

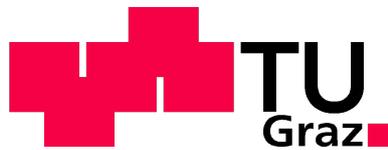
Caroline BAYR, Dipl.Ing.

Parametrisierte Varianzstrukturen für Generalisierte Lineare Modelle

MASTERARBEIT

**zur Erlangung des akademischen Grades einer
Diplom-Ingenieurin**

**Masterstudium Technische Mathematik:
Operations Research und Statistik**



Graz University of Technology

Technische Universität Graz

Betreuer:

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

Institut für Statistik

Graz, im März 2011

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am.....

.....

(Unterschrift)

Vorwort

Generalisierte Lineare Modelle wurden 1972 von Nelder und Wedderburn als eine Erweiterung der Linearen Modelle eingeführt. Sie dienen zur Modellierung einer Funktion des Erwartungswertes einer Zufallsvariablen, die aus einer Verteilung aus einer einparametrischen linearen Exponentialfamilie stammen. Anstatt jedoch die gesamte Verteilung anzunehmen diskutierte Wedderburn 1974 erstmals die sogenannte Quasi-Likelihood Funktion - eine Funktion, deren Score Funktion ähnliche Eigenschaften wie die Score Funktion der Log-Likelihood Funktion besitzt. Man geht bei diesem Ansatz lediglich von einer Erwartungswert-Varianz-Beziehung aus, der sogenannten Varianzfunktion. Dadurch können auch Zufallsvariablen, deren Verteilung nicht aus der einparametrischen linearen Exponentialfamilie stammen, modelliert werden. Um verschiedene Varianzfunktionen miteinander vergleichen zu können, betrachten wir die Extended-Quasi-Likelihood Funktion, die 1987 von Nelder und Pregibon definiert wurde. Es können nun auch unbekannte Parameter in der Varianzfunktion zusätzlich zu den Modellparametern geschätzt werden. Extended-Quasi-Likelihood Schätzung findet in der Praxis oft Anwendung in der Modellierung von Daten mit Überdispersion, d.h. wenn die Datenvariabilität größer als die Modellvariabilität ist.

An dieser Stelle möchte ich mich bei Ao. Univ.-Prof. Dipl.-Ing. Dr.techn. Herwig Friedl für die freundliche und engagierte Betreuung sowie die vielen hilfreichen Ratschläge bedanken. Die Besprechungen mit ihm haben mich immer einen großen Schritt vorwärts gebracht. Des Weiteren bedanke ich mich bei meinen Eltern, die mich sowohl finanziell als auch mit guten Zusprüchen während meiner gesamten Studienzeit unterstützt haben. Ebenso bedanke ich mich bei meiner Schwester Sandra für ihren moralischen Beistand. Besonderer Dank gilt auch meinem Freund Andreas Angermann, der mich immer wieder aufgebaut und mir mit seinen motivierenden Worten weitergeholfen hat.

Inhaltsverzeichnis

1	Einleitung	1
2	Generalisierte lineare Modelle	3
2.1	Die Verallgemeinerung der Modellannahmen	3
2.1.1	Die Exponentialfamilie	4
2.1.2	Die Linkfunktion	8
2.2	Maximum Likelihood Schätzung	8
2.3	Güte der Modellanpassung	11
2.3.1	Deviance	11
2.3.2	Pearson-Statistik	13
2.3.3	Analyse der Deviance	14
3	Quasi-Likelihood Funktion	16
3.1	Definition	16
3.2	Eigenschaften der QL Funktion	18
3.2.1	Güte der Modellanpassung	22
3.3	Quasi-Dichte	23
4	Extended-Quasi-Likelihood Funktion	25
4.1	Definition	25
4.1.1	Schätzer des Dispersionsparameter	26
4.2	Eigenschaften der EQL Funktion	27
4.3	Unbekannte Parameter in der Varianzfunktion	29
4.3.1	Tweedmodell	31
	Kumulantenerzeugende Funktion	34
	Dichtefunktion der Tweedieverteilung	36
4.3.2	EQL Funktion mit Power-Varianz	38
4.3.3	Binomialähnlicher Ansatz	40
5	Beispiele	46
5.1	Regenvorhersage	46
5.2	Aberrant Crypt Foci	59
5.3	Schiffsschaden	72
5.4	Stoffdaten	83
5.5	Textildaten	92
	Literatur	99

1 Einleitung

Die Basis für diese Masterarbeit bildet die Modellierung von Daten mittels *Generalisierten Linearen Modellen* (GLMe), welche eine Erweiterung der Lineare Modelle (LMe) sind.

Bei einem GLM gehen wir von der Annahme aus, dass die zu modellierende Variable (Responsevariable) aus einer Verteilung aus einer einparametrischen linearen Exponentialfamilie stammt. Die Varianz wird hier als konstant angenommen. Im Gegensatz zu einem LM beschreibt man den Erwartungswert der Responsevariable durch erklärende Variablen nicht direkt, sondern modelliert eine Funktion desselbigen, die sogenannte *Linkfunktion*. So werden im 2. Kapitel dieser Masterarbeit diese erweiterten Annahmen eines GLMs erläutert und deren Eigenschaften diskutiert.

Im 3. Kapitel gewinnen wir die Erkenntnis, dass die Score Funktion der Log-Likelihood Funktion einer Stichprobe aus der einparametrischen linearen Exponentialfamilie nur von den ersten beiden Momenten abhängt, was uns zur *Quasi-Likelihood Funktion* (QL Funktion) führt. Somit können auch Zufallsvariablen modelliert werden, die aus keiner Verteilung aus einer einparametrischen linearen Exponentialfamilie stammen. Hier muss die Varianzfunktion, die die Beziehung zwischen dem Erwartungswert und der Varianz beschreibt, bis auf eine multiplikative Konstante genau angegeben werden. Eine Verbesserung dieser Annahme finden wir im nächsten Kapitel.

Im 4. Kapitel befassen wir uns mit der sogenannten *Extended-Quasi-Likelihood Funktion* (EQL Funktion), die auch Varianzfunktionen mit unbekanntem Parameter erlaubt. Als Schätzer für den Dispersionsparameter ergibt sich die mittlere Deviance. Die unbekanntem Parameter der Varianzfunktion können ebenfalls simultan mit den Modellparametern geschätzt werden. Dadurch besteht die Möglichkeit, verschiedene Varianzfunktionen miteinander zu vergleichen. Wir erläutern zwei spezielle Varianzfunktionen, das Tweedie Modell und den binomialähnlichen Ansatz. Im Besonderen wird die Varianzfunktion zur Tweedieverteilung genauer diskutiert sowie deren spezielle Struktur erklärt. Schließlich leiten wir die EQL Funkti-

on nach den unbekanntem Parametern ab und können diese dadurch direkt schätzen.

Im 5. Kapitel wird die besprochene Theorie anhand praktischer Beispiele angewandt. Es werden die Ergebnisse aller Ansätze angeführt und verglichen.

Diese Masterarbeit bildet somit eine umfangreiche Diskussion von Generalisierten Linearen Modellen über die Quasi-Likelihood Funktion bis hin zur Extended-Quasi-Likelihood Funktion, die speziell mit zwei verschiedenen Ansätzen mit unbekanntem Parametern in der Varianzfunktion erklärt werden.

2 Generalisierte lineare Modelle

Dieses Kapitel orientiert sich am Buch von McCullagh und Nelder (1989), Kapitel 2, sowie an der Masterarbeit von Thaler (2009), Kapitel 2, und am Vorlesungsskriptum von Friedl (2000), Kapitel 2.

2.1 Die Verallgemeinerung der Modellannahmen

Generalisierte lineare Modelle (GLMe) sind eine Erweiterung der klassischen linearen Modelle (LMe), bei welchen man von n unabhängig normalverteilten Response-Variablen Y_i und einer konstanten Varianz $\text{var}[Y_i] = \sigma^2$ ausgeht. Der Erwartungswert $\mathbb{E}[Y_i] = \mu_i$ wird als Linearkombination der erklärenden Variablen $x_i \in \mathbb{R}^p$ und den Modellparametern $\beta \in \mathbb{R}^p$ modelliert

$$\mu_i = x_i' \beta. \quad (2.1)$$

Da diese Annahmen naturgemäß oft nicht erfüllt sind, verallgemeinert man sie: Anstatt von normalverteilten Response-Variablen auszugehen, trifft man die Annahme, dass diese aus einer Verteilung der linearen einparametrischen Exponentialfamilie stammen. Die konstante Varianz wird durch eine Varianzfunktion, abhängig vom Erwartungswert der Response-Variablen, ersetzt. Schließlich wird nicht der Erwartungswert selbst wie in (2.1) linear modelliert, sondern eine Funktion des selbigen, welche als *Linkfunktion* bezeichnet wird.

Nun können wir die verallgemeinerten Annahmen des GLM zusammenfassend darstellen. Sei $y = (y_1, \dots, y_n)'$ der Vektor der Realisierung des Zufallsvektors $Y = (Y_1, \dots, Y_n)'$, dann lautet das GLM:

$$\begin{aligned} Y_i &\sim F(\theta_i), \\ F &\in \text{Exponentialfamilie}, \\ \mathbb{E}[Y_i] &= \mu_i = \mu(\theta_i), \\ \eta_i &= x_i' \beta, \\ g(\mu_i) &= \eta_i. \end{aligned} \quad (2.2)$$

Bemerkung 2.1. Der p -dimensionale Vektor der bekannten Prädiktorvariablen $x_i = (x_{i1}, \dots, x_{ip})'$ wird in der sogenannten Designmatrix $X = (x_1, \dots, x_n)' \in \mathbb{R}^{n \times p}$ zusammengefasst.

Diese erweiterten Annahmen werden nun in den folgenden Abschnitten genauer erläutert.

2.1.1 Die Exponentialfamilie

Da wir von der Annahme ausgehen, dass die Response-Variablen aus einer Exponentialfamilie stammen, wollen wir nun die Dichte einer solchen definieren.

Definition 2.2. (*Exponentialfamilie*)

Sei θ der kanonische Parameter, so ist die Dichtefunktion der Exponentialfamilie definiert als

$$f_Y(y, \theta) = h(y)c(\theta) \exp \left\{ \sum_{i=1}^k w_i(\theta)t_i(y) \right\}$$

mit bekannten Funktionen $h(\cdot)$, $c(\cdot)$, $w_i(\cdot)$ und $t_i(\cdot)$.

Bemerkung 2.3. Für den Fall $w_i(\theta) = \theta$ spricht man von der Exponentialfamilie in kanonischer Form.

Für einen bekannten Dispersionsparameter ϕ gilt für die einparametrische lineare Exponentialfamilie folgende Definition.

Definition 2.4. (*Einparametrische lineare kanonische Exponentialfamilie*)

Die Dichtefunktion einer einparametrischen linearen Exponentialfamilie mit bekanntem Dispersionsparameter ϕ hat die Form:

$$f_Y(y, \theta) = \exp \left\{ \frac{y\theta - b(\theta)}{a \cdot \phi} + c(y, \phi) \right\}, \quad (2.3)$$

wobei $b(\cdot)$, $c(\cdot)$ bekannte Funktionen sind, a ein Gewicht und θ ein kanonischer Parameter ist.

Bemerkung 2.5. Der Dispersionsparameter ϕ wird hier als bekannt angesehen, wodurch ϕ in Definition 2.4 keine besondere Beachtung geschenkt wird.

Bemerkung 2.6. Ist ϕ ein unbekannter Dispersionsparameter, so wird Definition 2.2 als zweiparametrische Exponentialfamilie aufgefasst.

Zu beachten ist an dieser Stelle auch, dass die Dichtefunktion der Exponentialfamilie linear im kanonischen Parameter θ ist.

Um den Erwartungswert und die Varianz von Y bestimmen zu können, betrachten wir die Eigenschaften der Score Funktion.

Satz 2.7. *Sei $f_Y(y, \theta)$ eine beliebige differenzierbare Dichtefunktion von Y mit unbekanntem Parameter θ . Dann gilt unter der Voraussetzung der Vertauschbarkeit von Integration und Differentiation für die Score Funktion:*

$$\mathbb{E} \left[\frac{\partial \log f_Y(Y, \theta)}{\partial \theta} \right] = 0, \quad (2.4)$$

$$\mathbb{E} \left[\frac{\partial \log f_Y(Y, \theta)}{\partial \theta} \right]^2 = \mathbb{E} \left[-\frac{\partial^2 \log f_Y(Y, \theta)}{\partial \theta^2} \right]. \quad (2.5)$$

Beweis. Mit der ersten Ableitung der Log-Likelihood Funktion

$$\frac{\partial \log f_Y(Y, \theta)}{\partial \theta} = \frac{1}{f_Y(Y, \theta)} \frac{\partial f_Y(Y, \theta)}{\partial \theta} \quad (2.6)$$

erhalten wir (2.4)

$$\begin{aligned} \mathbb{E} \left[\frac{\partial \log f_Y(Y, \theta)}{\partial \theta} \right] &\stackrel{(2.6)}{=} \mathbb{E} \left[\frac{1}{f_Y(Y, \theta)} \frac{\partial f_Y(Y, \theta)}{\partial \theta} \right] \\ &= \int_{\mathbb{R}} f_Y(y, \theta) \frac{1}{f_Y(y, \theta)} \frac{\partial f_Y(y, \theta)}{\partial \theta} dy \\ &= \frac{\partial}{\partial \theta} \underbrace{\int_{\mathbb{R}} f_Y(y, \theta) dy}_{=1} = 0. \end{aligned}$$

Gleichung (1.5) ergibt sich durch Anwendung der Kettenregel

$$\begin{aligned} \mathbb{E} \left[-\frac{\partial^2 \log f_Y(Y, \theta)}{\partial \theta^2} \right] &= \mathbb{E} \left[-\frac{\partial^2 f_Y(Y, \theta)}{\partial \theta^2} \frac{1}{f_Y(Y, \theta)} + \left(\frac{\partial f_Y(Y, \theta)}{\partial \theta} \right)^2 \frac{1}{f_Y(Y, \theta)^2} \right] \\ &= -\frac{\partial^2}{\partial \theta^2} \underbrace{\int_{\mathbb{R}} f_Y(y, \theta) dy}_{=1} + \int_{\mathbb{R}} f_Y(y, \theta) \underbrace{\left(\frac{\partial \log f_Y(y, \theta)}{\partial \theta} \right)^2}_{\text{aus (2.6)}} dy \\ &= \mathbb{E} \left[\frac{\partial \log f_Y(Y, \theta)}{\partial \theta} \right]^2. \end{aligned}$$

□

Für die Exponentialfamilie resultiert aus Definition 2.4 und Gleichung (2.3)

$$\mathbb{E} \left[\frac{\partial \log f_Y(Y, \theta)}{\partial \theta} \right] = \frac{1}{a \cdot \phi} \mathbb{E}[Y - b'(\theta)] = 0.$$

Somit ergibt sich der Erwartungswert der Zufallsvariable Y als

$$\mathbb{E}[Y] = \mu = b'(\theta).$$

Die Varianz erhalten wir durch Verwendung von (1.5) und der Lösung folgender Gleichung:

$$\mathbb{E} \left[\frac{\partial^2 \log f_Y(Y, \theta)}{\partial \theta^2} \right] + \mathbb{E} \left[\frac{\partial \log f_Y(Y, \theta)}{\partial \theta} \right]^2 = -\frac{1}{a \cdot \phi} b''(\theta) + \frac{1}{a^2 \cdot \phi^2} \text{var}[Y] = 0. \quad (2.7)$$

Somit folgt aus (2.7) die Varianz der Zufallsvariable Y

$$\text{var}[Y] = a \cdot \phi \cdot b''(\theta) = a \cdot \phi \cdot V(\mu). \quad (2.8)$$

Gleichung (2.8) zeigt, dass der erste Faktor der Varianz vom kanonischen Parameter θ , und somit auch vom Erwartungswert μ , abhängt und wird als *Varianzfunktion* $V(\mu)$ bezeichnet. Hingegen wird ϕ Dispersionsparameter genannt.

Bemerkung 2.8. Die Funktion $b(\theta)$ nennt man *Kumulantenfunktion*.

Um Kumulanten höherer Ordnung bestimmen zu können, benötigen wir die Momenterzeugende Funktion

$$M(t) = \mathbb{E} [e^{tY}]$$

und die daraus resultierende Kumulantenerzeugende Funktion

$$K(t) = \log M(t).$$

Die k -te Kumulante ist nun definiert durch

$$\kappa_k(\theta) = \left. \frac{\partial^k K(t)}{\partial t^k} \right|_{t=0}.$$

Somit ergibt sich für Verteilungen aus der einparametrischen Exponentialfamilie

die Momenterzeugende Funktion zu

$$\begin{aligned}
 M(t) &= \mathbb{E} [e^{tY}] \\
 &= \int_{\mathbb{R}} e^{ty} \cdot \exp \left\{ \frac{y\theta - b(\theta)}{a \cdot \phi} + c(y, \phi) \right\} dy \\
 &= \exp \left\{ -\frac{b(\theta)}{a \cdot \phi} \right\} \int_{\mathbb{R}} \exp \left\{ \frac{y(a\phi t + \theta)}{a \cdot \phi} + c(y, \phi) \right\} dy \\
 &= \exp \left\{ -\frac{b(\theta)}{a \cdot \phi} \right\} \exp \left\{ \frac{a\phi t + \theta}{a \cdot \phi} \right\} \\
 &= \exp \left(\frac{b(a\phi t + \theta) - b(\theta)}{a \cdot \phi} \right).
 \end{aligned}$$

Daraus erhalten wir die Kumulantenerzeugende Funktion

$$K(t) = \frac{b(a\phi t + \theta) - b(\theta)}{a \cdot \phi} \tag{2.9}$$

und die k -te Kumulante

$$\kappa_k(\theta) = (a \cdot \phi)^{k-1} \frac{\partial^k b(\theta)}{\partial \theta^k}.$$

Beispiel 2.9. Schließlich wollen wir noch einige Beispiele für Mitglieder der Exponentialfamilie anführen:

- Die Normalverteilung $Y \sim N(\mu, \sigma^2)$ mit $\theta = \mu$ und $\phi = \sigma^2$:

$$a = 1, \quad b(\theta) = \frac{\theta^2}{2}, \quad c(y, \phi) = -\frac{y^2}{2\phi} - \frac{1}{2} \log(2\pi\phi).$$

- Die Poissonverteilung $Y \sim P(\mu)$ mit $\theta = \log \mu$ und $\phi = 1$:

$$a = 1, \quad b(\theta) = \exp(\theta), \quad c(y, \phi) = -\log y!.$$

- Die Gammaverteilung $Y \sim G(\mu, \nu)$ mit $\theta = -\frac{1}{\mu}$ und $\phi = \frac{1}{\nu}$:

$$a = 1, \quad b(\theta) = -\log(-\theta), \quad c(y, \phi) = \frac{1}{\phi} \log \frac{1}{\phi} + \left(\frac{1}{\phi} - 1 \right) \log y - \log \Gamma \left(\frac{1}{\phi} \right).$$

- Die Invers-Gaussverteilung $Y \sim IG(\mu, \sigma^2)$ mit $\theta = -\frac{1}{2\mu^2}$ und $\phi = \sigma^2$:

$$a = 1, \quad b(\theta) = -(-2\theta)^{1/2}, \quad c(y, \phi) = -\frac{1}{2} \left(\frac{1}{\phi y} + \log(2\pi\phi y^3) \right).$$

- Die standardisierte Binomialverteilung $mY \sim B(m, \mu)$ mit $\theta = \log \frac{\mu}{1-\mu}$ und $\phi = 1$:

$$a = \frac{1}{m}, \quad b(\theta) = \log \frac{1}{1-\mu}, \quad c(y, \phi) = \log \left(\frac{1}{\phi} \right).$$

2.1.2 Die Linkfunktion

Eine weitere verallgemeinerte Annahme entspricht der Modellierung einer Funktion des Erwartungswertes μ , der sogenannten Linkfunktion:

$$\eta := g(\mu) = x'\beta.$$

Diese Linkfunktion $g(\cdot)$ ist eine beliebige streng monotone, differenzierbare Funktion, deren Inverse existiert¹, die zur Modellierung des Erwartungswertes benötigt wird: $\mu = g^{-1}(\eta)$.

In LMe sind der lineare Prädiktor $\eta = g(\mu)$ und der Erwartungswert μ gleichgesetzt. Der Grund einer Modellierung durch eine Linkfunktion liegt darin, dass Einschränkungen des Wertebereichs (z.B. keine negativen Werte) für Erwartungswerte so eingehalten werden können. Geht man beispielsweise nur von ganzzahligen Response-Variablen und einer Poissonverteilung aus, so ist diese identische Linkfunktion nicht verwendbar, da hier der Erwartungswert größer Null und der lineare Prädiktor möglicherweise negativ ist. Dies wird im GLM durch den log Link dargestellt, d.h. $g(\mu) = \log \mu$, woraus sich $\mu = e^\eta$ ergibt.

Geht man von einer Binomialverteilung aus, so sollte die Linkfunktion die Bedingung erfüllen, dass nur Werte im Intervall $[0, 1]$ angenommen werden können. Hierfür verwendet man beispielsweise die inverse Verteilungsfunktion der Standardnormalverteilung $g(\mu) = \Phi^{-1}(\mu)$. Die bekanntesten Linkfunktionen für den Erwartungswert einer Binomialverteilung finden sich in folgender Aufzählung:

1. Logit: $g(\mu) = \log\{\mu/(1-\mu)\}$
2. Probit: $g(\mu) = \Phi^{-1}(\mu)$
3. Komplementärer log-log: $g(\mu) = \log\{-\log(1-\mu)\}$

2.2 Maximum Likelihood Schätzung

Ziel dieses Abschnittes ist unter Annahme eines GLMs den Vektor der unbekannt Parameter $\beta = (\beta_1, \dots, \beta_p)'$ des Modells zu schätzen. Dazu verwenden wir

¹Existenz der Inversen aufgrund der strengen Monotonie und Stetigkeit

die *Maximum Likelihood Schätzung*, die uns den Maximum Likelihood Schätzer (MLE) für β liefert. Hierfür benötigen wir die *Log-Likelihood Funktion*, die im folgenden erläutert wird.

Die Log-Likelihood Funktion der beobachteten Stichprobe $y = (y_1, \dots, y_n)'$ ist wegen der Unabhängigkeit der Y_i durch die Summe der Log-Likelihood Funktionen $\ell(\theta_i, y_i) = \log f(y_i, \theta_i)$ mit $\theta = (\theta_1, \dots, \theta_n)'$ als dem Vektor der unbekanntem Verteilungsparameter gegeben als

$$\mathcal{L}(\theta, y) = \log \left\{ \prod_{i=1}^n f(y_i, \theta_i) \right\} = \sum_{i=1}^n \ell(\theta_i, y_i). \quad (2.10)$$

Da die Response-Variablen Y_i aus einer Exponentialfamilie stammen, können wir Definition 2.4 der Dichtefunktion einer einparametrischen linearen Exponentialfamilie in (2.10) einsetzen und erhalten

$$\mathcal{L}(\theta, y) = \sum_{i=1}^n \left\{ \frac{y_i \theta_i - b(\theta_i)}{a_i \cdot \phi} + c(y_i, \phi) \right\} \quad (2.11)$$

mit dem Dispersionsparameter ϕ .

Da der Erwartungswert μ von β abhängt, d.h. $\mu_i = \mu_i(\beta)$, und die Linkfunktion des GLM (2.2) diesen Zusammenhang modelliert, kann (2.11) als Funktion von β gesehen werden. Um den MLE zu erhalten, wird nach dem Parameter β abgeleitet und dessen Schätzer aus den Score-Gleichungen

$$\frac{\partial \mathcal{L}(\theta, y)}{\partial \beta_j} = 0, \quad j = 1, \dots, p \quad (2.12)$$

erhalten.

Um die impliziten Abhängigkeiten $\theta_i = \theta_i(\mu_i)$, $\mu_i = \mu_i(\eta_i)$ und $\eta_i = \eta_i(\beta_j)$ darzustellen, benötigen wir die Kettenregel, womit (2.12) wie folgt dargestellt werden kann:

$$\frac{\partial \mathcal{L}(\theta, y)}{\partial \beta_j} = \frac{\partial \mathcal{L}(\theta, y)}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = 0.$$

Somit ergibt sich für alle $j = 1, \dots, p$:

$$\frac{\partial \mathcal{L}(\theta, y)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial}{\partial \theta_i} \left\{ \frac{y_i \theta_i - b(\theta_i)}{a_i \cdot \phi} + c(y_i, \phi) \right\} \frac{1}{V(\mu_i)} \frac{x_{ij}}{g'(\mu_i)} = \sum_{i=1}^n \frac{y_i - \mu_i}{a_i \cdot \phi V(\mu_i)} \frac{x_{ij}}{g'(\mu_i)} = 0. \quad (2.13)$$

Um (2.13) zu erhalten, wurden die Identitäten

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

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

verwendet.

Löst man (2.13) nach β_j auf, wobei die β_j implizit in $\mu_i = g^{-1}(\eta_i) = g^{-1}(x_i' \beta)$ stecken, erhält man den MLE $\hat{\beta}_j$. Da (2.13) nur für normalverteilte Response-Variablen mit der Identität als Linkfunktion linear in β ist, muss sie iterativ gelöst werden, wozu man (2.13) als Weighted Least-Squares (WLS) auffasst. Siehe hierfür McCullagh und Nelder (1989), Seite 41 ff., sowie Thaler (2009), Seite 15 ff.

Für asymptotische Eigenschaften des Schätzers $\hat{\beta}$ wird die Score Funktion um den wahren Parameter β entwickelt:

$$0 = \left. \frac{\partial \mathcal{L}(\mu, y)}{\partial \beta} \right|_{\hat{\beta}} \approx \left. \frac{\partial \mathcal{L}(\mu, y)}{\partial \beta} \right|_{\beta} + \left. \frac{\partial^2 \mathcal{L}(\mu, y)}{\partial \beta \partial \beta'} \right|_{\beta} (\hat{\beta} - \beta).$$

Die zweite Ableitung wird durch deren Erwartungswert ersetzt, wodurch wir die sogenannte *Fisher-Information* erhalten.

Definition 2.10. (*Fisher-Information*)

Sei $\mathcal{L}(\mu, y)$ die Log-Likelihood Funktion, so ist die Fisher-Information gegeben als

$$FI(y, \mu) = \mathbb{E} \left[-\frac{\partial^2 \mathcal{L}(\mu, Y)}{\partial \mu^2} \right] = \mathbb{E} \left[\left(\frac{\partial \mathcal{L}(\mu, Y)}{\partial \mu} \right)^2 \right] = \text{var} \left[\frac{\partial \mathcal{L}(\mu, Y)}{\partial \mu} \right].$$

Bemerkung 2.11. Aus Definition 2.10 ist ersichtlich, dass die Fisher-Information die Varianz der Score-Funktion ist.

Als erste Ableitung erhalten wir

$$\frac{\partial \mathcal{L}(\mu, y)}{\partial \beta} = X' DW(y - \mu)$$

mit den Diagonalmatrizen

$$\begin{aligned} D &= \text{diag}(d_i), & d_i &= g'(\mu_i), \\ W &= \text{diag}(w_i), & \frac{1}{w_i} &= a_i \cdot \phi \cdot V(\mu_i)(g'(\mu_i))^2. \end{aligned}$$

Daraus resultiert

$$\hat{\beta} - \beta \approx FI(y, \mu)^{-1} X' DW (y - \mu). \quad (2.14)$$

Aus (2.14) erhalten wir mit $\text{var}[y] = (DWD)^{-1}$ für den Erwartungswert des Schätzers $\hat{\beta}$

$$\mathbb{E}[\hat{\beta}] \approx \beta$$

und für die Varianz-Kovarianzmatrix

$$\text{var}[\hat{\beta}] \approx FI(y, \mu)^{-1},$$

woraus sich die Standardfehler

$$se[\hat{\beta}] \approx \text{diag}[FI(y, \mu)]^{-\frac{1}{2}}$$

ergeben.

2.3 Güte der Modellanpassung

Durch ein GLM werden die Erwartungswerte μ der beobachteten Response-Variablen beschrieben. Der Schätzer $\hat{\mu}$ weicht naturgemäß von den beobachteten Werten y ab. Um Aussagen über die Größenordnung der Abweichung zwischen y und $\hat{\mu}$ treffen zu können, betrachten wir nun zwei Maße für die Güte der Modellanpassung:

- die *Deviance*,
- die *Pearson-Statistik*.

