

# 1 Eine Erweiterung der Quasi-Likelihood-Funktion

Die Einführung der QL-Funktion für das GLM erlaubt die Spezifizierung von Varianzfunktionen, die keiner Verteilung aus der Exponentialfamilie zuzuordnen sind. Dabei muss die Varianzfunktion bis auf eine multiplikative Konstante exakt angegeben werden. Stammt die Varianzfunktion  $V(\mu)$  aus einer Familie von Funktionen  $\mathcal{F}_\theta$  mit einem *unbekannten* Parametervektor  $\theta$ , sind die bis jetzt vorgestellten Methoden *nicht* in der Lage, den Parametervektor  $\theta$  simultan mit dem Parametervektor  $\beta$  des linearen Modells zu schätzen. Das heißt mit anderen Worten, dass der direkte Vergleich unterschiedlicher Varianzfunktionen mit den bisher diskutierten Verfahren nicht möglich ist. Außerdem erlauben die QL-Methoden auch nicht die Schätzung von nicht konstanten und unbekanntem Dispersionsparametern.

Um auch verschiedene Varianzfunktionen miteinander vergleichen zu können, bedarf es einer weiteren Anpassung. Die **EQL**-Funktion geht auf Nelder and Pregibon [1987] zurück und erlaubt derartige Vergleiche.

**Definition 1.1** (EQL-Funktion, Nelder and Pregibon [1987]). Sei  $Y$  eine Zufallsvariable mit Erwartungswert  $\mathbb{E}(Y) = \mu$  und  $\text{var}(Y) = \phi V(\mu)$  mit bekannter Varianzfunktion  $V(\mu)$ . Bezeichne  $y$  eine Realisation dieser Zufallsvariable mit der QL-Funktion  $Q(\mu, y) = \int^\mu (y - t)/(\phi V(t)) dt$  und der Quasi-Deviance  $D(y, \mu) = -2\phi(Q(\mu, y) - Q(y, y))$ . Dann heißt die durch

$$Q^+(\mu, \phi, y) = -\frac{1}{2} \log \{2\pi\phi V(y)\} - \frac{1}{2} D(y, \mu)/\phi \quad (1.1)$$

definierte Funktion die *Extended-Quasi-Likelihood-Funktion*.

*Bemerkung 1.* Die EQL-Funktion beruht wie die QL-Funktion nur auf der Annahme der ersten beiden Momente von  $Y$ . Insbesondere muss die Verteilung von  $Y$  *nicht* vollständig spezifiziert sein.

*Bemerkung 2.* Für die  $n$  unabhängigen Zufallsvariablen  $Y_i$  ist die EQL-Funktion für die gesamte Stichprobe  $\mathbf{y} = (y_1, \dots, y_n)^\top$  durch

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

gegeben, wobei  $\mathbb{E}(Y_i) = \mu_i$  gilt.

*Bemerkung 3.* Manche Autoren bezeichnen  $-2Q^+(\boldsymbol{\mu}, \phi, \mathbf{y})$  auch als *Extended-Quasi-Deviance* [vgl. z.B Nelder and Lee, 1991].

Lässt sich der Erwartungswert  $\mu$  durch eine Funktion mit dem Parametervektor  $\beta$  beschreiben, kann ein Schätzer für diesen Vektor durch die Maximierung von  $Q^+$  bestimmt werden. Der erste Term von  $Q^+(\mu, \phi, \mathbf{y})$  ist unabhängig von  $\mu$  (und damit auch von  $\beta$ ), daher entspricht eine Maximierung von  $Q^+(\mu, \phi, \mathbf{y})$  der Minimierung der (Quasi-) Deviance  $D(\mathbf{y}, \mu)$ . Das heißt mit anderen Worten: derjenige Schätzer von  $\beta$ , der die EQL-Funktion  $Q^+(\mu, \phi, \mathbf{y})$  maximiert, ist der gleiche, den man erhält, wenn man die QL-Funktion  $Q(\mu, \mathbf{y})$  maximiert.

Einen Schätzer für den (gemeinsamen) Dispersionsparameter  $\phi$  erhält man durch die Maximierung von  $Q^+(\mu, \phi, \mathbf{y})$  nach  $\phi$ :

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

Da der wahre Parameter  $\mu$  natürlich nicht bekannt ist, verwendet man stattdessen den Schätzer  $\hat{\mu}$ , und man erhält somit als Schätzer für den Dispersionsparameter  $\phi$  die mittlere Deviance:

$$\hat{\phi} = D(\mathbf{y}, \hat{\mu})/n. \quad (1.2)$$

*Bemerkung.* Anstatt der mittleren Deviance findet auch die Bias korrigierte Variante Verwendung. Diese lautet:

$$\hat{\phi} = D(\mathbf{y}, \hat{\mu})/(n - p),$$

wobei  $p$  die Anzahl der Parameter im linearen Prädiktor beschreibt.

Für Verteilungen aus der Exponentialfamilie gilt für die Quasi-Score- und die Score-Funktion Äquivalenz. Für die EQL-Funktion ist das nicht zwingend der Fall. Es stellt sich daher die Frage, in welchen Fällen auch die EQL-Funktion mit einer log-Likelihood-Funktion übereinstimmt.

Nelder and Pregibon [1987] zeigen, dass für bestimmte Verteilungen aus der Exponentialfamilie tatsächlich Äquivalenz gilt und geben für die anderen Verteilungen an, inwieweit sich die log-Likelihood-Funktion von der EQL-Funktion unterscheidet:

- Für die *Normalverteilung* und die *Inverse-Gauß-Verteilung* gilt Äquivalenz:
 
$$\begin{aligned} V(\mu) = 1, \phi = \sigma^2 : \quad Q^+(\mu, \phi, \mathbf{y}) &= -\frac{1}{2} \log \{2\pi\phi V(y)\} - \frac{1}{2} D(y, \mu)/\phi \\ &= \log \frac{1}{\sqrt{2\pi \cdot \sigma^2 \cdot 1}} - \frac{1}{2\sigma^2} (y - \mu)^2 \\ &\hat{=} \log\text{-Likelihood einer Normalverteilung,} \\ V(\mu) = \mu^3, \phi = \sigma^2 : \quad Q^+(\mu, \phi, \mathbf{y}) &= -\frac{1}{2} \log \{2\pi\phi V(y)\} - \frac{1}{2} D(y, \mu)/\phi \\ &= \log \frac{1}{\sqrt{2\pi \cdot \sigma^2 \cdot y^3}} - \frac{1}{2\sigma^2} \frac{(y - \mu)^2}{\mu^2 y} \\ &\hat{=} \log\text{-Likelihood einer Inversen-Gauß-Verteilung.} \end{aligned}$$

Damit ist der Schätzer für  $\hat{\phi}$  in Gleichung (1.2) für diese beiden Verteilungen auch der ML-Schätzer.

- Für die *Poisson*-, die *standardisierte Binomial*- und die *Negativ-Binomialverteilung* erhält man die EQL-Funktion aus der jeweiligen log-Likelihood-Funktion in dem man die enthaltenen Fakultäten jeweils durch die Stirling-Approximation gemäß

$$n! \approx n^n e^{-n} \sqrt{2\pi n} \quad (*)$$

ersetzt. Exemplarisch gilt daher für die Poissonverteilung mit  $V(\mu) = \mu, \phi = 1$ :

$$\begin{aligned} Q^+(\mu, \phi, y) &= -\frac{1}{2} \log \{2\pi\phi V(y)\} - \frac{1}{2} D(y, \mu)/\phi \\ &= -\frac{1}{2} \log (2\pi \cdot 1 \cdot y) - y \log \frac{y}{\mu} + y - \mu. \end{aligned}$$

Betrachtet man die log-Likelihood-Funktion der Poissonverteilung erhält man:

$$\begin{aligned} \mathcal{L}(\mu, y) &= y \log \mu - \mu - \log y! \\ &\stackrel{(*)}{\approx} y \log \mu - \mu - \log \left( y^y e^{-y} \sqrt{2\pi y} \right) \\ &= y \log \mu - \mu - y \log y + y - \frac{1}{2} \log(2\pi y) \\ &= -\frac{1}{2} \log(2\pi y) - y \log \frac{y}{\mu} + y - \mu, \end{aligned}$$

was wiederum der EQL-Funktion entspricht.

