i(y_i^* - \check{\mu}_i^*) + \log(1 - y_i) + \psi(\phi_i) - \psi((1 - \mu_i)\phi_i) \right) \frac{z_{ij}}{g_2'(\phi_i)} \\ &= \sum_{i=1}^n \check{a}_i \frac{z_{ij}}{g_2'(\phi_i)}. \end{aligned} \quad (4.33)$$

Durch Hinzunahme der Diagonalmatrix

$$\check{\mathbf{D}}_2 = \text{diag} \left(\frac{1}{g_2'(\phi_1)}, \dots, \frac{1}{g_2'(\phi_n)} \right)$$

und des Vektors $\check{\mathbf{a}} = (\check{a}_1, \dots, \check{a}_n)^T$ resultiert schließlich der Scorevektor

$$\check{\mathbf{U}}_{\boldsymbol{\delta}}(\boldsymbol{\beta}, \boldsymbol{\delta}) = \frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\phi} | \mathbf{y})}{\partial \boldsymbol{\delta}} = \mathbf{Z}^T \check{\mathbf{D}}_2 \check{\mathbf{a}}. \quad (4.34)$$

Werden die beiden Scorevektoren aus (4.30) und (4.34) simultan Null gesetzt, können die Maximum Likelihood Schätzer $\hat{\boldsymbol{\beta}}$ und $\hat{\boldsymbol{\delta}}$ berechnet werden. Da das Gleichungssystem aber kein lineares System in $\boldsymbol{\beta}$ oder $\boldsymbol{\delta}$ darstellt, müssen numerische Verfahren zur Lösung herangezogen werden. Zum Beispiel kann wie in Abschnitt 4.2.3 das Newton-Verfahren in Kombination mit der Fisher-Scoring Methode verwendet werden. Da die Vorgehensweise in Abschnitt 4.2.3 bereits ausführlich behandelt wurde, wird sie für das Beta-Regressionsmodell zur gemeinsamen Modellierung von Erwartungswert und Dispersion nicht noch einmal besprochen. Es folgt lediglich die Herleitung der dazu benötigten Fisher-Informationsmatrix.

Bemerkung:

Wie zuvor in Abschnitt 4.2.1 gilt auch hier unter der Einhaltung gewisser Regularitätsbedingungen, dass die Erwartungswerte von (4.29) und (4.33) gleich Null sind. Daraus folgt

$$\begin{aligned} \mathbb{E}(y_i^*) &= \check{\mu}_i^*, \\ \mathbb{E}(\log(1 - y_i)) &= \psi((1 - \mu_i)\phi_i) - \psi(\phi_i). \end{aligned}$$

4.3.2 Fisher-Information

Aus Abschnitt 4.2.3 ist bekannt, dass für das vorgeschlagene numerische Lösungsverfahren zur Berechnung der Maximum Likelihood Schätzer die Fisher-Informationsmatrix benötigt wird. Die Vorgehensweise zur Herleitung der Fisher-Informationsmatrix ist vergleichbar mit jener aus Abschnitt 4.2.2 bezüglich des Grundmodells.

Aus Notationsgründen werden die Parameter β und δ zu einem Vektor $\theta = (\beta^T, \delta^T)^T$ zusammengefasst. Als Grundlage zur Herleitung der Fisher-Informationsmatrix dient wieder die aus (4.12) bekannte Identität

$$\mathbf{I}(\theta) = \mathbb{E} \left(\frac{\partial l(\theta|\mathbf{y})}{\partial \theta} \frac{\partial l(\theta|\mathbf{y})}{\partial \theta^T} \right) = -\mathbb{E} \left(\frac{\partial^2 l(\theta|\mathbf{y})}{\partial \theta \partial \theta^T} \right).$$

Die daraus resultierende Fisher-Informationsmatrix kann wieder in vier Submatrizen unterteilt werden,

$$\mathbf{I}(\beta, \delta) = \begin{pmatrix} \mathbf{I}_{\beta\beta}(\beta, \delta) & \mathbf{I}_{\beta\delta}(\beta, \delta) \\ \mathbf{I}_{\delta\beta}(\beta, \delta) & \mathbf{I}_{\delta\delta}(\beta, \delta) \end{pmatrix},$$

wobei $\mathbf{I}_{\beta\delta} = \mathbf{I}_{\delta\beta}^T$ gilt.

Für die Berechnung der ersten Submatrix $\mathbf{I}_{\beta\beta}(\beta, \delta)$ wird die zweifache Ableitung der Log-Likelihoodfunktion nach β benötigt. Zur Vereinfachung wird zunächst nach β_j und β_k abgeleitet. Vergleichbar mit Abschnitt 4.2.2 gilt unter Verwendung von (4.27)

$$\begin{aligned} \frac{\partial^2 l(\mu, \phi|\mathbf{y})}{\partial \beta_j \partial \beta_k} &= \sum_{i=1}^n \frac{\partial}{\partial \mu_i} \left(\frac{\partial l(\mu_i, \phi_i|y_i)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_j} \right) \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_k} \\ &= \sum_{i=1}^n \left(\frac{\partial^2 l(\mu_i, \phi_i|y_i)}{\partial \mu_i^2} \frac{\partial \mu_i}{\partial \eta_{1i}} + \frac{\partial l(\mu_i, \phi_i|y_i)}{\partial \mu_i} \frac{\partial}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \right) \frac{\partial \mu_i}{\partial \eta_{1i}} x_{ij} x_{ik}. \end{aligned}$$

Die zweite Ableitung der Log-Likelihoodfunktionen nach μ_i lässt sich mit (4.28) darstellen als

$$\frac{\partial^2 l(\mu_i, \phi_i|y_i)}{\partial \mu_i^2} = \frac{\partial}{\partial \mu_i} \phi_i (y_i^* - \check{\mu}_i^*) = -\phi_i^2 [\psi'(\mu_i \phi_i) + \psi'((1 - \mu_i)\phi_i)]. \quad (4.35)$$

4.3. Gemeinsame Modellierung von Erwartungswert und Dispersion

Die Eigenschaft $E\left(\frac{\partial l(\mu_i, \phi_i | y_i)}{\partial \mu_i}\right) = 0$ führt zusammen mit Ergebnis (4.35) zu dem Erwartungswert

$$E\left(\frac{\partial^2 l(\boldsymbol{\mu}, \boldsymbol{\phi} | \mathbf{y})}{\partial \beta_j \partial \beta_k}\right) = -\sum_{i=1}^n \phi_i^2 [\psi'(\mu_i \phi_i) + \psi'((1 - \mu_i)\phi_i)] \frac{x_{ij} x_{ik}}{(g'_1(\mu_i))^2}.$$

Wird zusätzlich die Diagonalmatrix $\mathbf{K}_{\beta\beta} = \text{diag}(k_1^{(1)}, \dots, k_n^{(1)})$ mit den Einträgen

$$k_i^{(1)} = \phi_i^2 [\psi'(\mu_i \phi_i) + \psi'((1 - \mu_i)\phi_i)] \frac{1}{(g'_1(\mu_i))^2}$$

eingeführt, so resultiert schließlich

$$\mathbf{I}_{\beta\beta}(\boldsymbol{\beta}, \boldsymbol{\delta}) = -E\left(\frac{\partial^2 l(\boldsymbol{\mu}, \boldsymbol{\phi} | \mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\right) = \mathbf{X}^T \mathbf{K}_{\beta\beta} \mathbf{X}.$$

Im nächsten Schritt soll die Matrix $\mathbf{I}_{\beta\delta}(\boldsymbol{\beta}, \boldsymbol{\delta})$ berechnet werden. Dazu wird zunächst der Ausdruck (4.29) nach δ_k abgeleitet,

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\mu}, \boldsymbol{\phi} | \mathbf{y})}{\partial \beta_j \partial \delta_k} &= \sum_{i=1}^n \frac{\partial}{\partial \phi_i} \left(\frac{\partial l(\mu_i, \phi_i | y_i)}{\partial \beta_j} \right) \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \delta_k} \\ &= \sum_{i=1}^n \frac{\partial}{\partial \phi_i} (\phi_i y_i^* - \phi_i \mu_i^*) \frac{x_{ij} z_{ik}}{g'_1(\mu_i) g'_2(\phi_i)}. \end{aligned}$$

Unter Berücksichtigung der Produktregel und unter Verwendung von

$$\frac{\partial \mu_i^*}{\partial \phi_i} = \psi'(\mu_i \phi_i) \mu_i - \psi'((1 - \mu_i)\phi_i) (1 - \mu_i)$$

kann die zweite Ableitung der Log-Likelihoodfunktion nach β_j und δ_k geschrieben werden als

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\mu}, \boldsymbol{\phi} | \mathbf{y})}{\partial \beta_j \partial \delta_k} &= \sum_{i=1}^n (y_i^* - \mu_i^* - \phi_i [\mu_i \psi'(\mu_i \phi_i) - (1 - \mu_i) \psi'((1 - \mu_i)\phi_i)]) \frac{x_{ij} z_{ik}}{g'_1(\mu_i) g'_2(\phi_i)} \\ &= \sum_{i=1}^n (y_i^* - \mu_i^* - \check{c}_i) \frac{x_{ij} z_{ik}}{g'_1(\mu_i) g'_2(\phi_i)}, \end{aligned}$$

mit

$$\check{c}_i = \phi_i [\mu_i \psi'(\mu_i \phi_i) - (1 - \mu_i) \psi'((1 - \mu_i)\phi_i)].$$

Da bereits bekannt ist dass $E(y_i^*) = \mu_i^*$ gilt, folgt für den Erwartungswert der zweiten Ableitung der Log-Likelihoodfunktion nach β_j und δ_k

$$E\left(\frac{\partial^2 l(\boldsymbol{\mu}, \boldsymbol{\phi} | \mathbf{y})}{\partial \beta_j \partial \delta_k}\right) = -\sum_{i=1}^n \check{c}_i \frac{x_{ij} z_{ik}}{g'_1(\mu_i) g'_2(\phi_i)}.$$

Die Definition der Diagonalmatrix $\mathbf{K}_{\beta\delta} = \text{diag}(k_1^{(2)}, \dots, k_n^{(2)})$ zusammen mit den Matrixeinträgen

$$k_i^{(2)} = \ddot{c}_i \frac{x_{ij} z_{ik}}{g_1'(\mu_i) g_2'(\phi_i)}$$

führt schließlich zu dem Ergebnis

$$\mathbf{I}_{\beta\delta}(\boldsymbol{\beta}, \boldsymbol{\delta}) = -\mathbb{E} \left(\frac{\partial^2 l(\boldsymbol{\mu}, \boldsymbol{\phi} | \mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\delta}^T} \right) = \mathbf{X}^T \mathbf{K}_{\beta\delta} \mathbf{Z}.$$

Um die Herleitung der Fisher-Informationsmatrix abzuschließen, fehlt nur noch die Submatrix $\mathbf{I}_{\delta\delta}(\boldsymbol{\beta}, \boldsymbol{\delta})$. Die zweite Ableitung der Log-Likelihoodfunktion nach δ_j und δ_k kann mit Hilfe von (4.31) und unter Berücksichtigung der Produktregel geschrieben werden als

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\mu}, \boldsymbol{\phi} | \mathbf{y})}{\partial \delta_j \partial \delta_k} &= \sum_{i=1}^n \frac{\partial}{\partial \phi_i} \left(\frac{\partial l(\mu_i, \phi_i | y_i)}{\partial \phi_i} \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \delta_j} \right) \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \delta_k} \\ &= \sum_{i=1}^n \left(\frac{\partial^2 l(\mu_i, \phi_i | y_i)}{\partial \phi_i^2} \frac{\partial \phi_i}{\partial \eta_{2i}} + \frac{\partial l(\mu_i, \phi_i | y_i)}{\partial \phi_i} \frac{\partial}{\partial \phi_i} \frac{\partial \phi_i}{\partial \eta_{2i}} \right) \frac{z_{ij} z_{ik}}{g_2'(\phi_i)}. \end{aligned}$$

Weiters kann unter Verwendung von (4.32) die Ableitung

$$\frac{\partial^2 l(\mu_i, \phi_i | y_i)}{\partial \phi_i^2} = - \left[\psi'(\mu_i \phi_i) \mu_i^2 + \psi'((1 - \mu_i) \phi_i) (1 - \mu_i)^2 - \psi'(\phi_i) \right]$$

hergeleitet werden, was dann zusammen mit der Eigenschaft $\mathbb{E} \left(\frac{\partial l(\mu_i, \phi_i | y_i)}{\partial \phi_i} \right) = 0$ zu dem Erwartungswert

$$\mathbb{E} \left(\frac{\partial^2 l(\boldsymbol{\mu}, \boldsymbol{\phi} | \mathbf{y})}{\partial \delta_j \partial \delta_k} \right) = - \sum_{i=1}^n \ddot{v}_i \frac{z_{ij} z_{ik}}{(g_2'(\phi_i))^2}$$

führt, wobei \ddot{v}_i gegeben ist durch

$$\ddot{v}_i = \psi'(\mu_i \phi_i) \mu_i^2 + \psi'((1 - \mu_i) \phi_i) (1 - \mu_i)^2 - \psi'(\phi_i).$$

Wird zusätzlich die Diagonalmatrix $\mathbf{K}_{\delta\delta} = \text{diag}(k_1^{(3)}, \dots, k_n^{(3)})$ zusammen mit den Matrixeinträgen

$$k_i^{(3)} = \ddot{v}_i \frac{1}{(g_2'(\phi_i))^2}$$

eingeführt, so resultiert für die Submatrix

$$\mathbf{I}_{\delta\delta}(\boldsymbol{\beta}, \boldsymbol{\delta}) = -\mathbb{E} \left(\frac{\partial^2 l(\boldsymbol{\mu}, \boldsymbol{\phi} | \mathbf{y})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} \right) = \mathbf{Z}^T \mathbf{K}_{\delta\delta} \mathbf{Z}.$$

Zusammengefasst ist die Fisher-Informationsmatrix $\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\delta})$ gegeben durch

$$\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\delta}) = \begin{pmatrix} \mathbf{X}^T \mathbf{K}_{\beta\beta} \mathbf{X} & \mathbf{X}^T \mathbf{K}_{\beta\delta} \mathbf{Z} \\ (\mathbf{X}^T \mathbf{K}_{\beta\delta} \mathbf{Z})^T & \mathbf{Z}^T \mathbf{K}_{\delta\delta} \mathbf{Z} \end{pmatrix} = \mathbf{S}^T \mathbf{K} \mathbf{S},$$

zusammen mit den Matrizen

$$\mathbf{S} = \begin{pmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z} \end{pmatrix}$$

und

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_{\beta\beta} & \mathbf{K}_{\beta\delta} \\ (\mathbf{K}_{\beta\delta})^T & \mathbf{K}_{\delta\delta} \end{pmatrix}.$$

Diese Ergebnisse decken sich mit jenen von Simas et al. (2010).