2.3.1 Deviance

Das einfachste Modell besteht aus nur einem Parameter

$$g(\mu_i) = \beta_0$$

und wird als *Null-Modell* bezeichnet. Daraus ergibt sich, dass jedes y_i denselben Erwartungswert μ hat (Zufallsstichprobe). Somit wird die Varianz der Y_i dieses Modells vollständig durch die stochastischen Komponente $Y_i \sim F(\theta_i)$ mit einer Verteilung $F \in$ Exponentialfamilie eines GLMs erklärt. Im Gegensatz dazu wird als Basis zur Berechnung der Deviance ein *saturiertes* Modell verwendet, d.h. es werden n Parameter in das Erwartungswert-Modell aufgenommen, einen für jede Beobachtung. Daraus folgt

$$g(\mu_i) = \beta_i, \quad i = 1, \dots, n.$$

Für dieses saturierte Modell gilt, dass die Beobachtungen y_i den Schätzern $\hat{\mu}_i$ entsprechen: $y_i = \hat{\mu}_i$. Hierbei wird die Varianz der Y_i ganz der systematischen Komponente $g(\mu_i) = \eta_i$ zugeschrieben und es bleibt keine zufällige Komponente übrig. In der Praxis ist das Null-Modell jedoch zu einfach und das saturierte Modell nicht informativ, da es die Daten nur wiedergibt. Aus diesem Grund wird das saturierte Modell mit dem aktuellen Modell mit p Parametern verglichen, wobei dieses Modell mit möglichst wenig Parametern möglichst viel Information enthält und somit zwischen dem Null-Modell und dem saturierten Modell liegt. Man betrachtet eine Likelihood-Quotienten Teststatistik (LRT), wobei die Nullhypothese dem aktuellen Modell und die Alternativhypothese einem anderem als dem aktuellen Modell entspricht.

Hiefür drücken wir die Likelihood Funktion in Termen des Erwartungswertes μ anstatt des kanonischen Parameters θ aus und definieren das Maß der Güte.

Definition 2.12. (*Skalierte Deviance*)

Sei $\mathcal{L}(\hat{\mu}, y)$ die Log-Likelihood Funktion und ϕ der Dispersionsparameter. Dann ist die skalierte Deviance definiert als

$$D^*(y, \hat{\mu}) = \frac{D(y, \hat{\mu})}{\phi} = -2[\mathcal{L}(\hat{\mu}, y) - \mathcal{L}(y, y)]. \quad (2.15)$$

Da $\mathcal{L}(y, y)$ unabhängig vom Schätzer $\hat{\mu}(\beta)$ ist, minimiert der Schätzer $\hat{\mu}$, der durch die Maximierung von $\mathcal{L}(\hat{\mu}, y)$ erhalten wird, auch die Deviance. Somit ist das Ziel bei der Maximum Likelihood Schätzung die skalierte Deviance zu minimieren.

Seien $\hat{\theta} = \theta(\hat{\mu})$ und $\tilde{\theta} = \theta(y)$ die Schätzer der kanonischen Parameter unter dem aktuellen Modell mit p Parametern und dem saturierten Modell mit n Parametern. Da Y_i aus einer Verteilung aus einer einparametrischen linearen Exponentialfamilie

stammt, gilt für die Differenz der Log-Likelihood Funktionen

$$\begin{aligned} -2[\mathcal{L}(\hat{\mu}, y) - \mathcal{L}(y, y)] &= -2 \left[\sum_{i=1}^n \frac{y_i \hat{\theta}_i - b(\hat{\theta}_i)}{a_i \cdot \phi} - \sum_{i=1}^n \frac{y_i \tilde{\theta}_i - b(\tilde{\theta}_i)}{a_i \cdot \phi} \right] \\ &= \sum_{i=1}^n 2 \left[\frac{y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)}{a_i \cdot \phi} \right] \\ &= \frac{D(y, \hat{\mu})}{\phi}, \end{aligned}$$

wodurch das Herausheben von ϕ aus der Differenz ersichtlich wird.

Abschließend zu diesem Abschnitt folgt eine Tabelle der Deviance-Terme verschiedener Verteilungen (siehe Thaler, 2009, Seite 21):

Verteilung	Deviance $D(y, \hat{\mu})$
Normal	$\sum_{i=1}^n (y_i - \hat{\mu}_i)^2$
Gamma	$2 \sum_{i=1}^n [-\log(y_i/\hat{\mu}_i) + (y_i - \hat{\mu}_i)/\hat{\mu}_i]$
Poisson	$2 \sum_{i=1}^n [y_i \log(y_i/\hat{\mu}_i) + (y_i - \hat{\mu}_i)]$
Binomial	$2 \sum_{i=1}^n m_i [(1 - y_i) \log[(1 - y_i)/(1 - \hat{\mu}_i)] + y_i \log(y_i/\hat{\mu}_i)]$

Tabelle 2.1: Deviance-Terme verschiedener Verteilungen

2.3.2 Pearson-Statistik

Ein weiteres Maß für die Größenordnung der Abweichung der Response-Variable y und dem Schätzer des Erwartungswertes $\hat{\mu}$ ist die sogenannte Pearson-Statistik X^2 , die wie folgt definiert ist:

Definition 2.13. (*Generalisierte Pearson-Statistik*)

Sei die Varianz der Y_i durch $\text{var}[y_i] = a_i \cdot \phi \cdot V(\mu_i)$ mit einem fixen Dispersionsparameter ϕ , beobachtungsspezifischen Gewichten a_i und der Varianzfunktion $V(\mu_i)$ gegeben, so lautet die generalisierte Pearson-Statistik X^2 :

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

Die Pearson-Statistik kann auch als Schätzer für den Dispersionparameter ϕ verwendet werden, welcher sich als gemittelte Pearson-Statistik ergibt:

$$\hat{\phi} = \frac{1}{n-p} X^2.$$

2.3.3 Analyse der Deviance

Um die erklärenden Variablen, die nur wenig zur Erklärung des Erwartungswertes beitragen, zu eliminieren, bedient man sich der Analysis-of-Deviance (analog zur Analysis-of-Variance, ANOVA). Hierfür testet man zwei ineinandergeschachtelte Modelle für $q = 1, \dots, p$:

$$\begin{aligned} \text{Modell 1: } \eta &= \beta_1 x_1 + \dots + \beta_q x_q + \beta_{q+1} x_{q+1} + \dots + \beta_p x_p, \\ \text{Modell 2: } \eta &= \beta_{q+1} x_{q+1} + \dots + \beta_p x_p. \end{aligned}$$

Nun wird die Differenz der Deviance-Werte dieser zwei Modelle bestimmt. Sei hierfür $\hat{\mu}_1$ der Schätzer für den Erwartungswert μ unter Modell 1 und $\hat{\mu}_2$ der Schätzer für den Erwartungswert μ unter Modell 2. Dann lässt sich die Differenz der skalierten Deviance schreiben als

$$\begin{aligned} \frac{1}{\phi} [D(y, \hat{\mu}_2) - D(y, \hat{\mu}_1)] &= -2[\mathcal{L}(\hat{\mu}_2, y) - \mathcal{L}(y, y)] + 2[\mathcal{L}(\hat{\mu}_1, y) - \mathcal{L}(y, y)] \\ &= -2[\mathcal{L}(\hat{\mu}_2, y) - \mathcal{L}(\hat{\mu}_1, y)]. \end{aligned}$$

Dies entspricht einer Likelihood-Quotienten-Teststatistik und ist somit unter geeigneten Regularitätsbedingungen asymptotisch χ_q^2 verteilt, wobei q gleich der Anzahl der Freiheitsgrade ist, die bei der Reduktion der Modell-Dimension erhalten wird.

Beispiel 2.14. *Betrachtet man ein Modell mit zwei Faktoren A und B, so kann das Modell mit dem Intercept 1, dem Faktor A, dem Faktor B und der Interaktion A : B gefittet werden, welches in folgender Tabelle zusammengefasst ist (siehe auch McCullagh und Nelder, 1989, Seite 25 f.):*

Modell	ANOVA				
	d.f.	Deviance	SS	d.f.	Term
1	11	1000			
A	8	500	500	3	nur A
A + B	6	200	300	2	A bei Hinzunahme von B
A + B + A : B	0	0	200	6	A, B bei Hinzunahme von A : B

Tabelle 2.2: Beispiel einer Analysis-of-Deviance-Tabelle

Die ersten drei Spalten von Tabelle 2.2 beschreiben die Sequenz der Modelle, die Freiheitsgrade (d.f.) sowie die Deviance. Zu erwähnen ist hier, dass die Deviance des ersten Modells mit nur dem Intercept 1 der totalen Quadratsumme entspricht. Das letzte Modell mit beiden Parametern und der Interaktion, in diesem Fall das saturierte Modell, enthält so viele Parameter wie Beobachtungen und hat somit keine Freiheitsgrade und eine Deviance von Null.

Die letzten drei Spalten von Tabelle 2.2 zeigen die Analyse der Varianz mit der Quadratsumme (SS), die als Differenz der Deviance-Werte erhalten wird. Diese Differenz ergibt sich aus den beiden Modellen, wobei beim zweiten Modell weitere Terme hinzugefügt wurden und die Zeilen unterhalb nicht beachtet werden. Von Zeile zu Zeile kann man so den Unterschied der Deviance durch Hinzunahme weitere Terme erkennen. Hierbei ist jedoch die „Sequenz“ der hinzugefügten Terme problematisch, da sich ein anderer Deviance-Wert ergibt, wenn man beispielsweise den Faktor B zuerst in das Modell aufnimmt. Das Modell, das die Daten bestmöglich fittet, kann auf diese Weise herausgefunden werden.

3 Quasi-Likelihood Funktion

Auch für dieses Kapitel ist eine der Grundlagen das Buch von McCullagh und Nelder (1989), Kapitel 9, sowie Wedderburn (1974) und Thaler (2009), Kapitel 3.

3.1 Definition

In diesem Abschnitt wollen wir die Auswirkungen auf das GLM betrachten, wenn wir die Modellannahmen schwächen. Sei die Zufallsvariable Y_i aus einer Verteilung der einparametrischen linearen Exponentialfamilie mit $\mathbb{E}[Y_i] = \mu_i$, $\mathcal{L}(\theta, y)$ die Log-Likelihood Funktion mit dem kanonischen Parameter θ und ϕ der Dispersionsparameter. Betrachten wir die Score Funktion

$$\frac{\partial \mathcal{L}(\theta, y)}{\partial \mu_i} = \frac{\partial}{\partial \theta_i} \left\{ \frac{y_i \theta_i - b(\theta_i)}{a_i \cdot \phi} + c(y_i, \phi) \right\} \frac{\partial \theta_i}{\partial \mu_i} = \frac{y_i - \mu_i}{a_i \cdot \phi \cdot V(\mu_i)} \quad (3.1)$$

so ist zu erkennen, dass sie nur von der Varianzfunktion $V(\mu_i)$, die den Zusammenhang zwischen dem Erwartungswert und der Varianz beschreibt, abhängt. Da es oft schwierig ist, die gesamte Verteilung der zu Grunde liegenden Zufallsvariablen zu spezifizieren, gehen wir nun nur von der Kenntnis der Erwartungswert-Varianz-Beziehung $V(\mu)$ aus, die nicht zu einer Verteilung aus der Exponentialfamilie korrespondieren muss.

Bemerkung 3.1. *Das beobachtungsspezifische Gewicht sei nun $a_i = 1$.*

Bemerkung 3.2. *Des Weiteren nehmen wir an, dass für die Gewichte $a_i = 1$ für alle $i = 1, \dots, n$ gilt.*

Definieren wir hierfür nun die sogenannte *Log-Quasi-Likelihood Funktion*.

Definition 3.3. (*Log-Quasi-Likelihood Funktion*)

Sei Y eine Zufallsvariable mit Realisierung y . Seien $\mathbb{E}[Y] = \mu$ und $\text{var}[Y] = \phi V(\mu)$ der Erwartungswert und die Varianz dieser Zufallsvariablen mit bekannter Varianzfunktion $V(\mu)$. Dann ist die Log-Quasi-Likelihood Funktion definiert als

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

oder äquivalent

$$\frac{\partial Q(\mu, y)}{\partial \mu} = \frac{y - \mu}{\phi V(\mu)}. \quad (3.3)$$

Bemerkung 3.4. Die Ableitung der Log-Quasi-Likelihood Funktion (3.3) wird Quasi-Score Funktion genannt.

Bemerkung 3.5. Zur Vereinfachung werden wir die Log-Quasi-Likelihood Funktion als Quasi-Likelihood Funktion bzw. QL Funktion bezeichnen.

McCullagh und Nelder (1989) definieren die QL Funktion im Unterschied zu (3.2) mit expliziter Angabe einer unteren Schranke als

$$Q(\mu, y) = \int_y^{\mu} \frac{y - t}{\phi V(t)} dt. \quad (3.4)$$

Da jedoch auch in (3.2) eine Funktion in y berücksichtigt wird und wir nach Auflösen des Integrals (3.4) ebenso eine Funktion in y erhalten, sind beide QL Funktionen identisch.

Wegen der Unabhängigkeit der Komponenten von Y ist die QL Funktion für alle Daten $i = 1, \dots, n$ die Summe der einzelnen QL Funktionen

$$Q(\mu, y) = \sum_{i=1}^n Q(\mu_i, y_i).$$

Wie auch im vorigen Abschnitt ist das Ziel, einen Schätzer für β durch die Beziehung zwischen dem Erwartungswert und den erklärenden Variablen, $g(\mu) = x'\beta$, zu bestimmen, der die QL Funktion (3.2) maximiert. Vergleicht man nun die Score Funktion (3.1) mit der Quasi-Score Funktion (3.3), so erkennt man dieselbe Struktur. Somit erhalten wir den Schätzer $\hat{\beta}_{QL}$ mit den gleichen Methoden, wie der Maximum Likelihood-Schätzer $\hat{\beta}_L$.

3.2 Eigenschaften der QL Funktion

Ähnlich zu Satz 2.7 über die Eigenschaften des Score der Log-Likelihood Funktion lassen sich diese Eigenschaften auch auf die Quasi-Score Funktion übertragen und in folgendem Satz zusammenfassen.

Satz 3.6. (*Eigenschaften der Quasi-Likelihood Funktion, Wedderburn, 1974*)
 Sei $Q(\mu, y)$ die Quasi-Likelihood Funktion einer Zufallsvariable Y mit Erwartungswert $\mathbb{E}[Y] = \mu$ und Varianz $\text{var}[Y] = \phi V(\mu)$. Unter gewissen Regularitätsbedingungen gelten für die Quasi-Likelihood Funktion folgende Eigenschaften:

$$(a) \quad \mathbb{E} \left[\frac{\partial Q(\mu, Y)}{\partial \mu} \right] = 0, \quad (3.5)$$

$$(b) \quad \mathbb{E} \left[\frac{\partial Q(\mu, Y)}{\partial \beta_i} \right] = 0, \quad (3.6)$$

$$(c) \quad \mathbb{E} \left[\frac{\partial Q(\mu, Y)}{\partial \mu} \right]^2 = -\mathbb{E} \left[\frac{\partial^2 Q(\mu, Y)}{\partial \mu^2} \right] = \frac{1}{\phi V(\mu)}, \quad (3.7)$$

$$(d) \quad \mathbb{E} \left[\frac{\partial Q(\mu, Y)}{\partial \beta_i} \frac{\partial Q(\mu, Y)}{\partial \beta_j} \right] = -\mathbb{E} \left[\frac{\partial^2 Q(\mu, Y)}{\partial \beta_i \partial \beta_j} \right] = \frac{1}{\phi V(\mu)} \frac{\partial \mu}{\partial \beta_i} \frac{\partial \mu}{\partial \beta_j}. \quad (3.8)$$

Beweis.

$$(a) \quad \mathbb{E} \left[\frac{\partial Q(\mu, Y)}{\partial \mu} \right] = \mathbb{E} \left[\frac{Y - \mu}{\phi V(\mu)} \right] = \frac{\mu - \mu}{\phi V(\mu)} = 0.$$

(b) folgt aus der Beziehung

$$\mathbb{E} \left[\frac{\partial Q(\mu, Y)}{\partial \beta_i} \right] = \mathbb{E} \left[\underbrace{\frac{\partial Q(\mu, Y)}{\partial \mu}}_{\mathbb{E}[\cdot] \stackrel{(a)}{=} 0} \frac{\partial \mu}{\partial \beta_i} \right] = 0.$$

(c) Für den ersten Teil gilt mit $\mathbb{E}[(Y - \mu)^2] = \text{var}[Y] = \phi V(\mu)$:

$$\mathbb{E} \left[\frac{\partial Q(\mu, Y)}{\partial \mu} \right]^2 = \mathbb{E} \left[\left(\frac{Y - \mu}{\phi V(\mu)} \right)^2 \right] = \frac{\phi V(\mu)}{\phi^2 V(\mu)^2} = \frac{1}{\phi V(\mu)}.$$

Für den zweiten Teil gilt:

$$\begin{aligned}
 -\mathbb{E} \left[\frac{\partial^2 Q(\mu, Y)}{\partial \mu^2} \right] &= -\mathbb{E} \left[\frac{\partial}{\partial \mu} \frac{Y - \mu}{\phi V(\mu)} \right] \\
 &= -\mathbb{E} \left[\frac{-\phi V(\mu) - (Y - \mu)\phi V'(\mu)}{\phi^2 V(\mu)^2} \right] \\
 &= \frac{1}{\phi V(\mu)} + \underbrace{\mathbb{E} \left[\frac{(Y - \mu)V'(\mu)}{\phi V(\mu)^2} \right]}_{=0} = \frac{1}{\phi V(\mu)}.
 \end{aligned}$$

(d) Da (c) ein Spezialfall von (d) ist, kann dieser nun hier verwendet werden. Für den ersten Teil gilt:

$$\begin{aligned}
 \mathbb{E} \left[\frac{\partial Q(\mu, Y)}{\partial \beta_i} \frac{\partial Q(\mu, Y)}{\partial \beta_j} \right] &= \mathbb{E} \left[\frac{\partial Q(\mu, Y)}{\partial \mu} \right]^2 \frac{\partial \mu}{\partial \beta_i} \frac{\partial \mu}{\partial \beta_j} \\
 &\stackrel{(c)}{=} \frac{1}{\phi V(\mu)} \frac{\partial \mu}{\partial \beta_i} \frac{\partial \mu}{\partial \beta_j}
 \end{aligned}$$

Der zweite Teil von (d) ergibt sich zu

$$\begin{aligned}
 -\mathbb{E} \left[\frac{\partial^2 Q(\mu, Y)}{\partial \beta_i \partial \beta_j} \right] &= -\mathbb{E} \left[\frac{\partial \mu}{\partial \beta_i} \frac{\partial}{\partial \mu} \left(\frac{Y - \mu}{\phi V(\mu)} \frac{\partial \mu}{\partial \beta_j} \right) \right] \\
 &\stackrel{(c)}{=} \frac{1}{\phi V(\mu)} \frac{\partial \mu}{\partial \beta_i} \frac{\partial \mu}{\partial \beta_j}.
 \end{aligned}$$

□

Korollar 3.7. (Wedderburn, 1974) Wenn die Log-Likelihood Funktion \mathcal{L} von y in Abhängigkeit vom Erwartungswert μ spezifiziert werden kann, so gilt:

$$-\mathbb{E} \left[\frac{\partial^2 Q(\mu, Y)}{\partial \mu^2} \right] \leq -\mathbb{E} \left[\frac{\partial^2 \mathcal{L}(\mu, Y)}{\partial \mu^2} \right]. \tag{3.9}$$

Beweis. Aus Satz 3.6 folgt, dass (3.9) äquivalent zu folgender Ungleichung ist:

$$\text{var} \left[\frac{\partial Q(\mu, Y)}{\partial \mu} \right] \geq \frac{1}{\mathbb{E} \left[\frac{\partial^2 \mathcal{L}(\mu, Y)}{\partial \mu^2} \right]}. \tag{3.10}$$

Ausdruck (3.10) ergibt sich direkt aus der sogenannten *Cramér-Rao Ungleichung*.¹

□

¹Die Cramér-Rao Ungleichung legt eine untere Schranke für die Varianz fest, siehe auch Merberg und Miller (2008).

Bemerkung 3.8. Ungleichung (3.9) zeigt somit eine untere Schranke für die Varianz der Score Funktion (Fisher-Information).

Somit können wir zusammenfassend feststellen, dass bezüglich der Score Funktion mit der Log-Likelihood Funktion und der Log-Quasi-Likelihood Funktion die selben Resultate erzielt werden können. Der Vorteil der QL Funktion ist jedoch, dass die Bestimmung der Beziehung zwischen dem Erwartungswert und der Varianz genügt, um Schätzer für die Parameter im GLM zu erhalten. Daher muss nicht die gesamte Verteilung, wie sie zur Evaluierung der Log-Likelihood Funktion verlangt wird, spezifiziert werden und wir müssen schwächere Annahmen treffen. Es genügt stattdessen die Auswahl einer Varianz- und Linkfunktion, um die Koeffizienten eines GLMs zu schätzen.

Aufgrund der selben Eigenschaften für die Score Funktionen (3.1) und (3.3) wollen wir nun herausfinden, wann die Log-Quasi-Likelihood Funktion mit der Log-Likelihood Funktion übereinstimmt. Die Antwort hierzu liefert folgender Satz von Wedderburn (1974), der besagt, dass eine Äquivalenz nur für Verteilungen aus einer einparametrischen Exponentialfamilie gilt.

Satz 3.9. Sei $\mathcal{L}(\mu, y)$ die Log-Likelihood Funktion einer Zufallsvariablen Y mit Erwartungswert $\mathbb{E}[Y] = \mu$ und Varianz $\text{var}[Y] = \phi V(\mu)$. Dann hat die Log-Likelihood Funktion die Eigenschaft

$$\frac{\partial \mathcal{L}(\mu, y)}{\partial \mu} = \frac{y - \mu}{\phi V(\mu)} \tag{3.11}$$

dann und nur dann, wenn die Verteilung der Zufallsvariablen Y aus einer einparametrischen Exponentialfamilie stammt, d.h. wenn sie eine Dichtefunktion der folgenden Form aufweist:

$$f_Y(y, \mu) = \exp \left\{ \frac{y\theta - b(\theta)}{a \cdot \phi} + c(y, \phi) \right\}.$$

Beweis. “ \Rightarrow “: Mit Gleichung (3.11) und Integration über μ ergibt sich:

$$\begin{aligned} \mathcal{L}(\mu, y) &= \int \frac{\partial \mathcal{L}(\mu, y)}{\partial \mu} d\mu \stackrel{(3.11)}{=} \int \frac{y - \mu}{\phi V(\mu)} d\mu \\ &= \frac{y}{\phi} \underbrace{\int \frac{1}{V(\mu)} d\mu}_{:=\theta} - \frac{1}{\phi} \underbrace{\int \frac{\mu}{V(\mu)} d\mu}_{:=b(\theta)} \\ &= \frac{y\theta - b(\theta)}{\phi} + c(y, \phi). \end{aligned}$$

Somit erhalten wir für die Dichtefunktion

$$f_Y(y, \mu) = \exp \left\{ \frac{y\theta - b(\theta)}{a \cdot \phi} + c(y, \phi) \right\}.$$

“ \Leftarrow “: Für eine Verteilung aus einer einparametrischen Exponentialfamilie gilt $\mathbb{E}[Y] = \mu = b'(\theta)$ und $\text{var}[Y] = \phi V(\mu) = \phi b''(\theta)$, wodurch sich folgende Identität ergibt:

$$\frac{\partial \mu}{\partial \theta} = \frac{\partial b'(\theta)}{\partial \theta} = b''(\theta) = V(\mu). \quad (3.12)$$

Mit (3.12) und unter Anwendung der Kettenregel erhalten wir für die Score Funktion

$$\begin{aligned} \frac{\partial \mathcal{L}(\mu, y)}{\partial \mu} &= \frac{\partial \theta}{\partial \mu} \frac{\partial}{\partial \theta} \left(\frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \right) \\ &= \frac{y - b'(\theta)}{\phi V(\mu)} = \frac{y - \mu}{\phi V(\mu)}. \end{aligned}$$

□

Bemerkung 3.10. Aus Satz 3.9 folgt, dass Ungleichung (3.9) für eine Verteilung aus der einparametrischen Exponentialfamilie zu einer Gleichung wird.

Beispiel 3.11. Betrachten wir nun in folgender Aufzählung Varianzfunktionen einiger Verteilungen mit dem dazugehörigen kanonischen Parameter $\theta = \int \frac{1}{V(\mu)} d\mu$ sowie der Quasi-Likelihood Funktion $Q(\mu, y) = \int^\mu \frac{y-t}{\phi V(t)} dt + \text{Funktion in } y \text{ und } \phi$.

1. $V(\mu) = 1, y, \mu \in \mathbb{R}$ (wie beispielsweise bei $Y \sim N(\mu, \sigma^2)$):

$$\theta = \mu, \quad Q(\mu, y) = -\frac{(y - \mu)^2}{2\sigma^2}$$

2. $V(\mu) = \mu, \mu > 0, y \geq 0$ (wie beispielsweise bei $Y \sim P(\mu)$):

$$\theta = \log \mu, \quad Q(\mu, y) = y \log \mu - \mu$$

3. $V(\mu) = \mu^2, \mu > 0, y \geq 0$ (wie beispielsweise bei $Y \sim G(\mu, 1)$):

$$\theta = -\frac{1}{\mu}, \quad Q(\mu, y) = -\log \mu - \frac{y}{\mu}$$

4. $V(\mu) = \mu^3, \mu > 0, y \geq 0$ (wie beispielsweise bei $Y \sim IG(\mu, 1)$):

$$\theta = -\frac{1}{2\mu^2}, \quad Q(\mu, y) = -\frac{y}{2\mu^2} + \frac{1}{\mu}$$

5. $V(\mu) = \mu^k, \mu > 0, y \geq 0, k \geq 3$:

$$\theta = -\frac{1}{(k-1)\mu^{k-1}}, \quad Q(\mu, y) = \frac{y\mu^{1-k}}{1-k} - \frac{\mu^{2-k}}{2-k}$$

6. $V(\mu) = \mu(1-\mu), 0 < \mu < 1, 0 \leq y \leq 1$ (wie beispielsweise bei $Y \sim B(1, \mu)$):

$$\theta = \log \frac{\mu}{1-\mu}, \quad Q(\mu, y) = y \log \frac{\mu}{1-\mu} + \log(1-\mu)$$

7. $V(\mu) = \mu^2(1-\mu)^2, 0 < \mu < 1, 0 \leq y \leq 1$:

$$\theta = 2 \log \frac{\mu}{1-\mu} - \frac{1}{\mu} + \frac{1}{1-\mu}, \quad Q(\mu, y) = (2y-1) \log \frac{\mu}{1-\mu} - \frac{y}{\mu} - \frac{1-y}{1-\mu}$$

8. $V(\mu) = \mu + \frac{\mu^2}{k}, \mu > 0, y \geq 0, k > 0$ (wie beispielsweise bei $Y \sim \text{NegB}(k, \mu)$):

$$\theta = \log \frac{\mu}{k+\mu}, \quad Q(\mu, y) = y \log \frac{\mu}{k+\mu} + k \log \frac{1}{k+\mu}$$

Bemerkung 3.12. Die Verteilung der fünften Varianzfunktion sowie die Negative Binomialverteilung sind nur für bekannte Parameter k Mitglieder der Exponentialfamilie. Für unbekanntes k ist eine Schätzung des selbigen mit dem QL-Ansatz nicht möglich.

3.2.1 Güte der Modellanpassung

Betrachten wir nun wie im vorigen Kapitel ein Maß für die Güte der Modellanpassung. Analog zur Deviance (2.15) kann auch eine sogenannte *Quasi-Deviance* definiert werden.

Definition 3.13. (*Quasi-Deviance*)

Seien der Erwartungswert und die Varianz der n unabhängigen Zufallsvariablen Y_i für $i = 1, \dots, n$ gegeben durch $\mathbb{E}[Y_i] = \mu_i$ und $\text{var}[Y_i] = \phi V(\mu_i)$ mit der Quasi-Likelihood Funktion $Q(\mu_i, y_i)$. Dann ist die Quasi-Deviance definiert als

$$D(y, \mu) = -2\phi[Q(\mu, y) - Q(y, y)] = -2 \sum_{i=1}^n \int_{y_i}^{\mu_i} \frac{y_i - t}{V(t)} dt. \quad (3.13)$$

Bemerkung 3.14. Verwendet man für die QL Funktion $Q(\mu, y)$ Definition (3.4), so ergibt sich für die Quasi-Deviance

$$D(y, \mu) = -2\phi \cdot Q(\mu, y). \quad (3.14)$$

Ebenso ist auch mit dem QL-Ansatz das Ziel, die Quasi-Deviance zu minimieren. Dieser Schätzer $\hat{\beta}_{QL}$ wird, analog zur skalierten Deviance (2.15), durch die Maximierung der QL Funktion erhalten. Des Weiteren ist auch die skalierte Quasi-Deviance durch Konstruktion eines Likelihood-Quotienten-Test asymptotisch χ_q^2 -verteilt mit q als der Anzahl der Freiheitsgrade gleich der Reduktion der Dimensionen der Parameter-Räume. Dies bedeutet, wenn unter der Nullhypothese H_0 ein Modell mit p Parametern angenommen wird und unter der Alternativhypothese H_1 ein Modell mit n Parametern, so ist die Differenz der Deviance-Werte

$$D(\hat{\mu}_1, \hat{\mu}_0) = D(y, \hat{\mu}_0) - D(y, \hat{\mu}_1)$$

asymptotisch χ_q^2 -verteilt mit $q = n - p$ gleich der Anzahl der Freiheitsgrade.

