Jun.-Prof. Dr. Zakhar Kabluchko WS 2010/2011
Sebastian Lück 24.11.2010

# Risikotheorie II
## Übungsblatt 6

**Aufgabe 1**
Die Verteilung der Zufallsvariable $Y$ gehöre zu einer Exponentialfamilie in $\theta_1 \in \mathbb{R}$ mit einer Dichte der Form

$$f_\theta(y) = e^{\frac{-b(\theta_1)}{\theta_2}} \exp\left(\theta_1 \frac{y}{\theta_2}\right) h(\theta_2, y),$$

wobei $\theta_2 \neq 0$ ein weiterer Parameter ist. Sei $b : \Theta \longrightarrow \mathbb{R}$ zweimal stetig differenzierbar, wobei $\Theta \subset \mathbb{R}$.

(a) Zeige, dass unter den Voraussetzungen von Aufg. 2 auf Blatt 1

$$\mathbb{E}Y = b'(\theta_1) \text{ und } \mathrm{Var}Y = \theta_2 b''(\theta_1).$$

(b) Falls $b'$ invertierbar ist, heißt die Funktion $V(\mu) = b''\big((b')^{-1}(\mu)\big)$ die Varianzfunktion der durch $f_\theta$ gegebenen Verteilung und $(b')^{-1}$ die natürliche Linkfunktion. Berechne diese beiden Funktionen für die Bernoulli-Verteilung mit $\theta_1 = \log \frac{p}{1-p}$.

(c) Berechne die natürliche Linkfunktion und die Varianzfunktion der Normalverteilung mit $\theta_1 = \mu$ und Varianz $\sigma^2 > 0$.

**Aufgabe 2**
Um einen neuen Gebäudeversicherungstarif zu kalkulieren, soll der Einfluss des Gebäudetyps und des Gebäudealters auf die erwarteten Schäden ermittelt werden. Dies soll mit Hilfe eines verallgemeinerten linearen Modells (GLM) geschehen, d.h. man nimmt an, dass

$$(g(\mathbb{E}Y_1), \ldots, g(\mathbb{E}Y_n))^T = X\beta,$$

wobei $X \in \mathbb{R}^{n \times m}$, $\beta \in \mathbb{R}^m$ und für unsere Aufgabe $g(x) = x$ für alle $x \in \mathbb{R}$. Die Zufallsvariablen $Y_1, \ldots, Y_n$ sind dabei unabhängig, im allgemeinen nicht identisch verteilt, man nimmt bei GLMs aber generell an, dass sie einer Exponentialfamilie des Typs aus Aufg. 1 entstammen. Es wurden folgende mittlere Schäden und Stichprobenvarianzen beobachtet:

| | Mittelwert | Stichprobenvarianz |
|---|---|---|
| Klasse 1 Alter 1 | 200 | 2500 |
| Klasse 1 Alter 2 | 100 | 2524 |
| Klasse 2 Alter 1 | 400 | 2575 |
| Klasse 2 Alter 2 | 300 | 2410 |

(a) Welche der in Aufg. 1 untersuchten Varianzfunktionen passt zu dieser Datenlage?

(b) Gib eine passende Designmatrix $X$ an und interpretiere die Einträge des Parametervektors $\beta$. Hat $X$ vollen Rang?

(c) Ermittle die zugehörige log-Likelihoodfunktion

$$\log L(\bar{Y}_1, \ldots, \bar{Y}_n, \mu_1, \ldots, \mu_n)$$

unter der Annahme, dass $Y_{ij} \sim N(\mu_i, \sigma^2)$, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$.

(d) Transformiere $\log L(\bar{Y}_1, \ldots, \bar{Y}_n, \mu_1, \ldots, \mu_n)$ in eine Funktion in Abhängigkeit von $\beta_1, \ldots, \beta_m$ und formuliere notwendige Bedingungen für die Existenz von Maximumstellen.

**Aufgabe 3**

Diese Aufgabe kann mit $R$ bearbeitet werden. Gegeben seien Inputvariable $x$ und zugehörige Outputvariable $y$, wobei

$$x = (0.20, 3.74, 2.64, 4.29, 5.97, 2.58, 5.79, 1.18, 2.80, 0.71, 4.24, 5.22, 5.37, 5.46, 5.81, 5.33, 3.96, 2.77, 4.42, 1.85)$$

und

$$y = (3.61, -2.86, -1.60, -1.59, 4.98, -2.76, 4.32, 1.18, -3.94, 1.07, -2.36, 1.38, 4.46, 2.35, 6.20, 2.46, -1.08, -3.30, -0.65, -2.35)$$

(a) Visualisiere die Daten in einem Scatterplot.

(b) Passe ein einfaches lineares Regressionsmodell $Y = \beta_1 + \beta_2 X + \varepsilon$ an die Daten an, plotte die Regressionsgerade und teste unter der Annahme, dass $\varepsilon \sim N(0, \sigma^2 I)$ die Hypothese $H_0 : \beta_1 = \beta_2 = 0$.

(c) Passe ein quadratisches Regressionsmodell $Y = \beta_1 + \beta_2 X^2 + \beta_3 X + \varepsilon$ an die Daten an. Wiederhole den Test aus (b). Teste zusätzlich, ob die einzelnen Faktoren $\beta_i$ jeweils signifikant von 0 verschieden sind.

(d) Wiederhole Aufgabenteil (c), wenn der Inputvektor $x$ durch $(x_1 - 3, \ldots, x_{20} - 3)$ ersetzt wird.