THE UNIVERSITY OF AUCKLAND

SECOND SEMESTER, 2012
Campus: City

STATISTICS

Statistical Inference

(Time allowed: TWO hours)

NOTE: Attempt all questions. The total mark is 100.

Be sure to show your working and define notation - full marks will NOT be awarded for answers that are not proved or argued, even if those answers are correct.

1. [9 marks] Two independent Poisson random samples were observed, both of size ten. The random sample in the first group were iid Poisson(λ_1) and the random sample in the second group were iid Poisson(λ_2). The maximum likelihood estimate was $(\widehat{\lambda}_1, \widehat{\lambda}_2) = (\overline{y}_1, \overline{y}_2) = (4, 6)$ and hence the approximate variances of $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ were 0.4 and 0.6, respectively.

   (a) Calculate the Wald test statistic for the hypothesis $H_0 : \lambda_1 = \lambda_2$.

   (b) Calculate the likelihood ratio test statistic for H_0.

   Recall that the density function of a Poisson(λ) random variable is

$$f(y; \lambda) = \frac{e^{-\lambda}\lambda^y}{y!} \quad , y = 0, 1, 2, \ldots$$

2. [9 marks]

   (a) Subject to regularity conditions, it has been shown that the likelihood is higher at the true parameter value, θ_0, than at any other fixed parameter value with probability that becomes arbitrarily close to 1 as the sample size increases. Using this fact, show that the maximum likelihood estimator $\widehat{\theta}$ is consistent. (You may assume $\theta \in \mathbb{R}$ and that the likelihood is unimodal.)

   (b) State and prove Jensen's inequality. You may use the following definition of a convex function defined on $D \subset \mathbb{R}$.

   Definition: $\phi : D \to \mathbb{R}$ is convex if for any $y_0 \in D$ there exists a $c \in \mathbb{R}$ such that

$$\phi(y_0) + c(y - y_0) \leq \phi(y) \quad \text{for all } y \in D.$$

3. [6 marks]

   (a) Suppose that the parameter vector $\boldsymbol{\theta} \in \mathbb{R}^s$ can be partitioned as $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\lambda})$. Define the profile likelihood function for $\boldsymbol{\psi}$.

   (b) Suppose that the MLE $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^s$ has approximate variance matrix $\widehat{\mathbf{V}}$ and that $g : \mathbb{R}^s \to \mathbb{R}^p$ is differentiable. Write down the formula for the approximate variance (matrix) of $\widehat{\boldsymbol{\zeta}} = g(\widehat{\boldsymbol{\theta}})$.

4. [8 marks]

   (a) As a general rule of thumb, it is often said that a fluorescent light bulb lasts ten times longer than an incandescent light bulb. Express this statement in terms of an accelerated failure time model.

   (b) Two groups of patients were used in a study of the effectiveness of a drug. The first group received the drug and the second group received a placebo. There were 4 patients in the drug group and 3 in the placebo group and the survival times were

Drug group:	4 (8) 11 (12)
Placebo group:	(3) 7 10

   where values in parentheses denote censored observations.

   Write down Cox's partial likelihood function for these data, assuming a proportional hazards model. Be sure to define your notation. (Do not attempt to simplify or maximize this partial likelihood).

5. [6 marks] Recall that for a scalar parameter $\theta \in \mathbb{R}$, the $(1-\alpha)100\%$ bootstrap percentile confidence interval is given by $(\widehat{\theta}^*_{\alpha/2}, \widehat{\theta}^*_{1-\alpha/2})$, where $\widehat{\theta}^*_\gamma$ denotes the empirical γ quantile of the bootstrapped values.

   Provide an argument to establish that this interval has (approximately) the desired coverage probability of $(1-\alpha)$. Be sure to state all assumptions used.

6. [10 marks] A zero-inflated Poisson model, denoted ZIP(p, λ) was fitted to iid count data $y_i, i = 1, ...100$. Consider the null hypothesis that the data are iid Poisson.

   (a) Write R code to conduct a bootstrap simulation of the distribution of the likelihood ratio test statistic. Specifically, your code should produce a vector `LRT` containing 1000 simulated likelihood ratio test statistic values. (For convenience, you may use the `glm` and `vglm` functions where appropriate. This will enable the required simulation to be implemented in about six lines of code.)

   (b) If the above code is correct then running it will show that the empirical 0.95 quantile of the bootstrapped LRT values is close to $\chi^2_{1,0.9}$ (the 0.90 quantile of a chi-square with 1 degree of freedom). Provide an intuitive explanation of why this is so.

