THE UNIVERSITY OF AUCKLAND

SECOND SEMESTER, 2013
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. [12 marks] One hundred iid observations $\boldsymbol{y} = (y_1, ..., y_{100})$ were observed from a $N(\mu, \sigma^2)$ distribution with density function $f(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2}$. The sum of the observations was $\sum_{i=1}^{100} y_i = 200$, and the residual sum-of-squares was $\sum_{i=1}^{100} (y_i - \overline{y})^2 = 10000$. Thus, the MLEs were calculated to be $\widehat{\mu} = \overline{y} = 2$ and $\widehat{\sigma} = \sqrt{\frac{\sum_{i=1}^{100}(y_i - \overline{y})^2}{100}} = 10$.

   (a) Calculate the value of the log-likelihood evaluated at the MLE, $l(\widehat{\mu}, \widehat{\sigma})$.

   (b) Determine the formula for the profile log-likelihood function

   $$l^*(\mu) = \max_{\sigma} l(\mu, \sigma) \ .$$

   (c) Using the profile log-likelihood function from above (or otherwise if you prefer), calculate the likelihood ratio test statistic for the null hypothesis $H_0 : \mu = 0$. You may use the fact that $\sum_{i=1}^{100} y_i^2 = 10400$.

2. [5 marks] Kullback-Leibler (KL) divergence is used to quantify the "distance" between two density functions. Specifically, for density functions f_1 and f_2, the KL-divergence between them is

   $$K = \int f_1(y) \log\left(\frac{f_1(y)}{f_2(y)}\right) dy \ .$$

   Use Jensen's inequality to show that $K \geq 0$.

   Hint: Let ϕ be the $-\log$ function.

3. [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.

CONTINUED

4. [6 marks] Suppose that $Y_i, i = 1...n$ are iid Poisson with mean μ. That is, with density function $f(y;\mu) = e^{-\mu}\mu^y/y!$ Which of the following estimators are consistent estimators of μ?

   (a) The sample mean.

   (b) The sample median.

   (c) The sample variance.

   (d) The estimator $\tilde{\mu} = -\log(p)$, where p is the proportion of Y_i values in the sample that are equal to zero.

5. [7 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. [14 marks] The unknown number, N, of animals in a closed population can be estimated by the method of removal. In this type of experiment, n_1 animals are captured and removed on the first removal occasion, leaving $N - n_1$ animals. A further n_2 are captured and removed on a second removal occasion. It will be assumed that animals are captured with probability p, and that

$$\begin{aligned} n_1 &\sim \text{Binomial}(N, p) \\ n_2 &\sim \text{Binomial}(N - n_1, p) \end{aligned}$$

where p is common to both removal occasions.

(a) Assume (for now) that p is known, leaving only N to be estimated. The likelihood for N is given by the product of the binomial likelihoods from the two removal occasions.

i. Show that
$$\widehat{N} = \left[\frac{n_1 + n_2}{2p - p^2}\right]$$
is an integer-valued MLE of N, where $[x]$ denotes the integer part of x.

ii. Is the above MLE unique?

(b) Write a short bit of R code to numerically optimize the log-likelihood for the removal experiment described above, in the case where both p and N are unknown parameters to be estimated. Also include the necessary code to provide the approximate 95% likelihood ratio confidence interval for N.

7. [6 marks] Let Y be the proportion of 1's from a Binomial(n, p) experiment. That is, for $y \in \{0, \frac{1}{n}, \frac{2}{n}, ..., \frac{n-1}{n}, 1\}$,

$$f(y; p) = \binom{n}{ny} p^{ny}(1-p)^{n(1-y)} .$$

Show that this density function can be expressed in exponential family form

$$f(y; \theta, \phi) = \exp\{(y\theta - b(\theta))/a(\phi) + c(y, \phi)\}$$

where $\theta, \phi \in \mathbb{R}$.

8. [11 marks] In your answers to the items below, be sure to define all relevant notation and terminology.

