Mid-Term Exam – ST 793 – Spring 2017

Put away and turn off all cell phones and electronic devices. You do not need a calculator. Put your answers on the sheets of paper handed out.

1. Let $X_1, \ldots, X_n$ be iid random variables, with mean μ and variance σ^2 and finite 3rd central moment $|\mu_3| < \infty$. The sample 3rd moment is $m_3 = n^{-1} \sum_{i=1}^{n} (X_i - \overline{X})^3$. Prove that $m_3 \xrightarrow{p} \mu_3$ as $n \to \infty$.

2. The multinomial density is

$$f(n_1, \ldots, n_k \,|\, \boldsymbol{p}) = \frac{n!}{n_1! n_2! \cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k},$$

where $\sum_{i=1}^{k} n_i = n$, $\sum_{i=1}^{k} p_i = 1$, and $0 < p_i < 1$, $i = 1, \ldots, k$. We are interested in Hardy-Weinberg equilibrium, which can be expressed as a null hypothesis in terms of p_A, the proportion of A alleles in the population,

$$H_0 : p_1 = p_A^2, p_2 = 2p_A(1 - p_A), p_3 = (1 - p_A)^2.$$

Assume we have multinomial counts (N_1, N_2, N_3) available, where N_1 is the number of AA alleles, N_2 is the number of Aa alleles, and N_3 is the number of aa alleles, $N_1 + N_2 + N_3 = n$. (The text uses $p_{AA} = p_1$ and $p_2 = p_{Aa}$, but I think the p_i are simpler notationally.) You may assume that the unrestricted maximum likelihood estimators are $\widehat{p}_i = N_i/n$, $i = 1, 2, 3$, and under H_0, the estimate of p_A is $(2N_1 + N_2)/(2n)$. Deleting p_3, $\boldsymbol{I}_{\mathrm{T}}(p_1, p_2)^{-1} = \{\mathrm{diag}(p_1, p_2) - (p_1, p_2)^T (p_1, p_2)\}/n$.

a. For testing, when $H_0 \cup H_a$ is the full unrestricted multinomial, Section 3.2.7 shows that the score statistic is the usual goodness-of-fit statistic

$$T_{\mathrm{S}} = \sum_{i=1}^{k} \frac{(N_i - n\widetilde{p}_i)^2}{n\widetilde{p}_i}.$$

Using this fact, give T_{S} for testing for Hardy-Weinberg equilibrium. What critical value should we use?

b. Give T_{LR} for testing H_0.

c. Give T_{W} for H_0. One way is to express H_0 as $H_0 : h(p_1, p_2) = 0$ using the definitions of p_1 and p_2 in H_0. Then set up the quadratic form for T_{W} in detail, but you do not need to multiply it out.

3. Suppose that we have n pairs $(Y_1, x_1), \ldots, (Y_n, x_n)$, where for simplicity we assume the x_i are fixed constants and positive. Consider the model

$$Y_i = \beta_0 + \beta_1 x_i^{\lambda} + e_i, \quad i = 1, \ldots, n,$$

where $\lambda > 0$ and $e_1, \ldots, e_n$ are iid N(0,σ^2). Our goal is to estimate $\boldsymbol{\theta}^T = (\beta_0, \beta_1, \lambda, \sigma^2)$.

a. Derive the log likelihood of $\boldsymbol{\theta}$.

b. Find the score vector of $\boldsymbol{\theta}$. (It helps to remember that $x^a = \exp\{a \log(x)\}$.)

c. Find the upper 3×3 portion of $\boldsymbol{I}_{\mathrm{T}}(\boldsymbol{\theta})$.

4. For $Y_i, \ldots, Y_n$ iid from a standard gamma(α, β) density with mean $\alpha\beta$. the log likelihood is

$$\ell(\alpha, \beta) = -n \log \Gamma(\alpha) - n\alpha \log \beta + (\alpha - 1) \sum \log Y_i - \frac{\sum Y_i}{\beta}.$$

Note that the first derivative of $\log \Gamma(\alpha)$ is $\psi(\alpha)$, called the digamma function, and the second derivative $\psi_1(\alpha)$ is the trigamma function.

a. Find $\boldsymbol{I}_{\mathrm{T}}(\alpha, \beta)$.

b. Give an approximate 95% confidence interval for α assuming you have $\widehat{\alpha}$ and $\widehat{\beta}$.

c. Find the profile log likelihood of α.

d. Find $\boldsymbol{I}_{\mathrm{T}}(\alpha)$ based on the profile log likelihood.