- Für die *Gammaverteilung* unterscheidet sich die EQL-Funktion von der log-Likelihood-Funktion um einen Faktor, der nur von  $\phi$  abhängt. Mit  $V(\mu) = \mu^2, \phi = 1/\nu$  gilt:

$$\begin{aligned} Q^+(\mu, \phi, y) &= -\frac{1}{2} \log \{2\pi\phi V(y)\} - \frac{1}{2} D(y, \mu)/\phi \\ &= -\frac{1}{2} \log (2\pi \cdot 1/\nu \cdot y^2) - \nu \left( -\log \frac{y}{\mu} + \frac{y - \mu}{\mu} \right) \\ &= -\log \sqrt{2\pi} + \log \sqrt{\nu} - \log y + \nu \log y - \nu \log \mu - \nu \frac{y}{\mu} + \nu \\ &= -\nu \frac{y}{\mu} - \nu \log \mu + (\nu - 1) \log y + \underbrace{\log \sqrt{\nu} + \nu - \log \sqrt{2\pi}}_{=: \hat{g}(\nu)}. \end{aligned}$$

Betrachtet man die log-Likelihood-Funktion der Gammaverteilung erhält man:

$$\mathcal{L}(\mu, \nu, y) = -\nu \frac{y}{\mu} - \nu \log \mu + (\nu - 1) \log y + \underbrace{\nu \log \nu - \log \Gamma(\nu)}_{=: \hat{g}(\nu)}$$

Tabelle 1.1: Modifizierte Varianz-Funktionen

Verteilung	$V(y)$	$V(y, c)$
Poisson	$y$	$y + c$
Binomial	$\frac{y(n-y)}{n}$	$\frac{(y+c)(n-y+c)}{n+c}$
Negativ-Binomial	$\frac{y(y+\nu)}{\nu}$	$\frac{(y+\nu)^2(y+c)(\nu+c)}{\nu^2(y+\nu+c)}$

und damit unterscheiden sich die beiden Funktionen nur um die Größe  $g(\nu) - \tilde{g}(\nu)$ . Verwendet man auch für die Gamma-Funktion die Stirling-Approximation

$$\Gamma(x) = x^x e^{-x} \sqrt{2\pi/x} (1 + \mathcal{O}(1/x)), \quad (**)$$

erhält man für die Differenz zwischen der log-Likelihood- und der EQL-Funktion:

$$\begin{aligned} g(\nu) - \tilde{g}(\nu) &= \log \sqrt{\nu} + \nu - \log \sqrt{2\pi} - (\nu \log \nu - \log \Gamma(\nu)) \\ &\stackrel{(**)}{\approx} \log \sqrt{\nu} + \nu - \log \sqrt{2\pi} - \left[ \nu \log \nu - \log \left( \nu^\nu e^{-\nu} \sqrt{2\pi/\nu} \right) \right] \\ &= \log \sqrt{\nu} + \nu - \log \sqrt{2\pi} - \left[ \nu \log \nu - \nu \log \nu + \nu - \log \sqrt{2\pi} + \log \sqrt{\nu} \right] \\ &= 0. \end{aligned}$$

*Bemerkung.* Für die Poisson-, die (nicht standardisierte) Binomial- und die Negativ-Binomialverteilung gilt insbesondere für  $y = 0$  :  $V(y) = 0$ . Die EQL-Funktion in Gleichung (1.1) ist aber für  $V(y) = 0$  nicht definiert, während die jeweiligen Likelihood-Funktionen in  $y = 0$  sehr wohl definiert sind. Die Ursache dafür liegt in der Verwendung der Stirling-Approximation. Diese ergibt für  $y = 0$  die Annäherung  $0! \approx 0$ . Zieht man stattdessen die modifizierte Form

$$n! \approx n^n e^{-n} \sqrt{2\pi(n+c)}$$

heran, erhält man für  $c = 1/6$  :  $0! \approx 1.023$ . Verwendet man die modifizierte Stirling-Approximation, muss man sich je nach Verteilung anderer Varianzfunktionen bedienen. Tabelle 1.1 zeigt, wie die Varianzfunktion modifiziert werden muss, wobei diese (lineare) Transformation die Schätzer für  $\beta$  und  $\phi$  nicht ändert [vgl. Nelder and Pregibon, 1987].

Nelder and Pregibon [1987] motivieren die Verwendung der EQL-Funktion mit deren „partial likelihood flavour“ (also in etwa „Likelihood-ähnlichem Verhalten“), das darauf beruht, dass man – wie auch schon für die QL-Funktion – für die EQL-Funktion eine Dichte finden kann, indem man  $\exp\{Q^+\}$  normalisiert.<sup>1</sup> Die Verwendung der nicht normalisierten EQL-Funktion, lässt sich dann dadurch rechtfertigen, dass sich der Normalisierungsfaktor (der in der Regel von  $\mu$ ,  $\phi$  und allfälligen weiteren Parametern  $\theta$  der

<sup>1</sup>Das heißt man bestimmt einen Faktor  $\omega$  für den  $\int_{\mathbb{R}} \omega \exp\{Q^+\} dy = 1$  gilt.

Varianzfunktion abhängt) in manchen Fällen auch für große Änderung in den Parametern nur geringfügig ändert. Nelder and Lee [1992] vergleichen beispielsweise die Schätzer für die Dispersion, die bei ML- und EQL-Schätzung bei einer NB $\alpha$ -Verteilung<sup>2</sup> und einer Poisson-Inverse-Gauß-Mischverteilung resultieren und kommen zu dem Ergebnis, dass der EQL-Schätzer in den meisten von ihnen getesteten Parameterkonfigurationen sogar einen kleineren mittleren quadratischen Fehler aufweist als der ML-Schätzer.

Eine andere Herleitung findet man bei McCullagh and Nelder [1989, S. 349 ff.] und ergibt sich aus der Idee, die QL-Funktion  $Q(\mu, y)$  derart zu einer Funktion  $Q^+(\mu, \phi, y)$  zu erweitern, dass  $Q^+$  einerseits für einen bekannten Dispersionsparameter  $\phi$  die gleichen Eigenschaften besitzt wie die Quasi-Likelihood-Funktion selbst und andererseits für unbekanntes  $\phi$  die Eigenschaften einer gewöhnlichen log-Likelihood-Funktion bezüglich  $\phi$  „erbt“. Für die folgenden Überlegungen wollen wir die Existenz der Momente bis zur vierten Ordnung annehmen. Außerdem legen wir folgende Definition der Quasi-Deviance zugrunde:

$$D(\mathbf{y}, \boldsymbol{\mu}) = -2\phi \cdot Q(\boldsymbol{\mu}, \mathbf{y}).$$

Für die Erweiterung der QL-Funktion wählen wir den Ansatz:

$$\begin{aligned} Q^+(\mu, \phi, y) &= Q(\mu, y) + h(\phi, y) \\ &= -\frac{D(y, \mu)}{2\phi} + h(\phi, y), \end{aligned}$$

für eine Funktion  $h(\phi, y) = -h_1(\phi)/2 - h_2(y)$ . Damit sich  $Q^+$  bezüglich  $\phi$  wie eine log-Likelihood-Funktion verhält, muss für die erwartete Score-Funktion bezüglich  $\phi$

$$\mathbb{E} \left[ \frac{\partial Q^+(\mu, \phi, y)}{\partial \phi} \right] = \frac{1}{2\phi^2} \mathbb{E} [D(y, \mu)] - \frac{1}{2} \frac{\partial h_1(\phi)}{\partial \phi} \stackrel{!}{=} 0$$

gelten. Daraus folgt unmittelbar

$$\phi^2 \frac{\partial h_1(\phi)}{\partial \phi} = \mathbb{E} [D(y, \mu)].$$

Eine Taylorentwicklung für  $D = D(y) = D(y, \mu)$  in der ersten Komponente liefert für den Entwicklungspunkt  $y = \mu$  die Abschätzung:

$$\begin{aligned} D \approx D(\mu) + (y - \mu) \frac{\partial D(y)}{\partial y} \Big|_{y=\mu} &+ \frac{1}{2}(y - \mu)^2 \frac{\partial^2 D(y)}{\partial y^2} \Big|_{y=\mu} + \frac{1}{6}(y - \mu)^3 \frac{\partial^3 D(y)}{\partial y^3} \Big|_{y=\mu} \\ &+ \frac{1}{24}(y - \mu)^4 \frac{\partial^4 D(y)}{\partial y^4} \Big|_{y=\mu}. \end{aligned} \tag{1.3}$$

---

<sup>2</sup>Die NB $\alpha$ -Verteilung ist eine Variante der Negativ-Binomialverteilung. Letztere kann man als stetige Mischung einer Poissonverteilung mit Parameter  $\mu$  auffassen, wobei  $\mu$  selbst einer Gammaverteilung mit den Parametern  $\alpha$  und  $\nu$  folgt. Für ein GLM betrachtet man  $\nu$  als fix, während  $\mu$  variiert. Somit ergibt sich für die Negativ-Binomialverteilung ein Erwartungswert von  $\mathbb{E}(Y) = \mu = \alpha\nu$  und eine Varianz von  $\text{var}(Y) = \mu + \mu^2/\nu$ . Nimmt man hingegen an, dass  $\mu$  und  $\nu$  variieren, während  $\alpha$  konstant bleibt, erhält man  $\text{var}(Y) = \mu(1 + \alpha)$ , was einer Poissonvarianz mit Überdispersion entspricht. Diese Verteilung wird als NB $\alpha$ -Verteilung bezeichnet [vgl. Nelder and Lee, 1992].