In the context of generalized linear models (GLMs):

(a) Define link function.

(b) Define deviance.

(c) What is an *offset* term?

(d) When using quasi-likelihood in the GLM context,

i. How is the estimated over-dispersion $\widehat{\phi}$ calculated?

ii. Define quasi-AIC.

iii. How do the estimated regression coefficients change (compared to a standard GLM)?

iv. How do the estimated standard errors of the regression coefficients change (compared to a standard GLM)?

v. How do likelihood-ratio tests change (compared to a standard GLM)?

9. [9 marks] True or false?

   (a) For the MLE $\widehat{\theta}$ to have a distribution that is well approximated by a normal distribution, it is necessary that $\widehat{\theta}$ has finite mean and variance.

   (b) In the GLM context, the Pearson chisquare statistic is generally preferred to the deviance as an omnibus test statistic for goodness of fit because it is more robust to sparseness of the data.

   (c) In the GLM context, if model A (having p_A parameters) is nested within model B (having $p_B > p_A$) parameters, then (under usual regularity conditions) the difference in the Pearson chi-square statistics of the two models, $P_{\chi^2,A} - P_{\chi^2,B}$, is approximately chi-square distributed with $p_B - p_A$ degrees of freedom when model A is true.

   (d) The delta method (for calculating the variance of functions of random variables) is exact for linear functions.

   (e) Akaike's information criterion (AIC) can be used to select amongst models that are not nested.

   (f) Suppose that the 5-parameter bi-normal model is fitted to data that are actually iid $N(\mu, \sigma^2)$. The likelihood ratio test statistic for the null hypothesis that the data are iid normal will be approximately distributed χ^2_3 for sufficiently large sample size.

10. [22 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) State the formula for the Laplace approximation to the q-dimensional integral $\int_{\mathbb{R}^q} h(\boldsymbol{u})d\boldsymbol{u}$ where $\boldsymbol{u}$ is of dimension q.

(c) Suppose that U is distributed as lognormal $LN(\mu, \sigma^2)$, with density function

$$f(u; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma u} \exp\left(-\frac{(\log u - \mu)^2}{2\sigma^2}\right) \ , \quad u > 0 \ ,$$

and $Y|u$ is distributed Poisson(u). Then the joint density function of (y, u) is given by

$$f(y, u; \mu, \sigma^2) = \frac{u^{y-1}}{\sqrt{2\pi}\sigma y!} \exp\left(-\frac{(\log u - \mu)^2}{2\sigma^2} - u\right) \ .$$

Assume that $\mu = 2, \sigma^2 = 0.1$ and $y = 11$, and denote $h(u) = f(11, u; 2, 0.1)$.

i. Verify that $h(u)$ is maximized at $\widehat{u} = 8.546$.

ii. Calculate the Laplace approximation to $\int_{\mathbb{R}} h(u)du$. (You may use the fact that $h(\widehat{u}) = 0.01148$.)

iii. Provide R code to evaluate $\int_{\mathbb{R}} h(u)du$ using importance sampling. You may wish to use R functions `rnorm` and/or `dnorm`, `rlnorm`, `dlnorm`. (`rlnorm`(n, μ, σ) is used to generate $LN(\mu, \sigma^2)$ random variables, and `dlnorm`(y, μ, σ) calculates the $LN(\mu, \sigma^2)$ density function.)

(d) Consider the following random effects model,

$$\begin{aligned} v_i &\sim N(a, \sigma_v^2) && , i = 1, ..., m \\ u_{ij}|v_i &\sim LN(v_i, \sigma_u^2) && , i = 1, ..., m, \ j = 1, ..., n \\ Y_{ij}|v_i, u_{ij} &\sim \text{Poisson}(u_{ij}) && , i = 1, ..., m, \ j = 1, ..., n \ . \end{aligned}$$

Write down the formula for the joint density function $f(\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{u}; a, \sigma_v^2, \sigma_u^2)$.