3.3 Quasi-Dichte

Satz 3.9 besagt, dass die Log-Likelihood Funktion und die Quasi-Likelihood Funktion nur für Varianzfunktionen, deren Verteilung aus der einparametrischen Exponentialfamilie stammen, übereinstimmen. Ist dies nicht der Fall, wäre es somit nicht möglich, eine Dichtefunktion anzugeben. Es ist jedoch möglich durch Normalisieren der QL Funktion eine sogenannte *Quasi-Dichte* zu konstruieren.

Definition 3.15. (*Quasi-Dichte*)

Sei $Q(\mu, y)$ eine Log-Quasi-Likelihood Funktion und $\omega(\mu) = \int \exp\{Q(\mu, y)\} dy$ eine Normalisierungsfunktion. Dann ist die Quasi-Dichtefunktion definiert als

$$f_Q(y, \mu) = \frac{\exp\{Q(\mu, y)\}}{\omega(\mu)}. \quad (3.15)$$

Bemerkung 3.16. Gilt $\omega(\mu) \neq 1$, so ist eine Varianzfunktion angenommen, deren Verteilung nicht aus einer Exponentialfamilie stammt. Ist umgekehrt $\omega(\mu) = 1$, so geht man von einer Varianzfunktion aus, deren Verteilung zur Exponentialfamilie gehört.²

Die entsprechende Log-Likelihood Funktion zur Quasi-Dichte (3.15) lässt sich schreiben als

$$\mathcal{L}(\mu, y) = \log f_Q(y, \mu) = Q(\mu, y) - \log \omega(\mu).$$

²Diese Aussagen leiten sich aus Satz 3.9 ab.

Somit unterscheidet sich die Quasi-Score Funktion von der Score Funktion unter der Annahme der Vertauschbarkeit der Integration und Differentiation durch:

$$\begin{aligned}
 \frac{\partial Q(\mu, y)}{\partial \mu} - \frac{\partial \mathcal{L}_Q(\mu, y)}{\partial \mu} &= \frac{\partial \log \omega(\mu)}{\partial \mu} = \frac{1}{\omega(\mu)} \frac{\partial \omega(\mu)}{\partial \mu} \\
 &= \frac{1}{\omega(\mu)} \frac{\partial}{\partial \mu} \int \exp \{Q(\mu, y)\} dy \\
 &= \int \frac{1}{\omega(\mu)} \frac{\partial}{\partial \mu} \exp \{Q(\mu, y)\} dy \\
 &= \int \frac{y - \mu}{\phi V(\mu)} \underbrace{\frac{\exp \{Q(\mu, y)\}}{\omega(\mu)}}_{=f_Q(y, \mu)} dy \\
 &= \mathbb{E} \left[\frac{Y - \mu}{\phi V(\mu)} \right] = \frac{\mu_Q - \mu}{\phi V(\mu)}.
 \end{aligned}$$

Hierbei bezeichnet

$$\mathbb{E}_Q[Y] = \mu_Q = \int y \cdot f_Q(y, \mu) dy$$

den *Quasi-Mean*. Wenn die Differenz zwischen dem Quasi-Mean und dem Erwartungswert $\mu_Q - \mu$ verglichen mit der Differenz $y - \mu$ klein ist, so kann die Aussage getroffen werden, dass der Maximum-Likelihood-Schätzer bezüglich der Quasi-Dichte nahe beim Maximum-Likelihood-Schätzer liegt.

4 Extended-Quasi-Likelihood Funktion

Die Basis zu diesem Kapitel liefern Nelder und Pregibon (1987), Thaler (2009) und Jørgensen (1997), Kapitel 4.

4.1 Definition

Durch den QL-Ansatz ist es nun möglich die Koeffizienten β im linearen Prädiktor zu schätzen, wenn man nur die Link- und Varianzfunktion kennt. Dies ist auch möglich für Varianzfunktionen, deren korrespondierende Verteilung nicht aus einer Exponentialfamilie stammen. Doch auch diese Methode weist noch „Schwachstellen“ auf: Die Varianzfunktion muss bis auf eine multiplikative Konstante genau angegeben werden. Ist dies nicht der Fall, d.h. stammt die Varianzfunktion aus einer Familie von Funktionen \mathcal{F}_k mit einem unbekanntem Parameter k , so ist man bis jetzt nicht in der Lage die Parameter β und k simultan zu schätzen. Anders gesagt ist es also nicht möglich, verschiedene Varianzfunktionen miteinander zu vergleichen. Ebenso können wir mit den bisher diskutierten Verfahren keine nicht konstanten, unbekanntem Dispersionsparameter schätzen.

Aus diesen Gründen bedarf es einer Erweiterung des bisherigen QL-Ansatz. Um auch verschiedene Varianzfunktionen miteinander vergleichen zu können, definieren wir die sogenannte *Extended-Quasi-Likelihood Funktion* (kurz: EQL Funktion). Diese beruht auch wie die QL Funktion nur auf der Annahme der ersten beiden Momente ohne die gesamte Verteilung kennen zu müssen.

Definition 4.1. (*Extended-Quasi-Likelihood Funktion*)

Sei der Erwartungswert der Zufallsvariablen Y gegeben als $\mathbb{E}[Y] = \mu$ und die Varianz $\text{var}[Y] = \phi V(\mu)$, wobei $V(\mu)$ eine bekannte Varianzfunktion und ϕ den Dispersionsparameter darstellt. Mit der Quasi-Likelihood Funktion

$$Q(\mu, y) = \int^{\mu} (y - t)/(\phi V(\mu)) dt$$

und der Quasi-Deviance $D(y, \mu) = -2\phi [Q(\mu, y) - Q(y, y)]$ ist die Extended-Quasi-Likelihood Funktion definiert als

$$Q^+(\mu, \phi, y) = -\frac{1}{2} \log \{2\pi\phi V(y)\} - \frac{1}{2} \frac{D(y, \mu)}{\phi}. \quad (4.1)$$

Da die Zufallsvariablen Y_i unabhängig sind, ist die EQL Funktion für die gesamte Stichprobe $i = 1, \dots, n$ die Summe der EQL Funktionen der einzelnen Beobachtungen

$$Q^+(\mu, \phi, y) = \sum_{i=1}^n Q^+(\mu_i, \phi, y_i).$$

Betrachten wir die EQL Funktion $Q^+(\mu, \phi, y)$ in Definition 4.1, so ist zu erkennen, dass der erste Term unabhängig von μ und somit auch von β ist. Daher entspricht eine Maximierung von (4.1) einer Minimierung der Quasi-Deviance (3.13). Aus diesem Grund ist auch der Schätzer $\hat{\beta}$, der durch die Maximierung der EQL Funktion (4.1) erhalten wird, auch der Schätzer, der sich durch Maximierung der QL Funktion (3.2) ergibt.

4.1.1 Schätzer des Dispersionsparameter

Um einen Schätzer für den Dispersionsparameter ϕ zu erhalten, muss die EQL Funktion $Q^+(\mu, \phi, y)$ nach ϕ abgeleitet und Null gesetzt werden:

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

Daraus folgt

$$\hat{\phi} = \frac{D(y, \mu)}{n}.$$

Da der Parameter μ jedoch nicht bekannt ist, verwendet man den Schätzer $\hat{\mu}$. Somit ergibt sich für den Schätzer des Dispersionsparameter

$$\hat{\phi} = \frac{D(y, \hat{\mu})}{n},$$

die mittlere Deviance.

4.2 Eigenschaften der EQL Funktion

Im vorigen Abschnitt über die QL Funktion wurde mit Satz 3.9 festgestellt, dass die Score- und Quasi-Score Funktion äquivalent sind, sofern die Verteilung aus der einparametrischen Exponentialfamilie stammt. Für EQL Funktionen ist dies aber nur für gewisse Verteilungen der Fall. Diese werden nun, angelehnt an die Ergebnisse von Nelder und Pregibon (1987), im Folgenden angeführt:

- Für die *Normalverteilung* und *Invers-Gauß-Verteilung* entspricht die EQL Funktion der Log-Likelihood Funktion.
- Für die *Gammaverteilung* unterscheidet sich die EQL Funktion von der Log-Likelihood Funktion um einen Faktor, der nur von ϕ abhängt.
- Für die diskreten Poisson-, Binomial- und negative Binomialverteilung muss man $k!$ durch die Stirling Approximation

$$k! \approx \sqrt{2\pi k} k^k e^{-k}$$

ersetzen, um von der Log-Likelihood Funktion auf die EQL Funktion zu gelangen.

Mit der folgenden Herleitung einer approximierten EQL Funktion erhalten wir die Score Funktionen für den Erwartungswert μ und den Dispersionsparameter ϕ . Für einen bekannten Dispersionsparameter ϕ ist die EQL Funktion $Q^+(\mu, \phi, y)$ im Grunde die gleiche, wie die QL Funktion $Q(\mu, y)$. Der Unterschied liegt jedoch darin, dass die EQL Funktion nach ϕ abgeleitet werden kann. Um die EQL Funktion nun daraus herzuleiten, schreiben wir (siehe McCullagh und Nelder, 1989, Seiten 349 f.)

$$Q^+(\mu, \phi, y) = Q(\mu, y) + h(\phi, y) \\ \stackrel{(3.14)}{=} -\frac{1}{2\phi} D(y, \mu) + h(\phi, y),$$

wobei $h(\phi, y)$ eine Funktion der Form

$$h(\phi, y) = -\frac{1}{2} h_1(\phi) - h_2(y).$$

Damit sich Q^+ wie eine Log-Likelihood Funktion verhält, muss für den Erwartungswert der ersten Ableitung nach ϕ

$$\mathbb{E} \left[\frac{\partial Q^+(\mu, \phi, Y)}{\partial \phi} \right] = 0 \tag{4.2}$$

gelten. Aus (4.2) ergibt sich

$$\frac{1}{2\phi^2}\mathbb{E}[D(Y, \mu)] - \frac{1}{2}h_1'(\phi) = 0.$$

Dies impliziert wiederum

$$\mathbb{E}[D(Y, \mu)] = \phi^2 h_1'(\phi).$$

Durch eine Approximation der ersten Ordnung für den Erwartungswert der Deviance Funktion $\mathbb{E}[D(Y, \mu)] \approx \phi$ erhalten wir

$$h_1(\phi) = \log(\phi) + \text{const}$$

und daraus folgende Approximation für die EQL Funktion

$$Q^+(\mu, \phi, y) \approx -\frac{1}{2\phi}D(y, \mu) - \frac{1}{2}\log(\phi). \quad (4.3)$$

Aus (4.3) können wir somit die Score Funktion für μ und ϕ bilden:

$$\frac{\partial Q^+(\mu, \phi, y)}{\partial \mu} = \frac{y - \mu}{\phi V(\mu)}, \quad (4.4)$$

$$\frac{\partial Q^+(\mu, \phi, y)}{\partial \phi} = \frac{1}{2\phi^2}D(y, \mu) - \frac{1}{2\phi}. \quad (4.5)$$

Bemerkung 4.2. Die Score Funktion für μ (4.4) entspricht der bereits bekannten Quasi-Score Funktion (3.3).

Bemerkung 4.3. Die Score Funktion für ϕ (4.5) gilt nur für hinreichend kleine ϕ , damit sich die Approximation $\mathbb{E}[D(Y, \mu)] \approx \phi$ rechtfertigen lässt.

Für den Erwartungswert beider ScoreFunktionen gilt:

$$\mathbb{E}\left[\frac{\partial Q^+(\mu, \phi, Y)}{\partial \mu}\right] = \mathbb{E}\left[\frac{\partial Q^+(\mu, \phi, Y)}{\partial \phi}\right] = 0.$$

Die Varianz von $\frac{\partial Q^+(\mu, \phi, Y)}{\partial \mu}$ ergibt sich analog zu b) aus Satz 3.6:

$$\text{var}\left[\frac{\partial Q^+(\mu, \phi, Y)}{\partial \mu}\right] = \frac{1}{\phi V(\mu)}.$$

Um die Varianz von $\frac{\partial Q^+(\mu, \phi, Y)}{\partial \phi}$ zu erhalten, benötigen wir eine Approximation für die Varianz der Deviance Funktion (siehe auch Thaler, 2000, Seiten 52 f.):

$$\text{var}[D(Y, \mu)] \approx 2\phi^2.$$

Mit dieser Abschätzung resultiert die Varianz der Score Funktion für ϕ zu

$$\text{var} \left[\frac{\partial Q^+(\mu, \phi, Y)}{\partial \phi} \right] \approx \frac{2\phi^2}{4\phi^4} = \frac{1}{2\phi^2}.$$

Berechnen wir die zweiten Ableitungen der EQL Funktion, so erhalten wir:

$$\begin{aligned} \frac{\partial^2 Q^+(\mu, \phi, y)}{\partial \mu^2} &= \frac{-\phi V(\mu) - (y - \mu)\phi V'(\mu)}{\phi^2 V(\mu)^2}, \\ \frac{\partial^2 Q^+(\mu, \phi, y)}{\partial \phi^2} &= -\frac{1}{\phi^3} D(y, \mu) + \frac{1}{2\phi^2}, \\ \frac{\partial^2 Q^+(\mu, \phi, y)}{\partial \mu \partial \phi} &= \frac{\partial^2 Q^+(\mu, \phi, y)}{\partial \phi \partial \mu} = -\frac{y - \mu}{\phi^2 V(\mu)}. \end{aligned}$$

Bilden wir nun den negativen Erwartungswert über diese zweiten Ableitungen, so können wir das Ergebnis in einer Matrix zusammenfassen:

$$\begin{pmatrix} \frac{1}{\phi V(\mu)} & 0 \\ 0 & \frac{1}{2\phi^2} \end{pmatrix}. \quad (4.6)$$

Bemerkung 4.4. Die Diagonalelemente dieser Matrix entsprechen genau den Varianzen der Score Funktionen (4.4) und (4.5).

Bemerkung 4.5. Matrix (4.6) wird analog zur Fisher-Information als Quasi-Fisher-Information bezeichnet.

Somit können wir zusammenfassend festhalten, dass für einen hinreichend kleinen Dispersionsparameter ϕ die EQL Funktion dieselben Eigenschaften bezüglich der Parameter μ und ϕ besitzt wie die QL Funktion. Hinzukommt, dass die Matrix der Quasi-Fisher-Information des Erwartungswertes μ und des Dispersionsparameters ϕ in (4.6) diagonal ist, wodurch sich eine approximative Unabhängigkeit der Schätzer feststellen lässt.

4.3 Unbekannte Parameter in der Varianzfunktion

Bei der QL Funktion geht man bis auf einen multiplen Faktor von einer bekannten Varianzfunktion aus. Diese spezielle Struktur wird bei der EQL Funktion abgeschwächt, sodass neben μ noch ein oder mehrere unbekannte Parameter vorliegen

können. Sei der unbekannte Parameter mit k bezeichnet, so wird für die Varianz angenommen

$$\text{var}[Y] = \phi V_k(\mu)$$

mit der Varianzfunktion abhängig vom unbekanntem Parameter k aus einer parametrisierten Familie \mathcal{F}_k . Mit der Deviance Funktion

$$D_k(y, \mu) = -2 \int_y^\mu \frac{y-t}{V_k(\mu)} dt,$$

von der Varianzfunktion $V_k(\mu)$ und somit von k selbst abhängig, ergibt sich die EQL Funktion zu

$$Q_k^+(\mu, \phi, y) = -\frac{1}{2} \sum_{i=1}^n \left\{ \log(2\pi\phi V_k(y_i)) - \frac{D_k(y_i, \mu_i)}{\phi} \right\}. \quad (4.7)$$

Ziel ist es nun, einen Parameter-Schätzer $(\hat{\beta}, \hat{k})$ zu bestimmen, der die EQL Funktion (4.7) maximiert.

In Nelder und Pregibon (1987) findet sich eine Methode zur Konstruktion dieses Parameter-Schätzers. Hierfür wird ein sogenanntes *Profil-Quasi-Likelihood-Intervall* I verwendet, das auch höherdimensional sein kann. Man maximiert die EQL Funktion (4.7) in β und ϕ für einen festen Wert von k

$$Q_{max}^+ = \max_{k \in I} Q_k^+,$$

woraus man Schätzer $\hat{\beta}(k)$ und $\hat{\phi}(k)$ erhält. Für diese Schätzer lassen wir schließlich k im selbst gewählten Intervall I variieren um den Maximalwert zu erhalten. Das Intervall, worüber die Differenz

$$Q_{PL}(k) = Q_{max}^+ - Q_k^+$$

kleiner ist als eine vordefinierte Schranke d wird als *Profil-Quasi-Likelihood-Intervall* bezeichnet. Eine mögliche Wahl für d ist beispielsweise das $\frac{1}{2}\chi_\nu^2$ -Quantil, wobei ν der Dimension von k entspricht.

Nun wollen wir zwei spezielle Familien von Varianzfunktionen diskutieren und deren QL Funktionen $Q_k(\mu, y)$ berechnen, welche in die Deviance Funktion $D_k(y, \mu)$ und somit in die EQL Funktion $Q_k^+(\mu, \phi, y)$ einfließen.

4.3.1 Tweediemodell

Unter Verwendung einer Varianzfunktion mit einem Potenz-Ansatz

$$V_k(\mu) = \mu^k, \quad \forall k \in \mathbb{R}^+ \quad (4.8)$$

erhalten wir die Familie der *Tweedieverteilungen*. Diese stammen aus der Klasse der *exponentiellen Dispersionsmodelle*, welche eine zweiparametrische Familie von Verteilungen darstellen, bestehend aus einer einparametrischen linearen Exponentialfamilie und einem zusätzlichen unbekanntem Dispersionsparameter ϕ . Somit können exponentielle Dispersionsmodelle wie folgt definiert werden.

Definition 4.6. (*Exponentielle Dispersionsmodelle*)

Die Dichtefunktion eines exponentiellen Dispersionsmodells ist definiert als

$$f_Y(y, \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \right\}, \quad (4.9)$$

mit unbekanntem Dispersionsparameter ϕ , bekannten Funktionen $b(\cdot)$, $c(\cdot)$ und dem kanonischen Parameter θ .

Bemerkung 4.7. Äquivalent zur Definition 4.6 ist die Schreibweise

$$f_Y(y, \theta, \phi) = a(y, \phi) \exp \left\{ \frac{y\theta - b(\theta)}{\phi} \right\}, \quad (4.10)$$

mit unbekanntem Dispersionsparameter ϕ , bekannten Funktionen $a(\cdot)$, $b(\cdot)$ und dem kanonischen Parameter θ .

Bemerkung 4.8. Der Unterschied von Definition 4.6 zu Definition 2.4 der einparametrischen linearen Exponentialfamilie ist der zusätzliche unbekanntes Dispersionsparameter ϕ .

Somit stellen exponentielle Dispersionsmodelle eine Erweiterung der Exponentialfamilie dar. Das exponentielle Dispersionsmodell, das zu (4.8) gehört, wird mit $Y \sim Tw_k(\mu, \phi)$ bezeichnet. Abbildung 4.1 zeigt Varianzfunktionen der Tweedieverteilung mit verschiedenen Werten für k .

Es stellt sich allerdings heraus, dass k nicht alle möglichen reellen Werte annehmen darf, damit wir auch ein exponentielles Dispersionsmodell erhalten (siehe Jørgensen, 1997, Seiten 129 ff.):

$$k \in (-\infty, 0] \cup [1, \infty),$$

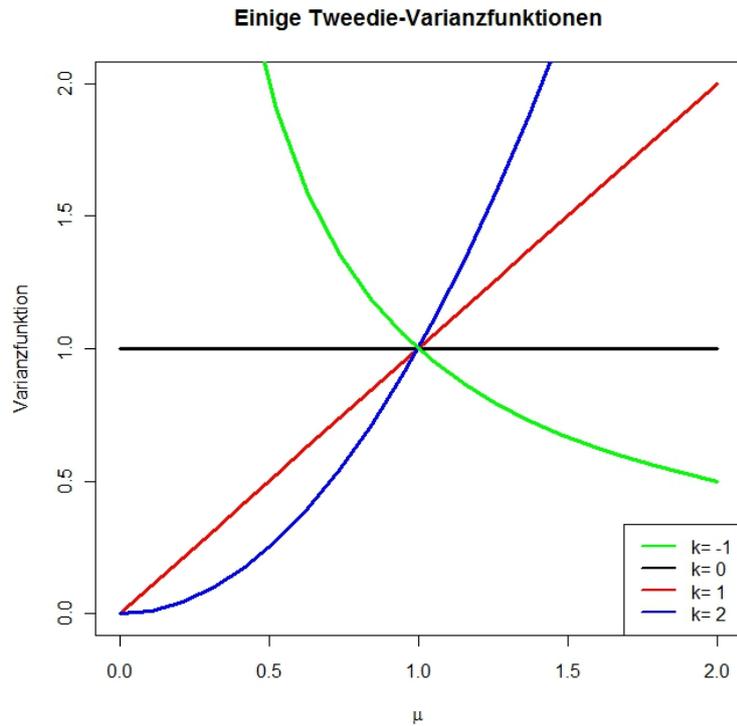


Abbildung 4.1: Varianzfunktionen der Tweedieverteilung mit $k = -1, 0, 1, 2$.

d.h. für Werte zwischen $(0, 1)$ gehört die Tweedieverteilung nicht zur Klasse der exponentiellen Dispersionsmodelle. Tabelle 4.1 fasst die Klasse der Tweediemodelle zusammen.

Verteilung	k	y	μ
Normal	$k = 0$	$y \in \mathbb{R}$	$\mu > 0$
[Existiert nicht]	$0 < k < 1$	–	–
Poisson	$k = 1$	$y \in \mathbb{N}_0$	$\mu > 0$
Zusammengesetzte Poisson	$1 < k < 2$	$y \geq 0$	$\mu > 0$
Gamma	$k = 2$	$y > 0$	$\mu > 0$
Inverse Gauß	$k = 3$	$y > 0$	$\mu > 0$

Tabelle 4.1: Exponentielle Tweedie-Dispersionsmodelle

Für $k = 0, 1, 2$ führt die Varianzfunktion $V_k(\mu) = \mu^k$ zu einer Verteilung aus der Exponentialfamilie. Im Folgenden werden diese speziellen Verteilungen mit deren

Score und Deviance Funktionen beschrieben:

- Mit $k = 0$ erhalten wir die *Normalverteilung*: $Tw_0(\mu, \phi) = N(\mu, \sigma^2)$ mit $\phi = \sigma^2$:

Somit entspricht die Varianzfunktion $V(\mu) = 1$ und die Score Funktion lautet für $y, \mu \in \mathbb{R}$:

$$\frac{\partial \mathcal{L}(\mu, y)}{\partial \mu} = -\frac{(y - \mu)^2}{2\sigma^2}.$$

Die daraus resultierende Deviance ergibt sich zu

$$D(y, \mu) = (y - \mu)^2.$$

- Für $k = 1$ ergibt sich die *Poissonverteilung*: $Tw_1(\mu, \phi) = P(\mu)$ mit $\phi = 1$: Die Varianzfunktion lautet daher $V(\mu) = \mu$ und die Score Funktion lässt sich für $\mu > 0, y \geq 0$ schreiben als:

$$\frac{\partial \mathcal{L}(\mu, y)}{\partial \mu} = y \log \mu - \mu,$$

woraus sich eine Deviance von

$$D(y, \mu) = 2 \left[y \log \left(\frac{y}{\mu} \right) - (y - \mu) \right]$$

ergibt.

- Mit $k = 2$ entsteht die *Gammaverteilung*: $Tw_2(\mu, \phi) = G(\mu, \phi)$ mit $\phi = 1$: Daraus ergibt sich eine quadratische Varianzfunktion $V(\mu) = \mu^2$. Die Score Funktion ergibt sich für $\mu > 0, y \geq 0$ zu

$$\frac{\partial \mathcal{L}(\mu, y)}{\partial \mu} = -\log \mu - \frac{y}{\mu}.$$

und die Deviance zu

$$D(y, \mu) = 2 \left[\log \left(\frac{\mu}{y} \right) + \frac{y}{\mu} - 1 \right].$$

Abbildung 4.2 zeigt Deviance Funktionen für $\mu = 1$ der Tweedieverteilung mit verschiedenen Werten von k .

Allgemein ergibt sich für $k \in \mathbb{R}^+ \setminus \{0, 1, 2\}$ mit der Varianzfunktion $V_k(\mu) = \mu^k$ die Score Funktion für $\mu > 0, y \geq 0$

$$\frac{\partial \mathcal{L}(\mu, y)}{\partial \mu} = \frac{y\mu^{1-k}}{1-k} - \frac{\mu^{2-k}}{2-k}.$$

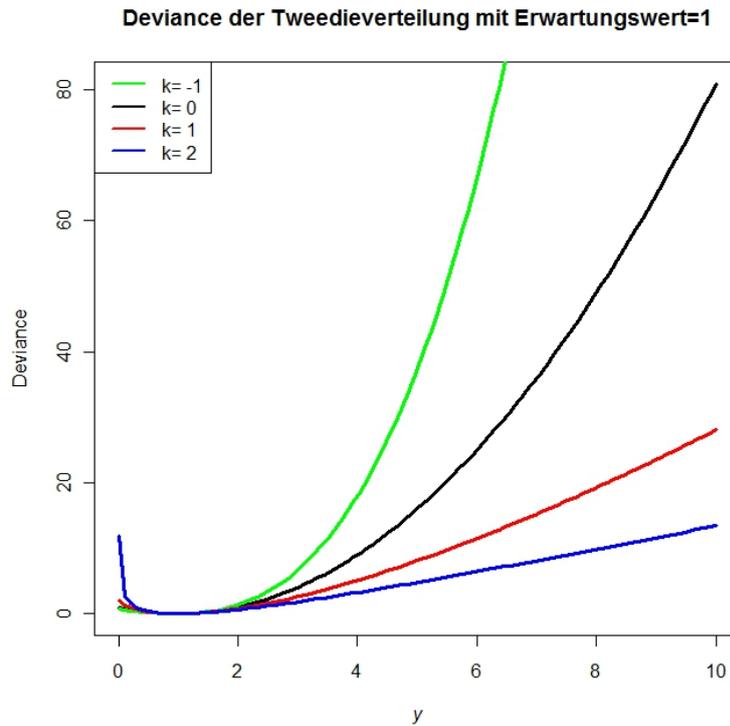


Abbildung 4.2: Deviance Funktionen der Tweedieverteilung mit $k = -1, 0, 1, 2$.

Die dazugehörige Deviance lautet (vgl. Jørgensen, 1997, Seiten 133 ff.)

$$D_k(y, \mu) = 2 \left[\frac{y^{2-k}}{(1-k)(2-k)} - \frac{y\mu^{1-k}}{1-k} + \frac{\mu^{2-k}}{2-k} \right].$$

Kumulantenerzeugende Funktion

Das exponentielle Dispersionsmodell ist eindeutig durch die Varianzfunktion charakterisiert. Wir wollen nun die Form der Momenterzeugenden Funktion finden, wobei die folgenden Formeln allgemein für $k \in \mathbb{R}$ hergeleitet werden. Siehe dazu auch Jørgensen (1997), Seiten 130 ff., sowie Dunn und Smyth (2007). Es ist zwar keine geschlossene Form der Dichtefunktion der Tweedieverteilung bekannt - außer für die speziellen Fälle in Tabelle 4.1 - aber wir können die einfache Kumulantenerzeugende Funktion analog zur Kumulantenerzeugenden Funktion der einparametrischen linearen Exponentialfamilie (2.9) als

$$K(t) = \frac{b(\theta + \phi t) - b(\theta)}{\phi} \tag{4.11}$$

mit der Kumulantenfunktion $b(\theta)$ darstellen.

Bemerkung 4.9. Die Funktion $b(\theta)$ wird Kumulantenfunktion genannt, da sich die k -te Kumulante der Verteilung proportional mit ϕ^{k-1} zur k -ten Ableitung $b^{(k)}(\theta)$ der Funktion $b(\theta)$ verhält.

Definieren wie die Deviance als

$$D(y, \mu) = 2\phi \{ \mathcal{L}(y, y) - \mathcal{L}(\mu, y) \},$$

so kann sie als ein Maß der Distanz gesehen werden, das $D(y, y) = 0$ und $D(\mu, y) > 0$ für $y \neq \mu$ erfüllt. Damit kann die Dichtefunktion in Termen der Deviance umgeschrieben werden:

$$f_Y(y, \mu, \phi, k) = f_Y(y, y, \phi, k) \exp \left\{ -\frac{1}{2\phi} D(\mu, y) \right\}. \quad (4.12)$$

Jørgensen (1997) bezeichnet dies als die Form des *Dispersionsmodells* der Dichtefunktion.

Die Kumulantenfunktion $b(\theta)$ kann für exponentielle Dispersionsmodelle der Tweedieverteilung durch Lösen der Gleichung

$$b''(\theta) = \frac{\partial b'(\theta)}{\partial \theta} = \frac{\partial \mu}{\partial \theta} = V(\mu) = \mu^k$$

mit dem Erwartungswert $\mu = b'(\theta)$ und der Varianzfunktion $V(\mu) = b''(\theta)$ berechnet werden. Ohne Beschränkung der Allgemeinheit kann an der Stelle $\theta = 0$: $b(\theta) = 0$ und $\mu = 1$ angenommen werden. Daraus resultiert für den Parameter θ

$$\theta = \int \frac{1}{V(\mu)} d\mu = \begin{cases} \frac{\mu^{1-k}-1}{1-k} & \text{für } k \neq 1, \\ \log \mu & \text{für } k = 1. \end{cases} \quad (4.13)$$

Als Inverse ergibt sich der Erwartungswert

$$\mu = \tau(\theta) = \begin{cases} [\theta(1-k) + 1]^{1/(1-k)} & \text{für } k \neq 1, \\ e^\theta & \text{für } k = 1, \end{cases} \quad (4.14)$$

mit einer monoton steigenden Funktion $\tau(\theta)$, die den Zusammenhang zwischen dem kanonischen Parameter θ und dem Erwartungswert μ beschreibt (siehe Jørgensen, 1997, Seite 46).

