MATH3091: Statistical Modelling II

Problem Sheet 5 (Solution)

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.

Solution:

a. The likelihood is

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} \lambda_i(\beta) e^{-\lambda_i(\beta) y_i} .$$

The log-likelihood is

$$\ell(\beta) = \sum_{i=1}^{n} \log \lambda_i(\beta) - \sum_{i=1}^{n} \lambda_i(\beta) y_i .$$

Differentiating with respect to β_j, $j = 1, 2$, gives

$$\frac{\partial}{\partial \beta_j} \ell(\beta) = \sum_{i=1}^{n} \frac{1}{\lambda_i(\beta)} \frac{\partial \lambda_i}{\partial \beta_j} - \sum_{i=1}^{n} \frac{\partial \lambda_i}{\partial \beta_j} y_i ,$$

where

$$\frac{\partial \lambda_i}{\partial \beta_1} = 1; \qquad \frac{\partial \lambda_i}{\partial \beta_2} = x_i \, .$$

So

$$\frac{\partial}{\partial \beta_1} \ell(\beta) = \sum_{i=1}^{n} \frac{1}{\lambda_i(\beta)} - \sum_{i=1}^{n} y_i \, ,$$

and

$$\frac{\partial}{\partial \beta_2} \ell(\beta) = \sum_{i=1}^{n} \frac{x_i}{\lambda_i(\beta)} - \sum_{i=1}^{n} x_i y_i \, .$$

So the MLE $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)^\top$ satisfies

$$\sum_{i=1}^{n} \frac{1}{\lambda_i(\hat{\beta})} = \sum_{i=1}^{n} y_i; \qquad \sum_{i=1}^{n} \frac{x_i}{\lambda_i(\hat{\beta})} = \sum_{i=1}^{n} x_i y_i \, ,$$

or

$$\sum_{i=1}^{n} \frac{1}{\hat{\beta}_1 + \hat{\beta}_2 x_i} = \sum_{i=1}^{n} y_i; \qquad \sum_{i=1}^{n} \frac{x_i}{\hat{\beta}_1 + \hat{\beta}_2 x_i} = \sum_{i=1}^{n} x_i y_i \, .$$

b. We have

$$\begin{aligned} H_{jk}(\beta) &= \frac{\partial^2}{\partial \beta_j \partial \beta_k} \ell(\beta) \\ &= \sum_{i=1}^{n} \frac{\partial}{\partial \beta_k} \left(\frac{1}{\lambda_i(\beta)} \frac{\partial \lambda_i}{\partial \beta_j} - \frac{\partial \lambda_i}{\partial \beta_j} y_i \right) \\ &= \sum_{i=1}^{n} \frac{-1}{\lambda_i(\beta)^2} \frac{\partial \lambda_i}{\partial \beta_j} \frac{\partial \lambda_i}{\partial \beta_k} + \frac{1}{\lambda_i(\beta)} \frac{\partial^2 \lambda_i}{\partial \beta_j \partial \beta_k} - \frac{\partial^2 \lambda_i}{\partial \beta_j \partial \beta_k} y_i \\ &= -\sum_{i=1}^{n} \frac{1}{\lambda_i(\beta)^2} \frac{\partial \lambda_i}{\partial \beta_j} \frac{\partial \lambda_i}{\partial \beta_k} \end{aligned}$$

as $\frac{\partial^2 \lambda_i}{\partial \beta_j \partial \beta_k} = 0$.

So the observed information is

$$-H(\beta) = \begin{pmatrix} \sum_{i=1}^{n} \frac{1}{(\beta_1 + \beta_2 x_i)^2} & \sum_{i=1}^{n} \frac{x_i}{(\beta_1 + \beta_2 x_i)^2} \\ \sum_{i=1}^{n} \frac{x_i}{(\beta_1 + \beta_2 x_i)^2} & \sum_{i=1}^{n} \frac{x_i^2}{(\beta_1 + \beta_2 x_i)^2} \end{pmatrix} ,$$

and the Fisher information matrix $\mathcal{I}(\beta) = -H(\beta)$, as the observed information matrix does not depend on y. The Newton-Raphson and Fisher scoring methods will be identical for this problem, because the observed information matrix and the Fisher information matrix are identical.

2. Suppose $Y_i \sim \text{Geometric}(p_i)$, 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 $\hat{\beta}$, the MLE of β.