7. [6 marks] Show that the gamma density function

$$f(y; \alpha, \beta) = \frac{y^{\alpha-1} e^{-y/\beta}}{\beta^\alpha \Gamma(\alpha)} \qquad , y > 0$$

   can be written in exponential dispersion family form

$$\log f(y; \psi) = \frac{y\psi - b(\psi)}{\phi / w} + c(y, \phi, w) \ . \tag{1}$$

   Be sure to identify ψ, ϕ and w.

8. [11 marks] In the context of a generalized linear model:

   (a) Define

      i. Link function.

      ii. Saturated model.

      iii. Deviance residual.

   (b) As an extreme example of sparse binomial data, suppose that $y_i, i = 1, ..., n$ are iid $\text{Bin}(1, p)$, that is, iid Bernoulli with $\text{Prob}(Y_i = 1) = p$. The MLE is of course $\widehat{p} = \sum_{i=1}^{n} y_i/n$.

      i. Calculate the Pearson chi-square statistic for the fitted model.

      ii. Calculate the model deviance.

      iii. Use the results of the above calculations to argue in favour of preferring the Pearson chi-square (to the model deviance) as an omnibus test of model fit when the data are sparse.

9. [15 marks] Two independent random samples of count data were observed, both of size 20. Vector y contains $(y_1, ..., y_{40})$ where $y_i, i = 1, ..., 20$ were observed from group 1, and $y_i, i = 21, ..., 40$ were observed from group 2.

To begin, an iid Poisson model was fitted to all 40 observations using the `glm` function in R. The results of this fit are shown below, and the value of the Pearson chi-square statistic is also provided.

```
                 MODEL 1
Call:
glm(formula = y ~ 1, family = poisson)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.71113    0.02472   150.1   <2e-16 ***

    Null deviance: 174.8  on 39  degrees of freedom
Residual deviance: 174.8  on 39  degrees of freedom
AIC: 396.66

Pearson chi-square is 174.0733
```

A second model was fitted, with the observations in group i modeled as Poisson$(\lambda_i), i = 1, 2$. Here `grp` is a factor variable indicating the group.

```
                 MODEL 2
Call:
glm(formula = y ~ grp, family = poisson, data = Df)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.77506    0.03386 111.476  < 2e-16 ***
grp2        -0.13222    0.04955  -2.668  0.00763 **

    Null deviance: 174.80  on 39  degrees of freedom
Residual deviance: 167.67  on 38  degrees of freedom
AIC: 391.52

Pearson chi-square is 163.2949
```

(a) Define Akaike's information criterion, AIC.

(b) What is the log-likelihood of Model 1?

(c) What is the log-likelihood of the saturated model?

(d) What is the sample mean of the observations in group 2?

If quasi-likelihood is used here,

(e) What is the estimated over-dispersion, $\widehat{\phi}$?

(f) What is the quasi-likelihood Wald test statistic for $H_0 : \lambda_1 = \lambda_2$?

(g) What is the quasi-likelihood likelihood ratio test statistic for $H_0 : \lambda_1 = \lambda_2$?

(h) What is the quasi-AIC of model 1?

(i) Why is quasi-likelihood not a true likelihood?

10. [20 marks]

(a) The Laplace approximation for a one dimensional integral is

$$\int_{\mathbb{R}} h(u)du \approx h(\widehat{u})\sqrt{\frac{2\pi}{c}} ,$$

where c is the negative of the curvature of $\log h(u)$ evaluated at the maximizing value $\widehat{u}$.

Provide a derivation of this approximation.

(b) Describe the method of importance sampling for numerical evaluation of the above integral.

(c) If random variable U is distributed as standard Cauchy, and $Y|u$ is distributed $N(u, \sigma^2)$ then the joint density function of (y, u) is given by

$$f(y, u; \sigma) = \frac{k}{\sigma} \exp\left(-\frac{(y-u)^2}{2\sigma^2}\right) \times \frac{1}{1+u^2} ,$$

where k denotes constant terms.

Assume that $y = 0$ is observed and denote $h(u) = f(0, u; \sigma)$.

Verify that $h(u)$ is maximized at $\widehat{u} = 0$, and hence calculate the Laplace approximation to $\int_{\mathbb{R}} h(u)du$.

(d) Consider the following Poisson-normal random effects model,

$$U_i \sim N(0, \sigma^2) \qquad , i = 1, ..., m$$

$$Y_{ij}|u_i \sim \text{Pois}\,(e^{u_i}) \qquad , i = 1, ..., m,\; j = 1, ..., n \,.$$

i. Write down the formula for the joint density function $f(\boldsymbol{y}, \boldsymbol{u}; \sigma)$.

ii. Explain how separability can be used to reduce the computational burden of numerically integrating $h(\boldsymbol{u}) = f(\boldsymbol{y}, \boldsymbol{u}; \sigma)$.