Bemerkung:

Simas et al. (2010) weisen darauf hin, dass $\mathbf{K}_{\beta\delta} \neq 0$ hält und die Parametervektoren β und δ somit nicht orthogonal sind.

Bemerkung:

Ebenfalls kann gezeigt werden, dass die Maximum Likelihood Schätzer $\hat{\beta}$ und $\hat{\delta}$ beide asymptotisch erwartungstreu sind und in Verteilung gegen die Normalverteilung konvergieren, das heißt

$$\sqrt{n} \begin{pmatrix} \hat{\beta} \\ \hat{\delta} \end{pmatrix} \xrightarrow{D} \mathcal{N}_{p+q} \left(\begin{pmatrix} \beta \\ \delta \end{pmatrix}, n\mathbf{I}^{-1}(\beta, \delta) \right).$$

4.3.3 Modelldiagnostik

Nach der Berechnung der Maximum Likelihood Schätzer $\hat{\beta}$ und $\hat{\delta}$ mit Hilfe eines geeigneten numerischen Verfahrens können wieder die Schätzer der Erwartungswerte und der Dispersionen abgeleitet werden durch

$$\hat{\mu}_i = g_1^{-1}(\mathbf{x}_i^T \hat{\beta})$$

$$\hat{\phi}_i = g_2^{-1}(\mathbf{z}_i^T \hat{\delta}).$$

Um zu entscheiden ob diese Größen ausreichend gute Schätzer darstellen und somit die Wahl des Modells bestärkt wird, sollte wieder die Anpassungsgüte überprüft werden. Auch für das Beta-Regressionsmodell zur gemeinsamen Modellierung von Erwartungswert und Dispersion können dazu die aus Abschnitt 4.2.4 bekannten Werkzeuge verwendet werden.

Die Modelldiskrepanz beziehungsweise Deviance kann übernommen werden. Sie wird nach wie vor als doppelte Differenz zwischen der Log-Likelihoodfunktion des saturierten Modells und der Log-Likelihood des betrachteten Modells berechnet. Es sei zu beachten, dass sich die Log-Likelihood aus diesem Abschnitt von jener des Grundmodells unterscheidet, gut zu erkennen im direkten Vergleich von (4.5) und (4.26).

Auch die Residuen bleiben im Grunde unverändert. Hier gilt ebenfalls, dass die konstante Dispersion ϕ lediglich durch die zum Index passende Größe ϕ_i ersetzt werden muss. Wie dies für die bisher vorgestellten Residuen genau aussieht zeigen Rocha und Simas

(2011).

Die aus Abschnitt 2.6 bekannte generalisierte Hebelwirkung kann wieder mit Hilfe der in Lemma 2.7 beschriebenen Formel bzw. dem aus (2.25) bekannten, auf die Log-Likelihoodfunktion angepassten Ausdruck hergeleitet werden. Es sei anzumerken, dass die generalisierte Hebelwirkung eine von $\hat{\boldsymbol{\alpha}} = (\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\delta}}^T)^T$ abhängige Größe darstellt, da in den geschätzten Erwartungswert $\hat{\boldsymbol{\mu}}$ sowohl $\hat{\boldsymbol{\beta}}$ als auch $\hat{\boldsymbol{\delta}}$ mit einfließt. Wie stark die daraus resultierenden Ergebnisse von jenen des Grundmodells abweichen kann bei Rocha und Simas (2011) nachgeschlagen werden.

Schließlich folgt noch die Cook-Distanz. Auch diese Größe gibt Auskunft über die Anpassungsgüte eines Beta-Regressionsmodells zur gemeinsamen Modellierung von Erwartungswert und Dispersion. Wie die aus Abschnitt 4.2.4 bekannte Form adaptiert werden kann, zeigen erneut Rocha und Simas (2011).

4.4 Korrektur und Reduktion des Bias

Bisher wurden in diesem Kapitel zwei Beta-Regressionsmodelle eingeführt, die korrespondierenden Log-Likelihoodfunktionen und Scorevektoren diskutiert und mögliche Verfahren zur Berechnung der Maximum Likelihood Schätzer betrachtet. Ergänzend wurden mehrere Techniken zur Analyse der Anpassungsgüte vorgestellt. In diesem Abschnitt werden nun Biaskorrektur und -reduktion der Schätzer behandelt.

Schon Ospina et al. (2006) und Simas et al. (2010) wiesen darauf hin, dass die vorgestellten Maximum Likelihood Schätzer einer Verzerrung unterliegen könnten. Für einen Stichprobenumfang der Größe n entspricht die Ordnung des Bias asymptotisch $\mathcal{O}(n^{-1})$. Das heißt, dass für große Stichproben keine unmittelbaren Probleme zum Vorschein treten. Da in der Praxis jedoch meist auch mit einer geringeren Anzahl an Beobachtungen gearbeitet wird, darf der Bias nicht einfach vernachlässigt werden, da er zu einer Unterschätzung der Varianz und der Standardfehler führen kann.

Kosmidis und Firth (2010) entwickelten einen Algorithmus, der nicht nur zur Korrektur sondern auch zur Reduktion des Bias angewendet werden kann. Dieser Algorithmus beruht auf der Fisher-Scoring Methode. Er kann sowohl auf das Grundmodell als auch auf das Beta-Regressionsmodell zur gemeinsamen Modellierung von Erwartungswert und Dispersion angewendet werden.

4.4.1 Algorithmus

Sei $\mathbf{S}(\boldsymbol{\theta}) = (s_1(\boldsymbol{\theta}), \dots, s_d(\boldsymbol{\theta}))^T$ ein Scorevektor, $\boldsymbol{\theta}$ der d -dimensionale Parametervektor und $\mathbf{0}_d$ ein Vektor bestehend aus d Nullen. Firth (1993) zeigte, dass die Lösung des Gleichungssystems

$$\mathbf{S}^*(\boldsymbol{\theta}) := \mathbf{S}(\boldsymbol{\theta}) + \mathbf{A}(\boldsymbol{\theta}) \stackrel{!}{=} \mathbf{0}_d \quad (4.36)$$

einen kleineren asymptotischen Bias aufweist als der Maximum Likelihood Schätzer von $\boldsymbol{\theta}$. Hierbei stellt $\mathbf{A}(\boldsymbol{\theta}) = (a_1(\boldsymbol{\theta}), \dots, a_d(\boldsymbol{\theta}))^T$ einen d -dimensionalen Vektor dar, bestehend aus den Elementen

$$a_j(\boldsymbol{\theta}) = \frac{1}{2} \text{trace} \left(\mathbf{I}^{-1}(\boldsymbol{\theta}) (\mathbf{P}_j(\boldsymbol{\theta}) + \mathbf{Q}_j(\boldsymbol{\theta})) \right), \quad j = 1, \dots, d,$$

mit der Fisher-Informationsmatrix $\mathbf{I}(\boldsymbol{\theta})$ und den Matrizen

$$\mathbf{P}_j(\boldsymbol{\theta}) = \mathbf{E} \left(\mathbf{S}(\boldsymbol{\theta}) \mathbf{S}^T(\boldsymbol{\theta}) s_j(\boldsymbol{\theta}) \right),$$

$$\mathbf{Q}_j(\boldsymbol{\theta}) = -\mathbf{E} \left(\mathbf{F}(\boldsymbol{\theta}) s_j(\boldsymbol{\theta}) \right),$$

wobei $s_j(\boldsymbol{\theta})$ den j -ten Eintrag des Scorevektors $\mathbf{S}(\boldsymbol{\theta})$ und $\mathbf{F}(\boldsymbol{\theta})$ die negative Hessematrix der Log-Likelihoodfunktion bezeichnet. Es sei anzumerken, dass $\mathbf{F}(\boldsymbol{\theta})$ auch als beobachtete Fisher-Information bekannt ist, wohingegen $\mathbf{I}(\boldsymbol{\theta})$ als erwartete Fisher-Information bezeichnet wird, basierend auf dem Zusammenhang

$$\mathbf{E}(\mathbf{F}(\boldsymbol{\theta})) = \mathbf{I}(\boldsymbol{\theta}).$$

Um jenen Vektor $\tilde{\boldsymbol{\theta}}$ zu finden, welcher das Gleichungssystem (4.36) löst, entwickelten Kosmidis und Firth (2010) die Quasi Fisher-Scoring Iteration. Dabei wird die Iterierte $\boldsymbol{\theta}^{(t+1)}$ berechnet durch

$$\begin{aligned} \boldsymbol{\theta}^{(t+1)} &= \boldsymbol{\theta}^{(t)} + \mathbf{I}^{-1}(\boldsymbol{\theta}^{(t)}) \mathbf{S}^*(\boldsymbol{\theta}^{(t)}) \\ &= \boldsymbol{\theta}^{(t)} + \mathbf{I}^{-1}(\boldsymbol{\theta}^{(t)}) \mathbf{S}(\boldsymbol{\theta}^{(t)}) - \mathbf{b}(\boldsymbol{\theta}^{(t)}). \end{aligned} \tag{4.37}$$

Der Vektor $\mathbf{b}(\boldsymbol{\theta}^{(t)})$ ist hierbei definiert durch

$$\mathbf{b}(\boldsymbol{\theta}^{(t)}) = -\mathbf{I}^{-1}(\boldsymbol{\theta}^{(t)}) \mathbf{A}(\boldsymbol{\theta}^{(t)})$$

und stellt den asymptotischen Bias der Ordnung $\mathcal{O}(n^{-1})$ des Maximum Likelihood Schätzers dar.

Wird der Vektor $\mathbf{b}(\boldsymbol{\theta}^{(t)})$ in Ausdruck (4.37) vernachlässigt, ist die Iterationsvorschrift äquivalent zu einem Newton Schritt in Kombination mit der Fisher-Scoring Methode, auch bekannt als Fisher-Scoring Iteration. Wie bereits mehrmals gesehen liefert dieses Verfahren bei Konvergenz den Maximum Likelihood Schätzer $\hat{\boldsymbol{\theta}}$.

Sollte der Maximum Likelihood Schätzer als Startwert $\boldsymbol{\theta}^{(0)}$ für das durch Iterationsvorschrift (4.37) definierte Verfahren gewählt werden, so stellt die erste Iterierte $\boldsymbol{\theta}^{(1)}$ den biaskorrigierten Schätzer $\hat{\boldsymbol{\theta}}^C$ dar. Dieser ist mittels (4.37) gegeben durch

$$\hat{\boldsymbol{\theta}}^C = \hat{\boldsymbol{\theta}} + \mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}) \mathbf{S}(\hat{\boldsymbol{\theta}}) - \mathbf{b}(\hat{\boldsymbol{\theta}}).$$

Da der Scorevektor ausgewertet im Maximum Likelihood Schätzer $\hat{\boldsymbol{\theta}}$ gleich Null ist, lässt sich der biaskorrigierte Schätzer vereinfachen zu

$$\hat{\boldsymbol{\theta}}^C = \hat{\boldsymbol{\theta}} - \mathbf{b}(\hat{\boldsymbol{\theta}}).$$

Wird das Verfahren nicht nach der ersten Iteration gestoppt, sondern bis zur Konvergenz durchgeführt, so resultiert daraus der biasreduzierte Schätzer $\hat{\theta}^R$.

Zusammenfassend bietet der von Kosmidis und Firth (2010) eingeführte Algorithmus demnach drei mögliche Varianten zur Berechnung der Parameterschätzer. Durch Weglassen des Vektors $\mathbf{b}(\theta^{(t)})$ in der Iterationsvorschrift ergibt sich der Maximum Likelihood Schätzer $\hat{\theta}$. Die Berechnung der ersten Iterierten führt zur Korrektur des Bias und bei Konvergenz des Verfahrens erfolgt die Biasreduktion. Wie bereits erwähnt kann der Algorithmus sowohl auf das Grundmodell als auch auf das Beta-Regressionsmodell zur gemeinsamen Modellierung von Erwartungswert und Dispersion angewendet werden. Das Verfahren bleibt unverändert, lediglich die unterschiedlichen Ergebnisse bezüglich des Scorevektors und der Informationsmatrix beeinflussen die Iteration. Eine genaue Berechnung der für das Verfahren relevanten Matrizen findet sich bei Kosmidis und Firth (2010) für das Grundmodell und bei Grün et al. (2012) für das Beta-Regressionsmodell zur gemeinsamen Modellierung von Erwartungswert und Dispersion.

4.5 Ausblick

Bisher wurde das Grundmodell eingeführt, eine erste Erweiterung zur gemeinsamen Modellierung von Erwartungswert und Dispersion betrachtet und eine Methode zur Korrektur und Reduktion des Bias diskutiert. Zusammen liefert dies umfangreiche Möglichkeiten zur Modellierung von Prozentwerten. Allerdings könnte es bei der praktischen Anwendung passieren, dass diese Techniken nicht ausreichen um die beobachteten Daten optimal zu reproduzieren. Aus diesem Grund werden nun noch zwei Erweiterungsmöglichkeiten vorgestellt, die im Folgenden kurz beschrieben werden und so eine Motivation zu möglichen weiteren Betrachtungsweisen liefern, die jedoch über den Rahmen dieser Arbeit hinausgehen würden.

4.5.1 Heterogenität bei Beta-Regressionsmodellen

Das Grundmodell und die Verwendung eines einzelnen linearen Prädiktors schien nicht immer passend um die tatsächlich vorliegende Variabilität zu beschreiben. Doch auch der Schritt zum Beta-Regressionsmodell zur gemeinsamen Modellierung von Erwartungswert und Dispersion und die damit einhergehende Einführung eines zweiten linearen Prädiktors ist möglicherweise nicht genug um die gesamte Heterogenität der Daten zu erfassen. Dies trifft vor allem dann zu, wenn mögliche unbeobachtete Gruppen oder Cluster der Daten existieren.

Um die mögliche Heterogenität zwischen den Gruppen zu erfassen schlagen Grün et al. (2012) zwei Methoden vor, wovon eine im nachfolgenden Abschnitt genauer behandelt wird. Bei der Methode, die nun kurz beschrieben werden soll, handelt es sich um sogenannte „beta regression trees“, welche sich auf Zeileis, Hothorn und Hornik (2008) zurückführen lassen. Die alternative Vorgehensweise zur Erfassung der Heterogenität

stellt die Verwendung der bei McLachlan und Peel (2000) beschriebenen endlichen gemischten Modelle dar.

Beta regression trees

Die beta regression trees wurden von Grün et al. (2012) eingeführt indem die Methode des „model-based recursive partitioning“ (MOB) von Zeileis et al. (2008) adaptiert und auf die Beta-Regressionsmodelle angepasst wurde. Wird vermutet, dass die beobachteten Daten in Gruppen unterteilt werden können, so kann die Heterogenität zwischen den Gruppen mit Hilfe dieser Bäume nachgewiesen oder widerlegt werden. Dabei ist es nicht notwendig jede Vermutung separat zu überprüfen, es können mehrere Partitionsvariablen gleichzeitig auf Heterogenität getestet werden.

