Statistics 583: Midterm Exam

Wellner; 5/1/2000

Instructions: This exam is to be done with closed notes and closed books.

1. **Define** the following terms, providing an appropriate context for your definition:
   A. A permutation test.
   B. Monotone likelihood ratio.
   C. An unbiased test.
   D. A locally most powerful test (for a one-parameter problem).

   Solution: see the chapter 6 notes.

2. State any two of the following four theorems, providing an appropriate context in each case: .br A. The Neyman-Pearson lemma.
   B. The generalized Neyman-Pearson lemma.
   C. The Wald - Wolfowitz - Noether - H'ajek finite sampling central limit theorem .
   D. The Karlin-Rubin theorem.

   Solution: see the chapter 6 notes.

3. Suppose that $Y_1, \ldots, Y_N$ are independent with $Y_i \sim \text{Binomial}(n_i, p_i)$, $i = 1, \ldots, N$, where
   $$p_i = \frac{1}{1 + \exp(-(\gamma + \delta x_i))}$$
   for given numbers $x_1, \ldots, x_N$. This is a model frequently used in bioassay, where x_i denotes the dose (or logarithm of the does) of a drug given to n_i experimental subjects, and Y_i is the number among these n_i subjects who respond to the drug at the level x_i.
   A. Show that the joint distribution of the Y's is an exponential family. Identify the natural parameters and the sufficient statistics for these parameters.
   B. Consider testing $H : \delta \le 0$ versus $K : \delta > 0$. Find the UMP unbiased size α test of H versus K. C. What is the distribution of $T \equiv \sum_1^N Y_i$ under $\delta = 0$?
   What is the conditional distribution of $(Y_1, \ldots, Y_N)$ given T under $\delta = 0$?

   Solution: A. First note that
   $$\frac{p_i}{1 - p_i} = \exp(\gamma + \delta x_i), \qquad i = 1, \ldots, N\,.$$
   Hence the joint distribution of $Y_1, \ldots, Y_N$ is given by
   $$\begin{aligned} p(y_1, \ldots, y_N; \gamma, \delta) &= \prod_{i=1}^N \binom{n_i}{y_i} \left(\frac{p_i}{1 - p_i}\right)^{y_i} (1 - p_i)^{n_i} \\ &= \left\{\prod_{i=1}^N \binom{n_i}{y_i} (1 - p_i)^{n_i}\right\} \exp(\sum_{i=1}^N (\gamma + \delta x_i) y_i) \\ &= c(\gamma, \delta) \exp(\delta \sum_{i=1}^N x_i y_i + \gamma \sum_{i=1}^N y_i)\,. \end{aligned}$$

Thus γ and δ are natural parameters for the exponential family and the associated sufficient statistics are $\sum_{i=1}^N Y_i$ and $\sum_{i=1}^N x_i Y_i$.

B. When $\delta = 0$, a sufficient statistic for γ is $\sum_{i=1}^N Y_i$, and for testing hypotheses about δ the natural identifications with our canonical exponential family problem are $\theta = \delta$, $U = \sum_{i=1}^N x_i Y_i$, $\xi = \gamma$, $T = \sum_{i=1}^N Y_i$. Thus the UMP unbiased test ϕ of $H : \delta \le 0$ versus $K : \delta > 0$ is of the form

$$\phi(U, T = t) = \begin{cases} 1 & \text{if} \quad \sum_1^N x_i Y_i > c(t) \\ \gamma(t) & \text{if} \quad \sum_1^N x_i Y_i = c(t) \\ 0 & \text{if} \quad \sum_1^N x_i Y_i < c(t) \end{cases}$$

where $c(t), \gamma(t)$ are chosen so that

$$E_{\delta=0}\left\{\phi(U,T)|T = t\right\} = \alpha\,.$$

C. Note that when $\delta = 0$, $p_i = 1/(1 + e^{-\gamma}) \equiv p_\gamma$ does not depend on i, and $T = \sum_1^N Y_i \sim \text{Binomial}(\sum_1^N n_i, p_\gamma)$. Thus the conditional distribution of $\underline{Y}$ given $T = t$ is multiple Hypergeometric:

$$P(\underline{Y} = \underline{y} \Big| T = t) = \frac{\prod_{i=1}^N \binom{n_i}{y_i}}{\binom{\sum_1^N n_i}{t}}, \qquad \text{if} \;\; \sum_1^N y_i = t\,,$$

and 0 otherwise. Thus the conditional distribution of $U = \sum_1^N x_i Y_i$ given $T = t$ is completely free of the nuisance parameter γ.

4. Suppose that under the null hypothesis H_c, $X_1, \ldots, X_n$ are i.i.d. $F \in \mathcal{F}_c$, while under the alternative hypothesis K_1 $X_1, \ldots, X_n$ have joint density h given by

$$h(\underline{x}) = \prod_{i=1}^n \lambda_i \exp(-\lambda_i x_i)$$

where $\lambda_i = \mu e^{\nu c_i}$ for some fixed numbers (covariates) $c_1, \ldots, c_n$, and constants $\mu > 0$ and $\nu \in R$.

