Statistics (Ph. D. Version)

Instructions to the Student

a. Answer all six questions. Each will be graded from 0 to 10.

b. Use a different booklet for each question. Write the problem number and your code number (NOT YOUR NAME) on the outside cover.

c. Keep scratch work on separate pages in the same booklet.

d. If you use a "well known" theorem in your solution to any problem, it is your responsibility to make clear which theorem you are using and to justify its use.

1. Consider the independent pairs \((X_1, Y_1), ..., X_n, Y_n)\), where the components \(X_i, Y_i, i = 1, ..., n\) are independent as well. Assume \(X_j, Y_j\) are \(N(\mu_j, \sigma^2)\). Define: \(Z_i = X_i - Y_i\).
   a. Find the MLE of \(\sigma^2\) from the differences \(Z_1, ..., Z_n\).
   b. Find the MLE of \(\sigma^2\) from the original \((X_1, Y_1), ..., X_n, Y_n)\) data.
   c. Which method a or b is better and why?

2. Let \(X\) be a Poisson random variable with parameter \(\theta\). Given a loss function \(L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2/\theta\) and a prior density \(\pi(\theta) = (\theta^2/2) \exp(-\theta), \theta > 0\), find the Bayes estimator of \(\theta\).

3. Consider a multinomial experiment with \(m\) categories and cell probabilities \(p_1, ..., p_m\) where \(\sum_{i=1}^{m} p_i = 1\). Let \(X_1, ..., X_m\) be the observed frequencies such that \(\sum_{i=1}^{m} X_i = n\). Suppose the statistician is interested in testing the hypothesis \(H_0\) that the cell probabilities depend on a \(k\)-dimensional parameter \(\theta\) versus the alternative \(H_1\) that the cell probabilities are free subject to
the fact they are nonnegative and sum to 1. Denote by $\Lambda$ the likelihood ratio test statistic.

a. Show that for suitable $O_i$ and $E_i$ we have

$$-2 \log \Lambda = 2 \sum_{i=1}^{m} O_i \log \left( \frac{O_i}{E_i} \right)$$

b. What is the asymptotic distribution of $-2 \log \Lambda$?

4. Let $X_1, \ldots, X_n$ be a random sample from the exponential distribution with density $f(x, \lambda) = \frac{1}{\lambda} \exp(-x/\lambda)$, $x > 0$.

a. Find the UMP test for testing $H_0 : \lambda = 7$ versus $H_1 : \lambda < 7$ at level $\alpha$.

b. Find a $1 - \alpha$ confidence interval for $\lambda$ by inverting the UMP test.

c. Find the expected length of the interval in part (b).

5. Let $X_1, \ldots, X_n$ be a random sample from a pdf

$$f(x, \theta) = \begin{cases} \theta f_1(x), & \text{if } x < 0 \\ (1 - \theta) f_2(x), & \text{if } x \geq 0 \end{cases}$$

where $f_1 \geq 0$, $f_2 \geq 0$, $0 < \theta < 1$, and

$$\int_{-\infty}^{0} f_1(x)dx = \int_{0}^{\infty} f_2(x)dx = 1$$

Prove or disprove that there exists a complete sufficient statistic for $\theta$.

6. Suppose a disease (D) is present in 4% of the population. A diagnostic test is available. The probabilities of $\overline{D}$ of no disease, and of "+" and "-" are given in the following table.

a. Suppose the test is given once. Compute $P(D|+)$.  
b. Suppose the test is given twice. Compute $P(D|++)$.

<table>
<thead>
<tr>
<th></th>
<th>+</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>0.0380</td>
<td>0.0020</td>
</tr>
<tr>
<td>\overline{D}</td>
<td>0.0384</td>
<td>0.9216</td>
</tr>
</tbody>
</table>