Die Deviance lässt sich zu

$$\begin{aligned} D(y, \mu) &= 2 \int_{\mu}^y \frac{y-t}{V(t)} dt = 2y \int_{\mu}^y \underbrace{\frac{1}{V(t)}}_{=:f_1(t)} dt - 2 \int_{\mu}^y \underbrace{t}_{=:f_2(t)} dt \\ &= 2y(F_1(y) - F_1(\mu)) - 2(F_2(y) - F_2(\mu)) \end{aligned}$$

umschreiben, wobei  $F_1(\cdot)$  und  $F_2(\cdot)$  die Stammfunktionen von  $f_1(\cdot)$  respektive  $f_2(\cdot)$  bezeichnen. Die Ableitungen der Deviance lauten mit  $V'(\mu) = \partial V(\mu)/\partial \mu$  und  $V''(\mu) = \partial^2 V(\mu)/\partial \mu^2$  daher:

$$\begin{aligned} \frac{\partial D(y)}{\partial y} &= 2(F_1(y) - F_1(\mu)) + \underbrace{2yf_1(y) - 2f_2(y)}_{=0} = 2(F_1(y) - F_1(\mu)), \\ \frac{\partial^2 D(y)}{\partial y^2} &= 2f_1(y) = \frac{2}{V(y)}, \\ \frac{\partial^3 D(y)}{\partial y^3} &= 2f_1'(y) = -\frac{2V'(y)}{V(y)^2}, \\ \frac{\partial^4 D(y)}{\partial y^4} &= 2f_1''(y) = \frac{2}{V(y)^3}(2V'(y)^2 - V(y)V''(y)). \end{aligned} \tag{1.4}$$

Setzt man die Identitäten (1.4) in die Approximation (1.3) ein, erhält man mit  $D(\mu) = 2\mu(F_1(\mu) - F_1(\mu)) - 2(F_2(\mu) - F_2(\mu)) = 0$  und  $D'(\mu) = 2(F_1(\mu) - F_1(\mu)) = 0$ :

$$D \approx (y - \mu)^2 \frac{1}{V(\mu)} - \frac{1}{3}(y - \mu)^3 \frac{V'(\mu)}{V(\mu)^2} + \frac{1}{12}(y - \mu)^4 \frac{2V'(\mu)^2 - V(\mu)V''(\mu)}{V(\mu)^3}. \tag{1.5}$$

Wir bezeichnen mit  $\mu_k = \mathbb{E}((y - \mu)^k)$  das  $k$ -te zentrale Moment; somit gilt  $\text{var}(Y) = \mu_2 = \phi V(\mu)$ . Für den Erwartungswert von (1.5) folgt daher für  $V = V(\mu)$ :

$$\begin{aligned} \mathbb{E}(D) &\approx \phi V \frac{1}{V} - \frac{1}{3} \frac{V'}{V^2} \mu_3 + \frac{1}{12} \frac{2V'^2 - VV''}{V^3} \mu_4 \\ &= \phi + \frac{1}{12V^2} (-4\mu_3 V' + 2\mu_4 V'^2/V - \mu_4 V''). \end{aligned} \tag{1.6}$$

Vernachlässigt man die Momente höherer Ordnung, erhält man aus Gleichung (1.6) die approximative Aussage  $\mathbb{E}[D(y, \mu)] \approx \phi$  und daraus folgt:

$$\begin{aligned} h_1(\phi) &= \log \phi + \text{Konstante}, \\ Q^+(\mu, \phi, y) &\approx -\frac{1}{2} D(y, \mu)/\phi - \frac{1}{2} \log \phi. \end{aligned} \tag{1.7}$$

Die Approximation (1.7) unterscheidet sich von Gleichung (1.1) also nur um eine additive Funktion in  $y$ .

Verzichtet man in Gleichung (1.6) nicht auf die höheren Momente, lassen sich noch genauere Ergebnisse erzielen. Über die Beziehungen  $\mu_4 = \kappa_4 + 3\kappa_2^2 = \kappa_4 + 3\phi^2 V^2$  und  $\mu_3 = \kappa_3$ , wobei  $\kappa_i$  die  $i$ -te Kumulante bezeichnet, erhält man aus der Approximation (1.6):

$$\mathbb{E}(D) \approx \phi + \frac{1}{12V^2} (-4\kappa_3 V' + 6\phi^2 V V'^2 - 3\phi^2 V^2 V'' + 2\kappa_4 V'^2/V - \kappa_4 V''),$$

und nach Weglassung der Kumulante vierter Ordnung resultiert:

$$\mathbb{E}(D) \approx \phi + \frac{1}{12V^2} (-4\kappa_3 V' + 6\phi^2 V V'^2 - 3\phi^2 V^2 V''). \quad (1.8)$$

Lässt sich zusätzlich die Beziehung

$$\kappa_{k+1} = \frac{\partial \kappa_k}{\partial \mu} \kappa_2 = \kappa'_k \kappa_2, \quad \text{für } k \geq 2, \quad (1.9)$$

für Kumulanten bis zur vierten Ordnung rechtfertigen, lässt sich die Approximation (1.8) noch weiter vereinfachen und wir erhalten mit der  $k$ -ten standardisierten Kumulante  $\rho_k = \kappa_k / \kappa_2^{k/2}$ :

$$\begin{aligned} \mathbb{E}(D) &\approx \phi + \frac{1}{12V^2} (-4\kappa_3 V' + 6\phi^2 V V'^2 - 3\phi^2 V^2 V'') \\ &= \phi + \frac{1}{12V^2} (-4\phi^2 V V'^2 + 6\phi^2 V V'^2 - 3\phi^2 V^2 V'') \\ &= \phi + (5\phi^2 V'^2 / V - 3\phi^2 V'' - 3\phi^2 V'^2 / V) / 12 \\ &= \phi + \phi \underbrace{(5\phi V'^2 / V)}_{=\rho_3^2} - 3\phi \underbrace{(V'' V + V'^2) / V}_{=\rho_4} / 12 \\ &= \phi \{1 + \phi(5\rho_3^2 - 3\rho_4) / 12\}. \end{aligned}$$

*Bemerkung 1.* Für Verteilungen aus der Exponentialfamilie trifft die Eigenschaft (1.9) stets zu. Für die  $k$ -te Kumulante einer Verteilung aus der Exponentialfamilie gilt nämlich:

$$\begin{aligned} \kappa_k &= a(\phi)^{k-1} \frac{\partial^k b(\theta)}{\partial \theta^k}, \\ &= \underbrace{a(\phi)^{k-2} \frac{\partial^k b(\theta)}{\partial \theta^k}}_{=\kappa'_{k-1}} \cdot \underbrace{\frac{\partial \theta}{\partial \mu}}_{=a(\phi)b'(\theta)=\text{var}(Y)=\kappa_2} \cdot \underbrace{a(\phi) \frac{\partial \mu}{\partial \theta}}_{=\kappa_2} \\ &= \frac{\partial \kappa_{k-1}}{\partial \mu} \kappa_2. \end{aligned}$$

*Bemerkung 2.* Trifft die Eigenschaft (1.9) zu, ergeben sich für die (standardisierten) Kumulanten folgende Abschätzungen:

$$\begin{aligned} \kappa_3 &= \mathcal{O}(\phi^2) & \rho_3 &= \mathcal{O}(\sqrt{\phi}), \\ \kappa_4 &= \mathcal{O}(\phi^3) & \rho_4 &= \mathcal{O}(\phi). \end{aligned}$$

Analog lassen sich nun auch die Varianz  $\text{var}(D)$  und die Kovarianz-Matrix  $\text{cov}(D, Y)$  approximieren:

$$\text{var}(D) \approx 2\kappa_2^2 / V^2 = 2\phi^2 \quad \text{cov}(D, Y) \approx (\kappa_3 - \kappa_2' \kappa_2) / V. \quad (1.10)$$

*Bemerkung.* Für Verteilungen aus der Exponentialfamilie folgt aus  $\kappa_3 = \kappa_2' \kappa_2$ , dass die Kovarianz  $\text{cov}(D, Y)$  verschwindet.

