Non-Life Insurance: Mathematics and Statistics

Exercise sheet 10

Exercise 10.1 Tariffication Methods (R Exercise)
Suppose that a car insurance portfolio of an insurance company has been divided according to two tariff criteria:

- vehicle type: {passenger car, delivery van, truck} = {1,2,3}
- driver age : {21-30 years, 31-40 years, 41-50 years, 51-60 years} = {1,2,3,4}

For simplicity, we set the number of policies $v_{i,j} = 1$ for all risk classes (i, j), $1 \leq i \leq 3$, $1 \leq j \leq 4$. Moreover, assume that we observed the following claim amounts

	21-30y	31-40y	41-50y	51-60y
passenger car	2'000	1'800	1'500	1'600
delivery van	2'200	1'600	1'400	1'400
truck	2'500	2'000	1'700	1'600

and that we work with a multiplicative tariff structure.

(a) Apply the method of Bailey & Simon in order to calculate the tariffs.
Hint: In order to get a unique solution, set $\mu = \chi_{1,1} = 1$.

(b) Apply the method of Bailey & Jung (i.e. the method of total marginal sums) in order to calculate the tariffs.
Hint: In order to get a unique solution, set $\mu = \chi_{1,1} = 1$.

(c) Determine the design matrix Z of the log-linear regression model and write an R-Code that applies the MLE method in the Gaussian approximation framework. Discuss the output. Is there statistical evidence that the classification into different types of vehicles could be omitted?
Hint: In order to get a unique solution, set $\beta_{1,1} = \beta_{2,1} = 0$.

(d) Compare the results found in parts (a) - (c).

Exercise 10.2 Claims Frequency Modelling with GLM (R Exercise)
Suppose that a motorbike insurance portfolio of an insurance company has been divided according to three tariff criteria:

- vehicle class: {weight over 60 kg and more than two gears, other} = {1,2}
- vehicle age: {at most one year, more than one year} = {1,2}
- geographic zone: {large cities, middle-sized towns, smaller towns and countryside} = {1,2,3}

Assume that we observed the following claim frequencies:

class	age	zone	volumes	number of claims	claim frequencies
1	1	1	100	25	0.250
1	1	2	200	15	0.075
1	1	3	500	15	0.030
1	2	1	400	60	0.150
1	2	2	900	90	0.100
1	2	3	7'000	210	0.030
2	1	1	200	45	0.225
2	1	2	300	45	0.150
2	1	3	600	30	0.050
2	2	1	800	80	0.100
2	2	2	1'500	120	0.080
2	2	3	5'000	90	0.018

(a) Write an R-Code that performs a GLM analysis for the claim frequencies.

(b) Plot the observed and the fitted claim frequencies against the vehicle class, the vehicle age and the geographic zone.

(c) Create a Tukey-Anscombe plot of the fitted claim frequencies versus the deviance residuals and a QQ plot of the deviance residuals versus the theoretical (estimated) quantiles.

(d) Is there statistical evidence that the classification into the geographic zones could be omitted?

Exercise 10.3 Tweedie's Compound Poisson Model

Let $S \sim \text{CompPoi}(\lambda v, G)$, where $\lambda > 0$ is the unknown claims frequency parameter, $v > 0$ the known volume and G the distribution function of a gamma distribution with known shape parameter $\gamma > 0$ and unknown scale parameter $c > 0$. Then S has a mixture distribution with a point mass of $\mathbb{P}[S = 0]$ in 0 and a density f_S on $(0, \infty)$.

(a) Calculate $\mathbb{P}[S = 0]$ and f_S.

(b) Show that S belongs to the exponential dispersion family with

$$
\begin{aligned}
w &= v, \\
\phi &= \frac{\gamma+1}{\lambda\gamma} \left(\frac{\lambda v \gamma}{c}\right)^{\frac{\gamma}{\gamma+1}}, \\
\theta &= -(\gamma+1) \left(\frac{\lambda v \gamma}{c}\right)^{-\frac{1}{\gamma+1}}, \\
\boldsymbol{\Theta} &= (-\infty, 0), \\
b(\theta) &= \frac{\gamma+1}{\gamma} \left(\frac{-\theta}{\gamma+1}\right)^{-\gamma}, \\
c(x, \phi, w) &= \log\left(\sum_{n=1}^{\infty} \left[\frac{(\gamma+1)^{\gamma+1}}{\gamma} \left(\frac{\phi}{w}\right)^{-\gamma-1}\right]^n \frac{1}{\Gamma(n\gamma) n!} x^{n\gamma-1}\right), \quad \text{if } x > 0, \text{ and} \\
c(0, \phi, w) &= 0.
\end{aligned}
$$