Math 322  Probability & Statistics II  
Spring 2013  
Final Exam

1. The reading on a voltage meter connected to a test circuit is uniformly distributed over the interval \([0, \theta + 1]\), where \(\theta\) is the true but unknown voltage of the circuit. Suppose that \(X_1, X_2, \ldots, X_n\) denotes a random sample of such readings. Show that \(\bar{X} = \frac{\sum X_i}{n}\) is a biased estimator of \(\theta\), and compute the bias. Find a function of \(\bar{X}\) that is an unbiased estimator of \(\theta\). [5]

Recall: If \(X \sim U[a, b]\) then \(E(X) = (a + b)/2\).

Here \(E(X) = \theta + \frac{1}{2}\); hence \(E(\bar{X}) = \mu = \theta + \frac{1}{2}\).

The bias is \(E(\bar{X}) - \theta = \frac{1}{2}\).

If \(\theta = e - \frac{1}{2}\) then \(E(\hat{\theta}) = \theta\).

2. Suppose \(X_1, X_2, X_3\) are i.i.d. \(N(\mu, \sigma = 4)\). Define two estimators for \(\mu\):

\[ T_1 = \frac{X_1 + X_2}{2} \quad \text{and} \quad T_2 = \frac{4X_1 + 3X_2 + X_3}{8}. \]

Both \(T_1\) and \(T_2\) is unbiased estimator for \(\mu\). Which estimator you prefer and why? [5]

\[ \text{Var}(T_1) = \text{Var} \left( \frac{X_1 + X_2}{2} \right) = \frac{1}{4} \left( \text{Var}(X_1) + \text{Var}(X_2) \right) = \frac{4^2 + 4^2}{4} = \frac{32}{4} = 8. \]

\[ \text{Var}(T_2) = \text{Var} \left( \frac{4X_1 + 3X_2 + X_3}{8} \right) \]

\[ = \frac{1}{8^2} \left[ 4^2 \text{Var}(X_1) + 3^2 \text{Var}(X_2) + \text{Var}(X_3) \right] \]

\[ = \frac{1}{64} \left[ (4^2)(4^2) + (3^2)(4^2) + (4^2) \right] \]

\[ = \frac{1}{64} \left[ 256 + 144 + 16 \right] = 6.5 \text{ min}. \]
3. The following is an observed sample from a uniform distribution on the interval [-\(\theta\), \(\theta\)], where \(\theta > 0\) is unknown: -6.9, 2.8, 3.4, 6.4, 6.7, 8.0
Find the method of moment estimate of \(\theta\) (Hint: The first moment is zero, which does not help, so you need to proceed to the second moment.)(5)

\[
\frac{1}{n} \sum_{i=1}^{n} X_i^2 = \mathbb{E}(X^2)
\]

\[
\mathbb{E}(X^2) = \int_{-\theta}^{\theta} x^2 \frac{1}{2\theta} \, dx = \frac{\theta^2}{3}
\]

Hence \(\frac{1}{n} \sum_{i=1}^{n} X_i^2 = \frac{\theta^2}{3} \Rightarrow \theta = \sqrt{\frac{3}{n} \sum_{i=1}^{n} X_i^2} = \sqrt{\frac{3}{6} \times 2.17} = 10.4
\]

4. Let \(X_1, X_2, \ldots, X_n\) be a random sample from the distribution

\[
f_X(x, \sigma) = \begin{cases} \frac{1}{2\sigma} e^{-\frac{|x|}{\sigma}}, & -\infty < x < \infty. \text{ where } \sigma \text{ is an unknown parameter.} \\ 0 & \text{otherwise} \end{cases}
\]

a. Find the maximum likelihood estimator (MLE) \(\hat{\sigma}\) of \(\sigma\). [10]

The likelihood function

\[
L(x; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \ldots \cdot f(x_n; \theta)
\]

\[
= \frac{1}{2^n \sigma^n} e^{\sum |x_i| / \sigma}
\]

The log-likelihood function

\[
\ln(L) = -n \ln(2) - n \ln(\sigma) - \sum \frac{|x_i|}{\sigma}
\]

\[
\frac{d \ln(L)}{d \sigma} = 0 \Rightarrow -n + \sum \frac{|x_i|}{\sigma^2} = 0 \Rightarrow \sigma = \frac{\sum |x_i|}{n}
\]

b. Show that \(\hat{\sigma}\) is an unbiased estimator. [5]

\[
\mathbb{E}(\hat{\sigma}) = \mathbb{E} \left( \frac{\sum |X_i|}{n} \right)
\]

\[
= \frac{1}{n} \left[ \mathbb{E}(|X_1|) + \mathbb{E}(|X_2|) + \ldots + \mathbb{E}(|X_n|) \right]
\]

\[
\mathbb{E}(|X_1|) = \frac{1}{2\sigma} \int_{-\sigma}^{\sigma} x_1 e^{-\frac{x_1}{\sigma}} \, dx_1 = \sigma
\]

