Statistics comprehensive exam. January 2021

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.*

Problem 1. Consider a linear model

$$Y = X\beta + e, \quad \mathrm{E}[e] = 0, \quad \mathrm{Cov}\,[e] = \sigma^2 I.$$

Let M be the perpendicular projection operator onto $C(X)$. Consider the parameter $\lambda'\beta$ where λ is known. For known vectors a and ρ, suppose $a'Y$ and $\rho'Y$ are both unbiased estimates of $\lambda'\beta$.

(a) Show that $\lambda'\beta$ is estimable.

(b) Show that $a'X = \rho'X$.

(c) Show that $\mathrm{Var}[a'Y] = \mathrm{Var}[a'Y - \rho'MY] + \mathrm{Var}[\rho'MY]$.

(d) State and prove the Gauss-Markov Theorem.

Problem 2. Consider conducting $\alpha = 0.10$ level Neyman-Pearson (N-P) tests of $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ for $\theta \in \{0, 1, \ldots, 100\}$. The null distribution is

$$\Pr[X = 0|\theta = 0] = .9 \quad \Pr[X = i|\theta = 0] = .001, i = 1, \ldots, 100.$$

and the alternative sampling distributions are

$$\Pr[X = 0|\theta = i] = .91 \quad \Pr[X = i|\theta = i] = .09, i = 1, \ldots, 100.$$

(a) Find the generalized likelihood ratio test and its power.

(b) Find the most powerful test for $H_0 : \theta = 0$ versus $H_1 : \theta = i$ for any particular simple alternative $i \neq 0$.

(c) If a UMP test exists, give it; or explain why it does not exist.

Problem 3. Consider a linear model

$$Y = X\beta + e, \qquad \mathrm{E}[e] = 0$$

with the (not necessarily estimable) linear constraint $\Lambda'\beta = d$. Consider two (known) solutions to the constraint, b_1 and b_2, so that $\Lambda' b_k = d$, $k = 1, 2$. Define $X_0 = XU$ with $C(U) = C(\Lambda)^{\perp}$.

(a) Define appropriate least squares fitted values $\hat{Y}_k$ from the model $Y = X_0\gamma + Xb_k + e$.

(b) Show that $\hat{Y}_1 = \hat{Y}_2$.

Hint: After finding $\hat{Y}_k$, show that $(I - M_0)X(b_1 - b_2) = 0$ where M_0 is the perpendicular projection operator onto $C(X_0)$.

Problem 4. Let $X_1, X_2, \ldots, X_n$ be a random sample from a scale-uniform distribution $X_i \sim$ Uniform $((1-k)\theta, (1+k)\theta)$, with unknown mean $\mathrm{E}[X_i] = \theta$ and known design parameter $k \in (0,1)$. Let $X_{(1)} \equiv \min_i\{X_i\}$ and $X_{(n)} \equiv \max_i\{X_i\}$. $\mathrm{E}\big[X_i|X_{(1)}, X_{(n)}\big] = \frac{X_{(1)}+X_{(n)}}{2}$.

(a) What is the mean and variance of X_i?

(b) Show that $(X_{(1)}, X_{(n)})$ is a sufficient statistic.

(c) Give a rigorous argument for why $(X_{(1)}, X_{(n)})$ is minimal sufficient.

(d) Why is $\mathrm{E}\big[X_i|X_{(1)}, X_{(n)}\big] = \frac{X_{(1)}+X_{(n)}}{2}$ a better estimate of θ than X_i.

(e) The variance of $\frac{X_{(1)}+X_{(n)}}{2}$ is $\frac{2k^2\theta^2}{(n+1)(n+2)}$. Establish whether this is larger or smaller than the variance of the sample mean $\overline{X}$.

(f) For some C there is an unbiased estimate of the form $C[(1-k)X_{(1)} + (1+k)X_{(n)}]$ with variance $\frac{2k^2\theta^2}{[(1+k^2)(n-1)+2](n+2)}$. Establish whether this is larger or smaller than the variance of $\frac{X_{(1)}+X_{(n)}}{2}$.

(g) Prove that $(X_{(1)}, X_{(n)})$ is not complete.

Problem 5. Let $X_1, X_2, \ldots$ be independent with X_n taking the values $\pm\sqrt{n-1}$ each with probability 1/2. Use a (not "the") Central Limit Theorem and other asymptotic results to show that the sample mean converges in distribution to a $\mathrm{N}(0, 0.5)$.
Hints: $\sum_{i=1}^{r} i = r(r+1)/2$; $\sum_{i=1}^{r} i^2 = r(r+1)(2r+1)/6$.

Problem 6. 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 an "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 $r\pi^{r-1}, 0 \le \pi \le 1$)

(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.