Das Ziel der Methode ist das Auffinden der Unterschiede in der Verteilung der Response-Variablen, die mit den bisher beschriebenen Techniken nicht entdeckt wurden. Dabei wird jedoch nicht die Verteilung selbst betrachtet, sondern die Unterschiede in jenen Parametern, welche die Verteilung beschreiben. Zusätzlich sei anzumerken, dass es sich um eine rekursive Vorgehensweise handelt, die aus vier grundlegenden Schritten besteht.

Sei c_{ij} die j -te Partitionsvariable der i -ten Beobachtung, mit $j = 1, \dots, l$. Da die Menge an Beobachtungen y_i während der Anwendung des Algorithmus immer wieder partitioniert wird, hängt der Bereich des Index i von der aktuellen Stichprobenmenge ab. Zu Beginn gilt $i = 1, \dots, n$. Die vier Schritte zur rekursiven Berechnung der beta regression trees sind laut Grün et al. (2012) gegeben durch:

1. Modellanpassung:

Im ersten Schritt erfolgt die Anpassung eines Beta-Regressionsmodells bezüglich jener Response-Variablen y_i , welche sich in der aktuell betrachteten Menge an Beobachtungen befinden. Der Parametervektor θ hängt von der Wahl des Modells ab. Beim Grundmodell besteht er aus dem Parametervektor β und dem Dispersionsparameter ϕ , wohingegen aus der Verwendung des Beta-Regressionsmodells zur gemeinsamen Modellierung von Erwartungswert und Dispersion $\theta = (\beta^T, \delta^T)^T$ folgt. Je nach Modellauswahl folgt die Maximum Likelihood Schätzung der Vorgehensweise aus Abschnitt 4.2.1 oder Abschnitt 4.3.1.

2. Stabilitätstest:

Danach wird für jede Partitionsvariable c_{ij} getestet ob der Parametervektor θ über die Partitionsvariable hinweg stabil ist.

3. Partitionierung:

Falls mindestens eine Partitionsvariable eine signifikante Instabilität aufweisen sollte, wird die aktuell betrachtete Stichprobenmenge geteilt. Dabei wird zunächst die Partitionsvariable mit der größten Instabilität gesucht und die Menge der Beobachtungen dann mittels dieser Variablen gespaltet. Mit anderen Worten wird genau jene Variable zur Partitionierung gewählt, welche die größte Verbesserung der Log-Likelihoodfunktion gewährleistet.

4. Wiederholung bzw. STOP:

Wurde im dritten Schritt eine Partitionierung durchgeführt, so werden die ersten drei Schritte solange rekursiv auf die entstandenen Teilmengen angewendet, bis keine signifikante Instabilität mehr nachgewiesen werden kann. Wurde die Menge in Schritt 3 nicht geteilt, heißt das, dass die Parameter über alle Partitionsvariablen hinweg stabil sind und der Algorithmus gestoppt werden kann.

Um die Stabilität der Parameter zu testen schlagen Zeileis et al. (2008) den von Zeileis und Hornik (2007) eingeführte M-Fluktuationstest vor. Dabei wird getestet ob die Scorefunktionen bezüglich der einzelnen Beobachtungen eine zufällige Schwankung um den Erwartungswert herum aufweist oder ob systematische Abweichungen von 0 über die Partitionsvariable hinweg beobachtet werden können.

Es sei anzumerken, dass die Scorefunktion bezüglich einer einzelnen Beobachtung y_i nicht von der gesamten Designmatrix \mathbf{X} , sondern nur von dem korrespondierenden Vektor \mathbf{x}_i der erklärenden Variablen abhängt.

4.5.2 Beta-Regression mit erweitertem Definitionsbereich

Bis jetzt wurde in allen Abschnitten stets darauf hingewiesen, dass die beobachteten Prozentwerte aus dem offenen Einheitsintervall $(0, 1)$ stammen müssen, da die Log-Likelihoodfunktion sonst gar nicht bestimmt werden kann. Diese Annahme ist jedoch recht theoretisch und lässt sich mit der Praxis oft nicht vereinbaren. Bei der praktischen Anwendung ist es sehr wahrscheinlich, dass die Response-Werte nicht nur aus dem Inneren des Intervalls stammen, sondern auch die Randwerte 0 und 1 annehmen. Um die kennengelernten Regressionsmodelle trotzdem zur Modellanpassung nutzen zu können, müssten die aus $\{0, 1\}$ stammenden Beobachtungen demnach vernachlässigt werden. Da dies in manchen Fällen aber zu einer starken Beeinflussung der Parameterschätzung führen könnte, wird nun eine Methode besprochen, die es erlaubt die Randwerte in das Modell mit aufzunehmen.

Eine Idee dieses Problem in den Griff zu bekommen stammt von Smithson und Verkuilen (2006). Dabei wird vorgeschlagen die Response-Werte um eine additive Konstante zu erhöhen bzw. zu verringern um die Beobachtungen aus $\{0, 1\}$ künstlich in das Intervall $(0, 1)$ zu verschieben. Bei dieser Methode stellt sich jedoch die Frage wie eine geeignete Wahl dieser Konstante aussehen sollte. Ospina und Ferrari (2012) hingegen empfehlen die Verwendung einer diskret-stetigen Mischung der Beta-Verteilung und Punktmassen auf 0 und 1. Eine dritte Möglichkeit stellt die von Kosmidis, Zou und Jagan (2015) stammende Idee der Erweiterung des Definitionsbereiches dar. Diese Methode wird nun genauer besprochen.

Beta-Verteilung mit erweitertem Definitionsbereich

Um das Beta-Regressionsmodell mit erweitertem Definitionsbereich einzuführen, muss zunächst eine Erweiterung der Beta-Verteilung betrachtet werden. Aus Notationsgründen muss in diesem Abschnitt eine Unterscheidung zwischen einer Zufallsvariablen und ihrer Realisierung möglich sein. Dazu werden Zufallsvariablen mit Großbuchstaben und Realisationen mit Kleinbuchstaben bezeichnet.

Sei W eine beta-verteilte Größe mit Erwartungswert μ , Dispersion ϕ und Dichtefunktion $f(\cdot)$ wie in (4.4). Zusätzlich sei die Zufallsvariable Y für beliebiges, festes $e > 0$ spezifiziert durch

$$Y = g(W|e) = -e(1 - W) + (1 + e)W.$$

Zur Berechnung der Dichtefunktion von Y wird folgender Satz benötigt.

Satz 4.2. (Transformationssatz für Dichtefunktionen, Casella und Berger, 2002)

Sei X eine Zufallsvariable definiert auf dem Bereich $S \subseteq \mathbb{R}$ mit Dichtefunktion $f_X(x)$ und Y eine Zufallsvariable mit der Eigenschaft $Y = g(X)$ für eine monotone Funktion $g(\cdot) : S \rightarrow T \subseteq \mathbb{R}$. Zusätzlich seien x und y die Realisierungen der beiden Variablen. Falls $f_X(x)$ stetig ist und $g^{-1}(y)$ stetig differenzierbar, so kann die Dichtefunktion von Y geschrieben werden als

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

für $y \in T$.

Beweis:

Siehe Casella und Berger (2002). □

Satz 4.2 kann nun dazu verwendet werden die Dichtefunktion von Y anzugeben, da die Transformation von W zu Y linear und somit monoton ist. Laut der Definition von Y kann die Umkehrung der Realisierung y geschrieben werden als

$$w(y, e) = g^{-1}(y|e) = \frac{y + e}{1 + 2e}.$$

Mit Hilfe von Satz 4.2 resultiert somit als Dichtefunktion von Y

$$f^*(y|\mu, \phi, e) = \frac{1}{1 + 2e} f(w(y, e)|\mu, \phi), \quad (4.38)$$

mit der Eigenschaft $-e < y < 1 + e$, da W eine beta-verteilte Größe ist und somit nur Werte aus dem Inneren aus $(0, 1)$ annimmt.

Durch die Einführung der Dichtefunktion $f^*(y|\mu, \phi, e)$ wurde der bisher zulässige Definitionsbereich der Beta-Verteilung, nämlich das Intervall $(0, 1)$, auf das Intervall $(-e, e+1)$ erweitert. Die hierbei zusätzlich eingeführte Größe e wird als Kontamination bezeichnet. Für $e = 0$ stimmt die Funktion $f^*(y|\mu, \phi, e)$ mit der Dichte-Funktion der Beta-Verteilung

überein. Die Motivation der Verwendung der Beta-Verteilung mit erweitertem Definitionsbereich beruht auf der Eigenschaft, dass $f^*(y|\mu, \phi, e)$ für beliebiges $e > 0$ auch an den Stellen $y = 0$ und $y = 1$ problemlos bestimmt werden kann und einen positiven Wert annimmt.

Modell mit zufälliger Kontamination

Da die Unterscheidung zwischen Zufallsvariablen und deren Realisierungen nicht länger notwendig ist, wird wieder zur ursprünglichen Notation gewechselt. Seien y_1, \dots, y_n n unabhängige Zufallsvariablen, die einer Beta-Verteilung mit erweitertem Definitionsbereich folgen und $w(y_i, e_i)$ die korrespondierenden beta-verteilten Größen mit Erwartungswerten μ_i und Dispersionsparametern ϕ_i . Wie bei der Beta-Regression zur gemeinsamen Modellierung von Erwartungswert und Dispersion aus Abschnitt 4.3 können auch hier die Erwartungswerte und Dispersionen mittels Linkfunktionen und linearen Prädiktoren modelliert werden.

Jeder Beobachtung y_i wird demzufolge eine eigene Kontamination e_i zugeschrieben. Da die Response-Werte aus dem geschlossenen Einheitsintervall $[0, 1]$ stammen, sollte auch der erweiterte Definitionsbereich der betrachteten Beta-Verteilung nicht zu weit über die Grenzen reichen. Dies sollte durch die Wahl der n Kontaminationen gewährleistet sein. Zusätzlich sei zu erwähnen, dass dieses Modell im Vergleich zu den bisher betrachteten nicht nur die Parameter zur Modellierung von Erwartungswert und Dispersion, sondern auch den Erwartungswert der n zufälligen Kontaminationen schätzen soll.

Um die eben beschriebenen Probleme in den Griff zu bekommen, schlagen Kosmidis et al. (2015) vor, die Kontaminationen e_1, \dots, e_n als unabhängig, identisch exponentialverteilte Größen anzusehen, mit bekanntem Erwartungswert ν . Weiters wird angenommen, dass y_1, \dots, y_n bedingt unabhängig sind, gegeben e_1, \dots, e_n und die Größen μ_i und ϕ_i der Beta-Verteilung mit erweitertem Definitionsbereich modelliert werden können mittels

$$g_1(\mu_i) = \eta_{1i} = \mathbf{x}_i^T \boldsymbol{\beta},$$

$$g_2(\phi_i) = \eta_{2i} = \mathbf{z}_i^T \boldsymbol{\delta},$$

mit Linkfunktionen $g_1(\cdot)$ und $g_2(\cdot)$.

Die marginale Likelihoodfunktion kann aufgrund der Unabhängigkeit bestimmt werden durch das Produkt

$$\prod_{i=1}^n L_i^*(\boldsymbol{\beta}, \boldsymbol{\delta}|\nu) = \prod_{i=1}^n \frac{1}{\nu} \int_0^\infty f^*(y_i|\mu_i, \phi_i, e_i) \exp\left(-\frac{e_i}{\nu}\right) de_i.$$

Zur Schätzung der Parameter $\boldsymbol{\beta}$ und $\boldsymbol{\delta}$ muss die marginale Likelihoodfunktion maximiert werden. Kosmidis et al. (2015) schlagen dazu eine Approximation durch die Gauß-Laguerre-Quadratur vor. Für die Wahl des Erwartungswertes ν empfehlen sie die Größe $10^{-\xi}$, wobei ξ die längste Anzahl der Dezimalstellen unter den Response-Werten

4.5. Ausblick

beschreibt.

Bemerkung:

Für die praktische Anwendung dieser Idee ist leider noch kein R Paket verfügbar, weshalb diese Arbeit nur den Grundgedanken dieser Erweiterung beschreibt.

5.1 Motivation

Im theoretischen Teil dieser Arbeit wurden verschiedene Möglichkeiten besprochen, um Beobachtungen aus dem Einheitsintervall $[0, 1]$ bzw. aus dem offenen Intervall $(0, 1)$ zu modellieren. Dazu gehört das Generalisierte Lineare Modell aus Kapitel 2, das Quasi-Likelihoodmodell aus Kapitel 3, das Extended Quasi-Likelihoodmodell aus Abschnitt 3.7 und das Beta-Regressionsmodell aus Kapitel 4. Im praktischen Teil sollen diese Regressionsmodelle nun veranschaulicht und miteinander verglichen werden. Dabei werden die Vor- und Nachteile der einzelnen Modelle besprochen und mögliche Unterschiede bezüglich der Signifikanz der Prädiktoren aufgezeigt.

Die dazu verwendeten Daten stammen von Cernko (2015) und wurden im Zuge einer Online-Umfrage erhoben. Der folgende Abschnitt beschreibt den Datensatz und die für das Beispiel ausgesuchten Variablen.

5.2 Datensatz

Der von Cernko (2015) stammende Datensatz wurde mittels einer Online-Umfrage erhoben, an der ausschließlich weibliche Angestellte der UniCredit Bank Austria AG teilnehmen durften. Er diente zur Untersuchung der Entwicklung der Rollenverteilung, sowie zum Auffinden jener Größen, welche möglicherweise einen Einfluss auf das Ausmaß der Hausarbeit ausüben. Zusätzlich wurde analysiert, wie sich die Arbeitszeit auf die Zufriedenheit dieser Frauen auswirkt.

An der Umfrage nahmen 301 Frauen teil, jedoch beantworteten nur 229 dieser Frauen alle Fragen vollständig. Aus den bestehenden 64 Kanälen wurden fünf ausgewählt, um sie

in dieser Arbeit zu analysieren. Diese fünf Variablen und deren möglichen Ausprägungen werden nun kurz erläutert.

5.2.1 Hausarbeit

Die Variable **Hausarbeit** lebt auf dem geschlossenen Intervall $[0, 100]$ und gibt Aufschluss über den von den Frauen erledigten Prozentsatz der im Haus geleisteten Arbeit. Die Frauen sollten dabei angeben, welchen Anteil der Haus- und Familienarbeit sie ihrer Meinung nach selbst bewerkstelligen.

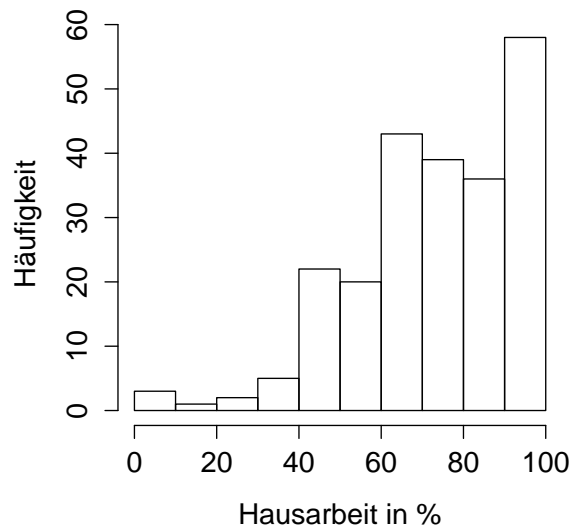


Abbildung 5.1: Histogramm der Haus- und Familienarbeit.