Aus der Approximation (1.7) lassen sich nun die  $\phi$ - und die  $\mu$ -Score-Funktion bestimmen:

$$\frac{\partial Q^+(\mu, \phi, y)}{\partial \mu} = \frac{y - \mu}{\phi V(\mu)}, \quad (1.11a)$$

$$\frac{\partial Q^+(\mu, \phi, y)}{\partial \phi} = \frac{D(y, \mu)}{2\phi^2} - \frac{1}{2\phi}, \quad (1.11b)$$

wobei Gleichung (1.11a) allgemein und Gleichung (1.11b) nur für hinreichend kleine  $\phi$  gilt.<sup>3</sup> Die Score-Funktion (1.11a) entspricht dabei der bereits bekannten (Quasi-) Score-Funktion für  $\mu$ . Beide Scores haben einen Erwartungswert von Null und für deren Varianz respektive deren Kovarianz gilt für  $Q_\mu^+ := \partial Q^+ / \partial \mu$  und  $Q_\phi^+ := \partial Q^+ / \partial \phi$  mit (1.10):

$$\begin{aligned} \text{var}(Q_\mu^+) &= \text{var}\left(\frac{y - \mu}{\phi V(\mu)}\right) = \frac{\phi V(\mu)}{\phi^2 V(\mu)^2} = \frac{1}{\phi V(\mu)}, \\ \text{var}(Q_\phi^+) &= \text{var}\left(\frac{D(y, \mu)}{2\phi^2} - \frac{1}{2\phi}\right) \approx \frac{2\phi^2}{4\phi^4} = \frac{1}{2\phi^2}, \\ \text{cov}(Q_\mu^+, Q_\phi^+) &= \text{cov}\left(\frac{y - \mu}{\phi V(\mu)}, \frac{D(y, \mu)}{2\phi^2} - \frac{1}{2\phi}\right) \approx \frac{\kappa_3 - \kappa_2' \kappa_2}{2\phi^3 V(\mu)^2}. \end{aligned} \quad (1.12)$$

Die zweiten Ableitungen der EQL-Funktion lauten außerdem:

$$\begin{aligned} \frac{\partial^2 Q^+(\mu, \phi, y)}{\partial \mu^2} &= \frac{-\phi V(\mu) - (y - \mu)\phi V(\mu)}{\phi^2 V(\mu)^2}, \\ \frac{\partial^2 Q^+(\mu, \phi, y)}{\partial \phi^2} &= -\frac{D(y, \mu)}{\phi^3} + \frac{1}{2\phi^2}, \\ \frac{\partial^2 Q^+(\mu, \phi, y)}{\partial \mu \partial \phi} &= -\frac{y - \mu}{\phi^2 V(\mu)}, \end{aligned} \quad (1.13)$$

womit sich schließlich der negative Erwartungswert der zu einer Matrix zusammengefassten Elemente in Gleichung (1.13) durch

$$\begin{pmatrix} \frac{1}{\phi V(\mu)} & 0 \\ 0 & \frac{1}{2\phi^2} \end{pmatrix} \quad (1.14)$$

ausdrücken lässt. Die Diagonalelemente der Matrix (1.14) entsprechen dabei genau den Varianzen der Score-Funktionen in Gleichung (1.12). In Analogie zu der Fisher-Information bezeichnen wir diese Größe als *Quasi-Fisher-Information*.

Gilt für die zweite und dritte Kumulante die Abschätzung

$$\kappa_3 - \kappa_2' \kappa_2 = \mathcal{O}(\phi^2),$$

---

<sup>3</sup>Widrigensfalls ließe sich die Abschätzung  $\mathbb{E}(D) = \phi$  nicht rechtfertigen.

folgt für die Korrelation der beiden Score-Funktionen  $Q_\mu^+$  und  $Q_\phi^+$ :

$$\begin{aligned} \text{cor} \left( Q_\mu^+, Q_\phi^+ \right) &= \frac{\text{cov} \left( Q_\mu^+, Q_\phi^+ \right)}{\sqrt{\text{var} \left( Q_\mu^+ \right) \text{var} \left( Q_\phi^+ \right)}} \\ &= \frac{\kappa_3 - \kappa_2' \kappa_2}{2\phi^3 V(\mu)^2} \cdot 2\phi \sqrt{\phi V(\mu)} \\ &= \mathcal{O}(\sqrt{\phi}), \end{aligned}$$

was eine zu vernachlässigende Größe darstellt. Zusammenfassend heißt das also, dass unter den Annahmen

1. Der Dispersionsparameter  $\phi$  ist hinreichend klein.
2. Die Kumulanten  $\kappa_k$  sind von der Größenordnung  $\mathcal{O}(\phi^{k-1})$

$Q^+(\mu, \phi, y)$  die Eigenschaften der QL-Funktion  $Q(\mu, y)$  bezüglich *beider* Parameter  $\mu$  und  $\phi$  besitzt.

*Bemerkung.* Eine weitere Herleitung ergibt sich aus der Anwendung der *Sattelpunkt-Approximation*.

In den folgenden Abschnitten werden nun verschiedene Einsatzmöglichkeiten der EQL-Funktion diskutiert. Dabei hilft die EQL-Funktion insbesondere die Parameter zu schätzen, die Nelder and Pregibon [1987] als „nicht lineare Parameter“ bezeichnen – also all jene Parameter, die nicht im linearen Prädiktor vorkommen.

## 1.1 Gemeinsame Modellierung von Erwartungswert und Dispersion

Bis jetzt sind wir von einem (unbekannten) konstanten Dispersionsparameter  $\phi$ , wenigstens aber von beobachtungsspezifischen aber bekannten Parametern  $\phi_i$  ausgegangen. Die EQL-Funktion erlaubt nun eine simultane Modellierung sowohl der Erwartungswerte  $\mu_i$  als auch der *unbekannten* Dispersionsparameter  $\phi_i$ .

Dazu betrachten wir zwei von einander *abhängige* Modelle:

$$\begin{aligned} \mathbb{E}(Y_i) &= \mu_i, & \mathbb{E}(t_i) &= \phi_i, \\ \text{var}(Y_i) &= \phi_i V(\mu_i), & \text{var}(t_i) &= \psi V_D(\phi_i), \\ g(\boldsymbol{\mu}) &= X\boldsymbol{\beta}, & h(\boldsymbol{\phi}) &= Z\boldsymbol{\gamma}, \end{aligned} \tag{1.15}$$

mit  $\boldsymbol{\beta} \in \mathbb{R}^p$ ,  $X \in \mathbb{R}^{n \times p}$  und  $\boldsymbol{\gamma} \in \mathbb{R}^q$ ,  $Z \in \mathbb{R}^{n \times q}$ . Dabei bezeichnen  $g(\cdot)$  und  $h(\cdot)$  die jeweiligen Linkfunktionen,  $X$  und  $Z$  die Designmatrizen der jeweiligen erklärenden Variablen,  $\boldsymbol{\beta}$  und  $\boldsymbol{\gamma}$  die Modellparameter,  $V(\cdot)$  und  $V_D(\cdot)$  die jeweiligen Varianzfunktionen und  $\psi$  den (festen) Dispersionsparameter des Dispersionsmodells. Wir nehmen an, dass die Dispersionsparameter  $\phi_i$  von den Erwartungswerten  $\mu_i$  funktional unabhängig sind. Das heißt, dass die Abhängigkeit der Varianz vom Erwartungswert vollständig durch die Varianzfunktion  $V(\mu)$  erklärt ist.

*Bemerkung.* Verwendet man für die Modellierung des Erwartungswertes und der Dispersion den gleichen Satz an erklärenden Variablen (gilt also  $X = Z$ ), können implizite Abhängigkeiten auftreten [vgl. Nelder and Pregibon, 1987].

Die Response-Variable des Dispersionmodells  $t_i$  muss derart gewählt sein, dass  $\mathbb{E}(t_i) = \phi_i$  und  $\text{var}(t_i) = \psi V_D(\phi_i)$  gilt. Eine mögliche Wahl ist die  $i$ -te Komponente der Deviance  $d_i$ , wobei dann  $\mathbb{E}(d_i) \approx \phi_i$  nur approximativ gilt, wie im vorherigen Abschnitt gezeigt wurde.

