3 Generalisierte lineare Modelle (IV)

Aufgabe 1

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 "working variance"

(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 2

Neben den Residuen $r_i = y_i - \hat{\mu}_i$ und den Pearsonresiduen kann man **Devianzresiduen** definieren:

$$r_{D_i} = \operatorname{sign}(y_i - \hat{\mu}_i)\sqrt{d_i},$$

wobei d_i die Devianzkomponenten in der Darstellung $D = \sum_{i=1}^n d_i$ sind.

Betrachten Sie nun ein generalisiertes lineares Modell mit Poisson-verteiltem Response, d.h. $y_i \sim \text{Po}(\mu_i)$ und $g(\mu_i) = \eta_i$; wählen Sie dabei für $g(\mu_i)$ den natürlichen Link.

(a) Bestimmen Sie Pearson- und Devianz-Residuen.

(b) Bei nicht-normalverteiltem Response haben Pearson-Residuen oft eine schiefe Verteilung. Daher betrachtet man bei den sog. Anscombe-Residuen nicht $y_i - \mu_i$ (bzw. $\hat{\mu}_i$), sondern $A(y_i) - A(\mu_i)$, wobei die Transformation $A(\mu)$ gegeben ist durch

$$A(\mu) = \int_{-\infty}^{\mu} \frac{1}{v(u)^{1/3}}\, du.$$

An Stelle von $\text{Var}(y_i)$ basiert die Skalierung der Residuen auf $\text{Var}(A(y_i))$, welche sich durch $A'(\mu_i)^2\text{Var}(y_i)$ approximieren lässt. Bestimmen Sie die Anscombe-Residuen für Poisson-verteilten Response.

(c) Veranschaulichen Sie Pearson-, Devianz- und Anscombe-Residuen mit Hilfe einer kleinen Simulation. Ziehen Sie hierzu Poisson-verteilte Response-Werte, wobei $\eta_i = \alpha + \beta x_i$, $i = 1, \ldots, n$. Wählen Sie dabei für die Koeffizienten z.B. $\alpha = 1$, $\beta = 2$. Der Prädiktor x könnte z.B. auf $[-1, +1]$ gleichverteilt sein.

Aufgabe 3

Im generalisierten linearen Modell ist die (generalisierte) Hat-Matrix gegeben durch

$$\mathbf{H} = \mathbf{W}^{T/2}\mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{1/2},$$

wobei $\mathbf{W} = \mathbf{W}(\hat{\boldsymbol{\beta}})$ die Gewichtungsmatrix und $\mathbf{X} = (1, \mathbf{x}_1, \ldots, \mathbf{x}_p)$ die Designmatrix darstellt. Man zeige:

(a) $\mathbf{H}$ ist symmetrisch und idempotent.

(b) Für die Diagonalelemente von $\mathbf{H}$ gilt: $0 \leq h_{ii} \leq 1$, $i = 1, \ldots, n$.

(c) Für das arithmetische Mittel der Diagonalelemente von $\mathbf{H}$ gilt: $\bar{h} = (p+1)/n$.