Abbildung 5.1 lässt gut erkennen, dass ein großer Anteil der befragten Frauen der Meinung ist, mehr als 90% der Hausarbeit selbst zu verrichten. Nur wenige Frauen hingegen geben an weniger als 10% der anfallenden Arbeit zu erledigen.

Da es in dieser Arbeit um die Modellierung von Prozentwerten geht, wird die Variable **Hausarbeit** als Response-Variable gewählt. Mit Abbildung 5.1 wird deutlich, warum die Lineare Regression keinen geeigneten Kandidaten zur Modellierung von Prozentwerten darstellt und bestärkt somit die Notwendigkeit der in dieser Arbeit hergeleiteten Regressionsmodelle. Die Annahme von normalverteilten Response-Werten lässt sich nicht mit den beobachteten Prozentsätzen vereinbaren.

5.2.2 Unterstützung

Die Variable **Unterstützung** wird als eine von fünf erklärenden Variablen zur Modellierung der Hausarbeit ausgewählt. Sie gibt an ob die Frauen bei der Haus- und Fami-

liendarbeit von ihren Ehepartnern bzw. Lebensgefährten unterstützt werden oder nicht und setzt sich aus den Faktorstufen **Ja** und **Nein** zusammen.

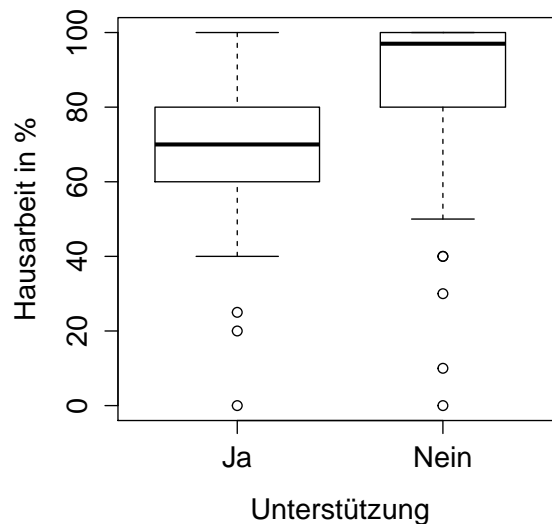


Abbildung 5.2: Boxplot: Hausarbeit gegen Unterstützung.

Abbildung 5.2 zeigt, dass jene Frauen, welche keine Unterstützung von ihren Ehepartnern oder Lebensgefährten erhalten, einen deutlich höheren Anteil der Haus- und Familienarbeit selbst erledigen als diejenigen, die von ihren Partnern unterstützt werden. Es sind nur wenige Ausnahmen dabei. Dass diese Variable einen Einfluss auf die Hausarbeit ausübt war vorhersehbar. Aus diesem Grund wird sie als erklärende Variable in das Modell aufgenommen.

5.2.3 Kinder

Die Variable **Kinder** enthält die Anzahl der Kinder der befragten Frauen.

Minimum	1. Quantil	Median	Mittelwert	3. Quantil	Maximum
0	0	1	1.105	2	6

Tabelle 5.1: Kenngrößen der Variablen **Kinder**.

Tabelle 5.1 gibt die wichtigsten Kenngrößen dieses Kanals an. Die minimale Kinderanzahl der 229 Frauen liegt demnach bei 0 und die maximale bei 6. Laut Median und Mittelwert haben die befragten Frauen im Durchschnitt ungefähr ein Kind.

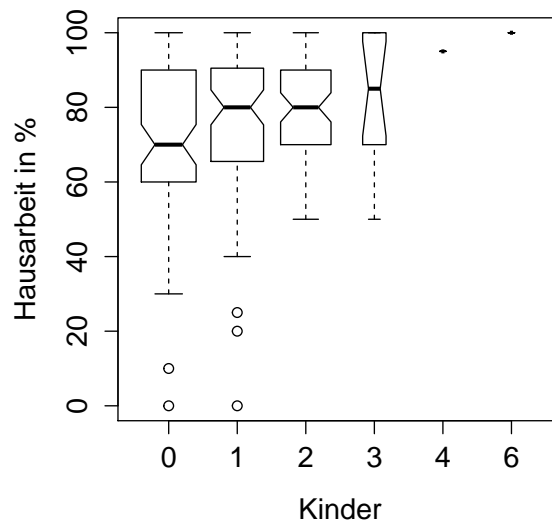


Abbildung 5.3: Boxplot: Hausarbeit gegen Kinder.

Abbildung 5.3 zeigt, dass der Median der von den Frauen selbst erledigten Haus- und Familienarbeit zusammen mit der Anzahl der Kinder steigt. Um zu untersuchen ob dieser Unterschied tatsächlich auf einen signifikanten Einfluss hinweist, wird `Kinder` als erklärende Variable in das Modell aufgenommen.

5.2.4 Anstellung

Die Variable `Anstellung` gibt die derzeitige Arbeitsposition der befragten Frauen an. Sie besteht aus folgenden zwei Faktorstufen:

- Führungskraft
- Mitarbeiterin

Abbildung 5.4 könnte darauf hinweisen, dass die erledigte Haus- und Familienarbeit von der Art der Anstellung abhängt. Einfache Mitarbeiterinnen scheinen einen höheren Anteil der Hausarbeit zu leisten als Frauen in Führungspositionen. Aus diesem Grund wird auch diese Variable in das Modell aufgenommen.

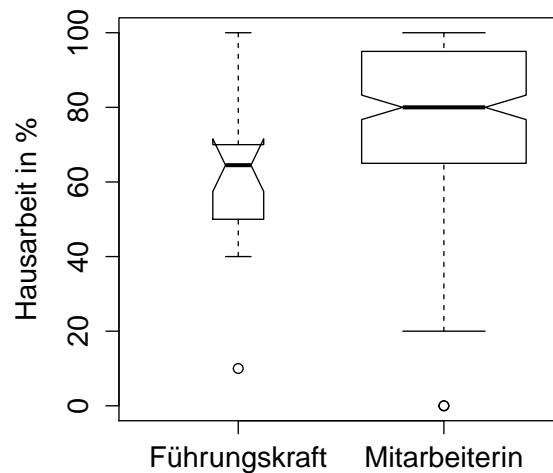


Abbildung 5.4: Boxplot: Hausarbeit gegen Art der Anstellung.

5.2.5 Beschäftigung

Die Variable *Beschäftigung* ist ein weiterer Kanal des Datensatzes von Cernko (2015) und gibt an in welchem Stundenausmaß die Frauen bei der UniCredit Bank Austria AG arbeiten.

Minimum	1. Quantil	Median	Mittelwert	3. Quantil	Maximum
9	28	38	33.53	39	70

Tabelle 5.2: Kenngrößen des Beschäftigungsausmaßes.

Tabelle 5.2 zeigt, dass das minimale Beschäftigungsausmaß bei 9 Stunden und das maximale bei 70 Stunden liegt. Mit einem Mittelwert von 33.53 Stunden arbeitet der Durchschnitt der Frauen weniger als 35 Stunden die Woche.

Um den Boxplot des Beschäftigungsmaßes betrachten zu können, wird die Variable *Beschäftigung* in drei Kategorien unterteilt. Abbildung 5.5 zeigt, dass der Prozentsatz der geleisteten Hausarbeit mit steigender Beschäftigung fällt. Vor allem jene Frauen, welche mehr als 50 Stunden die Woche arbeiten, scheinen einen großen Teil der Haus- und Familienarbeit abzugeben. Da dies auch aus praktischer Sicht Sinn macht, wird *Beschäftigung* zu den erklärenden Variablen hinzugefügt.

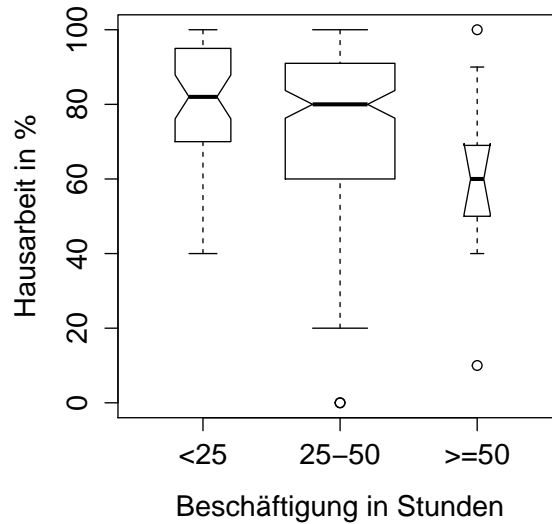


Abbildung 5.5: Boxplot: Hausarbeit gegen Beschäftigung.

5.2.6 Hausarbeit der Mutter

Die Variable **Mutter** enthält die Prozentwerte der erledigten Haus- und Familienarbeit der Mütter, bewertet aus Sicht der befragten Frauen. Tabelle 5.3 fasst die Kenngrößen dieses Kanals zusammen. Dabei fällt vor allem der hohe Median mit einem Wert von 90% auf.

Minimum	1. Quantil	Median	Mittelwert	3. Quantil	Maximum
0	50	90	70.4	100	100

Tabelle 5.3: Kenngrößen des Hausarbeitsanteils der Mütter.

Wird zusätzlich das Histogramm in Abbildung 5.6 betrachtet, so lässt sich erkennen, dass mehr als die Hälfte der 229 Mütter mehr als 80% der Hausarbeit selbst erledigen mussten. Im Vergleich zu Abbildung 5.1 scheint bei Abbildung 5.6 mehr Gewicht auf den niedrigen bzw. hohen Prozentwerten zu liegen.

Um mittels Boxplot zu untersuchen, ob die Variable **Mutter** eventuell einen Einfluss auf die Hausarbeit der befragten Frauen ausübt, wird dieser Kanal in vier Gruppen unterteilt. Abbildung 5.7 lässt vermuten, dass die verrichtete Haus- und Familienarbeit der Mütter keinen Rückschluss auf den erledigten Anteil der Töchter zulässt. Um dies zu testen wird die Variable **Mutter** in das Modell aufgenommen.

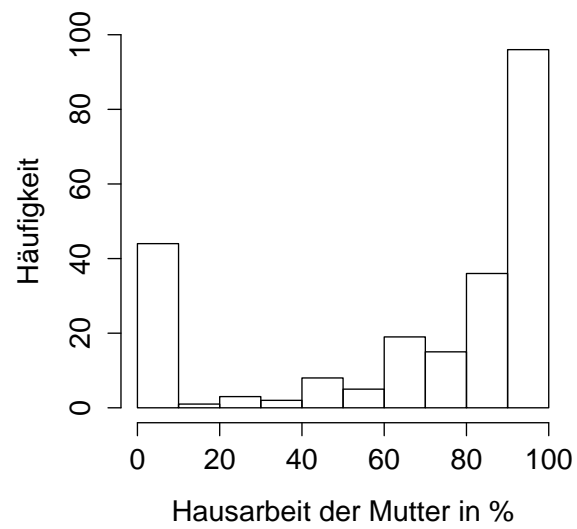


Abbildung 5.6: Histogramm: Hausarbeit der Mutter.

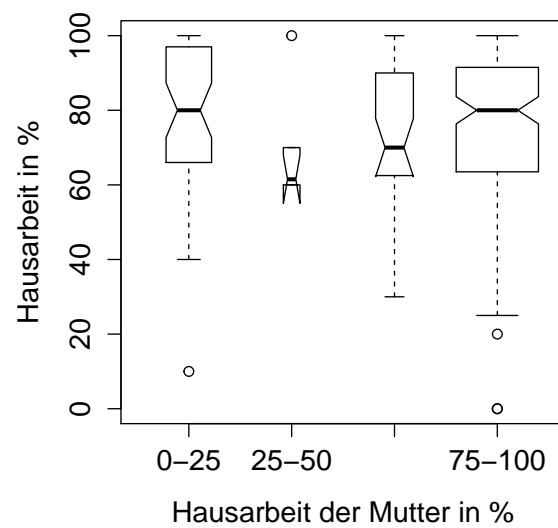


Abbildung 5.7: Boxplot: Hausarbeit der befragten Frauen gegen Hausarbeit der Mütter.

5.2.7 Modell

Im weiteren Verlauf dieser Arbeit werden die verschiedenen eingeführten Regressionsmodelle miteinander verglichen. Um die Unterschiede erkennbar zu machen und zu veranschaulichen, werden stets die gleichen fünf erklärenden Variablen in der selben Reihenfolge in die unterschiedlichen Regressionsmodelle aufgenommen. Zusätzlich wird der Einfachheit halber auf die Verwendung von Interaktionen verzichtet.

Zum Auswerten der Daten und zur Spezifikation der Modellstruktur wird zuerst der Datensatz unter Verwendung des Pakets **openxlsx** (Walker und Braglia, 2015) eingelesen. Anschließend werden die Prozentwerte der Haus- und Familienarbeit standardisiert auf das Intervall $[0, 1]$.

```
> library(openxlsx)
> setwd("C:/Users/amanda/Desktop/Masterarbeit/Hausarbeit")
> data <- read.xlsx("Hausarbeit.xlsx", sheet="Komplett", colNames = TRUE,
+                 rowNames = FALSE)
> attach(data)
> Hausarbeit <- Hausarbeit/100
```

Danach wird die Struktur des Modells über die ausgewählten erklärenden Variablen und die Response-Variable definiert.

```
> formula <- (Hausarbeit ~ Unterstuetzung + Kinder + Beschaeftigung +
+ Anstellung + Mutter)
```

5.3 GLM mit Binomialverteilung

Als erstes wird das Generalisierte Lineare Modell betrachtet. Die Bernoulli-Verteilung scheidet als Kandidat für die Verteilungsannahme aus, da sie nur Response-Werte aus der Menge $\{0, 1\}$ zulässt. Die Binomialverteilung hingegen ist für die Modellierung von Response-Werten aus $[0, 1]$ geeignet, vorausgesetzt es handelt sich um Raten oder Proportionen mit bekannter Versuchsanzahl. Da hier jedoch Prozentwerte vorliegen und 75% sowohl als 75 Erfolge von 100 Versuchen, als auch als 750 von 1000 gewertet werden können, erfolgt die Wahl des Gewichts willkürlich. Eine Modellierung der Prozentwerte aus $[0, 1]$ ohne die Verwendung eines Gewichts, das auf eine ganzzahlige Anzahl von Erfolgen zurückzuführen ist, ist bei der Binomialverteilungsannahme leider nicht möglich.

