Statistics comprehensive exam. August 2021

Instructions: *The exam has 5 equally weighted 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. Let $y_1, ..., y_n$ be a random sample modeled by a $\mathrm{N}(\mu, {}^1\!/_\tau)$ distribution.

(a) Consider the case where μ is known and τ is unknown. Find the score function for τ.

(b) Find the Fisher information for τ, still assuming that μ is known.

(c) Find the Jeffreys prior for τ when μ is known.

(d) Consider the case where μ is unknown and τ is known. Derive the Jeffreys prior for μ.

(e) Derive the Jeffreys prior in the case where μ and τ are both unknown.

(f) The Jeffreys prior is always a flat prior on some parameterization of the model it comes from. In the case where both μ and τ are unknown, find the parameterization for the Normal model that gives rise to a flat Jeffreys prior over $\mathbb{R}^2$.

Problem 2. For each n, let y_{ni}, $i = 1, \ldots, n$, be independent with mean 0 and variance σ_{ni}^2. Let $z_n \equiv \sum_{i=1}^n y_{ni}$ and $B_n^2 \equiv \mathrm{Var}[z_n] = \sum_{i=1}^n \sigma_{ni}^2$. We are going to look at how the Lindeberg Central Limit Theorem applies to exponentially weighted moving averages (EWMAs) of iid random variables x_i with $\mathrm{E}[x_i] = 0$, $\mathrm{Var}[x_i] = \sigma^2$. Define the EWMA as

$$\hat{\mu}_n \equiv (\alpha x_n + \alpha^2 x_{n-1} + \alpha^3 x_{n-2} + \cdots + \alpha^n x_1) \Big/ \sum_{i=1}^n \alpha^i = \frac{\sum_{i=1}^n \alpha^i x_{n-i+1}}{\sum_{i=1}^n \alpha^i},$$

for $0 < \alpha < 1$. Some algebra applied to power series establishes that

$$\mathrm{Var}[\hat{\mu}_n] = \sigma^2 \left(\frac{1+\alpha^n}{1-\alpha^n}\right)\left(\frac{1-\alpha}{1+\alpha}\right)$$

Relative to Lindeberg, define

$$z_n \equiv \hat{\mu}_n.$$

(a) State the Lindeberg Central Limit Theorem.

(b) What is y_{ni} in terms of the x_is?

(c) What is B_n^2 and what does it converge to?

(d) Does $\hat{\mu}_n$ converge in probability to $\mathrm{E}[x_i]$?

(e) Does $\mathrm{E}\left[|y_{nn}|^2 \mathcal{I}_{[\epsilon B_n, \infty)}(|y_{nn}|)\right]$ converge to 0?

(f) Does the Lindeberg condition hold?

Problem 3. Consider a random n-vector Y with $\mathrm{E}[Y] = \mu J$ and $\mathrm{Cov}\,[Y] = \sigma^2\,[(1-\rho)I + \rho JJ']$. Here J is a vector of 1s. Suppose ρ is known. Recall

$$\overline{y} = \frac{1}{n}J'Y \quad \text{and} \quad s^2 = \frac{1}{n-1}Y'[I-(1/n)JJ']Y.$$

(a) Show that $\mathrm{E}[s^2] = \sigma^2(1-\rho)$.

(b) Show that $\mathrm{Var}[\overline{y}] = (\sigma^2/n)[(1-\rho)+n\rho]$

(c) For sampling w/o replacement from a population of size N, $\mathrm{Cov}\,[y_i, y_j] = -\sigma^2/(N-1)$. Show that

$$\left(1-\frac{n}{N}\right)\frac{s^2}{n}$$

is an unbiased estimate of $\mathrm{Var}[\overline{y}]$.

Problem 4.

(a) State and prove the Gauss-Markov Theorem

(b) Give an example of a sequence of random variables that converge in probability but do not converge almost surely.

Problem 5. Let $y_1, ..., y_n$ be a random sample modeled by a $\mathrm{N}(\mu, {}^1\!/_\tau)$ distribution, where τ is known. We are interested in testing an hypothesis about μ, $H : \mu = \mu_0$, from both a Frequentist and a Bayesian approach. Let c be the prior probability that H is true, $\Pr[H] = c$, and let the remainder of the prior probability for values of μ be distributed uniformly from $\mu_0 - \theta$ to $\mu_0 + \theta$ for some known θ.

(a) Find an expression for the Frequentist p-value relative to H in terms of the absolute value of the test statistic and the cdf, Φ, of a standard normal. Call this $p \equiv p(y_1, \ldots, y_n)$.

(b) Find an expression for the posterior probability that H is true, $\Pr[H \mid y_1, ..., y_n]$. Assume that $\overline{y} = \sum_{i=1}^{n} y_i$ is well inside the interval $\mu_0 \pm \theta$, allowing you use the approximation $\int_{\mu_0-\theta}^{\mu_0+\theta} \exp\left(-\frac{n\tau}{2}[\overline{y}-\mu]^2\right) d\mu \simeq \int_{\mathbb{R}} \exp\left(-\frac{n\tau}{2}[\overline{y}-\mu]^2\right) d\mu = \left(\frac{2\pi}{n\tau}\right)^{1/2}$.

(c) Suppose the value of $\overline{y}$ is such that, on performing a standard Frequentist test of H, a p-value of p is obtained. Let λ_p be the test statistic associated with that p-value, that is let it be a number dependent only on p and the Normal cdf such that $\overline{y} = \mu_0 + \lambda_p/\sqrt{n\tau}$. Re-express the posterior probability for H in a way that depends on the data only through n and λ_p.

(d) Describe the limiting behavior of $\Pr[H \mid y_1, ..., y_n]$ as $n \to \infty$, assuming that the Frequentist p-value for the random sample of y's remains fixed at p.

(e) This result is known as the Jeffreys-Lindley Paradox. Explain what seems paradoxical about the results you have obtained.