Statistics comprehensive exam. January 2018

Instructions: *The exam has 6 problems. All parts of all problems will be graded. Write your code words on each of your answer sheets. Do not put your name or UNM ID on any of the sheets. Be clear, concise, and complete. All solutions should be rigorously explained.*

1. Consider two linear models

$$Y = X_1\gamma_1 + X_2\gamma_2 + X_3\gamma_3 + e, \qquad E(e) = 0,$$

and

$$Y = X_1\beta_1 + X_3\beta_3 + e, \qquad E(e) = 0,$$

where X_3 is a single column but X_1 and X_2 may have multiple columns. Let J be an $n \times 1$ vector of 1s. It is assumed that J is in $C(X_1)$. Knaeble and Dutter (2016) were interested in when least squares estimates have $\text{sign}(\hat{\beta}_3) \neq \text{sign}(\hat{\gamma}_3)$. Let M_1 be the perpendicular projection operator onto $C(X_1)$ and let M_{12} be the ppo onto $C(X_1, X_2)$.

(a) Use standard results from analysis of covariance to show that $\text{sign}(\hat{\beta}_3) \neq \text{sign}(\hat{\gamma}_3)$ if and only if $\text{sign}[X_3'(I - M_1)Y] \neq \text{sign}[X_3'(I - M_{12})Y]$.

(b) Write $M_{2|1}$, the ppo onto $C(X_1)^{\perp}_{C(X_1,X_2)}$, in terms of M_1 and M_{12}.

(c) Assuming that $\hat{\beta}_3$ is positive, show that $\text{sign}(\hat{\beta}_3) \neq \text{sign}(\hat{\gamma}_3)$ if and only if

$$X_3'(I - M_1)Y < X_3'M_{2|1}Y.$$

(d) To state their Proposition 2.1, Knaeble and Dutter (2016) define

$$r[\widehat{X_{3|X_1}}(X_{2|X_1}), \widehat{Y_{|X_1}}(X_{2|X_1})] = \frac{X_3'(I - M_1)M_{2|1}(I - M_1)Y}{\sqrt{X_3'(I - M_1)M_{2|1}(I - M_1)X_3}\sqrt{Y'(I - M_1)M_{2|1}(I - M_1)Y}}.$$

Show that

$$r[\widehat{X_{3|X_1}}(X_{2|X_1}), \widehat{Y_{|X_1}}(X_{2|X_1})] = \frac{X_3'M_{2|1}Y}{\sqrt{X_3'M_{2|1}X_3}\sqrt{Y'M_{2|1}Y}}.$$

2. Let $X_1, X_2, \ldots$ be independent with X_n taking the values $\sqrt{n}$, 0, and $-\sqrt{n}$ each with probability 1/3. Find the asymptotic distribution of $\bar{X}_n$.

3. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables (i.i.d.) with the probability density function,

$$f(x|\mu, \sigma) = \frac{1}{\sigma}\sqrt{\frac{2}{\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right); x > \mu$$

and 0 otherwise, where $-\infty < \mu < \infty$ and $\sigma > 0$ are unknown parameters.

(a) Describe or sketch this distribution.

(b) Identify a set of sufficient statistics for the unknown parameters (μ, σ^2).

(c) Derive the maximum likelihood estimator (MLE) of (μ, σ^2).

(d) Suppose that $\mu = 0$ is known. Derive the UMP test for $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma = \sigma_1$ for two fixed parameter values $\sigma_1 > \sigma_0 > 0$ at the $\alpha > 0$ level. Specify the rejection region.

4. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (iid) observations from a $Unif[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$ with probability density function $f(x|\theta) = I_{[\theta - \frac{1}{2}, \theta + \frac{1}{2}]}(x)$ where $-\infty < \theta < \infty$ is an unknown parameter. Consider the following estimator of θ,

$$\hat{\theta}_1 = \frac{X_{(1)} + X_{(n)}}{2}$$

where $X_{(1)}$ and $X_{(n)}$ denote the minimum and maximum order statistics.

(a) Is $\hat{\theta}_1$ unbiased for θ? Show your answer.

(b) Find the method of moment estimator of θ, and call it $\hat{\theta}_2$. Also identify the limiting distribution of $\hat{\theta}_2$ as $n \to \infty$.

(c) Without deriving their variance, state which of the two estimators you prefer ($\hat{\theta}_1$ or $\hat{\theta}_2$) and why you prefer it.

(d) Assume θ follows a prior distribution $Unif[-A, A]$ where $A > 0$. Find a Bayes estimator for θ.

5. A set of n counts $X = (X_1, X_2, \ldots, X_n)$ are modeled as

$$Pr(X_i = 0 \mid \gamma_i = 0, \pi, \lambda) = 1,$$

$$X_i \mid (\gamma_i = 1, \pi, \lambda) \sim Poisson(\lambda), \text{(indep. across i)}$$

where $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n)$ is a set of n latent binary counts, $\pi \in [0, 1]$ and $\lambda > 0$ are assigned as what is known as a "hierarchical prior":

$$\gamma_i|(\pi, \lambda) \sim \text{iid } Bernoulli(\pi), i = 1, \ldots, n;$$

$$\pi|\lambda \sim Beta(c\lambda, 1), \lambda \sim Gamma(a, b)$$

for some positive constants a, b and c. ($Beta(r, 1)$ has pdf $r\pi^{r-1}, 0 \leq \pi \leq 1$)

(a) Show that the conditional prior pdf of λ given π is $Gamma(a + 1, b - c\ log\ \pi)$.

(b) Write down the posterior conditional probability distributions of $\pi|(\gamma, \lambda, x)$, $\lambda|(\gamma, \pi, x)$, and $\gamma|(\pi, \lambda, x)$ given data $x = (x_1, x_2, \ldots, x_n)$ on X. Answer in terms of conditional distributions with explicit formulas for their parameters and with appropriate use of conditional independence.

6. Let $X_1, X_2, \ldots, X_n$ be a random sample from the Poisson (θ) distribution and let

$$Z_n = \frac{1}{n} \sum_{i=1}^{n} I(X_i = 0)$$

where $I_A(X)$ is the indicator function of X over the set A. Also consider that

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

(a) Find the joint asymptotic distribution of $\sqrt{n}(Z_n, \bar{X}_n)$ and provide all the parameters of this distribution.

(b) Find the asymptotic distribution of $\sqrt{n}(Z_n/\bar{X}_n)$. Justify your answer.

(c) Find the limit in probability of $Z_n/\bar{X}_n$. Justify your answer.