Die Darstellung der Deviance in (2.19) zeigt, dass sich die Wahl des Gewichts auf die Größenordnung der Deviance auswirkt. Wird ein Gewicht von 100 angenommen, so wird die Deviance ver Hundertfacht, bei einem Gewicht von 1000 wird sie tausendmal so groß. Um dies zu veranschaulichen wird das Generalisierte Lineare Modell nun mit beiden Gewichten spezifiziert.

```
> w1<-rep(100,229)
> mBin1<- glm(formula, weights = w1, family=binomial)
> summary(mBin1)
```


Deviance Residuals:

Min	1Q	Median	3Q	Max
-18.7378	-2.2971	0.7859	3.9992	9.2822

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.6871504	0.1023153	6.716	1.87e-11 ***
UnterstützungNein	0.9078692	0.0364648	24.897	< 2e-16 ***
Kinder	0.2080005	0.0188330	11.044	< 2e-16 ***
Beschäftigung	-0.0141851	0.0018697	-7.587	3.28e-14 ***
AnstellungMitarbeiterin	0.3878534	0.0542706	7.147	8.89e-13 ***
Mutter	0.0016604	0.0004436	3.743	0.000182 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 5787.1 on 228 degrees of freedom
 Residual deviance: 4674.9 on 223 degrees of freedom
 AIC: 5536.7

Number of Fisher Scoring iterations: 5

Die Nulldeviance mit einem Wert von 5787.1 und die Residualdeviance von 4674.9 werden durch die Wahl des Gewichts künstlich auf diese Größenordnung gebracht. Natürlich macht eine Interpretation der Ergebnisse daher wenig Sinn. Zum Vergleich werden nun noch die Ergebnisse unter Verwendung eines Gewichts von 1000 angeführt.

```
> w2<-rep(1000,229)
> mBin2<- glm(formula, weights = w2, family=binomial)
> summary(mBin2)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-59.254	-7.264	2.485	12.647	29.353

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.6871504	0.0323549	21.24	<2e-16 ***
UnterstützungNein	0.9078692	0.0115311	78.73	<2e-16 ***
Kinder	0.2080005	0.0059555	34.93	<2e-16 ***
Beschäftigung	-0.0141851	0.0005912	-23.99	<2e-16 ***
AnstellungMitarbeiterin	0.3878534	0.0171619	22.60	<2e-16 ***
Mutter	0.0016604	0.0001403	11.84	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

5.4. Quasi-Likelihoodmodelle

```
Null deviance: 57871 on 228 degrees of freedom
Residual deviance: 46749 on 223 degrees of freedom
AIC: 48038
```

```
Number of Fisher Scoring iterations: 5
```

Wie schon zuvor erklärt, wurde die Deviance erneut verzehnfacht. Die Standardfehler unter `mBin1` entsprechen natürlich dem $\sqrt{10}$ -fachen der hier angeführten Standardfehler unter Modell `mBin2`, bei den Residuen verhält es sich genau umgekehrt. Da das Gewicht keinen Einfluss auf die Parameterschätzer ausübt, bleiben diese unverändert.

Diese Resultate veranschaulichen die zuvor beschriebenen Probleme der Binomialverteilungsannahme bei vorliegenden Prozentwerten. Die Wahl des Gewichts wird willkürlich getroffen, was zu einer Verfälschung der Ergebnisse führt.

5.4 Quasi-Likelihoodmodelle

5.4.1 Varianzfunktion proportional zur Bernoulli-Varianz

In diesem Teil der Arbeit wird die aus Abschnitt 3.3 bekannte Varianzannahme

$$\text{Var}(\mathbf{y}) = \phi V(\boldsymbol{\mu}) = \phi \boldsymbol{\mu}(1 - \boldsymbol{\mu}) \quad (5.1)$$

und die daraus resultierende Quasi-Likelihoodfunktion betrachtet. Im Gegensatz zur Binomialverteilungsannahme ermöglicht dieser Ansatz die Verwendung eines Gewichts von 1, was natürlich von Beginn an die passendere Wahl bei vorliegenden Prozentwerten gewesen wäre. Die Resultate werden somit nicht länger durch die willkürliche Wahl des Gewichts verfälscht.

```
> mQB <- glm(formula, family = quasibinomial)
> summary(mQB)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.87378 -0.22971  0.07859  0.39992  0.92822
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.687150   0.458110   1.500   0.1350
UnterstützungNein 0.907869   0.163267   5.561 7.65e-08 ***
Kinder          0.208001   0.084323   2.467   0.0144 *
Beschäftigung  -0.014185   0.008371  -1.694   0.0916 .
AnstellungMitarbeiterin 0.387853   0.242994   1.596   0.1119
Mutter          0.001660   0.001986   0.836   0.4041
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for quasibinomial family taken to be 0.2004791)
```

```

Null deviance: 57.871 on 228 degrees of freedom
Residual deviance: 46.749 on 223 degrees of freedom
AIC: NA

```

```
Number of Fisher Scoring iterations: 4
```

Wie erwartet bleiben die Parameterschätzer unverändert und die Deviance beträgt genau ein Hundertstel bzw. ein Tausendstel der Ergebnisse von `mBin1` und `mBin2`. Bei einer geschätzten Dispersion von 0.2005 liegt die skalierte Deviance des Modells mit einem Wert von 233.2 ungefähr im Bereich des Freiheitsgrades. Die Standardfehler können unter Verwendung der Varianzstruktur des Generalisierten Linearen Modells (vgl. Gleichung (2.5)) aus den Standardfehlern von `mBin1` mittels des Zusammenhangs

$$SE(\text{mQB}) = \sqrt{\hat{\phi}} SE(\text{mBin1})\sqrt{100} \quad (5.2)$$

berechnet werden.

```
> anova(mQB, test = "F")
Analysis of Deviance Table
```

```
Model: quasibinomial, link: logit
```

```
Response: Hausarbeit
```

```
Terms added sequentially (first to last)
```

	Df	Deviance	Resid. Df	Resid. Dev	F	Pr(>F)
NULL			228	57.871		
Unterstützung	1	6.7798	227	51.092	33.8180	2.082e-08 ***
Kinder	1	2.6126	226	48.479	13.0316	0.0003782 ***
Beschäftigung	1	1.0093	225	47.470	5.0347	0.0258272 *
Anstellung	1	0.5822	224	46.887	2.9042	0.0897447 .
Mutter	1	0.1390	223	46.749	0.6932	0.4059778

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Wie schon aufgrund von Abbildung 5.7 vermutet wurde, übt die Variable `Mutter` mit einer Reduzierung der Deviance um 0.139 keinen signifikanten Einfluss auf die Haus- und Familienarbeit aus. Die Kanäle `Unterstützung` und `Kinder` scheinen hoch signifikant zu sein, während die Variable `Beschäftigung` mit einem p-Wert von 0.0258 einen etwas geringeren Einfluss ausübt und `Anstellung` mit einem p-Wert von 0.0897 schon nahe bei der 10%-Grenze liegt.

Die linke Grafik in Abbildung 5.8 zeigt, dass die Residuen für steigenden linearen Prädiktor betragsmäßig tendenziell kleiner werden. Wedderburn (1974) zeigte, dass die betrachtete Varianzfunktion zur Unterschätzung der Variabilität für Erwartungswerte nahe am Rand von (0, 1) neigt. Auch für den hier betrachteten Datensatz könnte dies

zutreffen.

Sowohl die QQ-Grafik als auch der Shapiro-Wilk-Test verwerfen die Normalverteilungsannahme der standardisierten Deviance-Residuen.

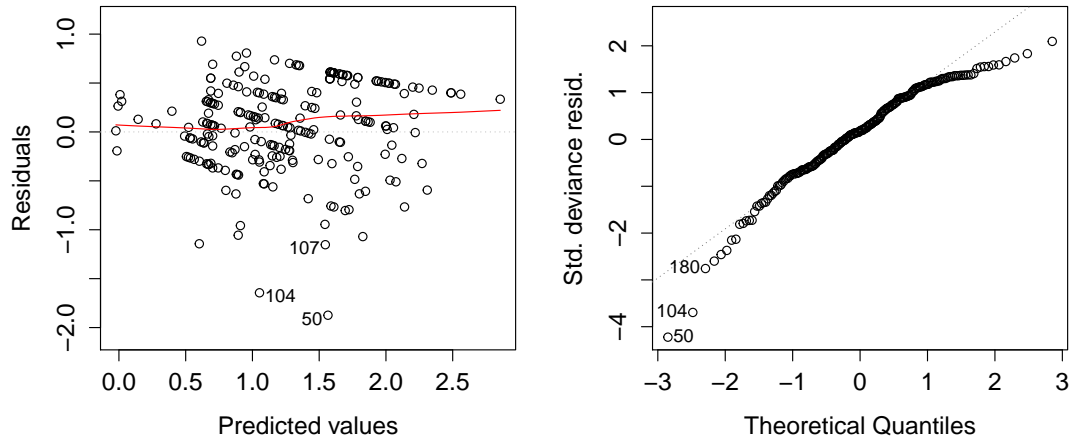


Abbildung 5.8: Residuengrafiken des Modells mQB.

```
> shapiro.test(rstandard(mQB))

      Shapiro-Wilk normality test

data:  rstandard(mQB)
W = 0.95192, p-value = 6.548e-07
```

5.4.2 Quadratische Varianzfunktion

Das zweite Quasi-Likelihoodmodell dieser Arbeit korrespondiert mit der in Abschnitt 3.3 erwähnten Varianzannahme

$$\text{Var}(\mathbf{y}) = \phi V(\boldsymbol{\mu}) = \phi \boldsymbol{\mu}^2 (1 - \boldsymbol{\mu})^2. \quad (5.3)$$

Zur Modellierung wird das R Paket **gnm** (Turner und Firth, 2015) benötigt. Die dort definierte Funktion **wedderburn** ist in der Lage das für die Modellspezifizierung benötigte Objekt der Varianzfamilie zu erzeugen. Dieses Objekt enthält die in (3.11) definierte und um ϵ korrigierte Formel zur Berechnung der Deviance, wodurch auch für dieses Modell eine Betrachtung der Analysis of Deviance möglich ist.

Eine Alternative zur Funktion **wedderburn** und zur Adjustierung um die Konstante ϵ stellt die bei Millar (2011) vorgestellte Varianzfunktion dar. Dabei werden die Deviance-Residuen durch die quadrierten Pearson-Residuen aus Gleichung (2.18) ersetzt. Diese

Variante hat jedoch den Nachteil, dass die Konvergenz der Modellierung nachhaltig beeinträchtigt wird.

```
> library(gnm)
> mQW <- glm(formula, family = wedderburn)
> summary(mQW)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.0938 -0.5137  0.2141  1.1269  3.6440

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.333805  0.492158  0.678 0.498318
UnterstützungNein 0.978790  0.160134  6.112 4.34e-09 ***
Kinder           0.292042  0.083241  3.508 0.000545 ***
Beschäftigung    -0.008492  0.008815 -0.963 0.336380
AnstellungMitarbeiterin 0.444520  0.288005  1.543 0.124141
Mutter           0.001747  0.002144  0.815 0.415936
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for wedderburn family taken to be 1.359547)

Null deviance: 685.76  on 228  degrees of freedom
Residual deviance: 616.47  on 223  degrees of freedom
AIC: NA
```

Number of Fisher Scoring iterations: 6

Dieses Modell schätzt die Dispersion auf einen Wert von 1.3595, welche somit mehr als zehnmal so groß ist als beim Modell `mQB`. Zusammen mit der Residualdeviance resultiert daraus eine skalierte Deviance von 453.4378, was verglichen mit den 223 Freiheitsgraden relativ hoch ist. Allerdings könnte der hohe Wert durch die Adjustierung der Deviance um die Konstante ϵ erklärt werden.

```
> anova(mQW, test = "F")
Analysis of Deviance Table

Model: wedderburn, link: logit

Response: Hausarbeit

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev      F    Pr(>F)
NULL                228      685.76
Unterstützung      1    40.077      227      645.69 29.4784 1.473e-07 ***
Kinder             1    20.803      226      624.88 15.3016 0.0001217 ***
```

5.4. Quasi-Likelihoodmodelle

Beschäftigung	1	3.649	225	621.23	2.6840	0.1027713
Anstellung	1	3.937	224	617.30	2.8961	0.0901888 .
Mutter	1	0.827	223	616.47	0.6082	0.4362909

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

Die Variablen *Unterstützung* und *Kinder* weisen, wie auch schon beim zuvor betrachteten Modell, eine hohe Signifikanz auf. Die Kanäle *Beschäftigung* und *Mutter* hingegen können laut Analysis of Deviance vernachlässigt werden. Die Hausarbeit der Mutter ist nur mit 1.2% an der Gesamtreduktion der Deviance beteiligt. Das Beschäftigungsausmaß liegt mit einem p-Wert von 0.1028 nur knapp über der Signifikanzgrenze, die Variable *Anstellung* hingegen mit einem Wert von 0.0902 knapp darunter.

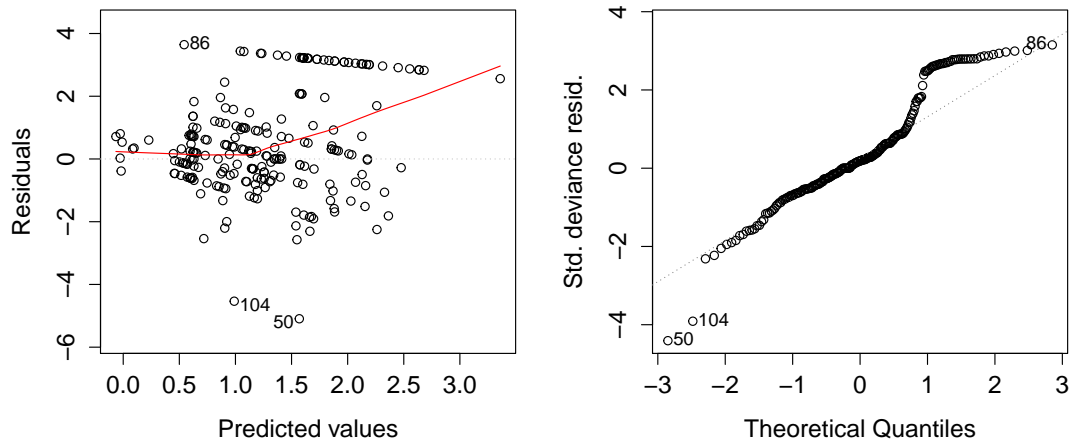


Abbildung 5.9: Residuengrafiken des Modells mQW.

Die linke Grafik in Abbildung 5.9 zeigt, dass die Residuen der geschätzten Erwartungswerte, die nahe bei 1 liegen, einen deutlich höheren Wert aufweisen als beim Modell mQB. Das deckt sich mit der von Wedderburn (1974) gemachten Beobachtung, dass das Quasi-Likelihoodmodell mit zur Bernoulli-Varianz proportionalen Varianzfunktion zur Unterschätzung der Variabilität bei Prozentwerten nahe am Rand von $(0, 1)$ neigt. Die quadratische Varianzfunktion hingegen scheint die Variabilität besser widerzuspiegeln.

