THE UNIVERSITY OF AUCKLAND

SECOND SEMESTER, 2014
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.
-partial marks WILL be awarded for incorrect answers if partial progress toward the correct answer is demonstrated.

1. [20 marks] Suppose that $y_1 = 4$ is observed from a Poisson(λ_1) distribution. Recall that a Poisson(λ) distribution has density function

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

   (a) Show that the MLE of λ_1 is $\widehat{\lambda}_1 = y_1 = 4$.

   (b) Calculate the curvature of the log-likelihood, and hence provide an estimate of the variance of $\widehat{\lambda}_1$.

   Suppose now that $y_2 = 12$ is observed from a Poisson(λ_2) distribution, and is independent of y_1.

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

   (d) Calculate the Wald test statistic for $H_0 : (\lambda_1, \lambda_2) = (10, 10)$.

   (e) Recalculate the Wald test statistic for the hypothesis in part (d), but using the alternative parameterization (ζ_1, ζ_2) where $\zeta_i = \log(\lambda_i), i = 1, 2$.

   (f) Calculate the likelihood ratio test statistic for $H_0 : \lambda_1 = \lambda_2$.

   (g) Calculate the likelihood ratio test statistic for $H_0 : (\lambda_1, \lambda_2) = (10, 10)$.

   For the final part of this question, consider prediction of the random variable $z^* = y_2^* - y_1^*$ where $y_1^* \sim$ Poisson(λ_1) and $y_2^* \sim$ Poisson(λ_2) are future realizations from their respective distributions.

   (h) Using your choice of either bootstrap prediction or pseudo-Bayesian prediction, write a short piece of R code to implement the generation of 1000 future realizations of z^*.

2. [8 marks] Consider a one-parameter statistical model $f(\boldsymbol{y};\theta)$ with MLE $\widehat{\theta}$.

   (a) Show that the Wald and likelihood ratio test statistics of $H_0 : \theta = \theta_0$ are equal if the log-likelihood function is quadratic. Assume that the Wald statistic uses the usual variance estimate, $\widehat{\text{var}}(\widehat{\theta}) = -1/l''(\widehat{\theta})$.

   (b) Use the result from part (a) to provide a justification for preferring likelihood ratio tests to Wald tests.

3. [10 marks]

   (a) 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 every $y_0 \in D$ there exists a $c \in \mathbb{R}$ such that

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

   (b) Subject to regularity conditions, it is known 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.)

4. [5 marks] Derive the Newton-Raphson algorithm for numerical optimization.

5. [10 marks] Consider a mark-recapture experiment where 100 animals are caught and marked on the first capture occasion. On the second capture occasion, 75 animals are caught, of which 25 were marked.

   (a) Arrange the data in the form of a partially completed 2-by-2 contingency table. making sure to clearly label the row and columns.

   (b) Write down the multinomial model log-likelihood for the above contingency table, and explain how profile likelihood can be employed to simplify the maximization of this log-likelihood. Recall that a s-cell multinomial with n trials and cell probabilities $p_1, ..., p_s$ has density function

$$f(y_1, ..., y_s; p_1, ..., p_s) = \frac{n!}{y_1!...y_s!} p_1^{y_1} ... p_s^{y_s} \ .$$

   Be sure to define your notation.

6. [8 marks] Show that the inverse-Gaussian density function

$$f(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma y^{3/2}} \exp\left(-\frac{(y-\mu)^2)}{2\sigma^2\mu^2 y} \right) \ ,$$

can be written in exponential dispersion family form

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

Be sure to clearly specify ψ, ϕ and $b(\psi)$.

7. [6 marks] In your answers to the items below, be sure to define all relevant notation and terminology.

   (a) Define AIC.

   (b) Define deviance residual.

   (c) The model deviance can not be used for assessing the fit of generalized linear mixed models. Why?

8. [12 marks] Count data were obtained from an experiment with two explanatory factors, A and B, each with 2 levels. Ten observations were measured for each of the four combinations of levels of the two factors, resulting in a sample size of $n = 40$.

   The attachment contains R code and output from fitting a selection of log-linear models to the data. Be sure to provide relevant working when answering the following questions:

   (a) What is the deviance of the null model?

   (b) Which model is preferred by AIC?

   (c) Are these data over-dispersed?

   (d) Which model is preferred by quasi-AIC?

   (e) For the model preferred by quasi-AIC, provide an over-dispersion corrected 95% Wald confidence interval for the intercept parameter.

   (f) Write the R code that would be required to obtain an over-dispersion corrected 95% likelihood ratio confidence interval for the intercept parameter.

9. [21 marks]

(a) Explain the concept of *separability* and how it can be used to ease the computational burden of fitting latent variable models.

