

MATH 523 - MIDTERM EXAMINATION 2011

Calculators may be used.

A Geometric model for random variable Y taking values on $\{0, 1, 2, \dots\}$ and parameter π has mass function

$$f_Y(y; \pi) = \pi^y(1 - \pi) \quad y = 0, 1, 2, \dots$$

for parameter $0 < \pi < 1$.

- (a) Write this model in Exponential or Exponential-Dispersion family form, and identify the canonical parameter, to be denoted θ , written as a function of π .

3 MARKS

- (b) Find the mean, to be denoted μ , and variance for this distribution, and identify the function $V(\mu)$ such that

$$\text{Var}[Y] = \phi V(\mu)$$

for dispersion parameter $\phi > 0$.

6 MARKS

- (c) Identify the canonical link function, that is, the function $g(\cdot)$ such that if η is the linear predictor, then

$$\eta = g(\mu) = \theta.$$

5 MARKS

- (d) Explain why the link function,

$$g(\mu) = \log \left(-\log \left(\frac{\mu}{1 + \mu} \right) \right)$$

may be more appropriate than the canonical link function in this problem.

4 MARKS

- (e) Find the maximum likelihood estimator of μ , $\hat{\mu}_0$, assuming that data $y_1, \dots, y_n$ are independent and identically distributed from this model.

3 MARKS

- (f) Now suppose that for each data point $i = 1, \dots, n$, values of a single predictor, X , are also measured. Denote the observed (scalar) predictor values $x_1, \dots, x_n$.

Suppose that X is a continuous predictor. If the linear predictor takes the form $\eta_i = \beta_0 + \beta_1 x_i$, **and a log link is used**, derive the maximum likelihood estimating equations for β_0 and β_1 .

5 MARKS

- (g) Derive the deviance for this model. Leave your answer in terms of the μ values under the fitted model, $\hat{\mu}_i, i = 1, \dots, n$.

4 MARKS

- (h) The R code and output below contains the fit of the GLM described above, **using a log link**, that is

$$\eta = \log(\mu)$$

to a data set recording the number of complete months that a moulded plastic component functions before failure ($y = 0$ corresponds failure **during** the first month), and the single predictor is the density (in g/cm^3) of the individual component measured after failure.

Comment on the analysis carried out and the model fitted, and interpret the results presented. Does the density of the plastic influence the probability of failure ? If so, how ? Justify your answers.

```
>summary(glm(Y ~ X, family=negative.binomial(theta=1,link="log")))
```

Call:
glm(formula = Y ~ X, family = negative.binomial(theta = 1, link = "log"))

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1269	-1.2531	-0.4940	0.1741	2.9832

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	8.07259	1.98763	4.061	0.000179	***
X	-0.15490	0.04434	-3.493	0.001035	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(1) family taken to be 1.541917)

Null deviance: 84.730 on 49 degrees of freedom
Residual deviance: 61.232 on 48 degrees of freedom
AIC: 224.73

Number of Fisher Scoring iterations: 5

10 MARKS