Statistics 405
Practice Final

Note. For the final examination, you may use a calculator and one page of notes (two-sided). You may not refer to any other source. Tables will be provided.

Important Terms-Probability: probability models, conditional probability and independence, important distributions, properties of expectation, the law of large numbers, and central limit theorem.

Important Terms-Statistics: Populations and samples; the method of moments; bias, variance and mean squared error; the likelihood function and maximum likelihood estimation; information; the Cramer Rao Inequality; the logic of hypothesis testing; the Neyman Pearson Lemma; likelihood ratio tests; the one and two sample problems; simple linear regression, prior and posterior distributions.

1. Suppose that $X$ has mean $\theta$ and variance $\sigma^2$ and that $Y$ has mean $\theta$ and variance $2\sigma^2$, where $\theta$ is unknown and $\sigma$ may be known or unknown. Show that any estimator of the form $\hat{\theta} = cX + (1 - c)Y$ is unbiased, where $c$ is a constant. Find the value of $c$ that minimizes the variance of $\hat{\theta}$.
   
   Ans: $c = 2/3$

2. Let $X_1, \ldots, X_n$ be independent with common density
   
   $$f_\theta(x) = \frac{\theta}{x^{\theta+1}}, \quad 1 \leq x < \infty,$$

   where $\theta > 0$ is an unknown parameter. Find the maximum likelihood estimator, $\hat{\theta}$ say, as a function of $X_1, \ldots, X_n$ and describe its approximate sampling distribution for large $n$.
   
   Ans: $\hat{\theta} = n/\sum_{i=1}^n \log(x_i)$ is approximately normal with mean $\theta$ and variance $\theta^2/n$ in large samples.

3. In the previous problem suppose that $n = 100$ and $\hat{\theta} = 5$. Find an approximate 95% confidence interval for $\theta$.
   
   Ans: $4.02 \leq \theta \leq 5.98$

4. Let $X_1, \ldots, X_n \sim \text{ind Poisson}(\theta)$. Find an unbiased estimator that attains the Cramer Rao Lower Bound for variances. If you wanted to estimate the variance of the $X_i$ would you use the sample mean or the sample variance? Why?
   
   Ans: $\bar{X}$ and $\bar{X}$; because $\bar{X}$ is efficient.

5. Let $X_1, \ldots, X_n \sim \text{ind Normal}(\theta, \theta)$, where $\theta > 0$ is an unknown parameter. (The mean and variance are both equal to $\theta$.) Find the method of moments and maximum likelihood estimators of $\theta$.
   
   Ans: $\bar{X}$ and $(\sqrt{n^2 + \frac{4T}{n}} - n)/2n$, where $T = X_1^2 + \cdots + X_n^2$.

6. Under $H_0$, $X$ has the standard normal distribution; under $H_1$, $X$ is normally distributed with mean $\mu = 1$ and variance $\sigma^2 = 1$. Find the most powerful level $\alpha = .05$ test of $H_0$ vs. $H_1$, and find the type II error probability of this test. What test minimizes the sum of the two error probabilities.
   
   Ans: reject if $X > 1.645$; .7405; reject if $X > 1/2$.

7. In order to compare two brands of tires, Brands 1 and 2 say, twelve tires of each brand were driven for 100,000 km. and the amount of treadwear $X$ in cm. was recorded. Assuming normality, describe a model for this experiment. If $\bar{X}_1 = 1.25, \bar{X}_2 = .75, S_1^2 = .03$ and $S_2^2 = .05$ find a 95% confidence interval for the difference between the two population means. Test the null hypothesis that the two population means are equal at the 5% level and estimate the attained significance level. In what way is the test a good test?
   
   Ans: $- .5 \pm .17$; reject; $p < .0005$; likelihood ratio test.

8. In a study of extra-sensory perception, were asked to guess the picture on tarrot cards in both a normal and hypnotized state with the following results:

<table>
<thead>
<tr>
<th>Brand</th>
<th>Std. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal.</td>
<td>18 19 16 21 16 20 20 14 11 22 19 29 16 27 15</td>
</tr>
<tr>
<td>Hypn.</td>
<td>25 20 26 26 20 23 14 18 18 20 22 27 19 27 21</td>
</tr>
</tbody>
</table>

Is the apparent difference statistically significant at the level $\alpha = .05$. What assumptions are you making here; why are these reasonable?
Ans: Yes; normality and independence between students; normality is suggested by the Central Limit
Theorem.

9. The data below list the average weight of varsity football players at the University of Texas for
selected years. Fit a linear relationship of the form weight = α + β × year; find a 95% confidence intervals
for β; and test the hypothesis that β = 0 at the 5% level.

<table>
<thead>
<tr>
<th>Year (x)</th>
<th>1905</th>
<th>1919</th>
<th>1932</th>
<th>1945</th>
<th>1955</th>
<th>1965</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight (y)</td>
<td>164</td>
<td>163</td>
<td>181</td>
<td>192</td>
<td>194</td>
<td>199</td>
</tr>
</tbody>
</table>

Ans: .6661 ± .2604; reject.

10. In the previous problem, provide a 95% prediction interval for the average weight in 1940.

Ans: 184.28 ± 15.40

11. If X₁, ···, Xₙ are independent with common density

\[ f_θ(x) = \begin{cases} 
  e^{-(x-θ)} & \text{if } x > θ \\
  0 & \text{if otherwise}, 
\end{cases} \]

where θ > 0 is unknown, then the maximum likelihood estimator is \( \hat{θ} = \min\{X₁, ···, Xₙ\} \), the smallest of
\( X₁, ···, Xₙ \). Show that \( \hat{θ} \) is a biased estimator of θ and find an unbiased estimator \( \tilde{θ} \) which has a smaller
mean squared error.

Ans: \( \tilde{θ} = \hat{θ} - 1/ₙ \).

12. A large population consists of 50% men and 50% women. Two experimenters are interested in the
average weight of people in the population. The first experimenter takes a simple random sample of size
n = 200 from the population; the second takes a stratified sample of 100 men and 100 women. Suggest
unbiased estimators of the population mean for both sampling experiments. Which is more efficient and
why? Compare the variances of the two estimators when the average weight of women is 120, the average
weight of men is 160, and the strata variances are 400 for both men and women.

Ans: In the first case, use the sample mean; in the second, use the average of the sample mean for men
and the sample men for women. Stratified sampling is more efficient, because the variance of the estimator
is smaller; in the special case, the variances are 4 and 2.