A. Find a most powerful similar test of size α for testing H_c versus K_1. Does your test depend on the values of the constants μ and/or ν?

B. If $n = 3$, $\lambda_1 = 1$, $\lambda_2 = 2$, and $\lambda_3 = 3$, and we observe $(X_1, X_2, X_3) = (3.5, 2.1, .9)$, carry out the test in A at level $\alpha = 1/6$.

C. Find a test which is a locally most powerful similar test of H_c versus $K : \nu > 0$. How does this test differ from the test you found in A?

Solution: A. To obtain a similar test of H_c, we condition on the order statistics $\underline{Z} \equiv (X_{(1)}, \ldots, X_{(n)})$ of the sample $X_1, \ldots, X_n$. A most powerful similar test of H_c versus K_1 rejects for those permutations $\underline{z}'$ of $\underline{Z} = \underline{z}$ which lead to large values of $h(\underline{z}')$, or equivalently for large values of

$$\log h(\underline{z}) = \sum_{i=1}^n \{\log \lambda_i - \lambda_i z_i\} = \sum_{i=1}^n \log \lambda_i - \sum_{i=1}^n \lambda_i z_i\,,$$

or, equivalently, for small values of

$$\sum_{i=1}^{n} \lambda_i z_i = \mu \sum_{i=1}^{n} \exp(\nu c_i) z_i \,,$$

or, equivalently, for small values of

$$\sum_{i=1}^{n} \exp(\nu c_i) z_i \,.$$

Thus we see that our test does not depend on the value of μ, but it does depend on the value of ν.
B. When $n = 3$ and $\lambda_i = i$, $i = 1, 2, 3$, and we observe $(X_1, X_2, X_3) = (3.5, 2.1, .9)$, we find the following $3! = 6$ values of the test statistic in part A:

$\underline{z}'/i$	1	2	3	$\sum_1^3 \lambda_i z_i'$
1	.9	2.1	3.5	15.6
2	2.1	.9	3.5	14.4
3	.9	3.5	2.1	14.2
4	2.1	3.5	.9	11.8
5	3.5	.9	2.1	11.6
6	3.5	2.1	.9	10.4

Since the observed value of the test statistic, 10.4, is the most extreme in the permutation distribution, we would reject H_c at level $\alpha = 1/6$.
C. Write $h = h_\nu$ to emphasize the dependence of h on ν. Recall that the conditional distribution of permutations $\underline{z}'$ of $\underline{z}$ under alternatives is

$$p_\nu(\underline{z}'; \underline{z}) = h_\nu(\underline{z}') / \sum_{\underline{z}''} h(\underline{z}'') \,.$$

Now the conditional power of a test ϕ as a function of ν is

$$\begin{aligned} \beta_\phi(\nu) &= E_\nu \left\{ \phi(\underline{X}) | \underline{Z} = \underline{z} \right\} \\ &= \sum_{z'} \phi(\underline{z}') p_\nu(\underline{z}') \,. \end{aligned}$$

Thus the slope of the power function at $\nu = 0$ is given by

$$\frac{d}{d\nu} \beta_\phi(\nu)|_{\nu=0} = \sum_{z'} \phi(\underline{z}') \frac{\partial}{\partial \nu} p_\nu(\underline{z}')|_{\nu=0} \,,$$

and by the generalized NP lemma, this leads us to reject H_c for those permutations yielding large values of

$$\left. \frac{\frac{\partial}{\partial \nu} p_\nu(\underline{z}')}{p_\nu(\underline{z}')} \right|_{\nu=0}$$

Thus we calculate

$$\begin{aligned}
\frac{\frac{\partial}{\partial\nu}p_\nu(\underline{z}')}{p_\nu(\underline{z}')} &= \frac{\partial}{\partial\nu}\log p_\nu(\underline{z}') \\
&= \frac{\partial}{\partial\nu}\left\{\log h_\nu(\underline{z}') - \log\sum_{z''} h_\nu(\underline{z}'')\right\} \\
&= \frac{\partial}{\partial\nu}\left\{\sum_1^n \log\lambda_i - \lambda_i z_i' - C_\nu(\underline{z})\right\} \\
&= \sum_{i=1}^n \{c_i - c_i\lambda_i z_i'\} - \frac{\partial}{\partial\nu}C_\nu(\underline{z})\,.
\end{aligned}$$

When this is evaluated at $\nu = 0$, we find that

$$\left.\frac{\frac{\partial}{\partial\nu}p_\nu(\underline{z}')}{p_\nu(\underline{z}')}\right|_{\nu=0} = \sum_{i=1}^n \{c_i - \mu c_i z_i'\} + C_0'(\underline{z})\,,$$

and hence the locally most powerful similar test rejects H for small values of $\sum_1^n c_i z_i'$.