ST 793 Final Exam – Spring 2017

Turn off and put away all cell phones and electronic devices. You may not use a calculator. Put your answers on the sheets of paper handed out.

1. One of the "big" ideas of the class has been to (i) assume a model in order to answer the main scientific question using the primary part, but then (ii) worry about secondary parts of the model when making inference via the asymptotic distributions. In the simplest setting, we have $Y_1, \ldots, Y_n$ iid from some unknown density g and distribution function G. We assume a parametric family density $f(y; \boldsymbol{\theta})$ in order to define a maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ and possibly test statistics about $\boldsymbol{\theta}$.

a. If the true density g is not of the form $f(y; \boldsymbol{\theta})$, then the first thing to worry about is whether $\widehat{\boldsymbol{\theta}}$ converges in probability to a meaningful parameter value, call it $\boldsymbol{\theta}_0$. Give an equation to show how this limit value is defined when $g(y) \neq f(y; \boldsymbol{\theta})$ for any $\boldsymbol{\theta}$.

b. Assuming that the limit in a. is meaningful, let us think next about estimating asymptotic variances. We have discussed a variety of methods: estimated information matrices, estimated M-estimation sandwich, jackknife, nonparametric bootstrap, and parametric bootstrap.

(i) Which of these methods (one or more) will in general be appropriate (consistent) only if the parametric model is correct?

(ii) Which of the methods (one or more) will generally be appropriate if the parametric model is correct or incorrect.

2. Suppose that $Y_1, \ldots, Y_n$ are iid with distribution function F. We will use two functions q_1 and q_2, and assume $\mathrm{E}\{q_1(Y_1)\} = 0$, $\mathrm{Var}\{q_1(Y_1)\} = \sigma_1^2 < \infty$, $\mathrm{E}\{q_2(Y_1)\} = \mu_2 \neq 0$, $\mathrm{Var}\{q_2(Y_1)\} = \sigma_2^2 < \infty$, $\mathrm{E}\{q_1(Y_1)q_2(Y_1)\} = \mu_{12}$, $\mathrm{Var}\{q_1(Y_1)q_2(Y_1)\} = \sigma_{12}^2 < \infty$. Consider the following three estimators:

$$\widehat{\theta}_1 = \left\{\frac{1}{n}\sum_{i=1}^n q_1(Y_i)\right\}\left\{\frac{1}{n}\sum_{i=1}^n q_2(Y_i)\right\}$$

$$\widehat{\theta}_2 = \frac{1}{n}\sum_{i=1}^n \{q_1(Y_i)q_2(Y_i)\}$$

$$\widehat{\theta}_3 = \left\{\frac{1}{n}\sum_{i=1}^n q_1(Y_i)\right\} \Big/ \left\{\frac{1}{n}\sum_{i=1}^n q_2(Y_i)\right\}.$$

a. Prove that each estimator converges in probability and give the limit.

b. Do the following for each estimator. If $\widehat{\theta}_i$ is an M-estimator, give the ψ function. If it is not an M-estimator by itself, give a set of ψ functions that will yield the estimator.

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

a. Give the approximation by averages h function for m_3.

b. The third moment skewness coefficient is Skew $= m_3/s_n^3$, where $s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2$. Give the approximation by averages h function for Skew.

4. Five patients were available to test whether a new treatment was better than the standard one. Three patients were randomly assigned to receive the new treatment, and the remaining two patients received the standard treatment. The new treatent patients obtained scores of 45, 38, 52 (higher is better), and the two patients receiving the standard treatment obtained scores of 32, and 25.

a. Compute the Wilcoxon Rank Sum statistic, W = sum of the new treatment ranks, for H_0 : treatments are identical versus the new treatment is better than the standard treatment.

b. Give the permutation distribution of W under H_0. (Hint: there are 10 unique splits into 2 samples.)

c. Give the permutation p-value = probability of greater than or equal to the observed W under the permutation distribution.

5. Suppose that $Y_1, \dots, Y_{n_1}$ are iid Poisson(λ_1) and $X_1, \dots, X_{n_2}$ are iid Poisson(λ_2), $H_0 : \lambda_1 = \lambda_2$ versus $H_a : \lambda_1 \neq \lambda_2$, and the samples are independent. The Wald test statistic is $T_{\mathrm{W}} = (\overline{Y} - \overline{X})^2 / \left(\overline{Y}/n_1 + \overline{X}/n_2\right)$.

a. Instead of a Poisson distribution, suppose that the Y sample is from a distribution with mean λ_1 and variance σ_1^2, and the X sample is from a distribution with mean λ_2 and variance σ_2^2. Under $H_0 : \lambda_1 = \lambda_2$, and assuming for simplicity that $n_1 = n_2 = n$, prove that $T_{\mathrm{W}} \xrightarrow{d} c_1 Z^2$, where Z is a standard normal random variable, and give the constant c_1. Use basic theorems from the large sample Chapter 5; do not use theorems from Ch. 8.

b. It takes some work to derive T_{GW}; instead just try and guess what the generalized Wald statistic T_{GW} would be for this situation by thinking about the obvious nonparametric estimator for $\mathrm{Var}(\overline{Y} - \overline{X})$.