(b) Application of Gauss-Hermite quadrature results in the exact formula

$$\int_{-\infty}^{\infty}(a+bu+cu^2)e^{-(u-\mu)^2}du = \sqrt{\pi}(a+b\mu+c(\mu^2+0.5)) \ . \qquad (2)$$

Deduce this formula by utilizing the known properties about the first and second order moments of a normal distribution with mean μ and variance $\frac{1}{2}$. This distribution has density function $f(u) = \frac{1}{\sqrt{\pi}}e^{-(u-\mu)^2}$.

(c) Write R code to numerically approximate the integral on the left-hand side of (2) using importance sampling, using random samples taken from a standard Cauchy distribution. The standard Cauchy has density function $f(u) = \frac{1}{\pi(1+u^2)}$. Function `rcauchy` randomly generates values from this distribution, and the density function is evaluated by the `dcauchy` function.

(d) Determine the form of the Laplace approximation to the integral

$$\int_0^{\infty} u^{\alpha-1}e^{-u/\beta}du \ .$$

[You can ignore the fact that this integral restricted to the non-negative reals - it will be assumed that α and β are such that the Laplace approximation will still be reasonable.]

(e) From part 9(d) above, evaluate your Laplace approximation for $\alpha = 10$ and $\beta = 0.25$. If you have done things correctly, you will have calculated a value that is close to the true value $\Gamma(10) \times 0.25^{10} = 0.34607$.

R code for Question 8

```
> #Summary function to provide parameter table, log-likelihood, Pearson chi-square,
> #and likelihood ratio confidence intervals.
> gSummary=function(gfit,level=0.95) {
+   list(pars=coef(summary(gfit)),logLik=logLik(gfit),
+         Pchisq=sum(resid(gfit,type="pearson")^2),LRCI=confint(gfit,level=level)) }
>
> #Null model
> g0=glm(y~1,family=poisson,data=df)
> #Effect of A only
> gA=glm(y~A,family=poisson,data=df)
> #Effect of B only
> gB=glm(y~B,family=poisson,data=df)
> #Additive effects of A and B
> gAB=glm(y~A+B,family=poisson,data=df)
> #Full model with interaction
> gFull=glm(y~A*B,family=poisson,data=df)
> #Saturated model
> gSatd=glm(y~as.factor(1:40),family=poisson,data=df)


> gSummary(g0)
$pars
            Estimate Std. Error  z value Pr(>|z|)
(Intercept) 2.844909 0.03812464 74.62127        0

$logLik
'log Lik.' -225.624 (df=1)

$Pchisq
[1] 279.6744

$LRCI
   2.5 %   97.5 %
2.769243 2.918712


> gSummary(gA)
$pars
             Estimate Std. Error   z value     Pr(>|z|)
(Intercept)  3.124565 0.04688072 66.649252 0.000000e+00
AYes        -0.669259 0.08055835 -8.307754 9.753029e-17

$logLik
'log Lik.' -189.1582 (df=2)

$Pchisq
[1] 202.2072

$LRCI
                 2.5 %    97.5 %
(Intercept)  3.0312495  3.215062
AYes        -0.8285793 -0.512626
```

```
> gSummary(gB)
$pars
              Estimate Std. Error   z value     Pr(>|z|)
(Intercept)  3.1201599 0.04698410 66.408848 0.000000e+00
BYes        -0.6563066 0.08039169 -8.163861 3.244824e-16

$logLik
'log Lik.' -190.4838 (df=2)

$Pchisq
[1] 221.1853

$LRCI
                2.5 %     97.5 %
(Intercept)  3.026635  3.2108529
BYes        -0.815270 -0.4999712


> gSummary(gAB)
$pars
              Estimate Std. Error   z value     Pr(>|z|)
(Intercept)  3.3998156 0.05433051 62.576547 0.000000e+00
AYes        -0.6692590 0.08055755 -8.307837 9.746229e-17
BYes        -0.6563066 0.08039037 -8.163996 3.241200e-16

$logLik
'log Lik.' -154.0179 (df=3)

$Pchisq
[1] 135.1952

$LRCI
                 2.5 %     97.5 %
(Intercept)  3.2916598  3.5046733
AYes        -0.8285793 -0.5126260
BYes        -0.8152700 -0.4999712
```

```
> gSummary(gFull)
$pars
              Estimate Std. Error  z value     Pr(>|z|)
(Intercept)  3.3499041 0.05923485 56.552922 0.000000e+00
AYes        -0.5285252 0.09726830 -5.433684 5.520229e-08
BYes        -0.5166907 0.09690625 -5.331862 9.721092e-08
AYes:BYes   -0.4328860 0.17529276 -2.469503 1.353009e-02

$logLik
'log Lik.' -150.8939 (df=4)

$Pchisq
[1] 126.3584

$LRCI
                 2.5 %      97.5 %
(Intercept)  3.2315120  3.46379538
AYes        -0.7208845 -0.33930271
BYes        -0.7082965 -0.32813191
AYes:BYes   -0.7806047 -0.09266573


> logLik(gSatd)
'log Lik.' -89.39534 (df=40)
```