Statistics comprehensive exam. August 2023

Instructions: The exam has 7 problems often with multiple parts. **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. Let $X_1, X_2, \ldots$ be independent with X_n taking the values $\pm\sqrt{n-2}$ with probability 3/8 and the values $\pm\sqrt{6}$ with probability 1/8. Show that $\bar{X}_n$ converges in distribution to a $N(0, 3/8)$.

2. (a) Let $y_1, \ldots, y_n$ be independent q vectors with Multinomial$(1, p)$ distributions. Show that $\sqrt{n}(\bar{y}_n - p)$ converges in distribution to a multivariate $N[0, \Sigma(p)]$ and find $\Sigma(p)$.

   (b) Let W be Multinomial(N, p). Find the asymptotic distribution of $[W - E(W)]/\sqrt{N}$.

   (c) Write $W = (w_{11}, w_{12}, w_{21}, w_{22})'$. Find a large sample approximation to the distribution of $\log(w_{11}w_{22}/w_{12}, w_{21})$. In particular, show that an estimated standard error can be taken as $\sqrt{\sum_i \sum_j 1/w_{ij}}$.

   (d) Consider a log-linear model, $\mathrm{E}(w_{ij}) = u + u_{1i} + u_{2j} + u_{12ij}$. Find a contrast in the interaction terms u_{12ij} and discuss how to test if it is different from 0.

3. Let Y be a n vector, the components of which are independent Poisson(λ) random variables.

   (a) Find the score function $S(Y; \lambda)$.

   (b) Find the variance of the score function as a function of λ.

   (c) For an unbiased estimate of λ, say $T(Y)$, find the covariance between it and the score function.

   (d) What can the Cauchy-Schwarz inequality tell you about the variance of $T(Y)$?

   (e) How does this result fit into the theory of statistical inference?

4. Suppose $Y \sim$ Pois(λ) and λ has a prior Gamma density, $(1/\Gamma(\alpha)\beta^\alpha)\lambda^{\alpha-1}\exp(-\lambda\beta)$.

   (a) Show that the marginal distribution of Y is negative binomial with parameters α and $p \equiv \beta/(1+\beta)$. (The negative binomial for integer α is the number of failures in an iid Bernoulli sequence prior to reaching a fixed number of successes.)

   (b) Find the posterior density of λ.

5. Consider testing the null hypothesis $Y \sim$ Pois(1) versus the alternative $Y \sim$ Pois(2). Use the Neyman-Pearson Lemma to find the most powerful size 0.05 test.

   Here is some R output that you might find useful.

```
> r
[1] 0 1 2 3 4 5
> dpois(r,1)
[1] 0.367879441 0.367879441 0.183939721 0.061313240 0.015328310 0.003065662
```

Why is this also the UMP test for $H_0 : \lambda \leq 1$ versus $H_1 : \lambda > 1$, when $Y \sim \text{Pois}(\lambda)$?

6. Let $y_1, \ldots, y_n$ be independent with $N(\mu, 1)$ distributions. Two unbiased estimates of μ are the sample mean $\bar{y}$ and the midrange $mr \equiv [y_{(1)} + y_{(n)}]/2$. Since

$$\text{E}(\bar{y}) = \text{E}[\text{E}(\bar{y}|mr)] = \mu$$

and

$$\text{Var}(\bar{y}) = \text{E}[\text{Var}(\bar{y}|mr)] + \text{Var}[\text{E}(\bar{y}|mr)] \geq \text{Var}[\text{E}(\bar{y}|mr)],$$

why do we not use $\text{E}(\bar{y}|mr)$ as an improved unbiased estimator of μ? To be more precise, an alternative unbiased estimate that is at least as good.

7. In a standard linear model $Y = X\beta + e$ we know that $\hat{\beta}$ is a least squares estimate if and only if $X\hat{\beta} = MY$ where M is the perpendicular projection operator onto $C(X)$, the column space of X. Show that $\hat{\beta}$ is a least squares estimate if and only if it is a solution to the normal equations $X'X\beta = X'Y$.