3 Generalisierte lineare Modelle

Aufgabe 6

Sei y eine eindimensionale Zufallsvariable, die von einer eindimensionalen Kovariable x abhänge. Nehmen Sie an, wir hätten den Erwartungswert der Verteilung $y|x$ durch

$$\mu = \exp(\beta_0 + \beta_1 x)$$

korrekt spezifiziert, $y|x$ sei poissonverteilt. Mangels Kenntnis der wahren Verteilung von $y|x$ bedienen wir uns der "Arbeits-Varianz"

(a) $v(\mu) = \mu$ bzw.

(b) $v(\mu) = \mu^2$,

jeweils mit Dispersion ϕ. Wie lauten in (a) und (b) die Quasi-Score-Funktion $s_Q(\beta_0, \beta_1)$ und die Quasi-Fisher-Information $F_Q(\beta_0, \beta_1)$?

Aufgabe 7

Der Datensatz `leafblotch` (McCullagh & Nelder, 1989; Download von der Veranstaltungshomepage) untersucht die Auswirkungen einer Blattkrankheit. Er enthält folgende Variablen:

`blotch`	Anteil des Befalls an der Oberfläche der Blätter
`site`	Anbau-Ort (kategorial, 9 Orte)
`variety`	Gerstensorte (kategorial, 10 Sorten)

Die Response-Variable `blotch` ist nicht binomial, aber auf das Intervall $[0, 1]$ beschränkt, da es sich um Anteile handelt. Daher erscheint eine Modellierung mit einem Quasi-Likelihood-Ansatz sinnvoll, bei dem die Strukturannahme für den Erwartungwert über den Logit-Link spezifiziert wird.

(a) Fitten Sie ein Haupteffektmodell mit den Prädiktoren `site` und `variety` über einen Quasi-Likelihood-Ansatz mit Logit-Link, der die gleiche Varianzstruktur wie das Logit-Modell verwendet, also $v(\mu) = \mu(1 - \mu)$. Was fällt bei der Betrachtung des Outputs auf?

(b) Erstellen Sie zur Diagnose Ihres Modells aus Teilaufgabe (a) einen Plot, in dem die durch das Modell geschätzten linearen Prädiktoren ($\hat{\eta}_i$) gegen die Residuen des Modells (r_i) abgetragen werden. Welche Konsequenzen hat die (offensichtliche) Fehlspezifikation der Varianzfunktion?

(c) Versuchen Sie die Varianz Ihres Parameter-Schätzers aus Teilaufgabe (a) zu schätzen. Fassen Sie hierzu die Varianzfunktion $v(\mu) = \mu(1 - \mu)$ lediglich als Arbeits-Varianz auf.

(d) Eine alternative Wahl für die Varianzfunktion wäre $v(\mu) = \mu^2(1 - \mu)^2$. Diese ist im `R`-Package `gnm` implementiert (setze `family=wedderburn`). Fitten Sie das entsprechende Modell und betrachten Sie den Output sowie einen Plot analog zu Teilaufgabe (b). Halten Sie die Varianzfunktion $v(\mu) = \mu^2(1 - \mu)^2$ für besser geeignet als $v(\mu) = \mu(1 - \mu)$ (Begründung)?

(e) Nehmen Sie nun an, dass es sich bei $v(\mu) = \mu^2(1 - \mu)^2$ um die wahre Varianzfunktion handelt. Zeigen Sie für das Modell aus Teilaufgabe (b), dass damit die Varianzen der einzelnen Komponenten von $\hat{\boldsymbol{\beta}}$, die sich aus der asymptotischen Verteilung von $\hat{\boldsymbol{\beta}}$ ergeben, unabhängig vom wahren Koeffizientenvektor $\boldsymbol{\beta}$ sind.