Die Kumulantenfunktion ergibt sich zu

$$b(\theta) = \int \frac{\mu}{V(\mu)} d\mu = \begin{cases} \frac{\mu^{2-k}-1}{2-k} & \text{für } k \neq 2, \\ \log \mu & \text{für } k = 2. \end{cases}$$

Mit (4.14) erhalten wir die Kumulantenfunktion

$$b(\theta) = \begin{cases} \frac{[\theta(1-k)+1]^{\frac{2-k}{1-k}} - 1}{2-k} & \text{für } k \neq 1, 2 \\ -\log(1-\theta) & \text{für } k = 2 \\ e^\theta - 1 & \text{für } k = 1. \end{cases} \quad (4.15)$$

Bemerkung 4.10. Definitionen (4.13) und (4.15) sind für θ und $b(\theta)$ stetig in k sowie in θ .

Bemerkung 4.11. Es können weder $a(\cdot)$ in (4.10) noch $b(\cdot)$ in (4.12) - mit Ausnahmen für die Fälle $k = 1, 2, 3$ - in geschlossener Form dargestellt werden.

Dichtefunktion der Tweedieverteilung

Die Fourier Transformation erlaubt die Dichtefunktion eines stetigen exponentiellen Dispersionsmodells in Termen der Kumulantenerzeugenden Funktion (4.11) mit $i^2 = -1$ zu schreiben als

$$f_Y(y, \mu, \phi, k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \{K(it) - ity\} dt. \quad (4.16)$$

Wir wollen nun die Dichtefunktion $f_Y(y, \mu, \phi, k)$ durch numerisches Berechnen des Integrals (4.16) erhalten. Durch die Eigenschaften der Tweedieverteilung können wir dieses Problem vereinfachen.

So sind Tweedie Modelle die einzigen exponentiellen Dispersionsmodelle, die unter der „Re-Scaling“ Transformation abgeschlossen sind. Betrachten wir hierzu die Transformation $Z = cY$ für $c > 0$ mit einer Zufallsvariablen Y , die aus einer Tweedieverteilung mit Erwartungswert μ und Varianzfunktion $V(\mu) = \mu^k$ stammt. Die Kumulantenerzeugende Funktion für Z zeigt, dass Z zu einer Tweedieverteilung mit demselben k und Erwartungswert $c\mu$ und Dispersion $c^{2-k}\phi$ gehört. Als Jacobi-Determinante¹ der Transformation ergibt sich $\frac{1}{c}, \forall y > 0$. Mit diesen Eigenschaften erhalten wir die sogenannte „Re-Scaling“ Identität (siehe dazu auch Jørgensen, 1997, Seiten 128f.)

$$f_Y(y, \mu, \phi, k) = cf_Y(cy, c\mu, c^{2-k}\phi, k), \quad \forall k > 0, y > 0, c > 0. \quad (4.17)$$

Diese Identität erlaubt uns, y -Werte zu wählen und die Dichtefunktion durch „Re-Scaling“ an anderen Stellen auszuwerten.

¹Die Jacobi-Determinante ist die Determinante der Jacobi-Matrix.

Des Weiteren ist es aufgrund der Eigenschaften der Tweedieverteilung nicht notwendig das Integral (4.16) für jede Kombination von y , μ und ϕ zu berechnen. Zum einen ist es ausreichend $a(y, \phi) = f_Y(y, 1, \phi, k)$ numerisch zu integrieren. Danach kann $f_Y(y, \mu, \phi, k)$ für jedes μ aus (4.10) berechnet werden. Es genügt daher die Kalkulation für $\mu = 1$, das zur Situation $\theta = b(\theta) = 0$ gehört. Zum anderen genügt es $f_Y(y, y, \phi, k)$ mit numerischer Integration zu berechnen, wonach $f_Y(y, \mu, \phi, k)$ aus (4.12) erhalten wird. Also genügt die Kalkulation an der Stelle $\mu = y$.

Es gibt viele Methoden mit denen $f_Y(y, \mu, \phi, k)$ für alle y, μ, ϕ und k durch invertieren der Kumulantenerzeugenden Funktion an der Stelle $\mu = 1$ berechnet werden kann, drei davon zeigen wir hier (siehe dazu auch Dunn und Smyth, 2007, Seiten 7ff.):

Methode 1:

- Berechne $c(y, \phi) = f_Y(y, 0, \phi)$ mittels der Fourier Transformation (4.16).
- Substituiere das Ergebnis in die Definition des exponentialen Dispersionsmodells (4.9).

Methode 2:

- Skalieren: $\mu = 1$.
- Berechne das Integral $A = a\left(\frac{y}{\mu}, \frac{\phi}{\mu^{2-p}}\right) = f_Y\left(\frac{y}{\mu}, 1, \frac{\phi}{\mu^{2-p}}\right)$ mit der Fourier Transformation (4.16).
- Verwende $f_Y(y, \mu, \phi, k) = \frac{A}{\mu}$.
Diese Methode verwendet die „Re-Scaling“ Identität (4.17) mit $c = \frac{1}{\mu}$.

Methode 3:

- Skalieren: $y = 1$ und kalkuliere $b(1, \phi)$.
- Berechne das Integral $A = a\left(1, \frac{\phi}{y^{2-p}}\right) = f_Y\left(1, 1, \frac{\phi}{y^{2-p}}\right)$ mit der Fourier Transformation (4.16).
- Verwende $f_Y(y, y, \phi, k) = \frac{A}{y}$ und substituiere dies in (4.12).
Diese Methode verwendet die „Re-Scaling“ Identität (4.17) mit $c = \frac{1}{y}$.

Alle drei Methoden sind äquivalent in der exakten Rechnung. Methode 3 hat jedoch theoretische Vorteile: Diese Methode ist aus der Sicht des Programmierers die einfachste Form zu implementieren, da sie die Fourier Transformation nur an der Stelle $y = \mu = 1$ benötigt. Die entscheidende Motivation für Methode 3 ist,

dass das Integral A , das mit der Fourier Transformation berechnet wird, fast immer größer als mit den anderen beiden Methoden ist. Denn unter Methode 3 wird das Integral immer zur Berechnung einer Dichtefunktion an der Stelle seines Erwartungswertes, d.h. bei $y = \mu$, kalkuliert, welcher nahe beim Maximalwert der Verteilung liegt. Somit sollten größere Werte des Integrals zu kleineren relativen Fehlern führen.

Mit Hilfe des R-Paketes `tweedie` plotten wir die Dichtefunktion der Tweedieverteilung für eine Varianzfunktion $V_k(\mu) = \mu^k$ mit $k = 2.5$. Dazu laden wir zusätzlich zum Paket `tweedie` noch das dazu benötigte Paket `statmod`. Mit dem R-Code aus der Hilfe für das Tweedie-Paket (siehe die Dokumentation des Tweedie-Paketes von Dunn, 2009)

```
> k <- 2.5
> mu <- 1
> phi <- 1
> y <- seq(0, 6, length=500)
> fy <- dtweedie.inversion(y=y, power=k, mu=mu, phi=phi)
> plot(y, fy, type="l", lwd=3, ylab="Dichtefunktion",
      main="Dichte der Tweedieverteilung mit k=2.5", col="red")
```

ergibt sich der Plot der Dichtefunktion, wie in Abbildung 4.3 ersichtlich. Hierbei verwendet die Funktion `dtweedie.inversion` die Fourier-Transformation zur Berechnung der Dichtefunktion.

4.3.2 EQL Funktion mit Power-Varianz

Betrachten wir nun die EQL Funktion (4.7) mit der Power-Varianfunktion $V_k(\mu) = \mu^k$. Im Unterschied zum vorhin betrachteten Tweediemodell wird bei der EQL Funktion mit der Power-Varianz nicht die gesamte Verteilung spezifiziert, sondern nur die ersten beiden Momente angenommen werden.

Bemerkung 4.12. *Da wir hier nur von der Annahme des Erwartungswertes und der Varianz ausgehen, erhalten wir kein numerisches Problem bei der Schätzung des unbekanntes Parameters, wenn dieser in die Nähe von 1 kommt bzw. kleiner als 1 wird (siehe dazu die Beispiele im nächsten Kapitel „Aberrant Crypt Foci“ und „Schiffdaten“).*

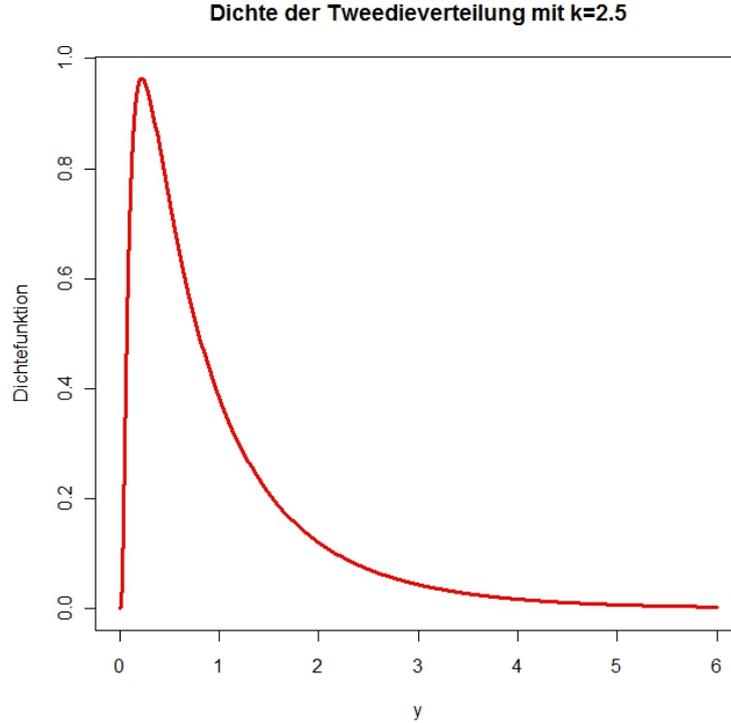


Abbildung 4.3: Dichtefunktion der Tweedieverteilung mit $V_k(\mu) = \mu^{5/2}$.

Unser Ziel ist nun die EQL Funktion (4.7) in allen Parametern zu maximieren und simultan Schätzer für den Parameter β und ϕ sowie für den unbekannt Parameter k in der Varianzfunktion zu konstruieren. Dafür leiten wir (4.7) nach μ und k ab und setzen diese Ableitungen gleich Null.

Für $k \in \mathbb{R} \setminus \{0, 1, 2\}$ erhalten wir mit der Varianzfunktion $V_k(\mu) = \mu^k$ die EQL Funktion

$$Q_k^+(\mu, y, \phi) = -\frac{1}{2} \log(2\pi\phi y^k) - \frac{1}{\phi} \left(\frac{y^{2-k}}{(1-k)(2-k)} - \frac{y\mu^{1-k}}{1-k} + \frac{\mu^{2-k}}{2-k} \right).$$

Die Ableitung nach μ führt uns zur Quasi-Score Funktion

$$\frac{\partial}{\partial \mu} Q_k^+(\mu, y, \phi) = \frac{1}{\phi} \left(\mu^{1-k} - \frac{y}{\mu^k} \right) = 0.$$

Die Ableitung nach dem unbekannt Parameter k der Varianzfunktion ergibt die

Score Funktion

$$\begin{aligned} \frac{\partial}{\partial k} Q_k^+(\mu, y, \phi) &= -\frac{\log(y)}{2} \\ &- \frac{1}{\phi} \left[\frac{y^{2-k}}{(1-k)(2-k)^2} + \frac{y^{2-k}}{(1-k)^2(2-k)} - \frac{y^{2-k} \log(y)}{(1-k)(2-k)} \right] \\ &- \frac{1}{\phi} \left[\frac{\mu^{2-k}}{(k-2)^2} + \frac{\mu^{1-k} y}{(k-1)^2} - \frac{\mu^{2-k} \log(\mu)}{k-2} + \frac{\mu^{1-k} y \log(\mu)}{k-1} \right] = 0. \end{aligned}$$

4.3.3 Binomialähnlicher Ansatz

In diesem Abschnitt betrachten wir eine Familie von Varianzfunktionen, welche die standardisierte Binomialverteilung mit $V(\mu) = \mu(1 - \mu)$ erweitert, indem zwei zusätzliche Potenz-Parameter $k, l \in \mathbb{N}$ eingeführt werden:

$$V_{k,l}(\mu) = \mu^k(1 - \mu)^l, \quad 0 < \mu < 1. \quad (4.18)$$

Ziel ist es wiederum die EQL Funktion zu dieser Varianzfunktion zu maximieren. Mit Darstellung (4.18), der QL Funktion

$$Q_{k,l}(\mu, y) = \int_0^\mu \frac{y-t}{\phi[t^k(1-t)^l]} dt$$

und der Quasi-Deviance

$$D_{k,l}(y, \mu) = -2\phi(Q_{k,l}(\mu, y) - Q_{k,l}(y, y)) = -2\phi \left(\int_y^\mu \frac{y-t}{\phi[t^k(1-t)^l]} dt \right)$$

ergibt sich für die EQL Funktion

$$Q_{k,l}^+(\mu, \phi, y) = -\frac{1}{2} \log \{ 2\pi\phi[y^k(1-y)^l] \} - \underbrace{\int_y^\mu \frac{y-t}{\phi[t^k(1-t)^l]} dt}_{:= (*)}. \quad (4.19)$$

Die Schwierigkeit dieses Ausdrucks liegt nun in der Lösung des Integrals (*). Eine Möglichkeit zeigt Thaler (2000) auf den Seiten 60ff. mit der Verwendung der Par-

tialbruchzerlegung. Somit erhalten wir die Lösung des Integrals für $k \geq 1, l \geq 1$:

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

mit

$$(A_i)_{i=0}^{k-1} := \binom{l-1+i}{l-1}, \quad (B_j)_{j=0}^{l-1} := \binom{k-1+j}{k-1}.$$

Die QL Funktion mit den Parametern der Varianzfunktion $k \geq 1, l \in \mathbb{N}$ und den Restriktionen $0 < \mu < 1, 0 \leq y \leq 1$ ergibt sich zu

$$Q_{k,l}(\mu, y) = \int_0^\mu \frac{y-t}{t^k(1-t)^l} dt = y \cdot I_{k,l}(\mu) - I_{k-1,l}(\mu). \quad (4.20)$$

Verwenden wir die standardisierte Binomialverteilung mit der Varianzfunktion $V_{1,1}(\mu) = \mu(1-\mu)$, so setzen wir $k = l = 1$ in (4.20) ein und bekommen

$$Q_{1,1}(\mu, y) = y \cdot I_{1,1}(\mu) - I_{0,1}(\mu) = y \log \left| \frac{\mu}{1-\mu} \right| + \log |1-\mu|.$$

Mit der Varianzfunktion $V_{2,2}(\mu) = \mu^2(1-\mu)^2$, also den Parametern $k = l = 2$, kann die Variabilität für Responses unter logistischen Regressionsmodellen beschrieben werden, die wir mit der standardisierten Binomialverteilung überschätzen. Die QL Funktion lässt sich mit diesen Parametern in geschlossener Form darstellen (siehe McCullagh und Nelder, 1989, Seite 330, sowie Thaler, 2000, Seite 61):

$$Q_{2,2}(\mu, y) = y \cdot I_{2,2}(\mu) - I_{1,2}(\mu) = (2y-1) \log \left| \frac{\mu}{1-\mu} \right| - \frac{y}{\mu} - \frac{1-y}{1-\mu}.$$

Lassen wir auch reelle Zahlen für $k, l \in \mathbb{R}$ zu, so kann das bestimmte Integral $\int_0^\mu \frac{1}{t^k(1-t)^l} dt$ mit Hilfe der Substitution $t = \mu p$ und somit $dt = \mu dp$ für $p \in (0, 1)$ und Umformen auf eine *hypergeometrische Funktion* ${}_pF_q(a_1, \dots, a_p; b_q, \dots, b_q; z)$

ebenfalls in geschlossener Form dargestellt werden:

$$\begin{aligned}
 \int_0^\mu \frac{1}{t^k(1-t)^l} dt &= \int_0^1 \frac{1}{(\mu p)^k(1-\mu p)^l} \mu \cdot dp \\
 &= \frac{1}{\mu^{k-1}} \int_0^1 p^{-k}(1-p)^{2-k-(1-k)-1}(1-\mu p)^{-l} dp \\
 &= \frac{1}{\mu^{k-1}} \frac{\Gamma(1-k)\Gamma(1)}{\Gamma(2-k)} {}_2F_1(1-k, l, 2-k; \mu) \\
 &= \frac{\mu^{1-k}}{1-k} {}_2F_1(1-k, l, 2-k; \mu),
 \end{aligned}$$

mit der allgemeinen Definition für die hypergeometrische Funktion als Reihendarstellung

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; z) = \sum_{k=0}^{\infty} \frac{\prod_{i=1}^p a_i^{\bar{k}}}{\prod_{j=1}^q b_j^{\bar{k}}} \cdot \frac{z^k}{k!}.$$

Hierbei bezeichnet $x^{\bar{k}}$ das steigende *Pochhammersymbol*, welches durch

$$x^{\bar{k}} = \frac{\Gamma(x+k)}{\Gamma(x)} = x \cdot (x+1) \cdot (x+2) \cdot \dots \cdot (x+k-1)$$

definiert ist. Da wir in unserer Verwendung für die hypergeometrische Funktion nur $p = 2, q = 1$ Parameter haben, betrachten wir nun diesen Spezialfall als Integraldarstellung

$${}_2F_1(a_1, a_2; b; z) = \frac{\Gamma(b)}{\Gamma(a_1)\Gamma(b-a_1)} \int_0^1 t^{a_1-1}(1-t)^{b-a_1-1}(1-zt)^{-a_2} dt,$$

für $b > a_1 > 0$ sowie als Reihendarstellung

$${}_2F_1(a_1, a_2; b; z) = \sum_{k=0}^{\infty} \frac{a_1^{\bar{k}} a_2^{\bar{k}} z^k}{b^{\bar{k}} k!}, \quad \text{für } b \notin \mathbb{Z} \setminus (\mathbb{N} \cup \{0\}).$$

Für die Deviance Funktion, die wir in der EQL Funktion benötigen, ergibt sich

$$\begin{aligned}
 -\frac{D_{k,l}(y, \mu)}{2\phi} &= Q_{k,l}(\mu, y) - Q_{k,l}(y, y) \\
 &= \int_y^\mu \frac{y-t}{\phi[t^k(1-t)^l]} dt \\
 &= \left\{ \int_0^\mu \frac{y}{\phi[t^k(1-t)^l]} dt - \int_0^\mu \frac{t}{\phi[t^k(1-t)^l]} dt \right\} \\
 &\quad - \left\{ \int_0^y \frac{y}{\phi[t^k(1-t)^l]} dt - \int_0^y \frac{t}{\phi[t^k(1-t)^l]} dt \right\}.
 \end{aligned}$$

Durch dieses Umschreiben können wir nun für jedes Integral obiges Ergebnis mit der Darstellung der hypergeometrischen Funktion verwenden. Aus Gründen der Übersichtlichkeit definieren wir neue Parameter für die Übergabeparameter der hypergeometrischen Funktion:

$$\begin{aligned}
 {}_2F_1(\alpha_1; \mu) &:= {}_2F_1(1 - k, l; 2 - k; \mu), \\
 {}_2F_1(\gamma_1; \mu) &:= {}_2F_1(2 - k, l; 3 - k; \mu), \\
 {}_2F_1(\delta; \mu) &:= {}_2F_1(3 - k, l; 4 - k; \mu), \\
 {}_2F_1(\nu; \mu) &:= {}_2F_1(4 - k, l; 5 - k; \mu), \\
 {}_2F_1(\rho; \mu) &:= {}_2F_1(5 - k, l; 6 - k; \mu), \\
 {}_2F_1(\omega; \mu) &:= {}_2F_1(6 - k, l; 7 - k; \mu), \\
 {}_3F_2(\alpha_2; \mu) &:= {}_3F_2(1 - k, 1 - k, l; 2 - k, 2 - k; \mu), \\
 {}_3F_2(\gamma_2; \mu) &:= {}_3F_2(2 - k, 2 - k, l; 3 - k, 3 - k; \mu).
 \end{aligned}$$

Somit erhalten wir

$$\begin{aligned}
 -\frac{D_{k,l}(y, \mu)}{2\phi} &= \frac{1}{\phi} \left\{ \frac{\mu^{1-k}}{1-k} [y \cdot {}_2F_1(\alpha_1; \mu) - \mu \cdot {}_2F_1(\gamma_1; \mu)] \right. \\
 &\quad \left. - \frac{y^{2-k}}{1-k} [{}_2F_1(\alpha_1; y) - {}_2F_1(\gamma_1; y)] \right\}.
 \end{aligned}$$

Daraus resultiert die EQL Funktion mit dem binomialähnlichen Ansatz (4.18) zu

$$\begin{aligned}
 Q_{k,l}^+(\mu, \phi, y) &= -\frac{1}{2} \log \{2\pi\phi[y^k(1-y)^l]\} \\
 &\quad - \frac{1}{\phi} \left\{ \frac{\mu^{1-k}}{1-k} [y \cdot {}_2F_1(\alpha_1; \mu) - \mu \cdot {}_2F_1(\gamma_1; \mu)] \right. \\
 &\quad \left. - \frac{y^{2-k}}{1-k} [{}_2F_1(\alpha_1; y) - {}_2F_1(\gamma_1; y)] \right\}.
 \end{aligned}$$

Ziel ist es nun diese EQL Funktion nach dem Erwartungswert μ bzw. β abzuleiten, um das Maximum bestimmen zu können und somit den Schätzer $\hat{\beta}$ zu bekommen. Um Schätzer für die unbekannt Parameter der Varianzfunktion k und l zu erhalten, betrachten wir die EQL Funktion (4.19) direkt.

Bemerkung 4.13. Da ${}_2F_1(a_1, a_2; b; z)$ für $b > a_1 > 0$ bzw. für $b > a_2 > 0$ (die Reihenfolge der Argumente ist nicht relevant) definiert ist, folgt für die hypergeometrische Funktion ${}_2F_1(1-k, l; 2-k; \mu)$ die Bedingungen $2-k > 1-k > 0 \Rightarrow 1 > k$

bzw. $2 - k > l \Rightarrow 2 > k + l$. Somit betrachten wir Modelle deren Variabilität größer als $\mu(1 - \mu)$ ist. Da wir in der Realität vorwiegend Überdispersion vorfinden und Unterdispersion nur selten auftritt, bilden diese Bedingungen jedoch keine starken Einschränkungen.

Die Ableitung nach μ liefert die Quasi-Score Funktion

$$\begin{aligned} \frac{\partial}{\partial \mu} Q_{k,l}^+(\mu, y, \phi) &= \frac{1}{\phi} \frac{\mu^{1-k}}{k-1} \left\{ \frac{(1-k)y}{\mu} \left[\frac{1}{(1-\mu)^l} - {}_2F_1(\alpha_1; \mu) \right] \right. \\ &\quad \left. - \frac{2-k}{(1-\mu)^l} + (1-k) {}_2F_1(\gamma_1; \mu) \right\} \\ &\quad - \frac{1}{\phi} \frac{1}{\mu^k} [y \cdot {}_2F_1(\alpha_1; \mu) - \mu \cdot {}_2F_1(\gamma_1; \mu)] = 0. \end{aligned}$$

Nun wollen wir den unbekanntem Parameter k der Varianzfunktion schätzen. Für den ersten Teil des Ausdrucks (4.19) ergibt sich als Ableitung nach k

$$-\frac{\partial}{\partial k} \frac{1}{2} \log \{2\pi\phi[y^k(1-y)^l]\} = -\frac{\log(y)}{2}. \quad (4.21)$$

Für den zweiten Teil betrachten wir die Ableitung der Quasi-Deviance, wobei wir durch das Vertauschen von Differentiation und Integration folgende Darstellung erhalten:

$$\begin{aligned} \frac{\partial}{\partial k} \int_y^\mu \frac{y-t}{\phi[t^k(1-t)^l]} dt &= \int_y^\mu \frac{\partial}{\partial k} \frac{y-t}{\phi[t^k(1-t)^l]} dt = \frac{1}{\phi} \int_y^\mu \frac{(y-t) \log(t)}{t^k(1-t)^l} dt \\ &= \mu^{1-k} \left\{ \frac{y \cdot {}_3F_2(\alpha_2; \mu)}{(k-1)^2} - \frac{\mu \cdot {}_3F_2(\gamma_2; \mu)}{(k-2)^2} + \log(\mu) \left[\frac{y \cdot {}_2F_1(\alpha_1; \mu)}{k-1} - \frac{\mu \cdot {}_2F_1(\gamma_1; \mu)}{k-2} \right] \right\} \\ &\quad - y^{2-k} \left\{ \frac{{}_3F_2(\alpha_2; y)}{(k-1)^2} - \frac{{}_3F_2(\gamma_2; y)}{(k-2)^2} + \log(y) \left[\frac{{}_2F_1(\alpha_1; y)}{k-1} - \frac{{}_2F_1(\gamma_1; y)}{k-2} \right] \right\}. \end{aligned}$$

Zusammen mit Teil (4.21) erhalten wir die Score Funktion, die Null gesetzt wird.