Hence \(\mathbb{E}(\hat{\sigma}) = \frac{\sum |X_i|}{n} = \sigma\)
5. Suppose $X_1, \ldots, X_6$ are i.i.d. Pois($\lambda$), and we observe the sample mean $\bar{X} = 2.0$
What is the maximum likelihood estimate of $P(X \leq 2)$? [5]

The likelihood estimator of $\lambda$ is

$$L(x_1, x_2, \ldots, x_6; \lambda) = \frac{e^{-\lambda} \lambda^{x_1}}{x_1!} \cdots \frac{e^{-\lambda} \lambda^{x_6}}{x_6!}$$

$$\ln L = -n \lambda + \sum_{i=1}^{n} x_i \ln(\lambda) - \ln(n! \lambda^n)$$

$$\frac{d\ln L}{d\lambda} = 0 \rightarrow \frac{\lambda}{\bar{X}} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X}$$

$$\therefore \lambda = 2.0 \rightarrow P(X \leq 2) = \sum_{\lambda=0}^{2} \frac{e^{-\lambda} \lambda^{x}}{x!} = 0.6767$$

6. Suppose that $\mu \in [-1.5, 3.5]$ is a 95% confidence interval for the mean cost incurred by a certain inventory policy. Further suppose that this interval was based on 5 independent normally distributed observations from the underlying inventory system. Unfortunately, the boss has decided that she wants a 90% confidence interval. So what is it?
[Hint: Notice that the confidence interval can also be written as $\mu \in [1.0 \pm 2.5]$] [5]

$$\mu \in \bar{X} \pm t_{0.025, 4} \frac{S}{\sqrt{n}}$$

$$1.0 \pm t_{0.025, 4} \frac{S}{\sqrt{5}} = 2.776$$

Thus $95\%$ C.I for $\mu$ is $2.132$.

$$2.5 = (2.776) \frac{S}{\sqrt{5}} \Rightarrow S = 4.055$$

$$\therefore S = 4.055$$

Thus $95\%$ C.I for $\mu$ is $2.132$.

$$1.0 \pm t_{0.025, 4} \sqrt{\frac{4.055}{5}} = 1.0 \pm 1.92$$

7. Suppose $X_1, \ldots, X_9$ are i.i.d. normal with unknown mean and known variance $\sigma^2 = 49$.
Further suppose that $\bar{X} = 50$ and $S^2 = 80$. Find a 95% two-sided confidence interval for $2\mu - 4$. [5]

$95\%$ C.I for $\mu$:

$$\bar{X} - Z_{0.025} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{0.025} \frac{\sigma}{\sqrt{n}}$$

$$50 - 1.96 \frac{\sqrt{49}}{\sqrt{9}} \leq \mu \leq 50 + 1.96 \frac{\sqrt{49}}{\sqrt{9}}$$

$$45.43 \leq \mu \leq 54.57$$

$$\Rightarrow 86.86 \leq 2\mu - 4 \leq 105.14$$
8. Consider i.i.d. normal observations $X_1, \ldots, X_5$ with unknown mean $\mu$ and unknown variance $\sigma^2$. What is the expected length of the usual 90% two-sided confidence interval for $\mu$. You can keep your answer in terms of $\sigma$.

[Hint: Find length of the interval first and then find the expected length.] [5]

$$\frac{(n-1)S^2}{\chi^2_{1-\frac{1}{2},n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{0.05,4}} \Rightarrow \frac{4S^2}{\chi^2_{0.05,4}} \leq \sigma^2 \leq \frac{4S^2}{\chi^2_{0.05,4}}$$

$$\text{Length} = \frac{4S^2}{\chi^2_{0.05,4}} - \frac{4S^2}{\chi^2_{0.05,4}} = \frac{4S^2}{0.711} - \frac{4S^2}{9.488} = 5.20S^2$$

$$\mathbb{E}(\text{Length}) = \mathbb{E}(5.20S^2) = 5.20 \mathbb{E}(S^2) = 5.20 \cdot \sigma^2$$