```
> shapiro.test(rstandard(mQW))

Shapiro-Wilk normality test

data:  rstandard(mQW)
W = 0.94122, p-value = 5.671e-08
```

Das Ergebnis des Shapiro-Wilk-Tests deckt sich mit den Ergebnissen des QQ-Plots der rechten Grafik in Abbildung 5.9. Die Annahme, dass die standardisierten Deviance-Residuen einer Normalverteilung folgen, kann erneut verworfen werden.

5.5 EQL-Modell

In Abschnitt 5.4 wurden die Quasi-Likelihoodmodelle ausgehend von zwei Varianzfunktionen betrachtet, die sich lediglich im Exponenten unterscheiden. Aufgrund der Interpretation der Residuengrafik scheint die quadratische Varianzstruktur die geeignetere Annahme für die vorliegenden Daten zu sein. Die Verwendung der EQL-Funktion ermöglicht eine optimale Wahl der Exponenten der binomialähnlichen Varianzstruktur

$$V_{k,l}(\boldsymbol{\mu}) = \boldsymbol{\mu}^k (1 - \boldsymbol{\mu})^l. \quad (5.4)$$

Das R Paket **EQL** (Thaler, 2009b) kann dazu verwendet werden diese Parameter zu finden und eine Modellanpassung durchzuführen.

Da die Varianzfunktion $V_{k,l}(\cdot)$ für Werte aus $\{0,1\}$ Null wird, ist die EQL-Funktion mit binomialähnlicher Varianzstruktur nur für Response-Werte aus dem Inneren von $(0,1)$ definiert. Aus diesem Grund werden 100% von nun an als 99% und 0% geleistete Hausarbeit als 1% gewertet. Die Beobachtungen können nicht einfach aus dem Datensatz entfernt werden, da die stärkere Gewichtung von 100% zu einer Verzerrung der Ergebnisse führen würde.

```
> ps1 <- list(seq(1, 2, length = 25), seq(1, 2, length = 25))
> mEQL1 <- eql(formula, param.space=ps1, family = extBinomialVarianceFamily())
> mEQL1
EQL-Max: 170.7455          at: k = 2.000          l = 1.917
EQL Summary:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  121.4  151.6   161.5   158.0  166.0   170.7
```

Werden beide Parameter aus dem Intervall $[1,2]$ ausgewählt, so erreicht die EQL-Funktion ihr Maximum bei einem Wert von 170.7455. Die optimale Wahl der Exponenten bezüglich dieser Parameterräume liegt bei $k = 2$ und $l = 1.917$. Auch der zu `mEQL1` korrespondierende Konturplot in Abbildung 5.10 spiegelt dieses Ergebnis wider. Da das Maximum am Rand angenommen wird, könnte eine Erweiterung der Parameterräume zu einer Verbesserung führen.

5.5. EQL-Modell

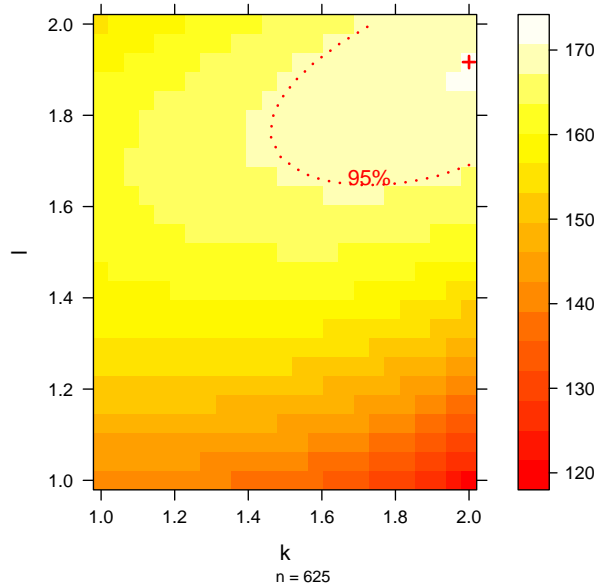


Abbildung 5.10: EQL-Konturplot für das Modell mEQL1.

```
> summary(mEQL1$model)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.2484	-0.4752	0.2017	1.0264	2.4165

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.436596	0.471106	0.927	0.35506
UnterstützungNein	0.924350	0.154695	5.975	9e-09 ***
Kinder	0.263185	0.080827	3.256	0.00131 **
Beschäftigung	-0.009811	0.008467	-1.159	0.24780
AnstellungMitarbeiterin	0.424096	0.271850	1.560	0.12017
Mutter	0.001669	0.002053	0.813	0.41714

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Extended-Binomial-Family family taken to be 1.099369)

Null deviance: 364.63 on 228 degrees of freedom
Residual deviance: 307.80 on 223 degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 7

Die Residualdeviance dieses Modells liefert zusammen mit der geschätzten Dispersion eine skalierte Deviance von 279.9788, was nur etwas größer ist als die Anzahl der Frei-

heitsgrade. Bezüglich der Größenordnung der Residuen lässt sich dieses Modell in etwa mit `mQW` vergleichen.

Da die Parameterräume offensichtlich nicht passend gewählt wurden, wird nun noch ein zweites Modell mit $k \in [1.6, 2.4]$ und $l \in [1.6, 2.2]$ betrachtet. Zusätzlich wird die Anzahl der Stützstellen erhöht.

```
> ps2 <- list(seq(1.6, 2.4, length = 50), seq(1.6, 2.2, length = 50))
> mEQL2 <- eql(formula, param.space=ps2, family = extBinomialVarianceFamily())
> mEQL2
EQL-Max: 170.7739          at: k = 2.024          l = 1.906
EQL Summary:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  152.2  165.9   168.1   167.2  169.5   170.8
```

Der maximale Wert der EQL-Funktion liegt nun bei 170.7739 und somit etwas höher als zuvor. Die optimale Wahl der Parameter ergibt hierbei $k = 2.024$ und $l = 1.906$, womit die Exponenten nur eine leichte Veränderung aufweisen. Auch der Konturplot in Abbildung 5.11 bestätigt, dass die Wahl der Parameterräume etwas besser ist als bei `mEQL1`. Das Ergebnis der Exponenten lässt darauf schließen, dass das Quasi-Likelihoodmodell in Kombination mit der quadratischen Varianzfunktion ein geeignetes Regressionsmodell für die Modellierung der vorliegenden Prozentwerte darstellt.

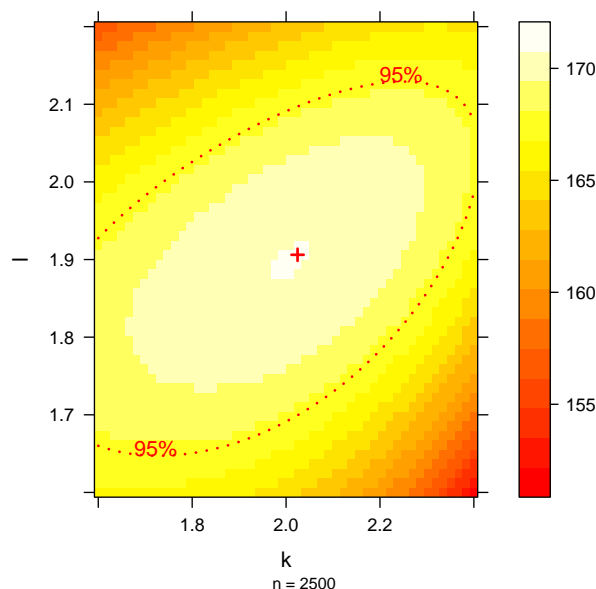


Abbildung 5.11: EQL-Konturplot für das Modell `mEQL2`.

```
> summary(mEQL2$model)
```

Deviance Residuals:

5.5. EQL-Modell

```

      Min       1Q   Median       3Q      Max
-4.2925 -0.4778  0.1998  1.0192  2.3927

```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.445412   0.470115   0.947  0.34443
UnterstützungNein 0.922627   0.154709   5.964 9.58e-09 ***
Kinder         0.261387   0.080873   3.232  0.00141 **
Beschäftigung -0.009940   0.008456  -1.176  0.24103
AnstellungMitarbeiterin 0.421850   0.270423   1.560  0.12019
Mutter         0.001671   0.002050   0.815  0.41583
---

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Extended-Binomial-Family family taken to be 1.085856)

```

Null deviance: 359.87 on 228 degrees of freedom
Residual deviance: 303.67 on 223 degrees of freedom
AIC: NA

```

Number of Fisher Scoring iterations: 7

Die Dispersion wird bei diesem Modell auf einen Wert von 1.085856 geschätzt, wodurch sich für die skalierte Deviance ein Wert von 279.6595 ergibt. Diese Ergebnisse weisen zwar auf eine Verbesserung hin, weichen jedoch nur leicht von den Resultaten des Modells `mEQL1` ab. Auch die Parameterschätzer und Standardfehler der beiden Modelle sind vergleichbar.

```

> anova(mEQL2$model, test = "F")
Analysis of Deviance Table

```

Model: Extended-Binomial-Family, link: logit

Response: Hausarbeit

Terms added sequentially (first to last)

```

              Df Deviance Resid. Df Resid. Dev      F    Pr(>F)
NULL                228      359.87
Unterstützung     1    32.573      227      327.30 29.9976 1.163e-07 ***
Kinder            1    16.184      226      311.12 14.9042 0.0001482 ***
Beschäftigung     1     3.612      225      307.50  3.3267 0.0695034 .
Anstellung        1     3.159      224      304.34  2.9097 0.0894433 .
Mutter            1     0.671      223      303.67  0.6184 0.4324871
---

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Die Signifikanzniveaus sind vergleichbar mit jenen des Modells `mEQL1`, deshalb wurden die Ergebnisse von `mEQL1` nicht zusätzlich angeführt. Laut Analysis of Deviance schaffen

es die beiden Variablen `Beschäftigung` und `Anstellung` nur knapp unter die 10%-Grenze. Diese Ergebnisse weichen leicht von der Analysis of Deviance des Modells `mQW` ab, dort lag der p-Wert des Beschäftigungsausmaßes über der Signifikanzgrenze.

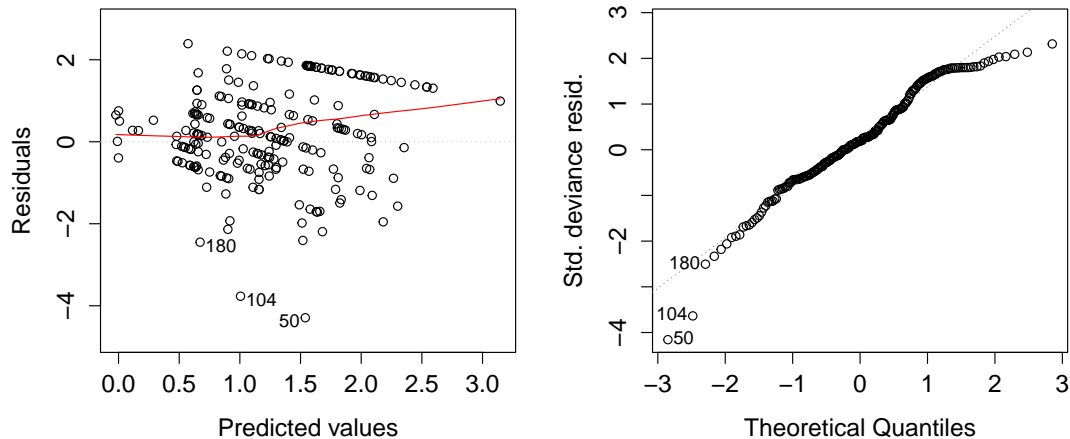


Abbildung 5.12: Residuengrafiken des Modells `mEQL2`.

```
> shapiro.test(rstandard(mEQL2$model))
```

Shapiro-Wilk normality test

```
data: rstandard(mEQL2$model)
W = 0.9678, p-value = 4.624e-05
```

Abbildung 5.12 ist vergleichbar mit Abbildung 5.9 des Modells `mQW`. Ein Sinken der Absolutbeträge der Residuen bei steigendem Prädiktor ist nur leicht erkennbar, es scheint keine Unterschätzung der Variabilität vorzuliegen. Die QQ-Grafik und der Shapiro-Wilk-Test verwerfen erneut die Normalverteilungsannahme der standardisierten Deviance-Residuen.

5.6 Beta-Regressionsmodelle

Als zweite große Klasse an Regressionsmodellen wird die Beta-Regression betrachtet. Das zur Modellierung geeignete R Paket `betareg` stammt von Zeileis et al. (2015) und wird bei Schalk (2015) und Cribari-Neto und Zeileis (2010) beschrieben. Da dieses Paket keine Analysis of Deviance bereitstellt, schlagen Cribari-Neto und Zeileis (2010) die Verwendung eines Wald-Tests mittels der Funktion `waldtest()` des Pakets `lmtest` (Hothorn et al., 2015) vor. Wie bereits bekannt, ist die Beta-Regression nur für Response-Werte

aus dem offenen Einheitsintervall $(0, 1)$ definiert, daher werden 100% Haus- und Familienarbeit erneut als 99% und 0% als 1% gewertet.

5.6.1 Grundmodell

Zuerst werden die nötigen Pakete geladen und das Modell mittels der Funktion `betareg()` spezifiziert.

```
> library(betareg)
> library(lmtest)
> mBeta <- betareg(formula)
```

Auch das Paket `betareg` enthält die Funktion `summary()`, die eine Zusammenfassung der Parameterschätzer und der Standardfehler liefert. Im Gegensatz zum Generalisierten Linearen Modell wird hier allerdings keine Deviance berechnet.

```
> summary(mBeta)
```

Standardized weighted residuals 2:

Min	1Q	Median	3Q	Max
-4.3895	-0.5447	-0.0564	0.6308	2.9914

Coefficients (mean model with logit link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.666446	0.395039	1.687	0.0916 .
UnterstützungNein	1.006286	0.134939	7.457	8.83e-14 ***
Kinder	0.104179	0.068175	1.528	0.1265
Beschäftigung	-0.010803	0.007132	-1.515	0.1298
AnstellungMitarbeiterin	0.428223	0.225203	1.902	0.0572 .
Mutter	0.000116	0.001728	0.067	0.9465

Phi coefficients (precision model with identity link):

	Estimate	Std. Error	z value	Pr(> z)
(phi)	4.0043	0.3645	10.98	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Type of estimator: ML (maximum likelihood)

Log-likelihood: 137.9 on 7 Df

Pseudo R-squared: 0.2696

Number of iterations: 16 (BFGS) + 2 (Fisher scoring)

Die geschätzte Dispersion liegt mit einem Wert von 4.004 deutlich höher als bei den bisher betrachteten Modellen. Die standardisierten gewichteten Residuen 2 (vgl. Abschnitt 4.2.4) entsprechen ungefähr der Größenordnung der berechneten Deviance-Residuen der EQL-Modelle. Anders als beim GLM enthält die Zusammenfassung hier auch den Wert der optimierten Log-Likelihoodfunktion, der bei 137.9 mit 7 Freiheitsgraden liegt. Es fällt auf, dass deutlich mehr Iterationsschritte zur Modellanpassung benötigt werden als bei den Generalisierten Linearen Modellen. Um die Parameter im Gesamten mittels

eines Wald-Tests auf ihre Signifikanz zu überprüfen, werden nun einige Hilfsmodelle spezifiziert. Dabei werden die Parameter sukzessive in der bisher betrachteten Reihenfolge aufgenommen.