Leiten wir den ersten Teil von (4.19) nach l ab, so ergibt sich

$$-\frac{\partial}{\partial l} \frac{1}{2} \log \{2\pi\phi[y^k(1-y)^l]\} = -\frac{\log(1-y)}{2}. \quad (4.22)$$

Für den zweiten Teil lässt sich wieder Differentiation und Integration vertauschen.

$$\frac{\partial}{\partial l} \int_y^\mu \frac{y-t}{\phi[t^k(1-t)^l]} dt = \int_y^\mu \frac{\partial}{\partial l} \frac{y-t}{\phi[t^k(1-t)^l]} dt = \frac{1}{\phi} \underbrace{\int_y^\mu \frac{(y-t) \log(1-t)}{t^k(1-t)^l} dt}_{:= (**)}.$$

Um das Integral (**) lösen zu können, approximieren wir den Ausdruck $\log(1-t)$ mit Taylor bis zur 4. Ordnung für $t < 1$:

$$\log(1-t) \approx -t - \frac{t^2}{2} - \frac{t^3}{3} - \frac{t^4}{4}. \quad (4.23)$$

Setzen wir (4.23) in (**) ein, erhalten wir approximativ für das Integral

$$\begin{aligned} & \mu^{2-k} \left\{ \frac{12y \cdot {}_2F_1(\gamma_1; \mu)}{k-2} + \mu \frac{6(y-2) {}_2F_1(\delta; \mu)}{k-3} - \frac{\mu^{2-k}}{k-4} [-2(2y-3) {}_2F_1(\nu; \mu)] \right\} \\ & - \frac{\mu^{4-k}}{k^2 - 11k + 30} [-(k-6)(3y-4) {}_2F_1(\rho; \mu) + 3(k-5) {}_2F_1(\omega; \mu)] \\ & - y^{3-k} \left\{ \frac{12 \cdot {}_2F_1(\gamma_1; y)}{k-2} + \frac{6(y-2) {}_2F_1(\delta; y)}{k-3} - \frac{y}{k-4} [2(2y-3) {}_2F_1(\nu; y)] \right\} \\ & - \frac{y^{5-k}}{k^2 - 11k + 30} [(k-6)(3y-4) {}_2F_1(\rho; y) - 3(k-5)y {}_2F_1(\omega; y)]. \end{aligned}$$

Hinzufügen der Ableitung des ersten Teils und Nullsetzen dieser Score Funktion liefert den Schätzer für l .

t	Regen $_t$										
1	0.00	19	0.00	37	0.13	55	0.05	73	0.00	91	0.27
2	0.01	20	0.00	38	0.00	56	0.00	74	0.00	92	0.00
3	0.38	21	0.81	39	0.00	57	0.22	75	0.00	93	0.00
4	0.14	22	0.00	40	0.00	58	0.00	76	0.00	94	0.00
5	0.00	23	0.00	41	0.00	59	0.93	77	0.00	95	0.00
6	0.07	24	0.02	42	0.00	60	0.00	78	0.00	96	0.00
7	1.19	25	0.67	43	0.43	61	0.56	79	0.00	97	1.09
8	1.06	26	0.36	44	0.92	62	0.00	80	0.00	98	0.00
9	0.00	27	0.00	45	0.21	63	0.29	81	0.36	99	0.00
10	0.00	28	0.00	46	0.18	64	1.70	82	0.25	100	0.00
11	0.00	29	0.00	47	0.72	65	0.13	83	0.06	101	0.00
12	0.00	30	0.18	48	0.12	66	0.00	84	1.20	102	0.00
13	0.00	31	0.97	49	1.13	67	0.12	85	0.00	103	0.00
14	0.00	32	0.00	50	0.00	68	0.00	86	0.00	104	1.37
15	0.00	33	0.00	51	0.00	69	0.00	87	0.00	105	0.26
16	0.00	34	0.00	52	0.00	70	0.00	88	0.72	106	1.01
17	0.00	35	0.00	53	0.00	71	0.00	89	3.02	107	0.00
18	0.00	36	0.06	54	2.62	72	0.67	90	0.50		

Tabelle 5.1: Regenmengen an 107 Tagen in inches.

```
y <- ifelse(Regen > 0, y <- 1, y <- 0)
```

Hier bezeichnet `Regen` die Regenmenge und `y` die binäre Variable 0 oder 1. Zur Veranschaulichung der Daten betrachten wir Abbildung 5.1, wobei die obere Graphik die beobachtete Regenmenge und die untere die binäre Größe zeigt. Der zugehörige Code lautet

```
par(mfrow=c(2,1))
plot(ts(Regen), xlab="Tag", ylab="Regenmenge in inches")
plot(ts(y), xlab="Tag", ylab="Regen ja/nein")
```

Definieren wir nun die Prädiktorvariablen Y_{t-3} , Y_{t-2} , Y_{t-1} sowie die Response Y_t , die möglicherweise vom Eintritt des Regenereignisses in den letzten drei Tagen abhängt.

```
> for (j in 1:104) {Yt_3[j] <- y[j]}
> for (j in 2:105) {Yt_2[j-1] <- y[j]}
```

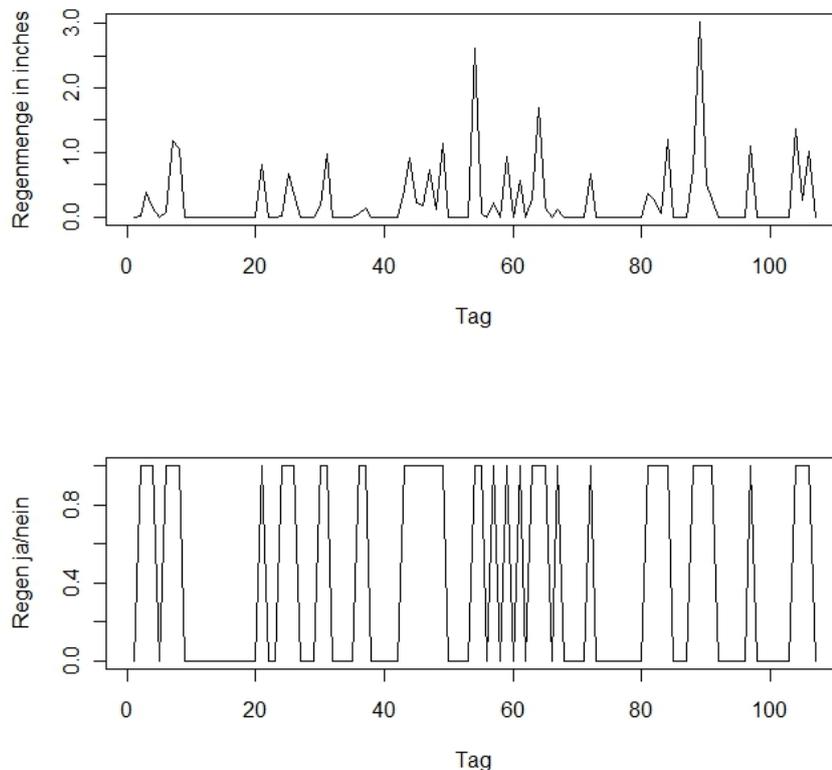


Abbildung 5.1: Beobachtete Regendaten.

```
> for (j in 3:106) {Yt_1[j-2] <- y[j]}
> for (j in 4:107) {Yt[j-3] <- y[j]}
```

Bezeichne $\mu_t(\beta)$ die bedingte Wahrscheinlichkeit des Eintretens des Ereignisses „Regen“:

$$\mu_t(\beta) = \mathbb{P}[Y_t = 1 | \mathcal{F}_{t-1}]$$

mit $\mathcal{F}_{t-1} = \sigma\{Y_{t-1}, Y_{t-2}, Y_{t-3}\}$ der von $Y_{t-1}, Y_{t-2}, Y_{t-3}$ erzeugten σ -Algebra, auch Filtration genannt. Um die bedingte Wahrscheinlichkeit

$$\mathbb{P}[Y_t = y_t | \mathcal{F}_{t-1}] = [\mu_t(\beta)]^{y_t} [1 - \mu_t(\beta)]^{1-y_t}, \quad y_t \in \{0, 1\}$$

modellieren zu können, bedienen wir uns der Definition der sogenannten *Partial-Likelihood-Funktion*.

Definition 5.1. Sei $\mu_t(\beta)$ die bedingte Wahrscheinlichkeit, dass ein bestimmtes Ereignis eintritt. So ist die Partial-Likelihood-Funktion definiert als

$$PL(\mu, y) = \prod_{t=1}^N [\mu_t(\beta)]^{y_t} [1 - \mu_t(\beta)]^{1-y_t}. \quad (5.1)$$

Um den dazugehörigen Maximum-Partial-Likelihood-Schätzer (MPLE) $\hat{\beta}$ zu erhalten, wird $\log PL(\mu, y) = \mathcal{L}(\mu, y)$ maximiert.

Nun testen wir den Einfluss der drei Prädiktorvariablen im logistischen Modell

$$\text{logit}(\mu_t(\beta)) = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3}$$

bzw. mit dem R-Code

```
> mod<-glm(Yt~Yt_1+Yt_2+Yt_3,family=binomial).
```

In der Funktion `glm` verwenden wir `family=binomial`¹, da die Responsevariable binomialverteilt ist und wir somit das Ereignis „Regen Ja/Nein“ modellieren. Mit dem Befehl `summary` können wir die Signifikanz der einzelnen erklärenden Variablen herausfinden.

```
> summary(mod)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6711  -0.8663  -0.6555   0.9537   2.0577

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.7867     0.3271  -2.405  0.0162 *
Yt_1          1.3386     0.4688   2.855  0.0043 **
Yt_2          0.5600     0.4887   1.146  0.2518
Yt_3         -1.2020     0.4998  -2.405  0.0162 *

(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 139.49  on 103  degrees of freedom
Residual deviance: 123.63  on 100  degrees of freedom
AIC: 131.63
Number of Fisher Scoring iterations: 4
```

¹Mit dem Befehl `family=binomial` wird als Default-Einstellung die logistische Linkfunktion herangezogen.

Ein p-Wert < 0.05 besagt, dass die Nullhypothese $H_0: \beta_j = 0$ verworfen werden kann und die dazugehörige Prädiktorvariable einen signifikanten Einfluss auf die Responsevariable hat. Als Ergebnis für unser betrachtetes Modell sehen wir, dass die Variable Y_{t-1} , also „Regen gestern“, mit einem p-Wert von 0.0043 einen signifikanten Einfluss auf die Response Y_t , „Regen heute“, hat. Auch die Variable Y_{t-3} scheint signifikanten Einfluss auf die Response zu nehmen. Als Schätzer für β erhalten wir die Werte $\hat{\beta} = (-0.7867, 1.3386, 0.5600, -1.2020)$. Das negative Vorzeichen für den Schätzer von Y_{t-3} bedeutet: Wenn vor drei Tagen kein Regen gefallen ist, so bewirkt dies eine höhere Regenwahrscheinlichkeit für heute.

Sei $Z_{t-1} = (1, Y_{t-1}, Y_{t-2}, Y_{t-3})$ der Kovariablen-Vektor, der Informationen über die Vergangenheit von Y_t enthält, so ist die Informationsmatrix definiert als

$$G_N(\beta) = \sum_{t=1}^N \text{var}[Z_{t-1}(Y_t - \mu_t(\beta)) | \mathcal{F}_{t-1}],$$

welche mit dem Aufruf

```
> info<-summary(mod)$cov.scaled
```

erhalten wird. Die inverse Informationsmatrix ist die geschätzte Varianz-Kovarianzmatrix des Parameterschätzers $\hat{\beta}$ und lautet

$$G_N^{-1} = \begin{pmatrix} 0.10702572 & -0.05986788 & -0.04875255 & -0.03616408 \\ -0.05986788 & 0.21980172 & -0.05153020 & -0.04642844 \\ -0.04875255 & -0.05153020 & 0.23880347 & -0.07765949 \\ -0.03616408 & -0.04642844 & -0.07765949 & 0.24983189 \end{pmatrix},$$

auf deren Diagonale sich die quadrierten Standardfehler der Schätzer für β_j befinden, die man als Ergebnis des Befehls

```
> sqrt(diag(info))
(Intercept)      Yt_1          Yt_2          Yt_3
  0.3271479    0.4688302    0.4886752    0.4998319
```

erhält.

Nun wollen wir durch sukzessives Hinzunehmen der einzelnen Prädiktorvariablen das beste Modell herausfinden. Beginnen wir hierfür mit dem Modell, bei dem die Response nur vom Regen am Vortag abhängt.

$$\text{Modell 1: } \text{logit}(\mu(\beta)) = \beta_0 + \beta_1 Y_{t-1}$$

```
> summary(mod1 <- glm(Yt ~ Yt_1, family=binomial))
Call:
glm(formula = Yt ~ Yt_1, family = binomial)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3018  -0.8006  -0.8006   1.0579   1.6087
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.9734     0.2847  -3.419 0.000628 ***
Yt_1          1.2611     0.4222   2.987 0.002818 **
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 139.49  on 103  degrees of freedom
Residual deviance: 130.20  on 102  degrees of freedom
AIC: 134.2
Number of Fisher Scoring iterations: 4
```

Die Summary liefert uns für die Prädiktorvariable Y_{t-1} einen p-Wert von 0.0028, was wiederum einen signifikanten Einfluss bedeutet und $\beta_1 \neq 0$ plausibel macht. Nehmen wir zusätzlich zu diesem Modell die erklärende Variable Y_{t-2} in das Modell auf und vergleichen daher die zwei Modelle

$$\text{Modell 1: } \text{logit}(\mu(\beta)) = \beta_0 + \beta_1 Y_{t-1}$$

$$\text{Modell 2: } \text{logit}(\mu(\beta)) = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2}.$$

Wir erkennen anhand der folgenden Analysis-of-Deviance-Tabelle, dass Y_{t-2} zusätzlich zu Y_{t-1} keinen Einfluss auf die Response hat:

```
> anova(mod2 <- update(mod1, .~. + Yt_2), test="Chisq")
Analysis of Deviance Table
Model: binomial, link: logit
Response: Yt
Terms added sequentially (first to last)
    Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                103    139.49
Yt_1  1     9.2854     102    130.20  0.00231 **
Yt_2  1     0.2231     101    129.98  0.63671
```

Somit kann das dazugehörige $\beta_2 = 0$ gesetzt werden, da wir hier einen p-Wert von 0.6367 erhalten, falls wir Y_{t-2} zusätzlich zu Y_{t-1} in das Modell aufnehmen. Der Aufruf `test=Chisq` bedeutet, dass die Deviance-Differenz der ineinandergeschachtelten Modelle mit einem Skalenparameter = 1 mit einem χ^2 -Quantil verglichen wird.

Vergleichen wir nun mit dem ersten Modell ein weiteres, in dem statt Y_{t-2} die Variable Y_{t-3} hinzugefügt wird:

$$\text{Modell 1: } \text{logit}(\mu(\beta)) = \beta_0 + \beta_1 Y_{t-1}$$

$$\text{Modell 3: } \text{logit}(\mu(\beta)) = \beta_0 + \beta_1 Y_{t-1} + \beta_3 Y_{t-3}.$$

Wir erhalten folgendes Ergebnis:

```
> anova(mod3 <- update(mod1, .~. + Yt_3), test="Chisq")
```

```
Analysis of Deviance Table
```

```
Model: binomial, link: logit
```

```
Response: Yt
```

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

	Df	Deviance	Resid. Df	Resid. Dev	P(> Chi)
NULL			103	139.49	
Yt_1	1	9.2854	102	130.20	0.00231 **
Yt_3	1	5.2511	101	124.95	0.02193 *

```
> summary(mod3)
```

```
Call:
```

```
glm(formula = Yt ~ Yt_1 + Yt_3, family = binomial)
```

```
Deviance Residuals:
```

Min	1Q	Median	3Q	Max
-1.5315	-0.9063	-0.5741	0.8606	1.9412

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.6776	0.3096	-2.188	0.02864 *
Yt_1	1.4800	0.4539	3.261	0.00111 **
Yt_3	-1.0418	0.4717	-2.209	0.02719 *

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 139.49 on 103 degrees of freedom
```

```
Residual deviance: 124.95 on 101 degrees of freedom
```

```
AIC: 130.95
```

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

Die Analysis-of-Deviance-Tabelle weist einen p-Wert von 0.0219 auf, der durch den p-Wert von 0.0272 in der Summary von `mod3` bestätigt wird. Daher wird die erklärende Variable Y_{t-3} in das Modell zusätzlich zu Y_{t-1} aufgenommen und $\beta_3 \neq 0$ gesetzt, da Y_{t-3} einen signifikanten Einfluss auf die Response hat. Für den Schätzer von β ergibt sich: $\hat{\beta} = (-0.6776, 1.4800, -1.0418)$. Die Deviance von 124.95 ist im Vergleich zu der Anzahl der Freiheitsgrade von 101 deutlich größer, woraus wir eine Überdispersion feststellen, d.h. die Datenvariabilität ist größer als die Modellvariabilität.

Testen wir nun zu diesem Modell 3 die Hinzunahme von Y_{t-2} , da die Reihenfolge der hinzugenommenen Terme eine entscheidende Rolle spielt. So könnte nun der Einfluss der Variable Y_{t-2} signifikant sein.

$$\text{Modell 3: } \text{logit}(\mu(\beta)) = \beta_0 + \beta_1 Y_{t-1} + \beta_3 Y_{t-3}$$

$$\text{Modell 4: } \text{logit}(\mu(\beta)) = \beta_0 + \beta_1 Y_{t-1} + \beta_3 Y_{t-3} + \beta_2 Y_{t-2}.$$

```
> anova(mod4 <- update(mod3, .~. + Yt_2), test="Chisq")
Analysis of Deviance Table
Model: binomial, link: logit
Response: Yt
Terms added sequentially (first to last)
      Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                103    139.49
Yt_1  1     9.2854     102    130.20  0.00231 **
Yt_3  1     5.2511     101    124.95  0.02193 *
Yt_2  1     1.3218     100    123.63  0.25026
```

Doch diese Tabelle zeigt, dass sich obiges Ergebnis nicht ändert und die Nullhypothese $H_0: \beta_2 = 0$ nicht verworfen werden kann. Verglichen mit der Summary von `mod`, bei dem die selben Parameter - jedoch in anderer Reihenfolge - ins Modell aufgenommen wurden, ergibt sich für Y_{t-2} ein p-Wert von 0.252, der in etwa dem p-Wert von `mod4` (= 0.250) entspricht.

Als Informationskriterien für die Auswahl des besten Modells verwenden wir das *Akaike-* und das *Bayessche-Informationskriterium*.

Definition 5.2. Sei $\mathcal{L}(\hat{\mu}, y)$ die Log-Likelihood Funktion mit dem Maximum an der Stelle $\hat{\mu}$. Sei p die Modellordnung, d.h. $p = \dim(\beta)$, und n der Stichprobenumfang. So ist das Akaike-Informations-Kriterium definiert als

$$AIC(p) = -2 \log \mathcal{L}(\hat{\mu}, y) + 2p,$$

sowie das Bayesian-Information-Kriterium gegeben als

$$BIC(p) = -2 \log \mathcal{L}(\hat{\mu}, y) + p \log n.$$

Der jeweils erste Term kann als Maß der Modellanpassung und der zweite als eine Art „Bestrafung“ für eine große Anzahl von Modellparametern interpretiert werden. Das beste Modell ergibt sich nun mit minimalen AIC bzw. BIC.

Definition 5.2 kann auch für die Partial-Likelihood Funktion $PL(\hat{\mu}, y)$ mit $n = N$ der Länge der Zeitreihe entsprechend unserem Beispiel angewandt werden.

Fassen wir die Ergebnisse der vier Modelle in Tabelle 5.2 zusammen, wobei $MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{\mu}_i)^2$ die mittlere Fehlerquadratsumme bezeichnet.

Modell	MSE	X^2	D	df	AIC	BIC
1	0.2175	104.00	130.20	102	134.20	139.49
2	0.2172	103.85	129.98	101	135.98	143.91
3	0.2057	105.35	124.95	101	130.95	138.88
4	0.2037	104.08	123.63	100	131.63	142.20

Tabelle 5.2: Zusammenfassung der getesteten Modelle.

Diese Ergebnisse wurden mit der Funktion `diagnostic` erhalten:

```
diagnostic <- function(model){
MSE      <- mean(residuals(model, type="response")^2)
Chisq    <- sum(residuals(model, type="pearson" )^2)
D        <- model$deviance
df       <- model$df.residual
n        <- length(model$y)
p        <- n-df
AIC      <- AIC(model)
BIC      <- D + p*log(n)
out      <- c(p,WR,MSE,Chisq,D,df,AIC,BIC)
names(out) <- c("p", "WR", "MSE", "Chisq", "D", "df", "AIC", "BIC")
return(out)}
```

Mit den Kriterien aus Tabelle 5.2 sowie obigen Analysen ergibt sich das beste Modell mit

$$\text{logit}(\mu(\hat{\beta})) = -0.6776 + 1.4800Y_{t-1} - 1.0418Y_{t-3}.$$

Gehen wir nun von der starken Annahme weg, dass die Verteilung vollkommen spezifiziert ist, und versuchen den QL-Ansatz, bei dem wir nur von der Kenntnis der ersten beiden Momente ausgehen. Ein Modellaufruf kann in R mit dem Befehl

```
> modql<-glm(Yt~Yt_1+Yt_3, family=quasibinomial)
```

gemacht werden, wobei `family=quasibinomial` die Quasi-Likelihood Funktion für Daten Y mit $\text{var}(y) = \phi\mu(1 - \mu)$ verwendet. Eine Zusammenfassung ergibt folgendes Ergebnis:

```
> summary(modql)
Call:
glm(formula = Yt ~ Yt_1 + Yt_3, family = quasibinomial)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5315 -0.9063 -0.5741  0.8606  1.9412
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.6776     0.3162  -2.143  0.03454 *
Yt_1          1.4800     0.4636   3.193  0.00188 **
Yt_3         -1.0418     0.4817  -2.163  0.03293 *
(Dispersion parameter for quasibinomial family taken to be 1.043116)
Null deviance: 139.49  on 103  degrees of freedom
Residual deviance: 124.95  on 101  degrees of freedom
AIC: NA
Number of Fisher Scoring iterations: 4
```

Dies zeigt kaum einen Unterschied zum Modell mit dem gewöhnlichen GLM-Ansatz. Zu erwähnen ist jedoch, dass beim QL-Ansatz kein AIC-Wert ermittelt werden kann, da hier nicht die Verteilung bekannt ist. Der geschätzte Dispersionsparameter ergibt $\hat{\phi} = 1.043 > 1$, welcher durch die gemittelte Pearson-Statistik gegeben ist:

$$\hat{\phi} = \frac{1}{N-p} X^2$$

mit

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

der Pearson-Statistik, wie sie in Definition 2.13 beschrieben ist. Dieser geschätzte Dispersionsparameter kann auch mit der vorhin definierten Funktion `diagnostic` in R berechnet werden, indem der dritte Wert (X^2) durch den fünften Wert (der Anzahl der Freiheitsgrade) dividiert wird.

```
> d <- diagnostic(modq1)
> d[3]/d[5]
  Chisq
1.043116
```

Durch Verwendung des QL-Ansatzes verringert sich die Überdispersion, denn wir erhalten nun $D(\mu, y)/\hat{\phi} = 119.785$ im Vergleich zu 101 Freiheitsgraden, was einer deutlichen Modellverbesserung entspricht. Der geschätzte Dispersionsparameter $\hat{\phi} > 1$ zeigt eine größere Varianz als beim gewöhnlichen GLM-Ansatz. Die Standardfehler ergeben sich durch Multiplikation der Standardfehler im gewöhnlichen Poissonmodell mit $\sqrt{\hat{\phi}} = 1.0213$. Die Parameterschätzer beider Modelle bleiben gleich.

Betrachten wir ein 95%-Konfidenzintervall für den Erwartungswert μ_i , das durch die Gleichung

$$\text{KIV}(\mu_i) = \hat{\mu}_i \pm t_{N-p} \cdot \sqrt{\hat{v}ar(\hat{\mu}_i)}, \quad (5.2)$$

mit t_{N-p} dem Quantil der t -Verteilung (vgl. Casella und Berger, 2002) ermittelt wird. Da es nur vier mögliche Situationen als linearen Prädiktor gibt, definieren wir eine neue Variable `new.Y`,

```
> (new.Y <- data.frame(Yt_1=c(0,0,1,1), Yt_3=c(0,1,0,1)))
  Yt_1 Yt_3
1     0     0
2     0     1
3     1     0
4     1     1
```

wofür wir unter `mod3` jeweils eine Schätzung des Erwartungswertes mit den zugehörigen Standardfehlern erhalten

```

> (pred <- predict.glm(mod3, newdata = new.Y, type = "response",
                        se.fit = T))
$fit
      1      2      3      4
0.3368046 0.1519529 0.6905005 0.4404494
$se.fit
      1      2      3      4
0.06915767 0.06000623 0.08604501 0.09760733
$residual.scale
[1] 1

```

Unter Verwendung der Gleichung (5.2) für ein 95%-Konfidenzintervall erhalten wir mit dem R-Code

```

> df <- summary(mod3)$df.residual
> fit <- pred$fit
> upper <- fit + qt(0.975, df)*pred$se.fit
> lower <- fit - qt(0.975, df)*pred$se.fit
> plot(fit, ylim=c(0,1),
       ylab=expression(paste("95% Konfidenzintervalle für ", mu)),
       xlab="Prädiktor-Situationen", axes=FALSE)
> axis(1,at=1:4,lab=c(expression(paste(Y[t-1]==0, " ", " ", Y[t-3]==0)),
                       expression(paste(Y[t-1]==0, " ", " ", Y[t-3]==1)),
                       expression(paste(Y[t-1]==1, " ", " ", Y[t-3]==0)),
                       expression(paste(Y[t-1]==1, " ", " ", Y[t-3]==1))))
> axis(2, at=seq(0,1,0.2))
> points(upper, col="red")
> points(lower, col="red")

```

Abbildung 5.2 mit den Schranken

```

> lower
      1      2      3      4
0.19961438 0.03291671 0.51981037 0.24682274
> upper
      1      2      3      4
0.4739948 0.2709891 0.8611907 0.6340761

```

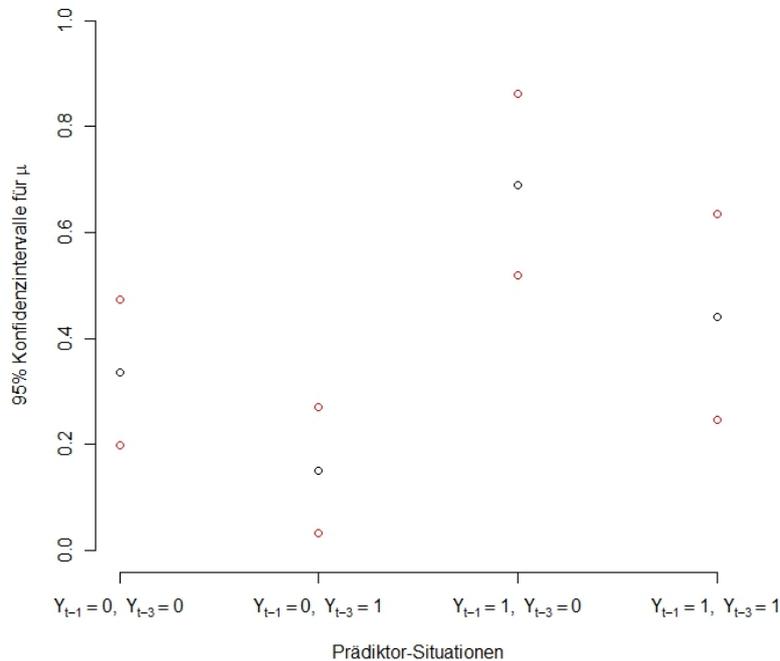


Abbildung 5.2: 95% Konfidenzintervall für Modell mit Regendaten.

Wir erkennen in Abbildung 5.2, dass die Wahrscheinlichkeit für das Eintreten von Regen am größten ist, wenn es gestern geregnet hat, vor drei Tagen aber kein Regen fiel, also mit der Situation $Y_{t-1} = 1, Y_{t-3} = 0$. Umgekehrt ist die Wahrscheinlichkeit von Regenfall heute für $Y_{t-1} = 0, Y_{t-3} = 1$ am kleinsten. Des Weiteren ist zu beobachten, dass die Konfidenzintervalle zu den Situationen $Y_{t-1} = 0, Y_{t-3} = 0$ und $Y_{t-1} = 1, Y_{t-3} = 0$ bzw. zu $Y_{t-1} = 0, Y_{t-3} = 1$ und $Y_{t-1} = 1, Y_{t-3} = 0$ disjunkt sind.

Wenden wir selben Code auf das QL-Modell `modq1` an, so ergeben sich nahezu die selben Schranken.

```
> lower
      1          2          3          4
0.19668807 0.03037763 0.51616949 0.24269262
> upper
      1          2          3          4
0.4769211 0.2735282 0.8648316 0.6382062
```

Zusammenfassend kann hier in diesem Fall somit kein großer Unterschied bezüglich der Varianzschätzer zwischen dem gewöhnlichen GLM und dem Modell mit dem QL-Ansatz festgestellt werden. Die QL Funktion modelliert die Überdispersion der Daten jedoch besser.

Wie bereits anfangs erwähnt, würde der EQL-Ansatz bei diesem Beispiel aufgrund des binären Datensatzes keinen Sinn machen. Denn für alle k und l in $V_{k,l}(\mu) = \mu^k(1-\mu)^l$ gilt für die Varianzfunktion $V_{k,l}(0) = 0$ und die EQL Funktion ist für $V_{k,l}(\cdot) = 0$ nicht definiert.

5.2 Aberrant Crypt Foci

„Aberrant Crypt Foci“ (ACF) sind die ersten Anzeichen, die im Dickdarm erkannt werden und zu Darmkrebs führen können. In einem Experiment wurden 22 Raten zu drei verschiedenen Zeitpunkten eine Dosis Karzinogen Azoxymethane (AOM) in den Dickdarm verabreicht. Die Anzahl der ACF wurde als Responsevariable Y in Abhängigkeit der Endzeit, der letzten Verabreichung von AOM (in Wochen) als erklärende Variable x modelliert.

Das erforderliche Paket in R ist `DAAG`, in dem die Daten enthalten sind, welche in Abbildung 5.3 veranschaulicht werden. Mit dem Aufruf

```
> str(ACF1)
'data.frame':  22 obs. of  2 variables:
 $ count  : num  1 3 5 1 2 1 1 3 1 2 ...
 $ endtime: num  6 6 6 6 6 6 6 12 12 12 ...
```

kann die Struktur des R Objektes gezeigt werden. Wir sehen, dass der Datensatz $n = 22$ Beobachtungen und die beiden Variablen `count` und `endtime` umfasst.

Betrachtet man Abbildung 5.3, so würde man von der Endzeit als dreistufigen Faktor ausgehen. Da wir jedoch am Einfluss der letzten Verabreichung von AOM auf die Anzahl der ACF nicht nur an den gegebenen drei Zeitpunkten 6, 12 und 18 Wochen, sondern kontinuierlich über die gesamte Zeit interessiert sind, betrachten wir ein Poissonmodell mit dem kanonischen Link $\log(\mu) = \eta$:

$$\text{Modell 1:} \quad \log(\mu) = \beta_0 + \beta_1 x.$$

Als Summary ergibt sich

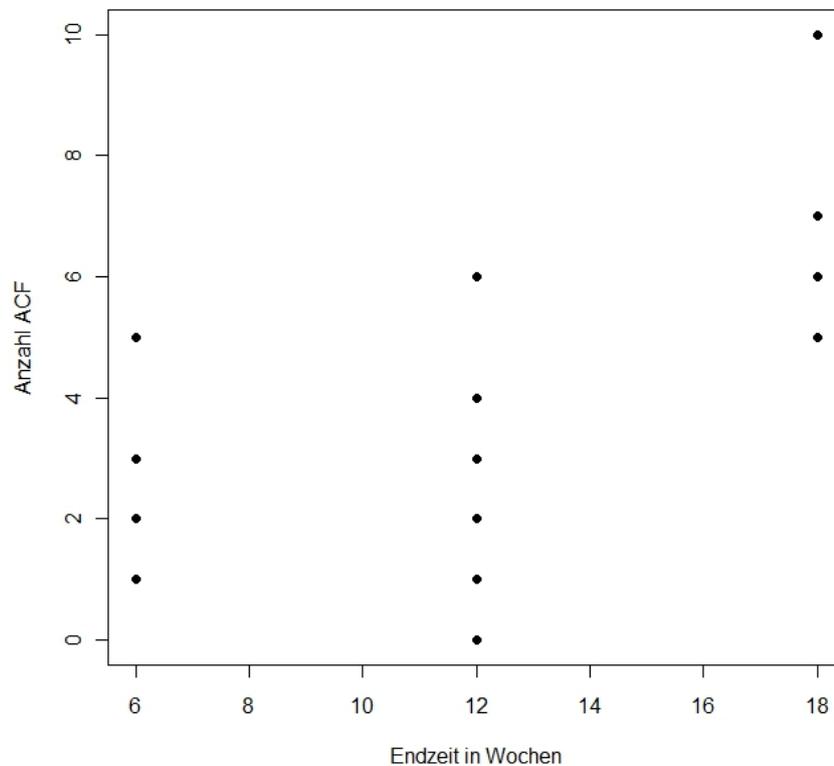


Abbildung 5.3: Anzahl von ACF in Abhängigkeit der Endzeit.