Solution:

a. With the canonical link, we always have
$$\theta_i = \eta_i = x_i^\top \beta,$$
for the canonical parameter θ_i. In this case, from Question 2 of Problem Sheet 4, we have $\theta_i = \log(1-p_i)$, or $p_i = 1 - \exp\{\theta_i\}$, so
$$p_i = 1 - \exp\{x_i^\top \beta\} \,.$$
This is not sensible, as $x_i^\top \beta$ could take any real value, so $1 - \exp\{x_i^T \beta\}$ could take any value in $(-\infty, 1)$, but we want $p_i \in (0, 1)$.

b. We have $p_i = \mu_i^{-1}$, and $\mu_i = g^{-1}(\eta_i)$, so
$$p_i = \frac{1}{g^{-1}(\eta_i)} \,.$$
So we need to choose $g(\cdot)$ such that
$$\frac{1}{g^{-1}(\eta)} = \text{logit}^{-1}(\eta) = \frac{\exp(\eta)}{1 + \exp(\eta)} \,.$$
This means that
$$g^{-1}(\eta) = 1 + \exp(-\eta),$$
so inverting this gives the required link function
$$g(\mu) = -\log(\mu - 1) \,.$$

c. From Problem Sheet 4, Question 2, we have

$$b(\theta) = -\log(e^{-\theta} - 1)\,, \qquad \mu_i = b'(\theta_i) = \frac{1}{1 - e_i^{\theta}}\,, \qquad \theta_i = \log\frac{\mu_i - 1}{\mu_i}\,.$$

So

$$b(\theta_i) = -\log(e^{-\theta_i} - 1) = -\log\left(\frac{\mu_i}{\mu_i - 1} - 1\right) = -\log\left(\frac{1}{\mu_i - 1}\right) = \log(\mu_i - 1)\,.$$

The scaled deviance is

$$\begin{aligned} L_{0s} &= 2\sum_{i=1}^{n} y_i[\hat{\theta}_i^{(s)} - \hat{\theta}_i^{(0)}] - [b(\hat{\theta}_i^{(s)}) - b(\hat{\theta}_i^{(0)})] \\ &= 2\sum_{i=1}^{n} y_i\left[\log\frac{\hat{\mu}_i^{(s)} - 1}{\hat{\mu}_i^{(s)}} - \log\frac{\hat{\mu}_i^{(0)} - 1}{\hat{\mu}_i^{(0)}}\right] - \left[\log(\hat{\mu}_i^{(s)} - 1) - \log(\hat{\mu}_i^{(0)} - 1)\right]. \end{aligned}$$

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.

Solution:

a. Let Y_i be the number of beetles killed and x_i for the dose in group i. In both cases, we have $Y_i \sim \text{binomial}(n_i, p_i)$, where $\text{logit}(p_i) = \eta_i$ and n_i is the number of beetles exposed in the i-th group. In `beetle_glm`, we have

$$\eta_i = \beta_1 + \beta_2 x_i \,.$$

In `beetle_glm_quad`, we have

$$\eta_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 \,.$$

We can compare these models by testing $H_0 : \beta_3 = 0$ against H_1 : "β_3 is unrestricted".

b. The scaled deviance for `beetle_glm` is 11.232.

c. We have $L_{01} = L_{0s} - L_{1s}$, where L_{0s} is the scaled deviance under H_0 (`mod_glm`), so $L_{0s} = 11.232$, and L_{1s} is the scaled deviance under H_1 (`mod_glm_quad`), so $L_{1s} = 3.1949$. So $L_{01} = 11.232 - 3.1949 = 8.04$.

Under H_0, $L_{01} \sim \chi^2_1$, as $p - q = 3 - 2 = 1$. So we should reject H_0 if L_{01} is greater than the 95% point of the χ^2_1 distribution, or

```r
qchisq(0.95, df = 1)
```

```
[1] 3.841459
```

Since $8.04 > 3.84$, we reject H_0, and prefer `beetle_glm_quad` to `beetle_glm`.

We could do this test in `R` with

```r
anova(beetle_glm, beetle_glm_quad, test = "LRT")
```

```
Analysis of Deviance Table

Model 1: prop_killed ~ dose
Model 2: prop_killed ~ dose + I(dose^2)
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1         6    11.2322
2         5     3.1949  1   8.0373 0.004582 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```