

Allgemeine GLM-Theorie

Aufgabe 1: Geometrisches Modell

Gegeben sei eine geometrisch verteilte Zufallsvariable y mit Dichte

$$f(y|p) = p(1-p)^y; \quad p \in (0,1), \quad y \in \{1, 2, 3, \dots\}$$

- (i) Zeigen Sie, dass diese Verteilung zu einer Exponentialfamilie gehört, und bestimmen Sie die Größen θ , $b(\theta)$ und den Dispersionsparameter ϕ .
- (ii) Geben Sie den natürlichen Link an und leiten Sie damit explizit die natürliche Linkfunktion $g(\cdot)$ als Funktion vom Erwartungswert μ sowie die Responsefunktion $h(\cdot)$ als Funktion des linearen Prädiktors η her.
- (iii) Der Einfluss der Kovariablen $\mathbf{x}_i^T = (1, x_{i1}, \dots, x_{ip})$ auf den Response y_i mit $i = 1, \dots, n$ soll nun mit Hilfe eines GLMs mit dem natürlichen Link aus b) modelliert werden. Stellen Sie die zugehörige Log-Likelihood-Funktion auf und leiten Sie daraus die Score-Funktion her. Überprüfen Sie anschließend Ihre Score-Funktion, indem Sie direkt die Formel aus der Formelsammlung verwenden. Was geschieht mit der Score-Funktion, wenn anstelle des natürlichen Links ein anderer verwendet wird?

Aufgabe 2: Zähldatenmodelle

Der Datensatz *aids.txt* (Download von der Tutoriumshomepage), beschreibt die AIDS-Toten pro Quartal in Australien für 14 Quartale des Zeitraumes Januar 1983 bis Juni 1986 und enthält die Variablen

<i>time</i>	Zeitindex für das Quartal
<i>deaths</i>	Anzahl der AIDS-Toten pro Quartal

- (i) Schätzen Sie mit R ein GLM für die Anzahl der AIDS-Toten pro Quartal, zunächst mit normalverteiltem Response und mit natürlicher Linkfunktion. Erstellen Sie einen Plot, in dem die beobachteten und die durch Ihr Modell geschätzten Werte abgebildet werden. Was fällt auf?
- (ii) Wählen Sie nun eine alternative Modellspezifikation (mit natürlichem Link), welche den Daten eventuell gerechter wird. Schätzen Sie das Modell, und vergleichen Sie den Plot der beobachteten gegen die geschätzten Werte mit dem aus Teilaufgabe (i).
- (iii) Berechnen Sie unter Verwendung des Attributs `coefficients()` einen Schätzwert für die Anzahl der AIDS-Toten im 3. Quartal 1986 basierend auf den in (i) und (ii) berechneten Modellen. Führen Sie nun die Berechnung mit der Funktion `predict()` durch und vergleichen Sie damit die Prognosen.