Um die Wahl von  $t_i = d_i$  zu motivieren, untersuchen wir die Eigenschaften der EQL-Funktion. Diese lautet für nicht konstante Dispersionsparameter:

$$Q^+(\boldsymbol{\mu}, \boldsymbol{\phi}, \mathbf{y}) = -\frac{1}{2} \sum_{i=1}^n \{\log(2\pi\phi_i V(y_i)) + D(y_i, \mu_i)/\phi_i\}. \quad (1.16)$$

Betrachtet man die  $\phi_i$  als bekannt, erhält man über die Ableitung von (1.16) nach  $\beta_j$ ,  $1 \leq j \leq p$  die bereits bekannte Score-Gleichung:

$$\sum_{i=1}^n \frac{y_i - \mu_i}{\phi_i V(\mu_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0.$$

Hält man umgekehrt  $\mu_i$  fest, liefert die Ableitung von (1.16) nach  $\gamma_k$ ,  $1 \leq k \leq q$  für  $d_i = D(y_i, \mu_i)$  die Gleichung:

$$\sum_{i=1}^n \frac{d_i - \phi_i}{2\phi_i^2} \frac{\partial \phi_i}{\partial \gamma_k} = 0. \quad (1.17)$$

Gleichung (1.17) entspricht also einem Quasi-Likelihood-Ansatz für die Response-Variable  $\mathbf{d} = (d_1, \dots, d_n)^\top$  mit Erwartungswert  $\mathbb{E}(d_i) = \phi_i$ , quadratischer Varianzfunktion  $V_D(\phi_i) = \phi_i^2$  und Dispersionsparameter  $\psi = 2$ .

Um nun die Parametervektoren  $\boldsymbol{\beta}$  und  $\boldsymbol{\gamma}$  zu bestimmen, bedient man sich des *Seesaw-Algorithmus*' (Sägezahn-Algorithmus). Dazu berechnet man zuerst das GLM für den Erwartungswert, wobei man mit  $1 \leq i \leq n : \phi_i = 1$  startet. Mit den geschätzten Erwartungswerten  $\hat{\boldsymbol{\mu}}$  erhält man auch die Deviance  $\mathbf{d}$ , die als Response im Dispersionsmodell dienen soll. Aus dieser bestimmt man dann eine Schätzung  $\hat{\boldsymbol{\phi}}$  für die Dispersionsparameter. Selbige verwendet man nun wiederum als a-priori Gewichte im GLM für den Erwartungswert. Dieser Vorgang wird wiederholt, wobei in der Regel vier bis fünf Iterationen ausreichend sind [vgl. Nelder and Lee, 1998]. Einen Algorithmus zur Bestimmung der Parametervektoren  $\boldsymbol{\beta}$  und  $\boldsymbol{\gamma}$  findet man bei Smyth [1989].

Für die aus der EQL-Funktion abgeleitete gemeinsame Modellierung von Erwartungswert und Dispersion verwendet man an zwei Stellen Approximationen. Zum einen nimmt man an, dass  $\mathbb{E}(d_i) = \phi_i$  gilt. Der Bias für  $d_i$  ist i. Allg. klein, außer für extreme Fälle, wie für Poisson-verteilte Fehler mit kleinen  $\mu$ -Werten. Da ein allfälliger Bias des Schätzers  $d_i$  allerdings nur als a-priori Gewicht mit  $\boldsymbol{\mu}$  interferiert, ist sein Einfluss auf die Schätzung von  $\boldsymbol{\beta}$  gering [vgl. Nelder and Lee, 1998].

Zum anderen nimmt man unabhängig von der zugrunde liegenden Verteilung für das Erwartungswertmodell eine Gamma-Varianz für die Dispersion an. Sind die  $Y_i$  normalverteilt, folgt, dass die  $d_i := D(y_i, \mu_i)$  eine  $\phi_i \chi_1^2$ -Verteilung aufweisen. Die  $\chi_n^2$ -Verteilung ist

äquivalent zu einer  $\Gamma(n/2, 2)$ -Verteilung und damit lässt sich für  $d_i/\phi_i \sim \chi_1^2 \Leftrightarrow d_i/\phi_i \sim \Gamma(1/2, 2)$ <sup>4</sup> die Varianz durch  $\text{var}(d_i/\phi_i) = 2 \Rightarrow \text{var}(d_i) = 2\phi_i^2$  ausdrücken und damit beschreibt die Gamma-Varianz die Varianz der Deviance-Komponenten bei normalverteilten Response-Variablen exakt.

Für nicht normalverteilte Response-Variablen trifft dies allerdings nicht mehr zu. Pierce and Schafer [1986] zeigen aber, dass die Berechnung der Deviance-Residuen  $r_i^D := \text{sign}(y_i - \hat{\mu}_i) \sqrt{D(y_i, \hat{\mu}_i)}$  der „best normalizing transformation“ (der besten Transformation auf Normalverteilung) der GLM-Verteilung sehr ähnlich ist, womit die  $i$ -te (skalierte) Deviance-Komponente (die dem  $i$ -ten quadrierten Deviance-Residuum entspricht) als Quadrat einer approximativ standardnormalverteilten Größe  $\chi_1^2$ -verteilt ist, was – wie bereits gezeigt – einer  $\Gamma(1/2, 2)$ -Verteilung entspricht. Somit lässt sich auch für nicht normalverteilte Response-Variablen die Verwendung der Gamma-Varianz rechtfertigen.

*Bemerkung 1.* Eine Variante der EQL-Funktion ist die **PL**-Funktion [Davidian and Carroll, 1988]. Dazu betrachtet man die log-Likelihood-Funktion einer Stichprobe  $\mathbf{y}$  unabhängig normalverteilter Zufallszahlen mit unterschiedlichen Erwartungswerten  $\mu_i$  und (unterschiedlichen) Varianzen  $\sigma_i^2$ , wobei man die Varianz als Funktion des Erwartungswertes über  $\text{var}(Y_i) = \phi_i V(\mu_i) = \sigma_i^2 \cdot 1$  ausdrückt:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \mathbf{y}) &= -\frac{1}{2} \sum_{i=1}^n \log(2\pi\sigma_i^2) - \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{2\sigma_i^2} \\ &= -\frac{1}{2} \sum_{i=1}^n \log(2\pi\phi_i V(\mu_i)) - \frac{1}{2\phi_i} \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{V(\mu_i)} \\ &= -\frac{1}{2} \sum_{i=1}^n \{ \log(2\pi\phi_i V(\mu_i)) + r_i^2/\phi_i \}, \end{aligned} \quad (1.18)$$

dabei bezeichnen die  $r_i$  die Pearson-Residuen. Die Gleichung (1.18) ist für normalverteilte Variablen eine exakte log-Likelihood-Funktion. Verwendet man sie jedoch auch für nicht normalverteilte Variablen (verwendet man also eine andere Varianzfunktion), trifft dies nicht mehr zu. Wir bezeichnen die Größe (1.18) als *Pseudo-Likelihood-Funktion*.

Die PL-Funktion unterscheidet sich von der EQL-Funktion also in der Verwendung der (quadrierten) Pearson-Residuen anstatt der Deviance und im Argument der Varianzfunktion im ersten Term. Verwendet man als Response-Variable im Dispersionsmodell (1.15) die quadrierten Pearson-Residuen,<sup>5</sup> erhält man die Schätzer für  $\boldsymbol{\gamma}$  über Maximierung der PL-Funktion.

*Bemerkung 2.* Da ML-Schätzer für die Varianz in Regressionsmodellen verzerrt sein können, greift man beispielsweise beim LM auf **REML**-Schätzer zurück.<sup>6</sup> Damit erreicht man zumindest asymptotische Unverzerrtheit. Smyth and Verbyla [1999] erweitern die Idee der REML-Schätzer auch für nicht normalverteilte Response-Variablen.

<sup>4</sup>Dabei ist die Parametrisierung derart gewählt, dass  $\mathbb{E}(d_i/\phi_i) = 2 \cdot 1/2 = 1$  und  $\text{var}(d_i/\phi_i) = 4 \cdot 1/2 = 2$  gilt.

<sup>5</sup>Für die Pearson-Residuen  $r_i$  gilt im Gegensatz zu den Deviance-Komponenten  $d_i : \mathbb{E}(r_i^2) = \phi_i$  exakt.

<sup>6</sup>REML wird auch als *Residual-Maximum-Likelihood* bezeichnet.

Tabelle 1.2: EQL-Modelle für die gemeinsame Modellierung von Erwartungswert und Dispersion

Komponente	Erwartungswertmodell GLM für $\mu$	Dispersionsmodell GLM für $\phi$
Response	$y$	$d$
Erwartungswert	$\mu$	$\phi$
Varianz	$\phi V(\mu)$	$2\phi^2$
Linkfunktion	$\eta = g(\mu)$	$\zeta = h(\phi)$
Linearer Prädiktor	$\eta = X\beta$	$\zeta = Z\gamma$
Deviance Komponente	$d$	$2 \{-\log(d/\phi) + (d - \phi)/\phi\}$
a-priori Gewichte	$1/\phi$	1

*Bemerkung 3.* Eine andere Zugangsweise zur Lösung von Problemen mit nicht konstanten Dispersionsparametern findet man bei Efron [1986]. Durch die Einführung der *Double Exponential Family* („Doppelte Exponentialfamilie“) – die bezüglich der Dispersionsparameter  $\phi$  ähnliche Eigenschaften wie für den Vektor der Erwartungswerte  $\boldsymbol{\mu}$  besitzt – wird auch eine Modellierung der Dispersionsparameter ermöglicht. Beide Ansätze liefern dabei die gleichen Resultate [vgl. Lee and Nelder, 2000].