6. A small subset of a table from a Monte Carlo study might look like the following:

Sample size	nMSE($\widehat{\theta}_1$)	nMSE($\widehat{\theta}_2$)	MSE($\widehat{\theta}_1$)/MSE($\widehat{\theta}_2$)
n=10	0.49	0.56	0.88
n=20	0.47	0.55	0.85
n=50	0.46	0.53	0.87

These estimates were calculated from $N = 1,000$ Monte Carlo samples. For each sample size n, picture 1,000 independent rows of data. In each row we have the basic estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ calculated from one Monte Carlo sample. Note that the two columns of data are correlated because $\widehat{\theta}_1$ and $\widehat{\theta}_2$ in a given row are calculated on the same generated data set. Each entry in the second column above is the average of N values of $(\widehat{\theta}_1 - \theta)^2$ times n, where θ is the known true value. Each entry in the third column above is the average of N values of $(\widehat{\theta}_2 - \theta)^2$ times n. The entries in column four are just the ratios of the entries in column 2 and column 3.

A good table should provide some idea of the standard errors (estimated standard deviations) of the entries given in the table, perhaps in the caption or in a footnote. I like to put a final row in the table with some summary of the standard errors for the entries such as the maximum standard error for each column's entries or the average standard error. In order to compute this summary, the researcher needs a standard error for each entry. There would be 9 such standard errors for the above table.

Give an expression for the standard error of each different type of estimate. That is, give an expression for computing a standard error for each entry in column 2 (no need do it for column 3 since it is similar), and a second expression for those in column 4. Note, although in practice I like to suggest the use of jackknife or bootstrap standard errors, for this question I want explicit expressions based on exact or asymptotic variances. (If you use M-estimation or Delta methods, you can just leave the expressions without multiplying them together.)

Theorem 5.19 (Delta Method - Real-Valued Function of a Vector) *Suppose that $\widehat{\boldsymbol{\theta}}$ is $AN_k(\boldsymbol{\theta}, b_n^2\boldsymbol{\Sigma})$ with $b_n \to 0$ and that g is a real-valued function with partial derivatives existing in a neighborhood of $\boldsymbol{\theta}$ and continuous at $\boldsymbol{\theta}$ with $g'(\boldsymbol{\theta}) = \partial g(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ not identically zero. Then as $n \to \infty$*

$$g(\widehat{\boldsymbol{\theta}}) \text{ is } AN\left[g(\boldsymbol{\theta}), b_n^2 g'(\boldsymbol{\theta})\boldsymbol{\Sigma} g'(\boldsymbol{\theta})^T\right].$$

Theorem 5.27. *For an iid sample $X_1, \ldots, X_n$, suppose that each component of $\widehat{\boldsymbol{\theta}}$ has the approximation by averages representation given by*

$$\widehat{\theta}_j - \theta_j = \frac{1}{n}\sum_{i=1}^{n} h_j(X_i) + R_{nj}, \quad j = 1, \ldots, k, \tag{1}$$

where $\sqrt{n}R_{nj} \xrightarrow{p} 0$ as $n \to \infty$, $E\{h_j(X_1)\} = 0$, and $Var\{h_j(X_1)\}$ is finite, $j = 1, \ldots, k$. Also assume that g is a real-valued function with partial derivatives existing in a neighborhood of the true value $\boldsymbol{\theta}$ and continuous at $\boldsymbol{\theta}$. Then

$$g(\widehat{\boldsymbol{\theta}}) - g(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{j=1}^{k} g_j'(\boldsymbol{\theta}) h_j(X_i)\right] + R_n, \tag{2}$$

where $\sqrt{n}R_n \xrightarrow{p} 0$ as $n \to \infty$.

Theorem 5.28. *For an iid sample $X_1, \ldots, X_n$, suppose that each component of $\widehat{\boldsymbol{\theta}}$ has the approximation by averages representation given by*

$$\widehat{\theta}_j - \theta_j = \frac{1}{n}\sum_{i=1}^{n} h_j(X_i) + R_{nj}, \quad j = 1, \ldots, b, \tag{3}$$

where $\sqrt{n}R_{nj} \xrightarrow{p} 0$ as $n \to \infty$, $E\{h_j(X_1)\} = 0$, and $Var\{h_j(X_1)\}$ is finite, $j = 1, \ldots, b$. Also assume that the real-valued function $q(X_i, \boldsymbol{\theta})$ has two partial derivatives with respect to $\boldsymbol{\theta}$, and

1. *$Var\{q(X_1, \boldsymbol{\theta})\}$ and $E\{q'(X_1, \boldsymbol{\theta})\}$ are finite;*
2. *there exists a function $M(x)$ such that for all $\boldsymbol{\theta}^*$ in a neighborhood of the true value $\boldsymbol{\theta}$ and all $j, k \in \{1, \ldots, b\}$, $|q''(x, \boldsymbol{\theta}^*)_{jk}| \leq M(x)$, where $E\{M(X_1)\} < \infty$.*

Then,

$$\frac{1}{n}\sum_{i=1}^{n} q(X_i, \widehat{\boldsymbol{\theta}}) - E\{q(X_1, \boldsymbol{\theta})\} = \frac{1}{n}\sum_{i=1}^{n} h_{\scriptscriptstyle T}(X_i) + R_n,$$

where

$$h_{\scriptscriptstyle T}(X_i) = q(X_i, \boldsymbol{\theta}) - E\{q(X_1, \boldsymbol{\theta})\} + \left[E\{q'(X_1, \boldsymbol{\theta})\}\right]\boldsymbol{h}(X_i),$$

and $\sqrt{n}R_n \xrightarrow{p} 0$ as $n \to \infty$.