```
> b0 <- betareg(Hausarbeit ~ 1)
> b1 <- betareg(Hausarbeit ~ Unterstützung)
> b2 <- betareg(Hausarbeit ~ Unterstützung + Kinder)
> b3 <- betareg(Hausarbeit ~ Unterstützung + Kinder + Beschäftigung)
> b4 <- betareg(Hausarbeit ~ Unterstützung + Kinder + Beschäftigung +
+ Anstellung)
```

Zur Betrachtung der Signifikanzniveaus können die eingebetteten Modelle nun gegeneinander getestet werden. Modell **b1** wird mit dem Nullmodell verglichen, **b2** danach mit **b1**. Diese Methode wird unter Verwendung des Wald-Tests mittels der Funktion `waldtest()` bis zum Modell `mBeta` fortgesetzt.

```
> waldtest(b0,b1, test = "F")
Wald test

Model 1: Hausarbeit ~ 1
Model 2: Hausarbeit ~ Unterstützung
  Res.Df Df      F    Pr(>F)
1      227
2      226  1 53.434 4.578e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> waldtest(b1,b2, test = "F")
Wald test

Model 1: Hausarbeit ~ Unterstützung
Model 2: Hausarbeit ~ Unterstützung + Kinder
  Res.Df Df      F Pr(>F)
1      226
2      225  1 5.847 0.0164 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Variable	Freiheitsgrade	F-Statistik	p-Wert	Signifikanz
Unterstützung	226	53.434	4.578e-12	***
Kinder	225	5.847	0.0164	*
Beschäftigung	224	3.8465	0.05109	.
Anstellung	223	3.7038	0.05556	.
Mutter	222	0.0045	0.9465	

Tabelle 5.4: Signifikanzniveaus der Variablen im Modell `mBeta`.

Die in Tabelle 5.4 dargestellten Ergebnissen sind vergleichbar mit den Resultaten des EQL-Modells `mEQL2`. Lediglich das Signifikanzniveau der Variable `Kinder` ist gesunken. Diese Resultate sind jedoch mit Vorsicht zu bewerten, da der Wald-Test lediglich die Submodelle gegeneinander testet ohne dabei das Gesamtmodell zu betrachten. Um zu testen welchen Einfluss die Parameter unter Berücksichtigung des Gesamtmodells auf die Haus- und Familienarbeit ausüben, können die Variablen einzeln aus dem Modell `mBeta` entfernt werden. So kann mittels Wald-Test geprüft werden, ob die Hinzunahme der Variable eine signifikante Verbesserung bringt, obwohl bereits alle anderen Variablen in das Modell aufgenommen wurden.

Variable	F-Statistik	p-Wert	Signifikanz
<code>Unterstützung</code>	55.612	1.966e-12	***
<code>Kinder</code>	2.3351	0.1279	
<code>Beschäftigung</code>	2.2945	0.1313	
<code>Anstellung</code>	3.6157	0.05853	.
<code>Mutter</code>	0.0045	0.9465	

Tabelle 5.5: Signifikanzniveaus der Variablen bei Entfernung aus dem Modell `mBeta`.

Laut Tabelle 5.5 scheinen die Variablen `Kinder`, `Beschäftigung` und `Mutter` ergänzend zu den übrigen Variablen keine zusätzliche Information für die Modellierung der Haus- und Familienarbeit zu liefern. Obwohl die Variable `Kinder` zuvor hoch signifikant war, fällt sie bei Berücksichtigung des Gesamtmodells unter die Signifikanzgrenze. Lediglich `Unterstützung` und `Anstellung` bleiben im gleichen Signifikanzniveau.

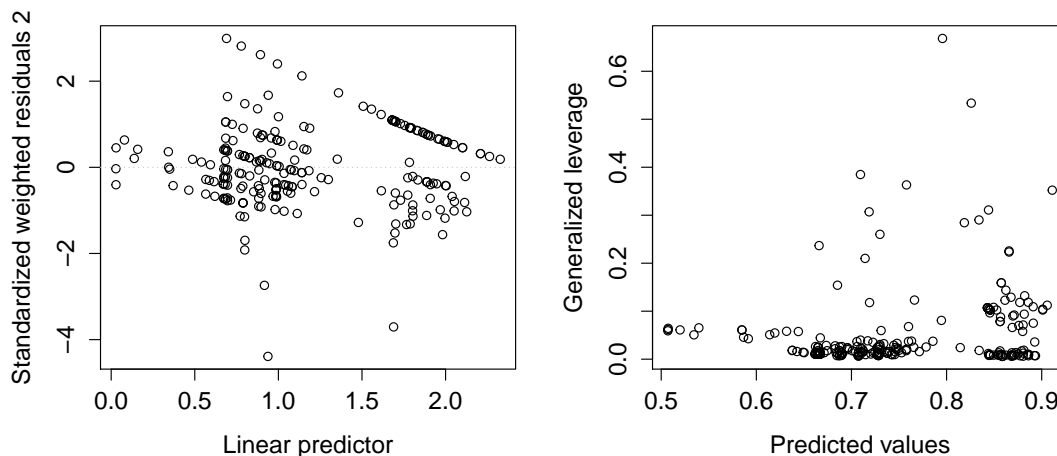


Abbildung 5.13: Residuengrafik und generalisierte Hebelpunkte des Modells `mBeta`.

Die linke Grafik von Abbildung 5.13 zeigt, dass die Residuen bei steigendem Prädiktor kleiner werden. Das Muster ist vergleichbar mit der Residuengrafik in Abbildung 5.8 des Modells `mQB`. Da dieses Ergebnis möglicherweise auf eine Unterschätzung der Variabilität hinweist, könnte die gemeinsame Modellierung von Erwartungswert und Dispersion zu einer Verbesserung führen. Einige wenige Punkte scheinen mit einem Wert von über 0.5 eine relativ große Hebelwirkung zu haben. Diese Beobachtungen y_i üben demnach einen großen Einfluss auf die Schätzung der linearen Prädiktoren η_i und somit auch auf die geschätzten Erwartungswerte μ_i aus.

Wird die generalisierte Hebelwirkung wie in Abbildung 5.14 gegen die Response-Werte aufgetragen, so wird deutlich, dass jene Punkte, welche am oberen Rand des Intervalls $(0, 1)$ liegen, eine relativ hohe Hebelwirkung im Vergleich zu den übrigen Punkten aufweisen. Diese Art der Darstellung der Hebelpunkte ist zwar ungewöhnlich, zeigt aber, dass die extremen Punkte offensichtlich eher dazu neigen einen Einfluss auf die Schätzung der Prädiktoren und Erwartungswerte auszuüben.

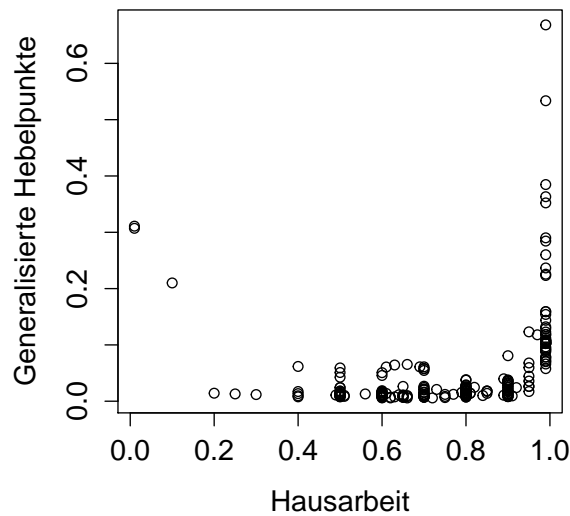


Abbildung 5.14: Generalisierte Hebelpunkte von `mBeta` gegen die Variable `Hausarbeit`.

5.6.2 Gemeinsame Modellierung von Erwartungswert und Dispersion

Als Erweiterung zum Grundmodell der Beta-Regression kann die gemeinsame Modellierung von Erwartungswert und Dispersion betrachtet werden. Zuerst werden sowohl für den Erwartungswert als auch für die Dispersion dieselben Parameter verwendet.

```
> mBetaD1 <- betareg(Hausarbeit ~ Unterstuetzung + Kinder + Beschaeftigung +
+ Anstellung + Mutter|Unterstuetzung + Kinder + Beschaeftigung + Anstellung +
+ Mutter)
```

5.6. Beta-Regressionsmodelle

```
> summary(mBetaD1)
```

```
Standardized weighted residuals 2:
```

```
      Min      1Q  Median      3Q      Max
-5.1879 -0.6151 -0.0549  0.5979  3.5027
```

```
Coefficients (mean model with logit link):
```

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	0.655370	0.368814	1.777	0.07557	.
UnterstützungNein	0.801718	0.160156	5.006	5.56e-07	***
Kinder	0.208869	0.068722	3.039	0.00237	**
Beschäftigung	-0.011966	0.006972	-1.716	0.08608	.
AnstellungMitarbeiterin	0.295658	0.182167	1.623	0.10459	
Mutter	0.001612	0.001768	0.912	0.36170	

```
Phi coefficients (precision model with log link):
```

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.771698	0.565879	3.131	0.001743	**
UnterstützungNein	-0.657332	0.195587	-3.361	0.000777	***
Kinder	0.310921	0.097213	3.198	0.001382	**
Beschäftigung	-0.001645	0.010085	-0.163	0.870437	
AnstellungMitarbeiterin	-0.710270	0.328272	-2.164	0.030490	*
Mutter	0.004516	0.002432	1.857	0.063305	.

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Type of estimator: ML (maximum likelihood)
```

```
Log-likelihood: 152.5 on 12 Df
```

```
Pseudo R-squared: 0.2503
```

```
Number of iterations: 22 (BFGS) + 2 (Fisher scoring)
```

Schon hier zeigt sich, dass die Modellierung der Dispersion zu einer Verbesserung der Maximum Likelihood Funktion führt. Es stellt sich allerdings die Frage, welche Variablen tatsächlich benötigt werden. Nach ausreichender Analyse wurde entschieden die Variablen `Unterstützung` und `Kinder` in das Dispersionsmodell aufzunehmen, da die Hinzunahme weiterer Variablen keine signifikante Verbesserung lieferte.

```
> mBetaD2 <- betareg(Hausarbeit ~ Unterstützung + Kinder + Beschäftigung +
+ Anstellung + Mutter | Unterstützung + Kinder)
> summary(mBetaD2)
```

```
Standardized weighted residuals 2:
```

```
      Min      1Q  Median      3Q      Max
-5.1465 -0.6459 -0.0168  0.6465  3.4411
```

```
Coefficients (mean model with logit link):
```

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	0.5966582	0.3708672	1.609	0.10766	
UnterstützungNein	0.8106431	0.1627469	4.981	6.33e-07	***
Kinder	0.2087199	0.0684977	3.047	0.00231	**

```

Beschäftigung          -0.0097507  0.0067035  -1.455  0.14579
AnstellungMitarbeiterin 0.4234770  0.2151002   1.969  0.04898 *
Mutter                 -0.0001836  0.0016062  -0.114  0.90899

```

```
Phi coefficients (precision model with log link):
```

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.36835    0.14628   9.355 < 2e-16 ***
UnterstützungNein -0.58878    0.19646  -2.997  0.00273 **
Kinder          0.28640    0.08947   3.201  0.00137 **
---

```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Type of estimator: ML (maximum likelihood)
```

```
Log-likelihood: 149.9 on 9 Df
```

```
Pseudo R-squared: 0.253
```

```
Number of iterations: 19 (BFGS) + 2 (Fisher scoring)
```

Das Modell `mBetaD2` scheint eine gute Wahl zu sein, was die Modellierung der Dispersion betrifft. Durch das Weglassen der drei Variablen wurde die Log-Likelihood lediglich um 2.6 verringert und der Algorithmus zur Modellanpassung benötigt nun 3 Iterationen weniger. Mittels eines Likelihood-Quotienten-Tests kann geprüft werden, ob die Aufnahme der beiden Variablen in das Dispersionsmodell tatsächlich zu einer signifikanten Verbesserung führt. Dabei werden die Variablen einzeln aus dem Modell `mBetaD2` entfernt.

```

> d1 <- betareg(Hausarbeit ~ Unterstützung + Kinder + Beschäftigung +
+ Anstellung + Mutter | Unterstützung)
> lrtest(d1, mBetaD2)
Likelihood ratio test

```

```
Model 1: Hausarbeit ~ Unterstützung + Kinder + Beschäftigung + Anstellung +
Mutter | Unterstützung
```

```
Model 2: Hausarbeit ~ Unterstützung + Kinder + Beschäftigung + Anstellung +
Mutter | Unterstützung + Kinder
```

```

#Df LogLik Df  Chisq Pr(>Chisq)
1   8 146.31
2   9 149.86  1  7.1097  0.007667 **
---

```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

> d2 <- betareg(Hausarbeit ~ Unterstützung + Kinder + Beschäftigung +
+ Anstellung + Mutter | Kinder)
> lrtest(d2,mBetaD2)
Likelihood ratio test

```

```
Model 1: Hausarbeit ~ Unterstützung + Kinder + Beschäftigung + Anstellung +
Mutter | Kinder
```

```
Model 2: Hausarbeit ~ Unterstützung + Kinder + Beschäftigung + Anstellung +
Mutter | Unterstützung + Kinder
```

```

#Df LogLik Df  Chisq Pr(>Chisq)
1   8 145.77

```

5.6. Beta-Regressionsmodelle

```
2 9 149.86 1 8.1838 0.004227 **
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Beide Tests auf unterschiedliche Varianzmodelle liefern ein signifikantes Ergebnis. Die Aufnahme der beiden Variablen **Unterstützung** und **Kinder** in das Dispersionsmodell führt zu einer Erhöhung der maximierten Log-Likelihoodfunktion und trägt somit zu einer Verbesserung des Modells bei. Somit wurde gezeigt, dass die Annahme der konstanten Dispersion signifikant zugunsten des variablen Dispersionsmodells verworfen werden kann.

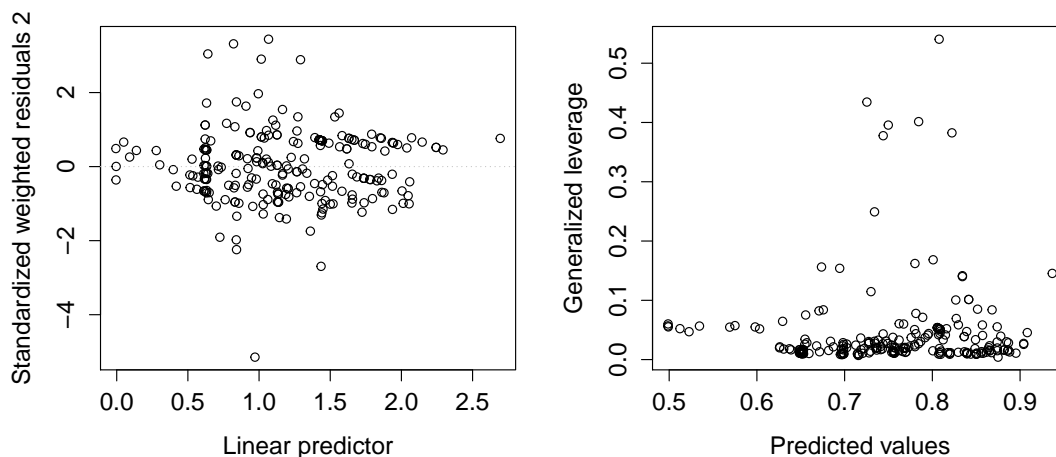


Abbildung 5.15: Residuengrafik und generalisierte Hebelpunkte des Modells `mBetaD1`.

Im Gegensatz zu Abbildung 5.13 weist die Residuengrafik in Abbildung 5.15 keinerlei Muster auf. Durch die Modellierung der Dispersion wird die Variabilität des Modells offensichtlich gut erfasst. Bei den generalisierten Hebelpunkten fällt auf, dass die Werte etwas niedriger liegen als zuvor, auch hier führt die Annahme der variablen Dispersion zu einer Verbesserung.

5.6.3 Korrektur und Reduktion des Bias

Ergänzend wird nun noch betrachtet, ob Korrektur und Reduktion der Verzerrung einen Einfluss auf die Parameterschätzer und deren Standardabweichungen liefern. Dazu wird die `summary` der drei Modelle betrachtet und die Werte für einen besseren Überblick in einer Tabelle zusammengefasst. Ausgangssituation bietet hier das Modell `mBeta`, die Modellierung der Dispersion wird für diese Veranschaulichung vernachlässigt.