Zusammenfassend lässt sich die Idee der gemeinsamen Modellierung von Erwartungswert und Dispersionsparameter also auf die EQL- bzw. die PL-Funktion zurückführen, da die jeweiligen Score-Funktionen nach  $\phi$  die Gestalt einer Score-Funktion eines GLM aufweisen. Die Tabelle 1.2 fasst die Komponenten der beiden, aus der EQL-Funktion abgeleiteten Modelle noch einmal zusammen [vgl. Nelder, 1998].

## 1.2 Parametrisierte Varianzfunktionen

Eine weitere Anwendungsmöglichkeit der EQL-Funktion besteht in der Verwendung von Varianzfunktionen aus einer parametrisierten Familie  $\mathcal{F}_{\boldsymbol{\theta}} \ni V_{\boldsymbol{\theta}}(\mu) := V(\mu, \boldsymbol{\theta})$ , wobei  $\boldsymbol{\theta}$  einen Vektor unbekannter Parameter der Varianzfunktion beschreibt. Die EQL-Funktion lautet dafür:

$$Q_{\boldsymbol{\theta}}^+(\boldsymbol{\mu}, \phi, \mathbf{y}) = -\frac{1}{2} \sum_{i=1}^n \{\log(2\pi\phi V_{\boldsymbol{\theta}}(y_i)) + D_{\boldsymbol{\theta}}(y_i, \mu_i)/\phi\}, \quad (1.19)$$

wobei die Deviance  $D_{\boldsymbol{\theta}}(y_i, \mu_i)$  über die QL-Funktion von der Varianzfunktion  $V_{\boldsymbol{\theta}}$  und damit von  $\boldsymbol{\theta}$  selbst abhängt. Das Ziel ist nun die Bestimmung jenes Parameter-Schätzers  $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$  der die EQL-Funktion (1.19) maximiert.

Um statistische Aussagen über die nicht linearen Parameter treffen zu können, schlagen Nelder and Pregibon [1987] vor, sogenannte *Profile-Quasi-Likelihood-Intervalle* zu verwenden. Dazu hält man den Parameter  $\boldsymbol{\theta}$  auf einem Wert  $\boldsymbol{\theta}_0$  fest und berechnet den Schätzer  $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)$  mittels der bereits vorgestellten Methoden. Diesen Vorgang wiederholt

man für verschiedene Werte aus einem zu definierendem Intervall  $I$  (das auch mehrdimensional sein kann) und bestimmt damit den Maximalwert der EQL-Funktion aus diesem Intervall:

$$Q_{\max}^+ = \max_{\theta \in I} Q_{\theta}^+. \quad (1.20a)$$

Definiert man weiters den Wert

$$Q_{\text{PL}}^+ = Q_{\max}^+ - Q_{\theta}^+, \quad (1.20b)$$

lässt sich über die Gleichungen (1.20) eine Menge  $\Theta$  finden, für dessen Werte die Differenz (1.20b) kleiner als eine vorgegebene Schranke  $q$  ist. Die Menge  $\Theta$  entspricht für den Fall, dass  $\theta \in \mathbb{R}$  gilt (die Varianzfamilie  $\mathcal{F}_{\theta}$  hängt also nur von dem Skalar  $\theta$  ab), einem Intervall. Dieses Intervall wird als Profile-Quasi-Likelihood-Intervall bezeichnet. Eine mögliche Wahl für  $q$  ist beispielsweise ein  $1/2 \cdot \chi_{\text{df}}^2$ -Quantil, wobei die Anzahl der Freiheitsgrade der Dimension von  $\theta$  entspricht. Bei Candy [2004] findet man eine praktische Anwendung der Profile-QL-Methode zur Bestimmung des Parameters  $\theta$  der Varianzfunktion  $V(\mu) = \mu^{\theta}$  im Zuge der Modellierung von Fischfangquoten.

Für Verteilungen aus der Exponentialfamilie findet man asymptotische Eigenschaften für Profile-QL-Intervalle bei Jørgensen [1983]. Trifft man nur Annahmen bezüglich der ersten beiden Momente, regen Nelder and Pregibon [1987] für bestimmte Datensätze Bootstrap-Verfahren an, um über die dadurch gewonnene „Resampling-Verteilung“ Intervalle für  $\theta$  zu bestimmen.

Im folgenden untersuchen wir unterschiedliche Familien von Varianzfunktionen und leiten uns die jeweiligen QL-Funktionen her, die in die Deviance einfließen.

### 1.2.1 Potenz-Ansatz

Verwenden wir für die Varianzfunktion einen Potenz-Ansatz  $V(\mu) = \mu^k$ , erhalten wir je nach Wahl von  $k$  unterschiedliche Resultate, wie sie im folgenden zusammengefasst sind.

- $k = 0$ : Die Varianzfunktion lautet somit  $V(\mu) = 1$  und entspricht der Annahme einer Normalverteilung. Es gilt also für  $y, \mu \in \mathbb{R}, \phi = \sigma^2$ :

$$Q(\mu, y) = \int^{\mu} \frac{y-t}{\sigma^2} dt + \text{Funktion in } y = -\frac{(y-\mu)^2}{2\sigma^2}.$$

- $k = 1$ : Der Ansatz von  $V(\mu) = \mu$  entspricht bei einem fixen Dispersionsparameter von  $\phi = 1$  einer Poisson-Verteilungsannahme. Die QL-Funktion ergibt damit für  $\mu > 0, y \geq 0$ :

$$Q(\mu, y) = \int^{\mu} \frac{y-t}{t} dt = y \log \mu - \mu.$$

- $k = 2$ : Die quadratische Varianzfunktion  $V(\mu) = \mu^2$  ergibt sich bei einer Gamma-Verteilung  $Y \sim \Gamma(\mu, 1)$ . Für  $\mu > 0, y \geq 0$  gilt:

$$Q(\mu, y) = \int^{\mu} \frac{y-t}{t^2} dt = -\frac{y}{\mu} - \log \mu.$$

- $k \in \mathbb{R} \setminus \{1, 2\}$ : Die Varianz-Annahme  $V(\mu) = \mu^k$  führt für  $k \in \mathbb{R} \setminus \{1, 2\}$  bei  $\mu > 0, y \geq 0$  zu der QL-Funktion

$$Q(\mu, y) = \int^{\mu} \frac{y-t}{t^k} dt = \frac{1}{(k-2)\mu^{k-2}} - \frac{y}{(k-1)\mu^{k-1}}.$$

*Bemerkung 1.* Die Potenzfunktion  $V(\mu) = \mu^k$  führt also für  $k \in \{0, 1, 2, 3\}$  zu einer Verteilung aus der Exponentialfamilie.<sup>7</sup>

*Bemerkung 2.* Die Familie der *Tweedie-Verteilungen* ist eine Familie von Verteilungen aus der Klasse der *Exponential Dispersion Models* (exponentielle Dispersionsmodelle),<sup>8</sup> die auf Jørgensen [1987] zurückgeht. Dabei charakterisiert eine Tweedie-Verteilung einen Erwartungswert von  $\mathbb{E}(Y) = \mu$  und eine Varianz von  $\text{var}(Y) = \phi\mu^k$  mit  $k \leq 0 \vee k \geq 1$  und damit sind insbesondere die Normal-, die Poisson-, die Gamma- und die Inverse-Gauß-Verteilung Mitglieder der Tweedie-Familie. Für den Potenz-Ansatz  $V(\mu) = \mu^k$  bei einem EQL-Modell existiert somit für  $k \leq 0 \vee k \geq 1$  eine Verteilung aus der Tweedie-Familie.

Das R-Paket *tweedie* [Dunn, 2007] stellt die Funktion *tweedie.profile* zur Verfügung, die den Schätzer für den Parameter  $k$  der Tweedie-Verteilung über einen Profile-Likelihood-Ansatz bestimmt. Der Dispersionsparameter  $\phi$  wird in diesem Paket über die Maximierung der log-Likelihood-Funktion bestimmt und ist somit ein ML-Schätzer. Diese Vorgehensweise unterscheidet sich von dem durch die EQL-Funktion induzierten Verfahren, indem sie stärkere Annahmen<sup>9</sup> trifft und somit einen anderen Zugang wählen kann.

---

<sup>7</sup>Dabei gehört die Varianz-Annahme  $V(\mu) = \mu^3$  zu einer Inversen-Gauß-Verteilung.

<sup>8</sup>Die Familie der „Exponential Dispersion Models“ selbst ist eine Erweiterung der Exponentialfamilie.

