MATH3091: Statistical Modelling II

Problem Sheet 5

1. The time to failure (Y) of a certain type of electrical component is thought to follow an exponential distribution, with probability density of the form

$$f_Y(y;\lambda) = \lambda \exp(-\lambda y), \quad y > 0; \quad \lambda > 0\,.$$

It is believed that the failure rate of a component λ is related to its electrical resistance (x) by the relationship

$$\lambda = \beta_1 + \beta_2 x\,.$$

Suppose that $y_1, \cdots, y_n$ are observations of the times to failure, $Y_1, \cdots, Y_n$ for n such components with corresponding resistances $x_1, \cdots, x_n$.

   a. Write down the likelihood in terms of β_1 and β_2 and hence derive a pair of simultaneous equations, the solutions of which are the maximum likelihood estimates.

   b. Calculate the observed and expected information matrices. Are the Newton-Raphson and the Fisher scoring methods identical for this problem? Justify your answer.

2. Suppose $Y_i \sim$ Geometric(p_i), the geometric distribution as studied in Question 2 of Problem Sheet 4. We want to model how p_i depends on explanatory variables x_i.

   a. Assuming a GLM with canonical link function, write down a formula for p_i in terms of x_i. Is this a sensible model?

   b. Suppose instead that

$$\text{logit}(p_i) = \log \frac{p_i}{1-p_i} = \mathbf{x}_i^\top \beta\,.$$

   Show that this is a GLM with a non-canonical link function, and write down the link function corresponding to this model.

   c. Derive an expression for the scaled deviance for this model, writing $\hat{\mu}_i$ for the estimate of $\mu_i = \mathbb{E}(Y_i)$ under the model from part (b). Write an expression for $\hat{\mu}_i$ in terms of $\widehat{\beta}$, the MLE of β.

3. We return to the `beetle` data studied in Computer Lab 5, with observations on $n = 8$ groups of beetles. There we considered the model:

```
beetle_glm <- glm(prop_killed ~ dose, data = beetle, family = binomial,
                  weights = exposed)
```

We could have also considered a model with quadratic dependence on `dose`

```
beetle_glm_quad <- glm(prop_killed ~ dose + I(dose^2), data = beetle,
                       family = binomial, weights = exposed)
```

   a. Write down mathematical expressions for the two models. Show that `beetle_glm` is nested with `beetle_glm_quad`, and write down the null hypothesis H_0 and the alternative hypothesis H_1 you would use for comparing the models.

   b. Consider the following output of a `summary()` call. What is the scaled deviance for `beetle_glm`?

```
summary(beetle_glm)
```

```
Call:
glm(formula = prop_killed ~ dose, family = binomial, data = beetle,
    weights = exposed)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -60.717      5.181  -11.72   <2e-16 ***
dose          34.270      2.912   11.77   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 284.202  on 7  degrees of freedom
Residual deviance:  11.232  on 6  degrees of freedom
AIC: 41.43

Number of Fisher Scoring iterations: 4
```

   c. The scaled deviance for `beetle_glm_quad` is 3.1949. Calculate the log likelihood ratio test statistic L_{01} for testing H_0 against H_1. Under H_0, what is the distribution of this statistic? Hence conduct a hypothesis test of H_0 against H_1, and make a conclusion about which model you prefer.