```
> summary(mod.poi1 <- glm(count ~ endtime, family=poisson))
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.46204 -0.47851 -0.07943  0.38159  2.26332
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.32152    0.40046  -0.803   0.422
endtime      0.11920    0.02642   4.511 6.44e-06 ***
(Dispersion parameter for poisson family taken to be 1)

Null deviance: 51.105  on 21  degrees of freedom
Residual deviance: 28.369  on 20  degrees of freedom
AIC: 92.21
Number of Fisher Scoring iterations: 5
```

Dieses Ergebnis zeigt einen signifikanten Einfluss der linearen Endzeit auf die zu erwartende Anzahl der ACF mit einer Deviance von 28.369 bei 20 Freiheitsgraden und einem AIC-Wert von 92.21. Als Schätzer für $\beta = (\beta_0, \beta_1)$ ergibt sich $\hat{\beta} = (-0.3215, 0.1192)$.

Als Nächstes nehmen wir auch den quadratischen Term in das Modell mit auf und betrachten somit

$$\text{Modell 2: } \log(\mu) = \beta_0 + \beta_1 x + \beta_2 x^2.$$

```
> summary(mod.poi2 <- glm(count ~ endtime + I(endtime^2),
  family=poisson))
```

Call:

```
glm(formula = count ~ endtime + I(endtime^2), family = poisson)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0616	-0.7834	-0.2808	0.4510	2.1693

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.722364	1.092494	1.577	0.115
endtime	-0.262356	0.199685	-1.314	0.189
I(endtime^2)	0.015137	0.007954	1.903	0.057 .

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 51.105 on 21 degrees of freedom

Residual deviance: 24.515 on 19 degrees of freedom

AIC: 90.354

Number of Fisher Scoring iterations: 5

```
> anova(mod.poi1, mod.poi2, test="Chisq")
```

Analysis of Deviance Table

Model 1: count ~ endtime

Model 2: count ~ endtime + I(endtime^2)

	Resid. Df	Resid. Dev	Df	Deviance	P(> Chi)
1	20	28.369			
2	19	24.515	1	3.8548	0.0496 *

Dies ergibt eine Deviancereduktion von $28.369 - 24.515 = 3.8548$ mit einem Verlust von 1 Freiheitsgrad und einem AIC-Wert von 90.354. Der Schätzer für $\beta = (\beta_0, \beta_1, \beta_2)$ ist $\hat{\beta} = (1.7224, -0.2624, 0.0151)$. Aufgrund der geringeren Deviance und des kleineren AIC-Wertes (90.35 im Vergleich zu 92.21.) scheint dieses Modell die Responsevariable besser zu beschreiben. Dies zeigt sich auch in Abbildung 5.4, in der die drei arithmetischen Mittelwerte und die Vorhersagen unter

beiden Modellen dargestellt sind. Es ist zu sehen, dass das grüne quadratische Modell genau die roten arithmetischen Mittel trifft und die mittleren Anzahlen der ACF somit sehr gut modelliert.

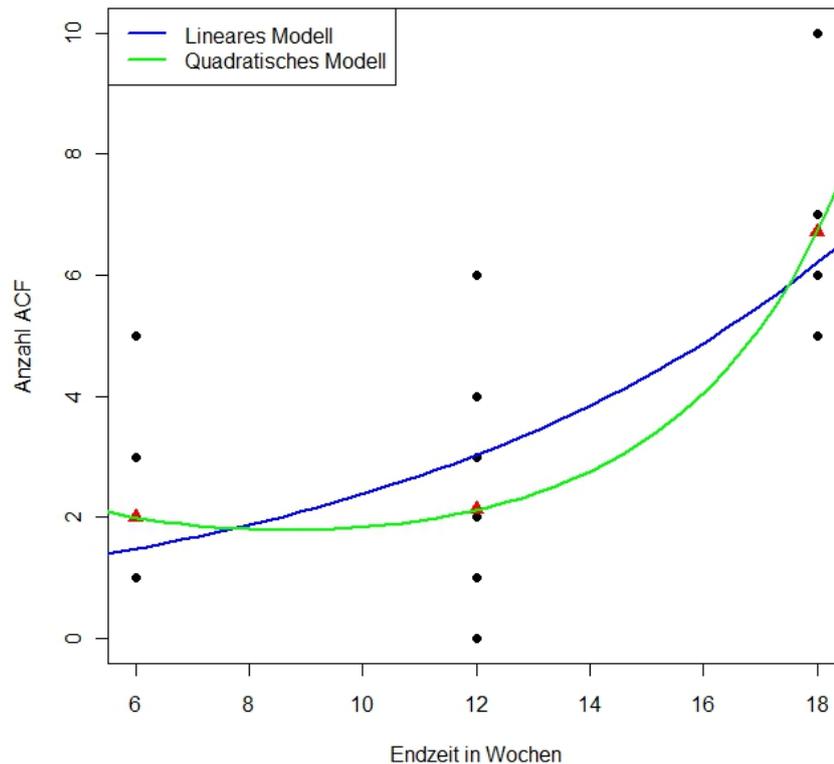


Abbildung 5.4: Arithmetische Mittel (rote Dreiecke) und Modellvorhersagen beider Poisson-Modelle.

Da die Deviance des Poissonmodells im Vergleich zur Anzahl der Freiheitsgrade deutlich größer ist und dies somit wiederum eine Überdispersion vermuten lässt, betrachten wir als nächstes den QL-Ansatz, der in R einfach mittels

```
> mod.q13<-glm(count~endtime+I(endtime^2), family=quasipoisson)
```

aufgerufen wird und als Ergebnis folgende Summary liefert.

```

> summary(mod.q13)
Call:
glm(formula = count ~ endtime + I(endtime^2), family = quasipoisson)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0616  -0.7834  -0.2808   0.4510   2.1693
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.722364   1.223433   1.408   0.175
endtime      -0.262356   0.223618  -1.173   0.255
I(endtime^2)  0.015137   0.008908   1.699   0.106
(Dispersion parameter for quasipoisson family taken to be 1.254071)
Null deviance: 51.105  on 21  degrees of freedom
Residual deviance: 24.515  on 19  degrees of freedom
AIC: NA
Number of Fisher Scoring iterations: 5

```

Zu beachten ist hierbei, dass der geschätzte Dispersionsparameter, der wiederum als gemittelte Pearson Statistik berechnet wird, $\hat{\phi} = X^2/(n - p) = 1.254 > 1$ ist, wodurch die Modellvariabilität mit dem QL-Ansatz größer als beim Poissonmodell erlaubt ist. Dass das QL-Modell die Daten besser beschreibt, zeigt auch der Vergleich $D(\mu, y)/\hat{\phi} = 19.55$ zur Anzahl der Freiheitsgrade bei 19. Die unskalierte Deviance ist beim zweiten und dritten Modell gleich, ebenso die Schätzer für β . Die Standardfehler werden jedoch größer (um den Faktor $\sqrt{1.25} = 1.12$ gestreckt):

```

> sqrt(summary(mod.q13)$dispersion)
[1] 1.119853

```

Deshalb scheint der Einfluss des quadratischen Terms mit dem QL-Ansatz mit einem p-Wert von 0.106 einen geringeren Einfluss zu haben. Hier kann wieder kein AIC-Wert berechnet werden, da keine exakte Verteilungsannahme getroffen wurde.

Die 95%-Konfidenzintervalle unter dem quadratischen Poissonmodell (schwarz) und dem Quasi-Poissonmodell (rot) zeigen sich in Abbildung 5.5. Zu sehen ist, dass das Konfidenzintervall mit dem QL-Ansatz etwas größer ist, was ein Resultat der um den Faktor 1.12 größeren Standardfehler ist.

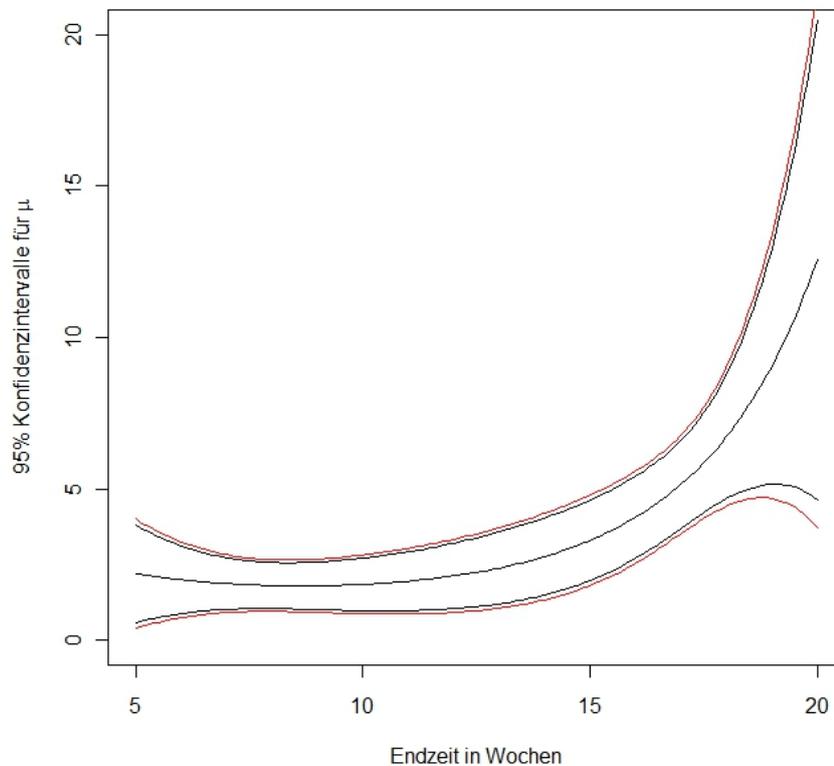


Abbildung 5.5: 95%-Konfidenzintervalle für das Poissonmodell (schwarz) und das Quasi-Poissonmodell (rot).

Verwenden wir schließlich die Power-Varianz $V_k(\mu) = \mu^k$, die in R mit zwei Funktionen umgesetzt werden kann:

- mit der Funktion `tweedie.profile`,
- mit der Funktion `eq1`.

Betrachten wir zunächst die erste Variante, bei der wir das R-Paket `tweedie` benötigen. Da der Datensatz zwei exakte Nullen enthält und die EQL Funktion für $V_k(0) = 0$ nicht definiert ist, setzen wir die Nullen auf NA („not available“).

```
count[count==0] <- NA
```

Des Weiteren haben wir das Problem, dass das Maximum für k -Werte, die näher als 1.1 bei 1 liegen (in unserem Fall 1.01), nach oben abreißt, da die Tweedie-

Dichte für $k \in (0, 1)$ nicht definiert ist (vgl. Kapitel 4, Tabelle 4.1). Dies ist in diesem Fall jedoch nur ein numerisches Problem und hat nicht wirklich etwas über das wahre Maximum der Profile-Likelihood Funktion zu sagen. Deswegen liegt das beobachtete scheinbare Maximum mit einem Wert von -36.55 bei etwa 1.1. Mit dem Aufruf

```
> tweedie1.01 <- tweedie.profile(count ~ endtime + I(endtime^2),
                               seq(1.01, 4, length=200), do.plot=T, do.smooth=T,
                               method="interpolation", fit.glm=T)
.....Done.
```

werden zwei Warnmeldungen (analog für $k = 1.1$)

Warnmeldungen:

```
1: In tweedie.profile(count ~ endtime + I(endtime^2), seq(1.01, 4, :
   True maximum possibly not detected.
2: In tweedie.profile(count ~ endtime + I(endtime^2), seq(1.01, 4, :
   Confidence interval cannot be found:
   insufficient data to find left CI.
```

ausgegeben. Die dazugehörigen Profile-Plots sind in der Abbildung 5.6 veranschaulicht, die auch gut erkennen lassen, dass keine linke Konfidenzschranke angegeben werden kann (wie bereits als Warnmeldung angekündigt).

Ein Plot der roten Tweedie-Dichte mit $k = 1.1$ im Vergleich zur schwarzen Poisson-Dichte zeigt eine sehr gute Anpassung (Abbildung 5.7).

Ein GLM mit der Tweedieverteilung mit $k = 1.1$ ergibt die Summary

```
> summary(mod.tweedie <- glm(count ~ endtime + I(endtime^2),
                             family=tweedie(var.power=1.1, link.power=0)))
Deviance Residuals:
   Min       1Q   Median       3Q      Max
-1.2138 -0.7650 -0.2558  0.2285  1.6902
```

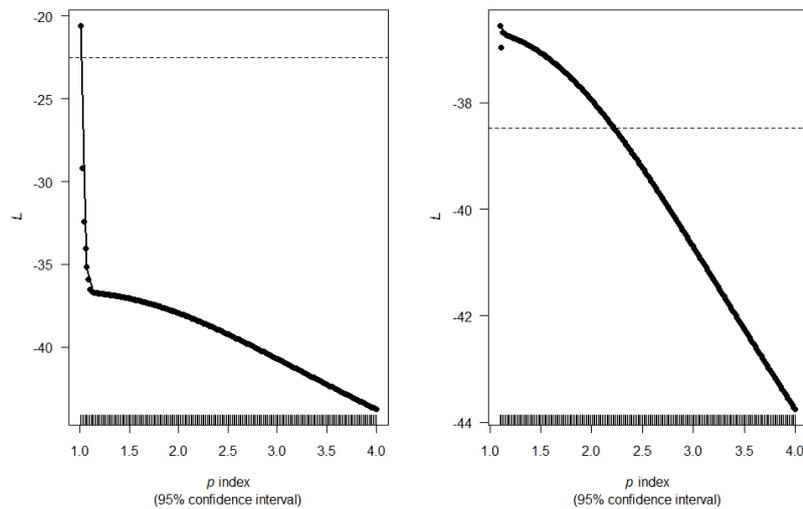


Abbildung 5.6: Profile-Plots für $k > 1.01$ (links) und $k > 1.1$ (rechts).

Coefficients:

```

      Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.859317   1.051020   0.818   0.425
endtime     -0.070568   0.193001  -0.366   0.719
I(endtime^2) 0.007146   0.007711   0.927   0.367
(Dispersion parameter for Tweedie family taken to be 0.8482529)
Null deviance: 32.020 on 19 degrees of freedom
Residual deviance: 13.278 on 17 degrees of freedom
(2 observations deleted due to missingness)
AIC: NA
Number of Fisher Scoring iterations: 5

```

Der AIC-Wert kann mittels

```

> AICtweedie(mod.tweedie)
[1] 82.14582

```

berechnet werden. Die Option `link.power=0` produziert den Log-Link. Der Parameterschätzer für β ergibt $\hat{\beta} = (0.859, -0.071, 0.0071)$. Um auch das Poisson- und Quasi-Poissonmodell mit dem Tweediemodell vergleichen zu können, betrachten wir die quadratischen Modelle ohne exakte Nullen im Datensatz. Dies liefert als Ergebnisse die zwei Summaries

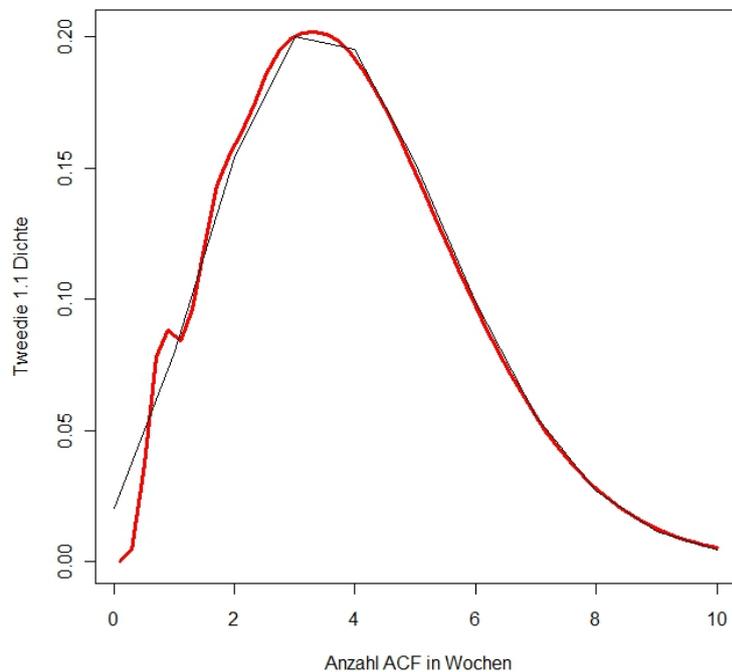


Abbildung 5.7: Dichte Funktionen der Tweedieverteilung mit $k = 1.1$ (rot) und Poisson-Dichte (schwarz).

```
> summary(mod.poisson <- glm(count ~ endtime + I(endtime^2),
  family=poisson))
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2585 -0.7834 -0.2808  0.2452  1.7785
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.859317  1.092495  0.787   0.432
endtime      -0.070568  0.199686 -0.353   0.724
I(endtime^2)  0.007146  0.007954  0.898   0.369
(Dispersion parameter for poisson family taken to be 1)
Null deviance: 36.237 on 19 degrees of freedom
Residual deviance: 14.733 on 17 degrees of freedom
(2 observations deleted due to missingness)
AIC: 80.573
Number of Fisher Scoring iterations: 5
```

```

> summary(mod.qpoisson <- glm(count ~ endtime + I(endtime^2),
                             family=quasipoisson))
Call:
glm(formula = count ~ endtime + I(endtime^2), family = quasipoisson)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2585  -0.7834  -0.2808   0.2452   1.7785
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.859317   1.058050   0.812   0.428
endtime      -0.070568   0.193390  -0.365   0.720
I(endtime^2)  0.007146   0.007704   0.928   0.367
(Dispersion parameter for quasipoisson family taken to be 0.9379371)
Null deviance: 36.237  on 19  degrees of freedom
Residual deviance: 14.733  on 17  degrees of freedom
(2 observations deleted due to missingness)
AIC: NA
Number of Fisher Scoring iterations: 5

```

mit den jeweiligen Standardfehlern

```

> sqrt(diag(summary(mod.poisson)$cov.scaled))
(Intercept)      endtime I(endtime^2)
1.092495076  0.199685681  0.007954336
> sqrt(diag(summary(mod.qpoisson)$cov.scaled))
(Intercept)      endtime I(endtime^2)
1.058050395  0.193389901  0.007703548
> sqrt(diag(summary(mod.tweedie)$cov.scaled))
(Intercept)      endtime I(endtime^2)
1.051020425  0.193000738  0.007711374

```

welche aus den Standardfehlern unter dem Poissonmodell wiederum durch Multiplikation mit $\sqrt{\hat{\phi}}$ des jeweiligen Modells erhalten werden und nahezu alle gleich sind. Dies zeigt sich auch in Abbildung 5.8, worin die 95%-Konfidenzintervalle für das Poisson- (schwarz), das Quasi-Poisson- (rot) und das Tweediemodell (blau) gleich sind.

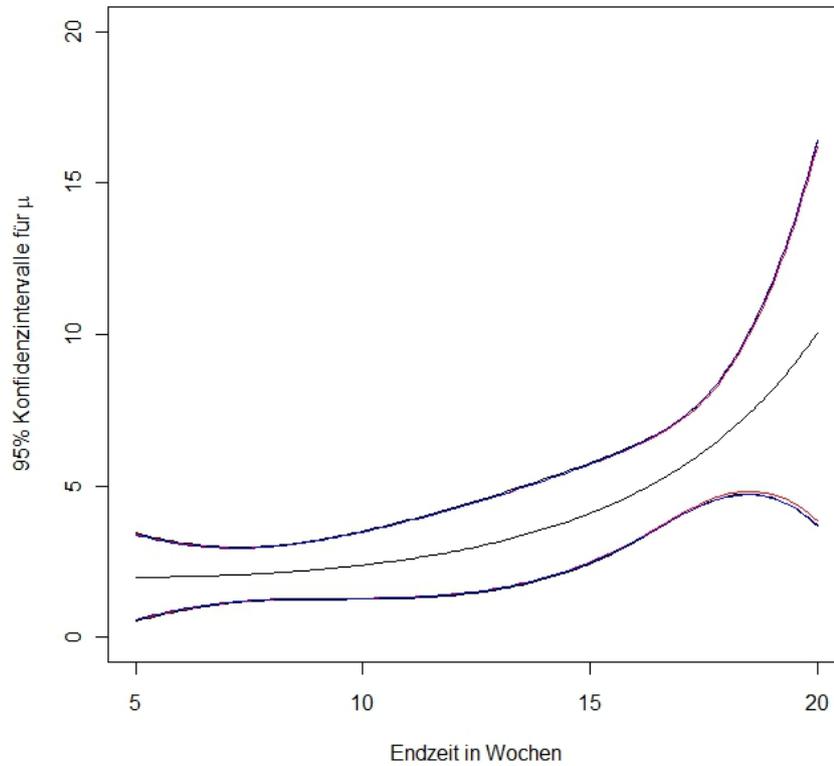


Abbildung 5.8: 95%-Konfidenzintervalle für das Poisson- (schwarz), Quasi-Poisson (rot) und Tweediemodell (blau).

Anhand Tabelle 5.3 sehen wir, dass sich mit dem Quasi-Poisson- sowie mit dem Tweedie-Ansatz eine deutliche Verbesserung der Modellanpassung ergeben.

	Poisson	Quasi-Poisson	Tweedie
$D(\mu, y)/\hat{\phi}$	14.73	15.71	15.66
df	17.00	17.00	17.00

Tabelle 5.3: Vergleich der skalierten Deviance Werte zu den Anzahlen der Freiheitsgraden (df).

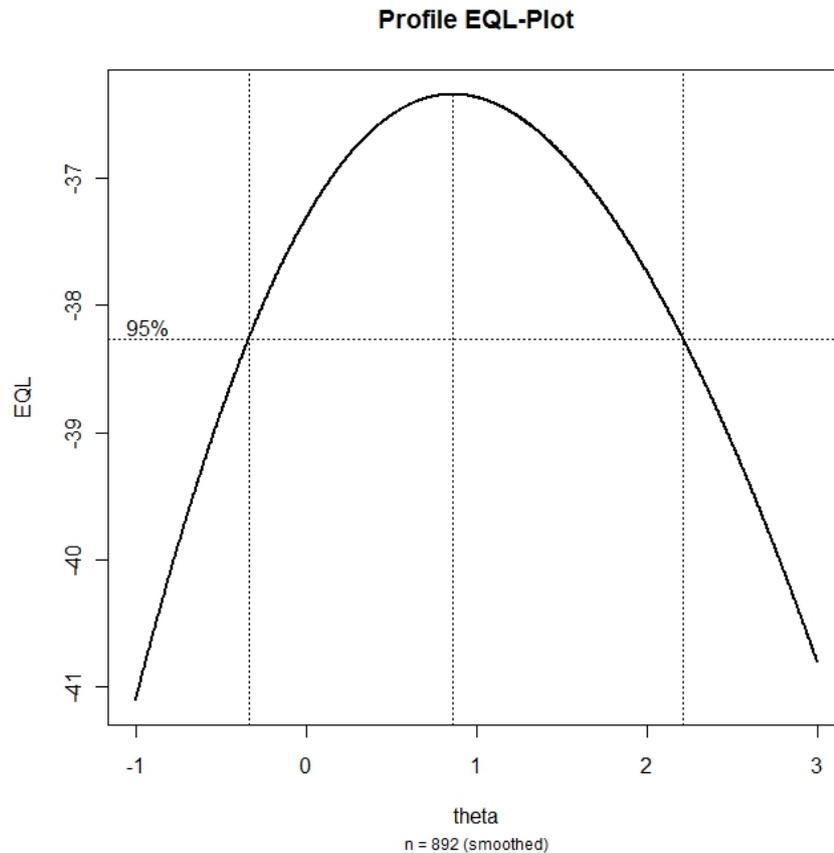


Abbildung 5.9: Profile-Plot für EQL Funktion mit $k \in [-1, 3]$.

Auffallend ist, dass egal welche Varianzstruktur angenommen wird, immer exakt dieselben Parameterschätzer erhalten werden.

```
> mod.poisson$coef
(Intercept)      endtime I(endtime^2)
0.859317370 -0.070568105  0.007145512
> mod.qpoisson$coef
(Intercept)      endtime I(endtime^2)
0.859317370 -0.070568105  0.007145512
> mod.tweedie$coef
(Intercept)      endtime I(endtime^2)
0.859317370 -0.070568105  0.007145512
```

Dies gilt sogar auch, wenn wir $k = 2$ setzen, wodurch wir diesselbe Varianzfunktion wie beispielsweise bei der Gammaverteilung erhalten.

```
> glm(count ~ endtime + I(endtime^2), family=Gamma(link="log"))
Coefficients:
 (Intercept)      endtime  I(endtime^2)
  0.859317      -0.070568      0.007146
Degrees of Freedom: 19 Total (i.e. Null);  17 Residual
(2 observations deleted due to missingness)
Null Deviance:      11.34
Residual Deviance:  5.761      AIC: 83.92
```

5.3 Schiffsschaden

Der Datensatz von McCullagh und Nelder (1989), Seiten 204f., ist ein bekanntes Beispiel für ein Quasi-Likelihood Modell. Als Responsevariable Y wird die Anzahl von Schiffsschäden (**Schaden**), die von 0 bis 58 reichen, abhängig von vier erklärenden Variablen, beschrieben:

- Schiffstyp (**Typ**): A-E,
- Baujahr (**Jahr**): 1960-64, 1965-69, 1970-74, 1975-79,
- Betriebsdauer (**Dauer**): 1960-1974, 1975-79,
- Gesamtmonate des Service (**Service**): 0-44.882.

Die Variablen **Typ**, **Jahr** sowie **Dauer** werden als Faktoren mit den Stufen 1 bis 5, 1 bis 4 und 1 bis 2 behandelt.

Der Datensatz selbst ist in der Tabelle 5.7 angeführt. Hierbei ist zu erwähnen, dass die mit * gekennzeichneten Zeilen, die keine Monate im Service und auch keinen Schaden aufweisen, einfach zu erklären sind, da diese Schiffe erst zwischen 1975 und 1979 gebaut wurden und somit in den Jahren 1960 bis 74 noch nicht in Betrieb genommen werden konnten.

5 Beispiele

Typ	Bau-jahr	Betriebs-dauer	Monate Service	Y	Typ	Bau-jahr	Betriebs-dauer	Monate Service	Y
A	1960-64	1960-74	127	0	C	1970-74	1960-74	783	6
A	1960-64	1975-79	63	0	C	1970-74	1975-79	1.948	2
A	1965-69	1960-74	1.095	3	C	1975-79	1960-74	0	0
A	1965-69	1975-79	1.095	4	C	1975-79	1975-79	274	1
A	1970-74	1960-74	1.512	6	D	1960-64	1960-74	251	0
A	1970-74	1975-79	3.353	18	D	1960-64	1975-79	105	0
A	1975-79	1960-74	0	0*	D	1965-69	1960-74	288	0
A	1975-79	1975-79	2.244	11	D	1965-69	1975-79	192	0
B	1960-64	1960-74	44.882	39	D	1970-74	1960-74	349	2
B	1960-64	1975-79	17.176	29	D	1970-74	1975-79	1.208	11
B	1965-69	1960-74	28.609	58	D	1975-79	1960-74	0	0*
B	1965-69	1975-79	20.370	53	D	1975-79	1975-79	2.051	4
B	1970-74	1960-74	7.064	12	E	1960-64	1960-74	45	0
B	1970-74	1975-79	13.099	44	E	1960-64	1975-79	0	0
B	1975-79	1960-74	0	0*	E	1965-69	1960-74	789	7
B	1975-79	1975-79	7.7117	18	E	1965-69	1975-79	437	7
C	1960-64	1960-74	1.179	1	E	1970-74	1960-74	1.157	5
C	1960-64	1975-79	552	1	E	1970-74	1975-79	2.161	12
C	1965-69	1960-74	781	0	E	1975-79	1960-74	0	0*
C	1965-69	1975-79	676	1	E	1975-79	1975-79	542	1

Tabelle 5.4: Daten für Schiffsschaden

In R werden die Daten mit dem Code

```
schiff.dat<-data.frame(Schaden=c(0,0,3,4,6,18,0,11,39,29,58,53,12,
                                44,0,18,1,1,0,1,6,2,0,1,0,0,0,0,2,
                                11,0,4,0,0,7,7,5,12,0,1),
                      Service=c(127,63,1095,1095,1512,3353,0,2244,
                                44882,17176,28609,20370,7064,13099,
                                0,7117,1179,552,781,676,783,1948,0,
                                274,251,105,288,192,349,1208,0,2051,
                                45,0,789,437,1157,2161,0,542) ,
                      Dauer=factor(rep(c(1,2),20,each=1)),
                      Jahr=factor(rep(c(1,2,3,4),5,each=2)),
                      Typ=factor(rep(c(1,2,3,4,5),1,each=8)))
```

eingesehen.