<sup>9</sup>Es wird die Kenntnis der Verteilung zugrunde gelegt, während sich der EQL-Ansatz mit der Existenz der ersten beiden Momente begnügt.

## 1.3 Beispiele

In diesem Abschnitt wollen wir die Funktionalität der Bibliothek *EQL* anhand einiger Beispiele vorstellen.

### 1.3.1 Potenzfamilie

#### Anschauungsbeispiel

Zuerst wollen wir in einem konstruierten Beispiel die Plausibilität der Funktion *eql* überprüfen. Dazu generieren wir zuerst je zwei Gruppen zu 100 poissonverteilten Zufallsvariablen  $Y_{ij} \sim \mathcal{P}(\mu_{ij})$ ,  $1 \leq i \leq 100, j = 1, 2$  mit den Erwartungswerten  $\mu_{i1} = 10$  und  $\mu_{i2} = 100$ . Die Poissonverteilung ist durch die Varianzfunktion  $V(\mu) = \mu$  charakterisiert und ist somit ein Mitglied der Potenzfamilie  $\mathcal{F}_\theta = \{V_\theta(\mu) = \mu^\theta\}$ . Wir wollen das Modell

$$1 \leq i \leq 100 : x_{ij} = \begin{cases} 1 & j = 1, \\ 2 & j = 2 \end{cases},$$

$$\log \mu_{ij} = \beta \cdot x_{ij} \quad \text{mit} \quad V_\theta(\mu) = \mu^\theta$$

an die Daten anpassen, wobei wir den EQL-Ansatz verwenden um den Parameter  $\theta$  zu schätzen. Da die Daten aus einer Poissonverteilung kommen, erwarten wir ein Ergebnis für  $\hat{\theta}$  in der Nähe von Eins. Der tatsächliche Wert des Parameters  $\beta$  liegt bei  $\beta = \log 10 = 2.3026$ , da gilt:

$$\begin{aligned} \log 10 = \beta &\Rightarrow \beta = &= 2.3026, \\ \log 100 = 2\beta &\Rightarrow \beta = \frac{1}{2} \log 100 = \log 10 = 2.3026. \end{aligned}$$

Als Suchgitter wählen wir ein äquidistantes Gitter im Intervall  $[0, 2]$ . Der folgende R-Code fasst die Berechnungen zusammen:

```
> library(EQL)
Lade nötiges Paket: ttutils

> set.seed(1234)

> y <- c(rpois(100, 10), rpois(100, 100))
> x <- rep(1:2, each=100)

> ps <- list(seq(0, 2, length=20))

> eq <- eql(y ~ x - 1, ps, powerVarianceFamily(), include.model=TRUE)
~~ Extended Quasi Likelihood ~~
* Dimension of parameter space: 20
* Starting loop:
```



Tabelle 1.3: Parameterschätzer für den Textildatensatz

Methode	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\theta}$	EQL-Wert
EQL	6.3477	0.8408	-0.6288	-0.3707	2.4906	-160.5579
Tweedie	6.3478	0.8407	-0.6286	-0.3718	2.4612	-160.5612
Direkte Optimierung	6.3487	0.8433	-0.6267	-0.3727	2.5360	-160.5726

### Textildaten

Den in Abschnitt ?? vorgestellten Datensatz über die Strapazierfähigkeit von Textilfasern von Box and Cox [1964] verwenden Nelder and Pregibon [1987] um mit Hilfe der EQL-Funktion den Parameter der Potenzfamilie zu bestimmen. Wir wollen diesen Datensatz ebenfalls analysieren und die erhaltenen Schätzer mit Schätzern vergleichen, die der Tweedie-Ansatz und eine direkte numerische Optimierung der EQL-Funktion liefert.

*Bemerkung 1.* Der Unterschied der direkten Optimierung zur Funktionalität von `eql` besteht darin, dass in der direkten Optimierung versucht wird, die linearen und die nicht linearen Parameter des Modells unter Zuhilfenahme eines numerischen Optimierungsverfahrens *direkt* über die Maximierung der EQL-Funktion zu bestimmen, während `eql` die linearen Parameter (also den Parametervektor  $\beta$ ) mit den Methoden des Generalisierten Linearen Modells bestimmt und den nicht linearen Parameter (also den Parameter  $\theta$  der Varianzfamilie) über die Gittersuche ermittelt.

*Bemerkung 2.* Im Allgemeinen verlangt die Verwendung eines numerischen Optimierungsverfahrens eine Spezifizierung eines Startvektors für die Parameter, über die optimiert werden soll. Daher erweist sich diese Methode zur Bestimmung der Parameter als wenig praktikabel, da sie eine ungefähre Kenntnis der Größerordnung aller im Modell vorkommenden Parameter voraussetzt.

Wie bereits in Abschnitt ?? betrachten wir das Modell

$$\log \mu = \alpha + x_1\beta_1 + x_2\beta_2 + x_3\beta_3, \quad V_\theta(\mu) = \mu^\theta,$$

wobei wir für den Tweedie-Ansatz noch zusätzlich  $Y \sim$  Tweedie-Verteilung annehmen.

Das Suchgitter für die Funktionen `tweedie.profile` und `eql` besteht in beiden Fällen aus einer äquidistanten Folge im Intervall  $[1.1, 4]$ . Die Startwerte für die direkte Optimierung setzen wir mit  $(\alpha^0, \beta_1^0, \beta_2^0, \beta_3^0, \theta^0) = (4, 0, 0, 0, 1)$  fest. Die resultierenden Schätzer sind in Tabelle 1.3 zusammengefasst. Man erkennt, dass alle Ansätze in etwa die gleichen Ergebnisse liefern, wobei der Ansatz über die Gittersuche den höchsten EQL-Wert liefern kann. Allerdings unterscheiden sich die Schätzer dieser Methode von den Schätzern der anderen Methoden nur kaum. Die Schätzer für die linearen Parameter beim Tweedie- und beim EQL-Ansatz sind nahezu identisch, wenngleich die EQL-Funktion einen leicht größeren Wert für  $\theta$  schätzt, der allerdings selbst wiederum kleiner als der von der direkten Optimierung bestimmte Wert ist.

Vergleicht man die log-Likelihood-Funktion der Tweedie-Verteilung in Abbildung ?? mit dem Profile-EQL-Plot in Abbildung 1.2 erkennt man, dass das Konfidenzintervall

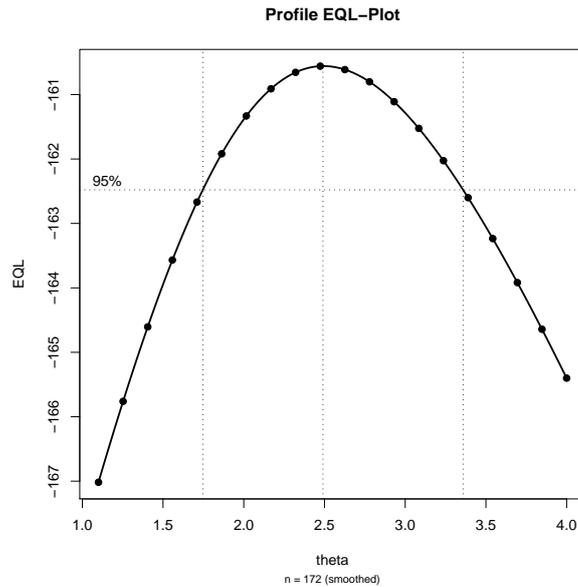


Abbildung 1.2: Profile-Plot für den Potenz-Ansatz für den Textildatensatz

bei letzterem größer ist, was sich damit erklären lässt, dass der Tweedie-Ansatz – wie bereits erläutert – restriktivere Annahmen trifft und damit stärkere Aussagen zu machen im Stande ist.

*Bemerkung.* Die direkte Optimierung reagiert äußerst sensitiv auf die Wahl der Startwerte. Verwendet man beispielsweise den Vektor  $(\alpha^0, \beta_1^0, \beta_2^0, \beta_3^0, \theta^0) = (0, 0, 0, 0, 1)$  erhält man die Schätzer  $(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\theta}) = (4.9780, 1.0205, -2.0455, -0.3083, 0.3881)$  mit einem zugehörigen EQL-Wert von  $-207.4769$  als Lösung. Daher ist die direkte Optimierung nur dann zuverlässig, wenn man die Größenordnung der zu schätzenden Parameter kennt.