```
> mBetaBK <- betareg(formula, type = "BC")
> mBetaBR <- betareg(formula, type = "BR")
```

Parameter	ML	BK	BR
Intercept	0.6664 (0.3950)	0.6653 (0.3990)	0.6639 (0.3990)
UnterstützungNein	1.0063 (0.1349)	1.0024 (0.1360)	1.0024 (0.1360)
Kinder	0.1042 (0.0682)	0.1049 (0.0688)	0.1051 (0.0688)
Beschäftigung	-0.0108 (0.0682)	-0.0107 (0.0072)	-0.0107 (0.0072)
Mitarbeiterin	0.4282 (0.2252)	0.4222 (0.2279)	0.4219 (0.2279)
Mutter	0.0001 (0.0017)	0.0001 (0.0017)	0.0001 (0.0017)
ϕ	4.0043 (0.3645)	3.8603 (0.3511)	3.860 (0.351)
Log-Likelihood	137.9	137.8	137.8

Tabelle 5.6: Vergleich der Parameterschätzer bei Korrektur und Reduktion des Bias.

Die Parameterschätzer und dessen Standardabweichungen variieren in Tabelle 5.6 nur sehr gering, dasselbe gilt für die maximierte Log-Likelihoodfunktion. Lediglich die Schätzung der Dispersion weist kleine Veränderungen auf. Auch die Signifikanzniveaus der Prädiktoren bleiben beinahe unverändert und werden deshalb nicht erneut angeführt. Dies lässt darauf schließen, dass die Maximum Likelihood Methode bei Modell `mBeta` zu keiner großen Verzerrung führt.

5.7 Interpretation der Ergebnisse

Zur Modellierung der vorliegenden Prozentwerte der Haus- und Familienarbeit wurde mit dem Generalisierten Linearen Modell in Kombination mit der Binomialverteilung begonnen. Durch das Vorliegen von Prozentwerten war der Rückschluss auf eine passende Gesamtanzahl bzw. auf ein geeignetes Gewicht nicht möglich. Das führte dazu, dass die Ergebnisse verfälscht wurden und eine Interpretation nicht möglich war. Eine erste Verbesserung boten die Quasi-Likelihoodmodelle. Die proportional zur Bernoulli-Varianz angesetzte Varianzstruktur lieferte zwar die gleichen Parameterschätzer wie die Annahme der Binomialverteilung, jedoch wurden die Resultate nicht länger durch die willkürliche Wahl des Gewichts beeinflusst und Deviance und Residuen konnten zur Modellanalyse herangezogen werden. Die Variable `Mutter` schien keinen Einfluss auf die Vorhersage der Response-Werte auszuüben. Ein großes Problem stellte allerdings die Unterschätzung der Variabilität jener Punkte dar, welche sich am Rand des Intervalls $(0, 1)$ befanden. Durch die Verwendung der Quasi-Likelihoodfunktion in Verbindung mit der quadratischen Varianzfunktion konnte diese Auffälligkeit weitestgehend beseitigt werden. Auch im Vergleich der Signifikanzniveaus zeigte sich, dass ein weiterer Kanal, nämlich `Beschäftigung`, unter die Signifikanzgrenze fiel. Der Ansatz der EQL-Funktion lieferte dann überraschenderweise das Ergebnis, dass die quadratische Varianzfunktion nur knapp von der optimalen Wahl der Exponenten abwich. Zur Verwendung des EQL-Modells mussten jedoch die Beobachtungen mit den Werten $\{0, 1\}$ als $\{0.1, 0.99\}$ interpretiert werden.

Die Beta-Regression lieferte von Beginn an sehr plausible und gut interpretierbare Resultate. Der einzige Nachteil besteht darin, dass auch hier nur Response-Werte aus dem Inneren von $(0, 1)$ zugelassen sind. Schon das Grundmodell wies mit dem EQL-Modell vergleichbare Signifikanzniveaus auf, allerdings schien auch hier die Variabilität unterschätzt zu werden. Auch bezüglich der generalisierten Hebelwirkung konnten einige wenige Ausreißer entdeckt werden. Durch die gemeinsame Modellierung von Erwartungswert und Dispersion wurde verdeutlicht, dass die Annahme einer konstanten Dispersion zugunsten der variablen Dispersion verworfen werden konnte. Durch die Aufnahme der Variablen **Unterstützung** und **Kinder** in das Dispersionsmodell wurde die Variabilität des Modells gut erfasst und die Ausreißer der generalisierten Hebelwirkung auf einen Punkt reduziert. Ergänzend wurden schließlich noch die Methoden zur Korrektur und Reduktion des Bias getestet. Die daraus resultierenden Ergebnisse ließen den Rückschluss zu, dass durch die Maximum Likelihood Schätzung keine nennenswerten Verzerrungen entstanden sind.

Motiviert durch den vorliegenden Datensatz beschreibt diese Arbeit zahlreiche, zur Modellierung von Prozentwerten geeignete Regressionsmodelle. Dabei wurde bei der Einführung des Generalisierten Linearen Modells begonnen, der Übergang zur Quasi-Likelihoodfunktion mit der Möglichkeit zum Vergleich unterschiedlicher Varianzfunktionen beschrieben und sogar die Extended-Quasi-Likelihoodfunktion diskutiert, die vor allem durch die Verwendung der binomialähnlichen Varianzstruktur und der Optimierung der korrespondierenden Exponenten überzeugte. Zusätzlich wurde eine zweite große Klasse von Regressionsmodellen eingeführt. Die Beta-Regression basiert auf der Annahme beta-verteilter Response-Variablen und scheint aufgrund der Bandbreite der Beta-Verteilung ein geeigneter Kandidat zur Modellierung von Werten aus dem offenen Einheitsintervall $(0, 1)$ zu sein. Aus diesem Grund wurden sowohl das Grundmodell, als auch die Erweiterung bezüglich der gemeinsamen Modellierung von Erwartungswert und Dispersion ausführlich beschrieben. Als zusätzliche Erweiterung wurde die Korrektur bzw. Reduktion der durch die Maximum Likelihood Schätzung erzeugten Verzerrung diskutiert.

Da ein rein theoretischer Vergleich nicht ausreicht, um alle Unterschiede in der Modellanpassung und Parameterschätzung aufzuzeigen, wurden die Regressionsmodelle anhand eines Beispiels getestet. Bei der Analyse des vorliegenden Datensatzes konnten vor allem das EQL-Modell und das Beta-Regressionsmodell überzeugen. Das Generalisierte Lineare Modell in Kombination mit der Annahme einer Binomialverteilung ist aufgrund der Fehlinformation der Gesamtanzahl an Versuchen nicht für die Modellierung von Prozentwerten geeignet. Das Quasi-Likelihoodmodell mit der zur Bernoulli-Varianz proportionalen Varianzannahme scheidet wegen der Unterschätzung der Variabilität am Rand von $(0, 1)$ aus. Die quadratische Varianzstruktur des Quasi-Likelihoodmodells lieferte plausible Ergebnisse, auch das EQL-Modell bestätigte dies. Genauso überzeugend wie die Resultate des EQL-Modells waren die Ergebnisse der Beta-Regression. Vor al-

lem die gemeinsame Modellierung von Erwartungswert und Dispersion konnte durch die Residuengrafiken und die Erfassung der Variabilität überzeugen.

Alle vorgestellten Regressionsmodelle bieten zahlreiche Möglichkeiten zur Anwendung. Jedes der Modelle hat sowohl Vor- als auch Nachteile, weshalb eine eindeutige Entscheidung nicht möglich ist. Da die Ergebnisse natürlich auch von den zu untersuchenden Daten abhängen, sollten stets mehrere Regressionsmodelle verglichen werden, um so eine optimale Wahl zu treffen und einen möglichen Irrtum auszuschließen.

Literaturverzeichnis

- Casella, G. und Berger, R. L. (2002). *Statistical Inference. Second Edition*. Boston: Thomson Learning.
- Cernko, G. (2015). *Die Entwicklung der Rollenverteilung in Familien und Partnerschaften unter Berücksichtigung der angebotenen Förderungsmaßnahmen für Mitarbeiterinnen am Beispiel der UniCredit Bank Austria AG* (Unveröffentlichte Diplomarbeit). IMC Fachhochschule Krens, Austria.
- Cribari-Neto, F. und Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, 34 (2), 1–24.
- Espinheira, P. L., Ferrari, S. L. P. und Cribari-Neto, F. (2008a). Influence diagnostics in beta regression. *Computational Statistics and Data Analysis*, 52 (9), 4417–4431.
- Espinheira, P. L., Ferrari, S. L. P. und Cribari-Neto, F. (2008b). On beta regression residuals. *Journal of Applied Statistics*, 35 (4), 407–419.
- Fahrmeir, L. und Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics*, 13 (1), 342–368.
- Ferrari, S. L. P. und Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31 (7), 799–815.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80 (1), 27–38.
- Green, P. J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society, Series B (Methodological)*, 46 (2), 149–192.
- Grün, B., Kosmidis, I. und Zeileis, A. (2012). Extended beta regression in R: Shaken, stirred, mixed and partitioned. *Journal of Statistical Software*, 48 (11), 1–25.
- Hothorn, T., Zeileis, A., Farebrother, R. W., Cummins, C., Millo, G. und Mitchell, D. (2015). **lmtest**: *Testing Linear Regression Models*. <http://CRAN.R-project.org/package=lmtest>. (R package version 0.9-34)
- Kosmidis, I. und Firth, D. (2010). A generic algorithm for reducing bias in parametric estimation. *Electronic Journal of Statistics*, 4, 1097–1112.

- Kosmidis, I., Zou, A. und Jagan, K. (2015). Beyond beta regression: modelling percentages and fractions in the presence of boundary observations. In H. Friedl und H. Wagner (Hrsg.), *Proceedings of the 30th international workshop on statistical modelling: Volume 1* (S. 267–272). Statistical Modelling Society.
- McCullagh, P. (1983). Quasi-likelihood functions. *The Annals of Statistics*, 11 (1), 59–67.
- McCullagh, P. und Nelder, J. A. (1989). *Generalized Linear Models. Second Edition*. London: Chapman & Hall.
- McLachlan, G. J. und Peel, D. (2000). *Finite Mixture Models*. New York: John Wiley & Sons.
- Millar, R. B. (2011). *Maximum Likelihood Estimation and Inference: With Examples in R, SAS and ADMB*. West Sussex: Wiley.
- Nelder, J. A. und Lee, Y. (1992). Likelihood, quasi-likelihood and pseudolikelihood: Some comparisons. *Journal of the Royal Statistical Society, Series B (Methodological)*, 54 (1), 273–284.
- Nelder, J. A. und Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika*, 74 (2), 221–232.
- Nelder, J. A. und Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A (General)*, 135 (3), 370–384.
- Ospina, R., Cribari-Neto, F. und Vasconcellos, K. L. P. (2006). Improved point and interval estimation for a beta regression model. *Computational Statistics & Data Analysis*, 51 (2), 960–981.
- Ospina, R. und Ferrari, S. L. P. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, 56 (6), 1609–1623.
- Pregibon, D. (1980). Goodness of link tests for generalized linear models. *Applied Statistics*, 28 (1), 15–24.
- Rocha, A. V. und Simas, A. B. (2011). Influence diagnostics in a general class of beta regression models. *Test*, 20 (1), 95–119.
- Schalk, A. (2015). *Das betareg Paket*. (TU-Graz, Institut für Statistik. Projekt-Arbeit)
- Simas, A. B., Barreto-Souza, W. und Rocha, A. V. (2010). Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis*, 54 (2), 348–366.
- Smithson, M. und Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11 (1), 54–71.
- Staudte, R. G. und Sheather, S. J. (1990). *Robust Estimation and Testing*. New York: Wiley.
- Thaler, T. (2009a). *Die Extended-Quasi-Likelihood-Funktion in Generalisierten Linearen Modellen* (Unveröffentlichte Diplomarbeit). Graz University of Technology, Austria. (Institut für Statistik.)
- Thaler, T. (2009b). **EQL**: *Extended-Quasi-Likelihood-Function*. <http://CRAN.R-project.org/package=EQL>. (R package version 1.0-0)
- Turner, H. und Firth, D. (2015). **gnm**: *generalized nonlinear models*. <http://CRAN.R-project.org/package=gnm>. (R package version 1.0-8)

- Walker, A. und Braglia, L. (2015). **openxlsx**: *read, write and edit xlsx files*. <http://CRAN.R-project.org/package=openxlsx>. (R package version 3.0.0)
- Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61 (3), 439–447.
- Wei, B.-C., Hu, Y.-Q. und Fung, W.-K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics*, 25 (1), 25–37.
- Zeileis, A., Cribari-Neto, F., Grün, B., Kosmidis, I., Simas, A. B. und Rocha, A. V. (2015). **betareg**: *beta Regression*. <http://CRAN.R-project.org/package=betareg>. (R package version 3.0-5)
- Zeileis, A. und Hornik, K. (2007). Generalized m-fluctuation tests for parameter instability. *Statistica Neerlandica*, 61 (4), 488–508.
- Zeileis, A., Hothorn, T. und Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, 17 (2), 492–514.