Zur Veranschaulichung werden die Daten einzeln mit der Responsevariable in Abbildung 5.10 geplottet.

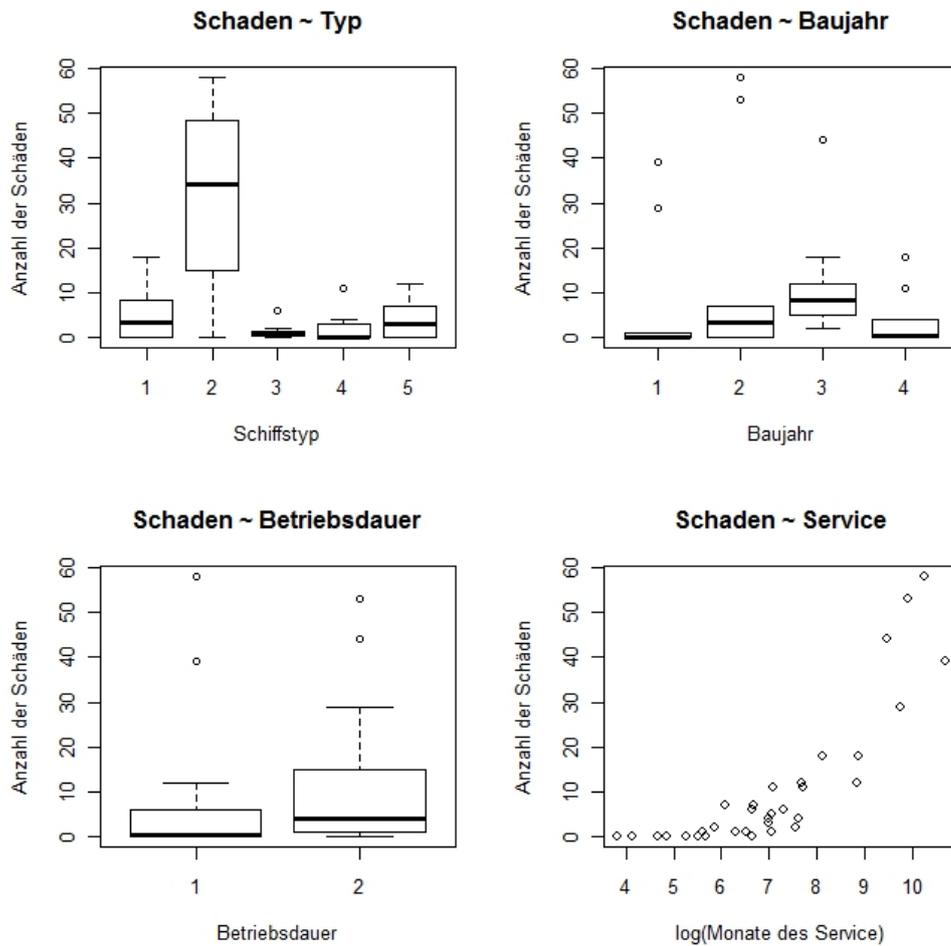


Abbildung 5.10: Schiffsdaten mit Merkmalen Typ, Baujahr, Betriebsdauer und Monate des Service mit Response „Anzahl der Schäden“.

McCullagh und Nelder (1989) verwenden als ungeordneten Kontrast „treatment“, welcher mit dem Code

```
> options(contrasts=c("contr.treatment", "contr.poly"))
```

eingestellt werden kann bzw. bereits die default Einstellung in R ist.

Unser Modell lautet nun

$$\log(\mu) = \beta_0 + \beta_1 \cdot \text{Typ} + \beta_2 \cdot \text{Jahr} + \beta_3 \cdot \text{Dauer} + 1 \cdot \log(\text{Service}),$$

wobei die letzte Komponente $\log(\text{Service})$ ein quantitatives Merkmal ist, deren Regressionskoeffizient als bekannt mit 1 angesehen wird. Daraus ergibt sich die zu oben äquivalente Formulierung:

$$\log\left(\frac{\mu}{\text{Service}}\right) = \beta_0 + \beta_1 \cdot \text{Typ} + \beta_2 \cdot \text{Jahr} + \beta_3 \cdot \text{Dauer}.$$

Dies kann in R mit dem Befehl `offset(log1p(Service))` direkt im Modellaufruf umgesetzt werden.

```
> mod<-glm(Schaden~Typ+Jahr+Dauer+offset(log1p(Service)),
           family=poisson)
```

Die Funktion `log1p(x)` berechnet $\log(1+x)$ für $|x| < 1$ exakt, jedoch für $x \approx -1$ weniger exakt, d.h. `log1p` geht mit einer Null als Argument so um, dass $\log(0) = 0$ gilt, was mit der gewöhnlichen Log Funktion $-\infty$ ergibt.

Als Ergebnis liefert die Summary

```
> summary(mod)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.67589 -0.65993 -0.09363  0.37456  2.79099

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.40691    0.21743  -29.467  < 2e-16 ***
Typ2        -0.54261    0.17758   -3.056  0.00225 **
Typ3        -0.68791    0.32904   -2.091  0.03656 *
Typ4        -0.07675    0.29057   -0.264  0.79166
Typ5         0.32499    0.23588    1.378  0.16826
Jahr2        0.69721    0.14964    4.659  3.17e-06 ***
Jahr3        0.81880    0.16975    4.824  1.41e-06 ***
Jahr4        0.45335    0.23313    1.945  0.05182 .
Dauer2       0.38463    0.11826    3.252  0.00114 **
```

```
(Dispersion parameter for poisson family taken to be 1)
Null deviance: 146.29 on 39 degrees of freedom
Residual deviance: 38.73 on 31 degrees of freedom
AIC: 154.60
Number of Fisher Scoring iterations: 5
```

Die Deviance von 38.73 deutet im Vergleich zur Anzahl der Freiheitsgrade von 31 wiederum auf eine Überdispersion hin. Wenn man einen Parameter zum Prädiktor $\log(\text{Service})$ schätzen würde, so erhält man 0.9057 mit einem Standardfehler von 0.1160.

Mit dieser Zusammenfassung des Modells ist jedoch der Einfluss der einzelnen Faktoren bzw. Faktorstufen auf die Responsevariable schwer zu interpretieren, da die einzelnen Stufen jeweils eine bestimmte Referenzstufe verwenden, wobei für verschiedene Vergleichsstufen, verschiedene Ergebnisse erhalten werden. Dies hat aber keine Aussage über den signifikanten Einfluss des Faktors auf die Response. Aus diesem Grund bedienen wir uns der Analysis-of-Deviance Tabelle, um nun den Einfluss der einzelnen Faktoren herauszufinden:

```
> anova(mod, test="Chisq")
Analysis of Deviance Table
Model: poisson, link: log
Response: Schaden
Terms added sequentially (first to last)
      Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                                39    146.292
Typ    4    55.321         35     90.971 2.782e-11 ***
Jahr   3    41.569         32     49.402 4.953e-09 ***
Dauer  1    10.672         31     38.730 0.001088 **
```

Diese ANOVA zeigt einen signifikanten Einfluss aller drei Faktoren Typ, Jahr und Dauer.

Bemerkung 5.3. *Da die Reihenfolge der Hinzunahme einzelner Faktoren entscheidend ist, wurde dies auch mit unterschiedlichen Sequenzen durchgeführt, was jedoch wieder zum selben Ergebnis geführt hat, dass alle drei Faktoren einen signifikanten Einfluss auf die Response haben.*

Aufgrund der Überdispersion versuchen wir die Modellanpassung zu verbessern und betrachten den QL-Ansatz.

```

> summary(modql<-glm(Schaden~Typ+Jahr+Dauer+offset(log1p(Service)),
  family=quasipoisson))
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.67589 -0.65993 -0.09363  0.37456  2.79099
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.40691    0.25392 -25.232 < 2e-16 ***
Typ2        -0.54261    0.20738  -2.616 0.013609 *
Typ3        -0.68791    0.38427  -1.790 0.083197 .
Typ4        -0.07675    0.33934  -0.226 0.822542
Typ5         0.32499    0.27547   1.180 0.247056
Jahr2        0.69721    0.17475   3.990 0.000376 ***
Jahr3        0.81880    0.19824   4.130 0.000254 ***
Jahr4        0.45335    0.27226   1.665 0.105966
Dauer2       0.38463    0.13811   2.785 0.009046 **
(Dispersion parameter for quasipoisson family taken to be 1.363865)
Null deviance: 146.29 on 39 degrees of freedom
Residual deviance: 38.73 on 31 degrees of freedom
AIC: NA
Number of Fisher Scoring iterations: 5

```

Dies ergibt einen geschätzten Dispersionsparameter $\hat{\phi} = 1.364$, wodurch sich eine skalierte Deviance von 28.40 ergibt (im Vergleich zu 31 Freiheitsgraden), wodurch wir eine bessere Modellanpassung erhalten. Die Standardfehler sind wiederum mit dem Faktor $\sqrt{1.364} = 1.168$ gestreckt. Die ANOVA zu diesem Modell liefert

```

> anova(modql, test="Chisq")
Analysis of Deviance Table
Model: quasipoisson, link: log
Response: Schaden
Terms added sequentially (first to last)
      Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                                39    146.292
Typ   4    55.321         35    90.971 3.312e-08 ***
Jahr  3    41.569         32    49.402 1.094e-06 ***
Dauer 1    10.672         31    38.730 0.005154 **

```

Wir erkennen, dass auch mit dem QL-Ansatz alle drei Faktoren einen signifikanten Einfluss auf die Response der Anzahl der Schiffsschäden hat.

Betrachten wir nun den Einfluss der einzelnen Faktoren auf die Anzahl der Schiffsdaten mit den Funktionen `tweedie.profile` und `eql`.

Für das Tweediemodell benötigen wir das entsprechende Paket und die exakten Nullen müssen auf NA gesetzt werden.

```
> library(tweedie)
> Schaden[Schaden==0] <- NA
```

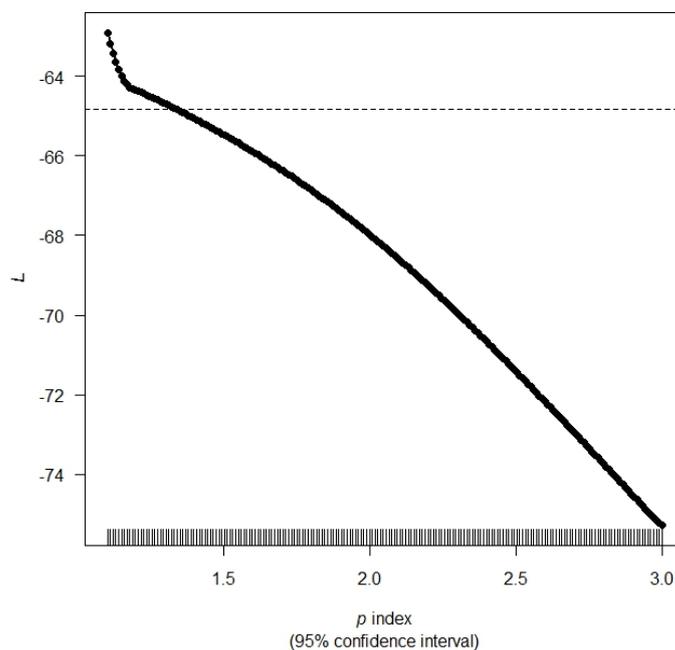
Mit der Annahme der Tweedieverteilung ergibt sich wiederum das Problem, dass k größer 1 gewählt werden muss und falls der Schätzer in die Nähe von 1 kommt, erhalten wir ein numerisches Problem, wodurch das Maximum der Profile-Likelihood Funktion nach oben geht. Deswegen wählen wir als untere Schranke für k den Wert 1.1 und erhalten mit dem Aufruf

```
> tweedie <- tweedie.profile(Schaden~Typ+Jahr+Dauer
+offset(log1p(Service)),
seq(1.1, 3, length=200), do.plot=T, do.smooth=T,
method="interpolation", fit.glm=T)
> (c(tweedie$p.max, tweedie$L.max))
[1] 1.10000 -62.90597
```

den maximalen Profile-Likelihood Wert bei -62.906 an der Randstelle $k = 1.1$. Den dazugehörigen Profile-Plot sehen wir in Abbildung 5.11.

Eine ANOVA und Summary liefern

```
> anova(mod.tweedie, test="Chisq")
Analysis of Deviance Table
Model: Tweedie, link: mu^0
Response: Schaden
Terms added sequentially (first to last)
      Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                25    107.348
Typ                4     48.792    21     58.556 1.001e-05 ***
Jahr               3     26.568    18     31.988 0.001434 **
Dauer              1      6.793    17     25.195 0.046486 *
```

Abbildung 5.11: Profile-Plot für Tweedie-Verteilung mit $k \in [1.1, 3]$.

```
> summary(mod.tweedie<-glm(Schaden~Typ+Jahr+Dauer
+  offset(log1p(Service)),
+  family=tweedie(var.power=1.1,
+  link.power=0)))
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-1.82708	-0.65840	0.01047	0.67641	2.39037

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-6.3430	0.3295	-19.248	5.59e-13	***
Typ2	-0.5815	0.2643	-2.201	0.04188	*
Typ3	-0.5404	0.4535	-1.192	0.24975	
Typ4	0.1007	0.4230	0.238	0.81465	
Typ5	0.3388	0.3423	0.990	0.33611	
Jahr2	0.7100	0.2336	3.039	0.00742	**
Jahr3	0.7751	0.2613	2.966	0.00866	**
Jahr4	0.3832	0.3548	1.080	0.29524	
Dauer2	0.3575	0.1807	1.979	0.06431	.

```
(Dispersion parameter for Tweedie family taken to be 1.713751)
  Null deviance: 107.348  on 25  degrees of freedom
Residual deviance: 25.195  on 17  degrees of freedom
  (14 observations deleted due to missingness)
AIC: NA
Number of Fisher Scoring iterations: 5
```

Um wiederum alle drei Modelle miteinander vergleichen zu können, überschreiben wir die exakten Nullen des Datensatzes mit NA und führen die Analyse mit dem Poisson- und Quasi-Poissonmodell erneut durch.

Poissonmodell:

```
> summary(mod<-glm(Schaden~Typ+Jahr+Dauer+offset(log1p(Service)),
  family=poisson))
Call:
glm(formula = Schaden ~ Typ + Jahr + Dauer
  + offset(log1p(Service)), family = poisson)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.87835 -0.72591  0.00641  0.72633  2.60505
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.35229    0.21926 -28.972 < 2e-16 ***
Typ2         -0.57965    0.17860  -3.245  0.00117 **
Typ3        -0.58809    0.32930  -1.786  0.07412 .
Typ4         0.08364    0.29306   0.285  0.77534
Typ5         0.31686    0.23584   1.344  0.17910
Jahr2        0.70460    0.14992   4.700 2.60e-06 ***
Jahr3        0.77738    0.17141   4.535 5.76e-06 ***
Jahr4        0.39339    0.23705   1.660  0.09701 .
Dauer2       0.36995    0.11840   3.125  0.00178 **
(Dispersion parameter for poisson family taken to be 1)
  Null deviance: 138.125  on 25  degrees of freedom
Residual deviance: 29.638  on 17  degrees of freedom
  (14 observations deleted due to missingness)
AIC: 145.50
Number of Fisher Scoring iterations: 5
```

Quasi-Poissonmodell:

```
> summary(modql<-glm(Schaden~Typ+Jahr+Dauer
+offset(log1p(Service)),family=quasipoisson))
Call:
glm(formula = Schaden ~ Typ + Jahr + Dauer
+ offset(log1p(Service)), family = quasipoisson)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.87835 -0.72591  0.00641  0.72633  2.60505
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.35229    0.31247 -20.329 2.29e-13 ***
Typ2        -0.57965    0.25453  -2.277  0.03598 *
Typ3        -0.58809    0.46930  -1.253  0.22713
Typ4         0.08364    0.41764   0.200  0.84365
Typ5         0.31686    0.33611   0.943  0.35901
Jahr2        0.70460    0.21366   3.298  0.00425 **
Jahr3        0.77738    0.24429   3.182  0.00545 **
Jahr4        0.39339    0.33782   1.164  0.26031
Dauer2       0.36995    0.16873   2.193  0.04255 *
(Dispersion parameter for quasipoisson family taken to be 2.031006)
Null deviance: 138.125 on 25 degrees of freedom
Residual deviance: 29.638 on 17 degrees of freedom
(14 observations deleted due to missingness)
AIC: NA
Number of Fisher Scoring iterations: 5
```

Die Standardfehler ergeben sich durch Multiplikation der Standardfehler des Poissonmodells mit dem entsprechenden geschätzten Dispersionsparameter (bei Quasi-Poissonmodell mit $\sqrt{2.031} = 1.425$ und beim Tweedmodell mit $\sqrt{1.714} = 1.309$.)

```
> sqrt(diag(summary(mod)$cov.scaled))
(Intercept)      Typ2      Typ3      Typ4      Typ5
 0.2192560  0.1786031  0.3293050  0.2930560  0.2358414
      Jahr2      Jahr3      Jahr4      Dauer2
0.1499205  0.1714140  0.2370453  0.1183978
> sqrt(diag(summary(modql)$cov.scaled))
(Intercept)      Typ2      Typ3      Typ4      Typ5
 0.3124691  0.2545333  0.4693036  0.4176440  0.3361055
```

```

      Jahr2      Jahr3      Jahr4      Dauer2
0.2136567  0.2442879  0.3378212  0.1687327
> sqrt(diag(summary(mod.tweedie)$cov.scaled))
(Intercept)      Typ2      Typ3      Typ4      Typ5
  0.3295428  0.2642605  0.4534844  0.4230309  0.3422528
      Jahr2      Jahr3      Jahr4      Dauer2
0.2336432  0.2613461  0.3548213  0.1806849

```

Ein Vergleich der Modelle mittels skaliertes Deviance zeigt sich in Tabelle 5.5.

	Poisson	Quasi-Poisson	Tweedie
$D(\mu, y)/\hat{\phi}$	29.638	14.593	14.702
df	17.00	17.00	17.00

Tabelle 5.5: Vergleich der skalierten Deviance Werte zu den Anzahlen der Freiheitsgraden (df).

Es zeigt sich eine deutliche Verbesserung der Modellanpassung mit dem Quasi-Poisson- bzw. dem Tweediemodell.

Verwenden wir nun die `eql` Funktion, so laden wir das dafür benötigte R-Paket `EQL`. Da wir hier nur die ersten beiden Momente anstatt der gesamten Verteilung spezifizieren, können wir für k beliebige Werte zulassen, was eine entscheidende Verbesserung bringt.

```

> k <- list(theta=seq(-1, 2, length = 50))
> eq.power <- eql(Schaden~Typ+Jahr+Dauer+offset(log1p(Service)),
                  param.space = k, family=powerVarianceFamily("log"),
                  smooth.grid=100)
~~ Extended Quasi Likelihood ~~
* Dimension of parameter space:  50
* Starting loop:
0%          50%          100%
+-----+-----+
|          |          |
* End loop
* Start smoothing:
Smoothing on an equidistant grid between EQL values with 100 nodes
* End smoothing

```

Betrachten wir die Ergebnisse, so sehen wir, dass der maximale EQL Wert von -63.5 an der Stelle $k = 0.349$ angenommen wird.

```
> eq.power$eql.max
[1] -63.4998
> eq.power$param.max
      theta
0.3487941
```

Die Schätzer der Faktoren ergeben sich zu

```
> eq.power$model$coef
(Intercept)      Typ2      Typ3      Typ4      Typ5
-6.41519495 -0.54679740 -0.93050205 -0.01355561  0.17329636
      Jahr2      Jahr3      Jahr4      Dauer2
0.69879739  0.83466339  0.50244466  0.38707943
```

Abbildung 5.12 zeigt den dazugehörigen Profile-Plot für die Power-Varianz mit $k \in [-1, 2]$.

5.4 Stoffdaten

Dieses Beispiel beinhaltet einen Datensatz von Hinde (1982) über die Länge eines Stoffes l , dessen Einfluss auf die Responsevariable, der Anzahl der Fehler eines Stoffes f , untersucht wird. Das Einlesen der Daten geschieht in R mit dem Code

```
l <- c(551,543,651,842,832,905,375,542,715,522,868,122,271,657,630,
      170,491,738,372,371,645,735,441,749,895,495,458,716,642,952,
      492,417)
f <- c( 6,  8,  4,  9, 17, 23,  9,  9, 14,  6,  8,  1,  5,  9,  7,
      4,  7,  9,  7, 14,  6, 17,  8, 10, 28,  7,  4,  3, 10,  9,
      4,  2)
```

Ein Plot der Daten lässt eine Abhängigkeit zwischen der Anzahl der erwarteten Fehler des Stoffes und der Länge erkennen (siehe Abbildung 5.13).

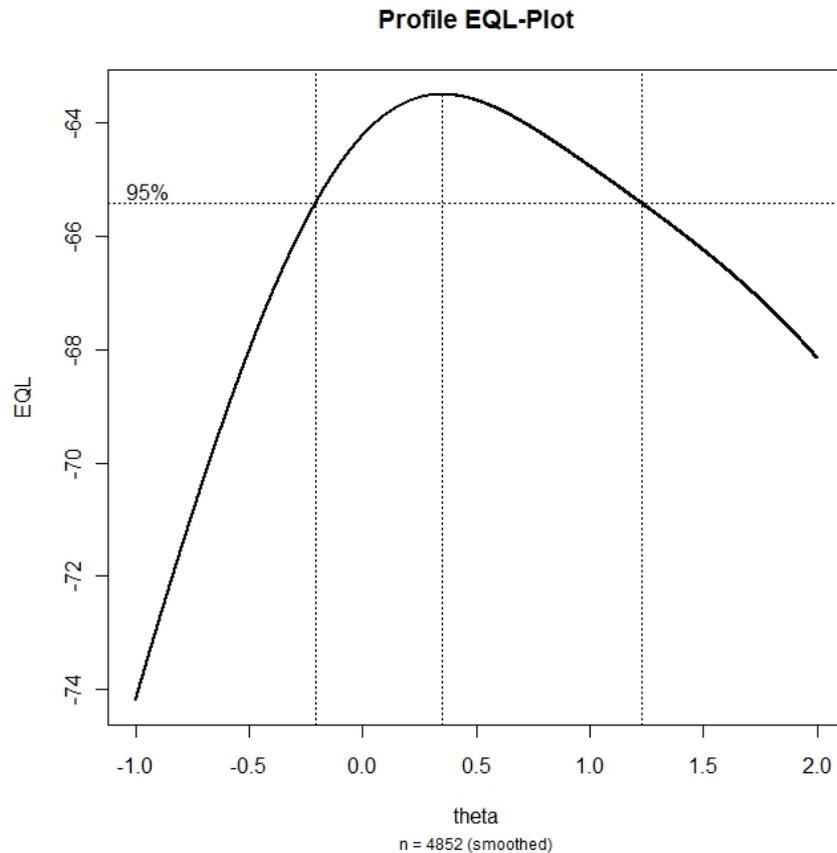


Abbildung 5.12: Profile-Plot für das EQL Modell mit Varianzfunktion $V_k(\mu) = \mu^k$.

Bei diesem Beispiel wird getestet, ob die erwartete Anzahl der Fehler proportional zur Länge ist. Betrachten wir daher ein Modell mit dem Log-Link und $\log(l)$ als Prädiktor und erhalten durch Umformen die Gleichung

$$\mu = e^{\beta_0} \cdot l^{\beta_1}.$$

Die erwartete Anzahl der Fehler μ ist daher genau dann mit der Konstante e^{β_0} proportional zur Länge l , wenn $\beta_1 = 1$ gilt. Da die Anzahl der Fehler modelliert wird, betrachten wir zunächst ein einfaches Poissonmodell:

$$\log(\mu) = \beta_0 + \beta_1 \cdot \log(l).$$

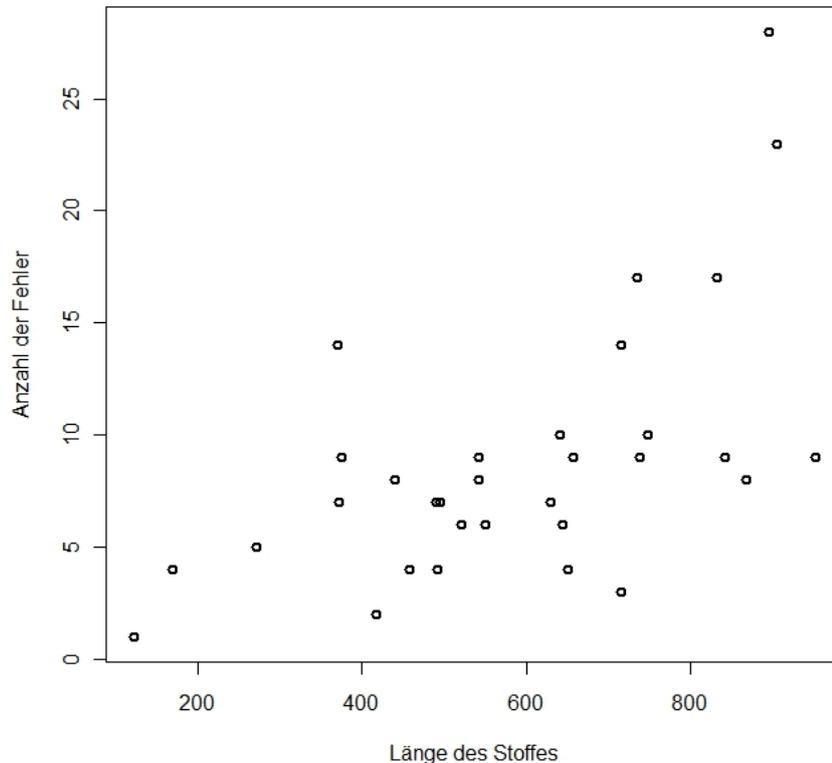


Abbildung 5.13: Beobachtete Anzahl der Fehler des Stoffes in Abhängigkeit der Länge.

```
> summary(mod <- glm(f ~ log(l), family=poisson(link=log)))
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8156 -1.1248 -0.2375  0.6227  3.4422
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.1730     1.1352  -3.676 0.000237 ***
log(l)         0.9969     0.1759   5.668 1.45e-08 ***
(Dispersion parameter for poisson family taken to be 1)
    Null deviance: 103.714  on 31  degrees of freedom
Residual deviance:  64.537  on 30  degrees of freedom
AIC: 191.84
Number of Fisher Scoring iterations: 4
```

Diese Summary liefert den Parameterschätzer $\hat{\beta} = (-4.1730, 0.9969)$, wobei der Schätzer für die Länge fast 1 ergibt. Somit erhalten wir das Modell

$$\log(\hat{\mu}) = -4.17 + 1 \cdot \log(l).$$

Der Plot der gefitteten Werte gegen die Länge des Stoffes zeigt sich in Abbildung 5.14, worin wir die gute Anpassung des Poissonmodells feststellen können.

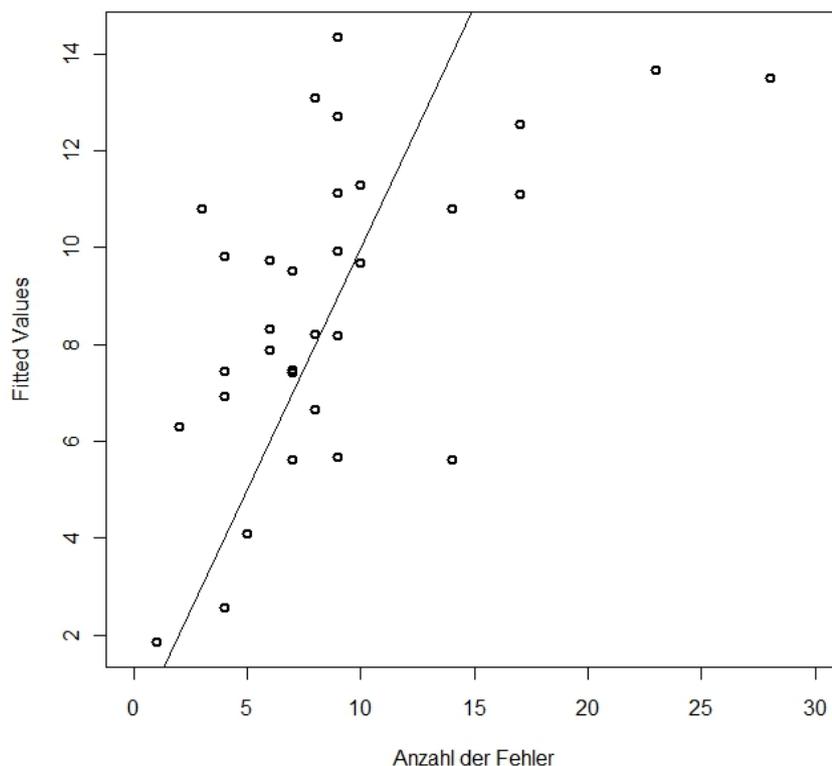


Abbildung 5.14: Fitted Values.

Jedoch ist die Deviance von 64.537 im Verhältnis zur Anzahl der Freiheitsgrade von 30 sehr groß, was wiederum auf eine Überdispersion der Daten schließen lässt. Verwenden wir deswegen das QL-Modell. Dies liefert uns folgende Zusammenfassung.