### 1.3.2 Erweiterte Binomialfamilie

In Tabelle 1.4 findet sich ein Datensatz [vgl. Wedderburn, 1974], der die von der Blattfleckenkrankheit (*Rhynchosporium secalis*) befallene Blattfläche verschiedener Gerstenarten für unterschiedliche Anbaugebiete beschreibt. McCullagh and Nelder [1989] untersuchen diesen Datensatz unter Zuhilfenahme des QL-Ansatzes  $V(\mu) = \mu^2(1 - \mu)^2$ , da die Binomialvarianz  $V(\mu) = \mu(1 - \mu)$  speziell für sehr kleine bzw. sehr große Flächenanteile zu groß sein scheint.

*Bemerkung.* Der Datensatz enthält exakte Nullen und für die erweiterte Binomialvarianz gilt  $\forall k, l : V(0) = 0$ . Da die EQL-Funktion für Varianzfunktionen mit  $V(\cdot) = 0$  nicht definiert ist, belegen wir die entsprechenden Zellen mit dem Wert NA. Das heißt im speziellen auch, dass die Ergebnisse von denen von McCullagh and Nelder abweichen.

Mit Hilfe des EQL-Ansatzes wollen wir nun untersuchen, welche Werte für  $k$  und  $l$  plausibel sind. Dazu passen wir ein Modell mit den beiden Faktoren „Anbaugebiet“ und

Tabelle 1.4: Anteil [in %] der mit der Blattfleckenkrankheit befallenen Blattfläche verschiedener Gerstenarten

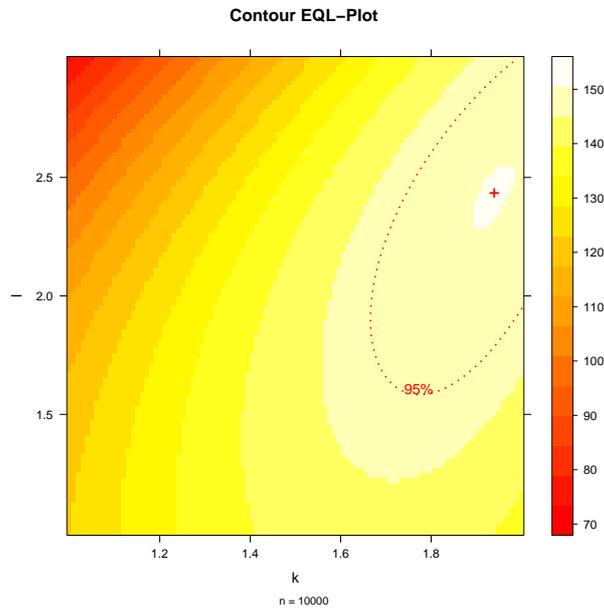
Gebiet	Gerstenart									
	1	2	3	4	5	6	7	8	9	10
1	0.05	0.00	0.00	0.10	0.25	0.05	0.50	1.30	1.50	1.50
2	0.00	0.05	0.05	0.30	0.75	0.30	3.00	7.50	1.00	12.70
3	1.25	1.25	2.50	16.60	2.50	2.50	0.00	20.00	37.50	26.25
4	2.50	0.50	0.01	3.00	2.50	0.01	25.00	55.00	5.00	40.00
5	5.50	1.00	6.00	1.10	2.50	8.00	16.50	29.50	20.00	43.50
6	1.00	5.00	5.00	5.00	5.00	5.00	10.00	5.00	50.00	75.00
7	5.00	0.10	5.00	5.00	50.00	10.00	50.00	25.00	50.00	75.00
8	5.00	10.00	5.00	5.00	25.00	75.00	50.00	75.00	75.00	75.00
9	17.50	25.00	42.50	50.00	37.50	95.00	62.50	95.00	95.00	95.00

„Gerstenart“ an und untersuchen mit dem EQL-Konturplot welche Kombinationen von Werten für  $k$  und  $l$  innerhalb des Konfidenzbereichs liegen.

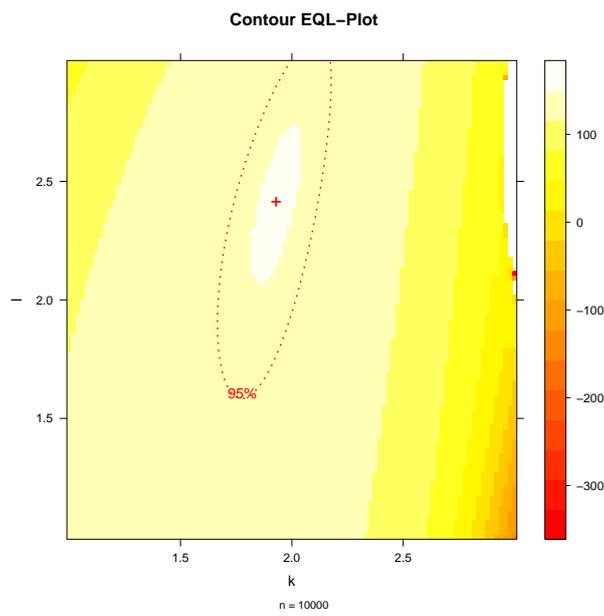
Für das Suchgitter bilden wir das kartesische Produkt der äquidistanten Folgen  $L_1 \subseteq [1, 2]$  und  $L_2 \subseteq [1, 3]$  mit  $|L_1| = |L_2| = 100$ . Damit ergibt sich ein Suchgitter  $\Theta$  mit  $|L_1| \cdot |L_2| = 10000$  Gitterpunkten.

Der maximale EQL-Wert im Gitter von 150.6194 wird für  $\hat{k} = 1.939$  und  $\hat{l} = 2.434$  erreicht. Die Abbildung 1.3(a) zeigt den EQL-Konturplot und man erkennt, dass die von McCullagh and Nelder gewählte Parameterwahl  $k = l = 2$  im Konfidenzbereich liegt, während die Binomialvarianz  $k = l = 1$  von den Daten ganz deutlich *nicht* unterstützt wird.

Die Deviance bei der erweiterten Binomialvarianz wird – wie bereits besprochen – numerisch bestimmt. Für gewisse Parameter-Kombinationen konvergiert das Integrationsverfahren allerdings nicht. Verschiebt man das Suchgitter etwas nach links und verwendet eine äquidistante Folge  $L_1 \subseteq [1, 3]$ , können gewisse Kombinationen nicht berechnet werden. In der Abbildung 1.3(b) erkennt man am rechten Rand einen weißen Bereich von Parameter-Kombinationen, für die kein EQL-Wert berechnet werden konnte.



(a)  $L_1 \subseteq [1, 2]$



(b)  $L_1 \subseteq [1, 3]$

Abbildung 1.3: EQL-Konturplots für den Blattflecken-Datensatz

# Literaturverzeichnis

- G. E. P. Box and D. R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26:211–252, 1964.
- S. G. Candy. Modelling catch and effort data using generalised linear models, the tweedie distribution, random vessel effects and random stratum-by-year effects. *CCAMLR Science*, 11:59–80, 2004.
- M. Davidian and R. J. Carroll. A note on extended quasi-likelihood. *Journal of the Royal Statistical Society. Series B (Methodological)*, 50:74–82, 1988.
- P. K. Dunn. *tweedie: Tweedie exponential family models*, 2007. R package version 1.5.2.
- B. Efron. Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, 81:709–721, 1986.
- B. Jørgensen. Maximum likelihood estimation and large-sample inference for generalized linear and nonlinear regression models. *Biometrika*, 70:19–28, 1983.
- B. Jørgensen. Exponential dispersion models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 49:127–162, 1987.
- Y. Lee and J. A. Nelder. The relationship between double-exponential families and extended quasi-likelihood families, with application to modelling geissler’s human sex ratio data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49: 413–419, 2000.
- P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Monographs on Statistics and Applied Probability. Chapman and Hall, 1989.
- J. A. Nelder. A large class of models derived from generalized linear models. *Statistics in Medicine*, 17:2747–2753, 1998.
- J. A. Nelder and Y. Lee. Generalized linear models for the analysis of taguchi-type experiments. *Applied Stochastic Models and Data Analysis*, pages 107–120, 1991.
- J. A. Nelder and Y. Lee. Likelihood, quasi-likelihood and pseudolikelihood: some comparisons. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54:273–284, 1992.
- J. A. Nelder and Y. Lee. Joint modeling of mean and dispersion. *Technometrics*, 40: 168–175, 1998.

- J. A. Nelder and D. Pregibon. An extended quasi-likelihood function. *Biometrika*, 74: 221–232, 1987.
- D. A. Pierce and D. W. Schafer. Residuals in generalized linear models. *Journal of the American Statistical Association*, 81:977–986, 1986.
- G. K. Smyth. Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society. Series B (Methodological)*, 51:47–60, 1989.
- G. K. Smyth and A. P. Verbyla. Adjusted likelihood methods for modelling dispersion in generalized linear models. *Environmetrics*, 10:695–709, 1999.
- R. W. M. Wedderburn. Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika*, 61:439–447, 1974.