```

> summary(mod.ql <- glm(f ~ log(1), family=quasipoisson))
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8156 -1.1248 -0.2375  0.6227  3.4422
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -4.1730     1.7094  -2.441 0.020752 *
log(1)         0.9969     0.2649   3.764 0.000727 ***
(Dispersion parameter for quasipoisson family taken to be 2.267506)
Null deviance: 103.714  on 31  degrees of freedom
Residual deviance:  64.537  on 30  degrees of freedom
AIC: NA
Number of Fisher Scoring iterations: 4

```

Die Schätzer sind wiederum dieselben wie beim Poissonmodell, jedoch ergibt sich ein deutlich größerer geschätzter Dispersionsparameter mit $\hat{\phi} = 2.2675$. Betrachten wir die skalierte Deviance $D(\mu, y)/\hat{\phi}$, so ergibt sich 28.4617 im Vergleich zu 30 Freiheitsgraden, wodurch die Überdispersion der Daten viel besser modelliert wird, da in diesem Modell eine größere Varianz erlaubt ist. Die Standardfehler des QL-Modells verhalten sich wiederum im selben Verhältnis wie die Wurzel der Dispersionsparameter des gewöhnlichen Poissonmodells zum Quasi-Poissonmodell. Mit anderen Worten: Multiplizieren wir die Standardfehler des Poissonmodells mit $\sqrt{2.268} = 1.506$, so erhalten wir die Standardfehler des QL-Modells.

Betrachten wir nun auch die Power-Varianz $V_k(\mu) = \mu^k$ und die Tweedie Funktion in R. Wir wählen das Intervall für $k \in [1.1, 3]$.

```

> library(tweedie)
Lade nötiges Paket: statmod
> tweedie <- tweedie.profile(f ~ log(1), seq(1.1, 3, length=200),
    do.plot=T,do.smooth=T,method="interpolation",fit.glm=T)
.....Done.

```

Das dazugehörige Modell liefert

```

> summary(mod.tweedie <- glm(f ~ log(1),
    family=tweedie(var.power=tweedie$p.max, link.power=0)))

```

```

Call:
glm(formula = f ~ log(1), family = tweedie(var.power = tweedie$p.max,
      link.power = 0))
Deviance Residuals:
      Min       1Q   Median       3Q      Max
-1.01936 -0.39785 -0.08176  0.20792  1.00000
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.7325     1.2258  -3.045  0.00481 **
log(1)         0.9281     0.1945   4.771 4.44e-05 ***
(Dispersion parameter for Tweedie family taken to be 0.2410838)
Null deviance: 11.8986 on 31 degrees of freedom
Residual deviance: 7.2613 on 30 degrees of freedom
AIC: NA
Number of Fisher Scoring iterations: 4
> AICtweedie(mod.tweedie)
[1] 179.6432

```

Der Parameterschätzer ergibt sich zu $\hat{\beta} = (-3.73, 0.928)$ mit den Standardfehlern um den Faktor $\sqrt{\hat{\phi}} = 0.491$ im Vergleich zu den Standardfehlern beim Poissonmodell gestreckt. Das globale Maximum ergibt sich an der Stelle $\hat{k} = 2.031$ mit dem Wert -86.811 . Somit scheint eine quadratische Varianzstruktur hierfür passend zu sein.

```

> (c(tweedie$p.max, tweedie$L.max))
[1] 2.030612 -86.810007

```

Der Profile-Plot ist in Abbildung 5.15 veranschaulicht.

Die Standardfehler der drei betrachteten Modelle (`mod`, `mod.q1` und `tweedie.mod`) ergeben sich in R mit dem folgenden Aufruf.

```

> sqrt(diag(summary(mod)$cov.scaled))
(Intercept)      log(1)
 1.1351658    0.1758873
> sqrt(diag(summary(mod.q1)$cov.scaled))
(Intercept)      log(1)
 1.7093598    0.2648553

```

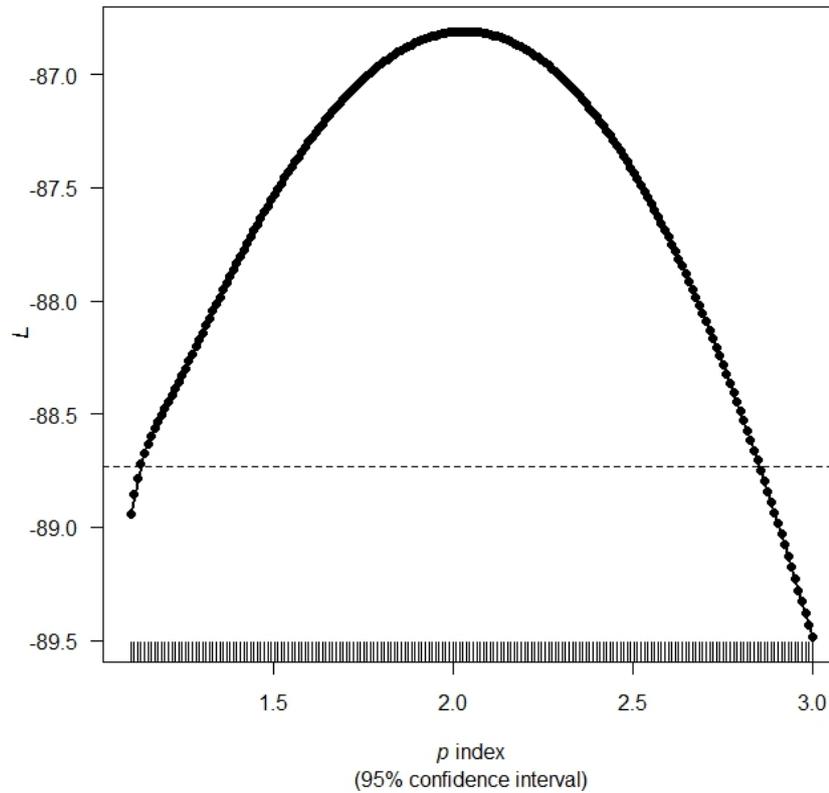


Abbildung 5.15: Profile-Plot für Tweedie-Verteilung mit $k \in [1.1, 3]$.

```
> sqrt(diag(summary(mod.tweedie)$cov.scaled))
(Intercept)      log(1)
  1.2258299    0.1945123
```

Ein Vergleich der skalierten Deviance Funktionen ist in Tabelle 5.6 ersichtlich.

	Poisson	Quasi-Poisson	Tweedie
$D(\mu, y)/\hat{\phi}$	64.54	28.46	30.12
df	30.00	30.00	30.00

Tabelle 5.6: Vergleich der skalierten Deviance Werte zu den Anzahlen der Freiheitsgraden (df).

Betrachten wir jedoch die unskalierte Deviance so ergibt sich für das Quasi-Poissonmodell ein Wert von 64.537 mit einem geschätzten Dispersionparameter von 2.268. Im Vergleich dazu erhalten wir beim Tweedmodell eine unskalierte Deviance von 7.261 mit einem geschätzten Dispersionsparameter von 0.241. Dadurch ergeben sich die kleineren Standardfehler beim Tweedmodell.

Bedienen wir uns schließlich der EQL Funktion.

```
> library(EQL)
> k <- list(theta=seq(-0.5, 3.5, length = 100))
> eq.power <- eql(f ~ log(1), param.space = k,
                 family=powerVarianceFamily("log"),smooth.grid=100)
~~ Extended Quasi Likelihood ~~
* Dimension of parameter space: 100
* Starting loop:
0%          50%          100%
+-----+-----+
|          |          |
* End loop
* Start smoothing:
Smoothing on an equidistant grid between EQL values with 100 nodes
* End smoothing
```

Als Ergebnis erhalten wir das Maximum -86.215 an der Stelle $\hat{k} = 1.984$.

```
eq.power$eql.max
[1] -86.21484
> eq.power$param.max
      theta
1.983828
```

mit den Parameterschätzern

```
> eq.power$model$coef
(Intercept)      log(1)
-3.7318638      0.9279866
```

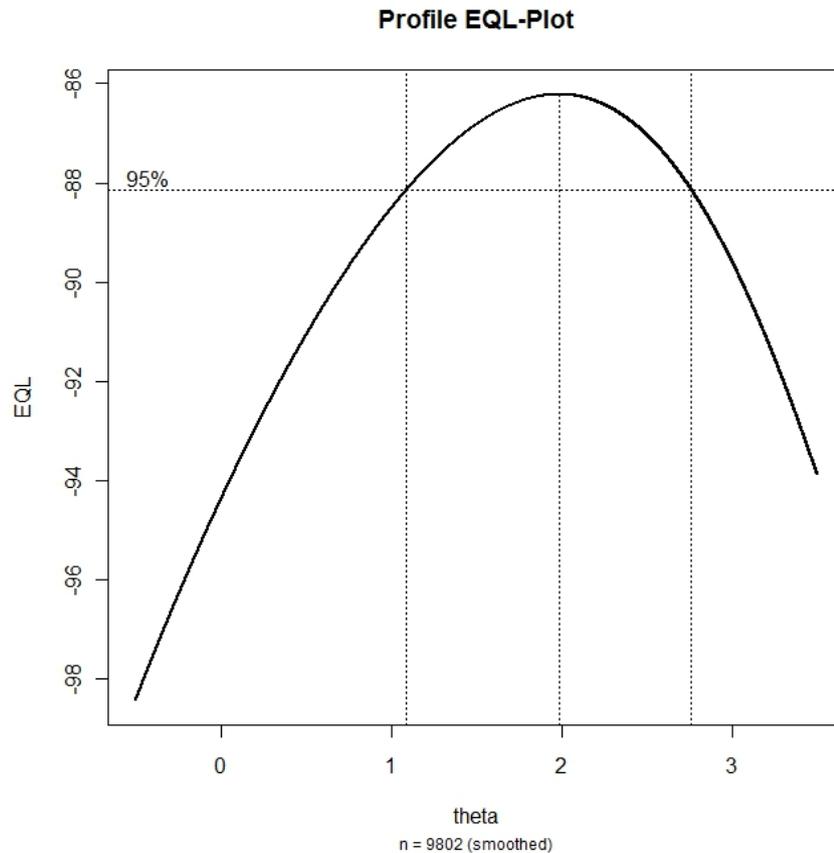


Abbildung 5.16: Profile-Plot mit EQL Funktion mit Varianzfunktion $V_k(\mu) = \mu^k$ und $k \in [-0.5, 3.5]$.

Den dazugehörigen Profile-Plot mit den unteren und oberen Konfidenzschranken für k sehen wir in Abbildung 5.16.

Da der unbekannte Parameter k der Varianzfunktion auf etwa 2 geschätzt wird, betrachten wir auch noch die Modellierung der Daten mittels eines Gammamodells:

```
> mod.gamma <- glm(f ~ log(1), family=Gamma(link="log"))  
> summary(mod.gamma)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.04947	-0.41305	-0.08447	0.21447	1.03149

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-3.7319	1.2383	-3.014	0.00521 **
log(1)	0.9280	0.1963	4.727	5.03e-05 ***

(Dispersion parameter for Gamma family taken to be 0.2569964)

Null deviance: 12.639 on 31 degrees of freedom

Residual deviance: 7.712 on 30 degrees of freedom

AIC: 179.65

Number of Fisher Scoring iterations: 4

Wir sehen, dass das Gammamodell mit dem Log-Link eine gute Anpassung liefert. Die skalierte Deviance ergibt sich zu 30.008 im Vergleich zu 30 Freiheitsgraden.

5.5 Textildaten

Nelder und Pregibon (1987) beschreiben ein Beispiel eines 3^3 -faktoriellen Versuchsplans über die Belastbarkeit von Wolle, dessen Datensatz aus Box und Cox (1964) stammt. Bei diesem Experiment zählt man die Wiederholungen, bei denen die Wolle belastet wird, bis diese reißt. Diese Anzahl der Wiederholungen (W) wird als Responsevariable bezeichnet. Die Länge des Fadens (L), die Amplitude der Belastung (A) sowie die Belastung selbst (B) beschreiben die erklärenden Variablen. Da es sich um einen faktoriellen Versuchsplan handelt, werden die Variablen als Faktoren $-1, 0, 1$ aufgefasst. Der Datensatz ist in Tabelle 5.7 ersichtlich und wird in R mittels

```
> textil.dat<-data.frame(W=c(674,370,292,338,266,210,170,118,90,
    1414,1198,634,1022,620,438,442,332,220,
    3636,3184,2000,1568,1070,566,1140,884,360),
    L=rep(c(-1,0,1),each=9),
    A=rep(c(-1,0,1),3,each=3),
    B=rep(c(-1,0,1),9))
> attach(textil.dat)
```

eingeladen.

<i>L</i>	<i>A</i>	<i>B</i>	<i>W</i>	<i>L</i>	<i>A</i>	<i>B</i>	<i>W</i>	<i>L</i>	<i>A</i>	<i>B</i>	<i>W</i>
-1	-1	-1	674	0	-1	-1	1.414	1	-1	-1	3.636
-1	-1	0	370	0	-1	0	1.198	1	-1	0	3.184
-1	-1	1	292	0	-1	1	634	1	-1	1	2.000
-1	0	-1	338	0	0	-1	1.022	1	0	-1	1.568
-1	0	0	266	0	0	0	620	1	0	0	1.070
-1	0	1	210	0	0	1	438	1	0	1	566
-1	1	-1	170	0	1	-1	442	1	1	-1	1.140
-1	1	0	118	0	1	0	332	1	1	0	884
-1	1	1	90	0	1	1	220	1	1	1	360

Tabelle 5.7: Textil Daten

Zur Modellierung dieser Daten verwenden wir zunächst ein GLM mit einer Gamma-Verteilung und dem Log-Link. Wir sind in diesem Beispiel jedoch nicht an Faktoren, sondern an erklärenden Variablen interessiert und modellieren die Response mittels linearer Prädiktoren, wodurch wir die **Summary**

```
> summary(mod<-glm(W~L+A+B, family=Gamma(link="log")))
Call:
glm(formula = W ~ L + A + B, family = Gamma(link = "log"))
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.433910  -0.115528  -0.009223   0.102596   0.253417

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.34891    0.03422 185.511 < 2e-16 ***
L             0.84251    0.04192  20.100 4.34e-16 ***
A            -0.63132    0.04192 -15.062 2.10e-13 ***
B            -0.38513    0.04192  -9.188 3.68e-09 ***
(Dispersion parameter for Gamma family taken to be 0.03162430)
Null deviance: 22.8861  on 26  degrees of freedom
Residual deviance:  0.7694  on 23  degrees of freedom
AIC: 332.76
Number of Fisher Scoring iterations: 5
```

mit dem Parameterschätzer $\hat{\beta} = (6.349, 0.843, -0.631, -0.385)$ erhalten. Der Dispersionsparameter wird durch $\hat{\phi} = 0.0316$ geschätzt. Die Summary liefert eine

Null Deviance von 22.886 mit 26 Freiheitsgraden und eine Deviance von 0.769 mit 23 Freiheitsgraden. Dies ergibt eine skalierte Deviance von 24.329 im Vergleich zu 23 Freiheitsgraden. Der AIC-Wert liegt bei 332.76.

Im Vergleich dazu betrachten wir eine Schätzung mit der Tweedieverteilung. Das Modell mit der Power-Varianz lautet:

$$\log(\mu) = \beta_0 + \beta_1 \cdot L + \beta_2 \cdot A + \beta_3 \cdot B, \quad V_k(\mu) = \mu^k. \quad (5.3)$$

Es bestehen nun wiederum zwei Möglichkeiten, Modell (5.3) in R umzusetzen:

- Mit Hilfe der Funktion `tweedie.profile`
- Mit Hilfe der Funktion `eq1`

Für die erste Variante erhalten wir mit dem R-Code

```
> tweedie <- tweedie.profile(W~L+A+B,
                             seq(1.01, 4, length=200), do.plot=T, do.smooth=T,
                             method="interpolation", fit.glm=T)
.....Done.
```

den maximalen EQL Wert an der Stelle $k = 2.474$ mit einem Wert von -160.553 . Die 95%-Konfidenzintervallschranken für k liegen bei 1.746 und 3.286.

```
> c(tweedie$p.max, tweedie$L.max, tweedie$ci)
[1] 2.474490 -160.552778 1.746322 3.286069
```

Als Summary ergibt sich

```
> summary(mod.tweedie<-glm(W~L+A+B,
                             family=tweedie(var.power=tweedie$p.max,link.power=0)))
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.091558 -0.032000  0.001333  0.021964  0.063732
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.34778    0.03441 184.494 < 2e-16 ***
L             0.84074    0.04017  20.929 < 2e-16 ***
A            -0.62865    0.03949 -15.919 6.54e-14 ***
B            -0.37129    0.03893  -9.537 1.85e-09 ***
```

(Dispersion parameter for Tweedie family taken to be 0.001448973)

Null deviance: 1.074594 on 26 degrees of freedom

Residual deviance: 0.035884 on 23 degrees of freedom

AIC: NA

Number of Fisher Scoring iterations: 5

```
> AICtweedie(mod.tweedie)
```

```
[1] 331.1056
```

Der Parameterschätzer für β ist ähnlich wie beim Gammamodell und lautet $\hat{\beta} = (6.348, 0.841, -0.629, -0.371)$. Es ergibt sich mit dem Tweediemodell ein geschätzter Dispersionsparameter $\hat{\phi} = 0.0014$. Betrachten wir die skalierte Deviance, so erhalten wir $0.0359/0.0014 = 24.765$. Der AIC-Wert von 331.106 ist etwas geringer als beim Gammamodell zuvor, was auch für eine etwas bessere Modellanpassung spricht. Abbildung 5.17 zeigt den dazugehörigen Profile-Plot.

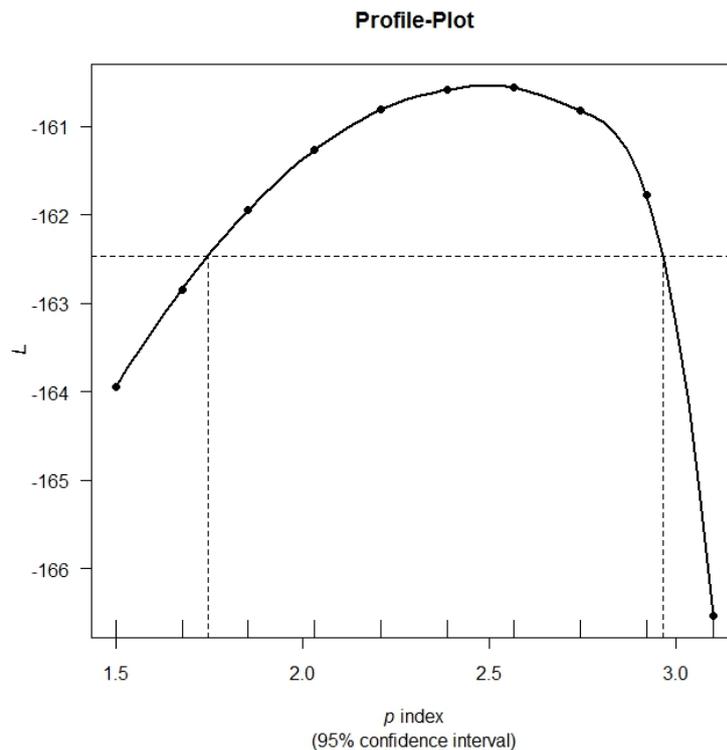


Abbildung 5.17: Profile-Plot der Stoffdaten mit Tweedie-Verteilung für $k \in [1.01, 4]$.

Vergleichen wir das 95% Konfidenzintervall für das GLM mit der Gammaverteilung und für das Tweediemodell, so erhalten wir Abbildung 5.18. Hier sind die einzelnen Stufen des 3^3 -Versuchplanes gut zu erkennen: Im Bereich $[1, 9]$ liegen die Werte mit Faktorstufe $L = -1$, in $[10, 18]$ die Werte mit Faktorstufe $L = 0$ und schließlich im Intervall $[19, 27]$ die Werte mit Faktorstufe $L = 1$. Die Zacken dazwischen deuten die Stufen für den Faktor A an.

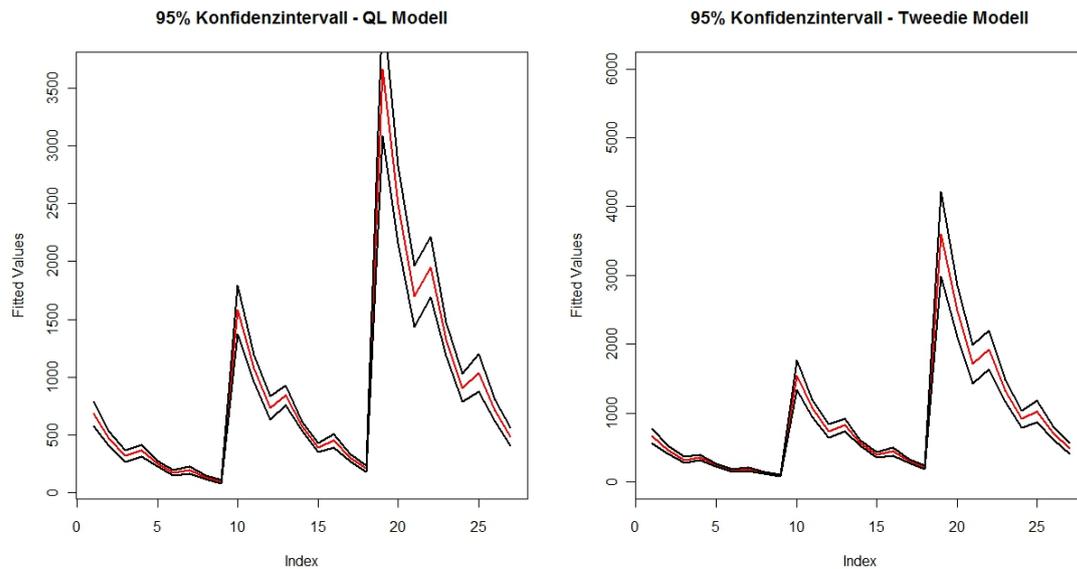


Abbildung 5.18: 95% Konfidenzintervall für GLM und Tweedie Modell.

Verwenden wir nun die R-Funktion `eq1` von Thaler (2009):

```
> k <- list(theta=seq(1.1, 4, length = 20))
> eq.power <- eq1(W~L+A+B, param.space=k,
                 family=powerVarianceFamily("log"), smooth.grid=100,
                 verbose=0)
```

Das Intervall für k wählen wir auch hier mit $[1.01, 4]$.

```
> library(EQL)
> k <- list(theta=seq(1.01, 4, length = 20))
> eq.power <- eq1(W~L+A+B, param.space = k,
                 family=powerVarianceFamily("log"), smooth.grid=100)
```

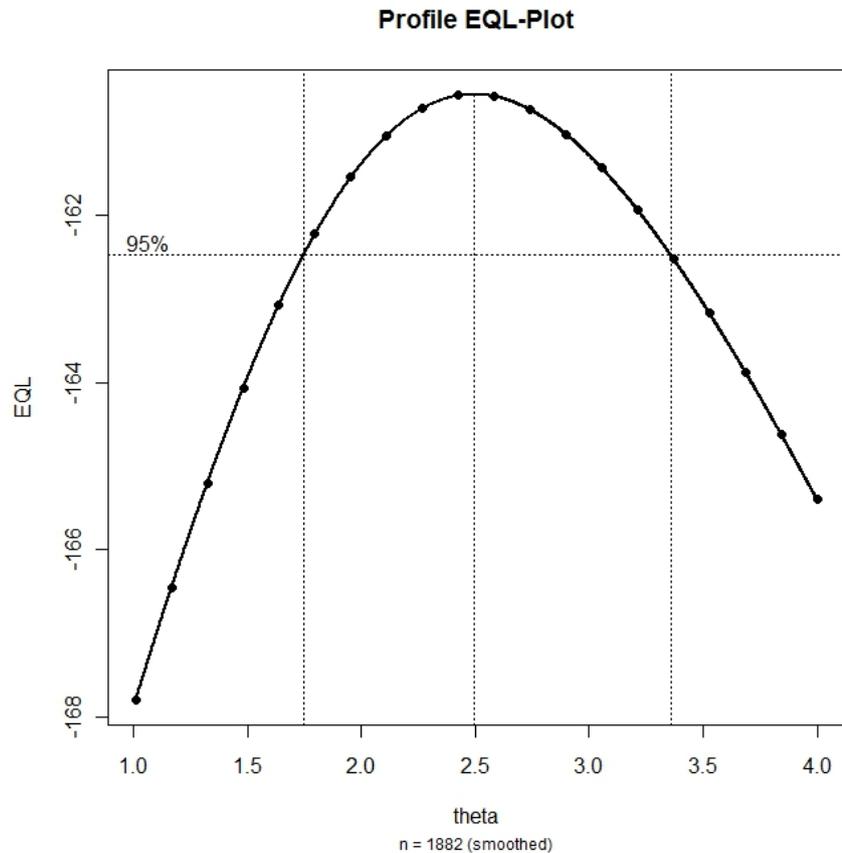



Abbildung 5.19: Profile-Plot EQL Funktion der Textildaten mit $k \in [1.01, 4]$.

Vergleichen wir den Profile-Plot der Tweedieverteilung (Abbildung 5.17) und den der EQL Funktion (Abbildung 5.19), so ist zu erkennen, dass das Tweediemodell ein kleineres 95%-Konfidenzintervall beschreibt. Dies ist damit zu erklären, dass beim Tweedie-Ansatz die gesamte Verteilung angenommen wird. Mit dieser restriktiveren Annahme kann auch eine stärkere Aussage getroffen werden. Ansonsten liefern beide Modelle nahezu die selben maximalen Werte. Da der Schätzer für k bei 2.47 (Tweedie) bzw. bei 2.49 (EQL) liegt, kann die Annahme der Gammaverteilung bei der Analyse mittels eines gewöhnlichen GLMs gerechtfertigt werden, da wir für $k = 2$ die Gammaverteilung erhalten (siehe dazu Kapitel 4).

Literatur

- Box, G. E. P. und Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26, 211–252.
- Casella, G. und Berger, R. L. (2002). *Statistical Inference*. Duxbury.
- Corpet, D. und Taché, S. (2002). *ACF: Aberrant crypt focus*. (<http://www.inra.fr/reseau-nacre/sci-memb/corpet/acf.html>)
- Dunn, P. K. (2007). tweedie: Tweedie exponential family models [Software-Handbuch]. (R package version 2.0.2)
- Dunn, P. K. und Smyth, G. K. (2001). Tweedie family densities: Methods of evaluation. *Statistical Modelling: Proceedings of the 16th International Workshop*, 155–162.
- Dunn, P. K. und Smyth, G. K. (2007). *Evaluation of tweedie exponential dispersion model densities by fourier inversion*. (http://eprints.usq.edu.au/3888/1/Dunn_Smyth_Stats_and_Comp_v18n1.pdf)
- Friedl, H. (2000). *Generalisierte Lineare Modelle*. Institut für Statistik, Universität Graz. (Vorlesungsskriptum)
- Hinde, J. und Demétrio, G. B. (1998). Overdispersion: Models and estimation. *Computational Statistics and Data Analysis*, 27, 151–170.
- Hinde, J. P. (1982). Compound Poisson regression models. *Gilchrist, R., GLIM82*, 109–121.
- Jørgensen, B. (1997). *The Theory of Dispersion Models*. London: Chapman and Hall.
- Kedem, B. und Fokianos, K. (2002). *Regression Models for Time Series Analysis*. New York: John Wiley.
- McCullagh, P. und Nelder, J. A. (1989). *Generalized Linear Models* (2. Aufl.). New York: Chapman and Hall.
- Mendlik, T. (2008). *Generalisierte lineare Mischmodelle zur Analyse von Longitudinaldaten*. Unveröffentlichte Diplomarbeit, Institut für Statistik an der Technischen Universität Graz, Institut für Medizinische Informatik, Statistik und Dokumentation an der Medizinischen Universität Graz.
- Merberg, A. und Miller, S. J. (2008). *Course Notes for Math 162: Mathematical Statistics. The Cramér-Rao Inequality*. (<http://www.williams.edu/go/math/sjmillier/public.html/BrownClasses/162/Handouts/CramerRaoHandout08.pdf>)
- Nelder, J. A. und Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika*, 74, 221–232.
- Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford: Clarendon Press.
- Thaler, T. (2009a). *Die Extended-Quasi-Likelihood-Funktion in Generalisierten*

- Linearen Modellen*. Unveröffentlichte Diplomarbeit, Institut für Statistik an der Technischen Universität Graz.
- Thaler, T. (2009b). Extended-quasi-likelihood-function (eql) [Software-Handbuch]. (R package version 1.0-0)
- Verani, S. (2010). *The Fisher Information*. (<http://www.econ.ucsb.edu/~verani/notes/info.pdf>)
- Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, *61*, 439–447.