9. A survey estimated that 20% of all Americans aged 16 to 20 drove under the influence of drugs or alcohol. A similar survey is planned for New Zealand. They want a 95% confidence interval to have a margin of error of 0.04. Find the necessary sample size if they expect to find results similar to those in the United States. [5]

$$\beta = 0.2$$
$$\alpha = 0.8$$
$$ME = 0.04$$
$$Z = 2 \times ME = 0.08$$
$$Z_{0.025} = 1.96$$

$$n = \frac{\left(\frac{4}{(1.96)^2}\right) \left(0.2 \times 0.8\right)}{(0.08)^2}$$

$$= 384.16$$

$$\approx 385$$

10. A tax assessor wants to assess the mean property tax bill for all homeowners in Madison, Wisconsin. A survey ten years ago got a sample mean and standard deviation of $1400 and $1000. How many tax records should be sampled for a 95% confidence interval to have a margin of error of $100? [5]

$$\bar{X} = 1400$$
$$S = 1000$$
$$Z = 2 \times ME = 200$$
$$Z_{0.025} = Z_{0.025} = 1.96$$

$$n = \left[\frac{2 \times (1.96)(1000)}{200}\right]^2$$

$$= 384.16$$

$$\approx 385$$
11. A mixture of pulverized fuel ash and Portland cement to be used in grouting should have a comprehensive strength of more than 1300 KN/m². The mixture will not be used unless experimental evidence indicates conclusively that the strength specification has been met. Suppose compressive strength for specimens of this mixture is normally distributed with $\sigma = 60$. Let $\mu$ denote the true average compressive strength.

a. What are the appropriate null and alternative hypotheses?

$$H_0 : \mu = 1300$$
$$H_a : \mu > 1300$$

b. Let $\bar{X}$ denote the sample average compressive strength for $n = 20$ randomly selected specimens. Consider the test procedure with test statistic $\bar{X}$ and rejection region $\{ \bar{X} \geq 1331.26 \}$. What is the probability of type I error for this procedure? [5]

$$\alpha = P\{ \text{Type I error} \}$$
$$= P\{ \text{Reject } H_0 \text{ when } H_0 \text{ is true} \}$$
$$= P\{ \bar{X} \geq 1331.26, \text{ when } \mu = 1300 \}$$
$$= P\{ Z \geq \frac{1331.26 - 1300}{60/\sqrt{20}} \} = 1 - \Phi(2.33)$$
$$\approx 1 - 0.9901 = 0.0099$$

c. In terms of this problem, if the actual comprehensive strength is 1280 KN/m², what is the probability of type II error? [5]

$$\beta = P\{ \text{Type II error} \}$$
$$= P\{ \text{fail to reject } H_0 \text{ when } H_0 \text{ is false} \}$$
$$\beta(1280) = P\{ \bar{X} \leq 1331.26, \text{ when } \mu = 1280 \}$$
$$= P\{ Z \leq 3.82 \} \approx 1$$

12. A sample of 12 radon detectors of a certain type was selected, and each was exposed to 100 pCi/L of radon. The resulting readings were as follows:

<table>
<thead>
<tr>
<th>104.3</th>
<th>89.6</th>
<th>89.9</th>
<th>95.6</th>
<th>95.2</th>
<th>90.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>98.8</td>
<td>103.7</td>
<td>98.3</td>
<td>106.4</td>
<td>102.0</td>
<td>91.1</td>
</tr>
</tbody>
</table>

Suppose that prior to the experiment, a value of $\sigma = 7.5$ had been assumed. How many determinations would then have been appropriate to obtain $\beta = 0.10$ for the alternative $\mu = 95$? [5]

$$n = \left[ \frac{\sigma [Z_{\alpha} + Z_{\beta}]}{\mu - \mu_0} \right]^2$$
$$= \left[ \frac{7.5 \cdot (1.645 + 1.28)}{100 - 95} \right]^2$$
$$= 19.25 \approx 20$$
13. A college is trying a new student registration system and would like to know if there is sufficient evidence to conclude that at least 60% of the students favor the new system. 

**Hypothesis**

\( H_0: p \leq 0.6 \)  \( \text{at most 60}\% \text{ of the students prefer the new system} \)

\( H_a: p > 0.6 \)  \( \text{more than 60}\% \text{ of the students prefer the new system} \)

A random sample 15 students, 10 favored the new system. Test the hypothesis at 4\% level of significance. [5]

\[
\alpha = 0.04 = P\{\text{Type I error}\} \\
= P\{\text{reject } H_0 \text{ when } H_0 \text{ is true}\} \\
= P\{ X > c \text{ when } p = 0.6\} \\
= P\{ X > c; X \sim \text{bin}(n=15, p=0.6) \} \\
= 1 - P\{ X \leq c; X \sim \text{bin}(15, 0.6) \} \\
= 1 - F(9,15;0.6)
\]

\( \Rightarrow c_{1-0.04} = 12 \& c = 13 > 10, \text{ we fail to reject } H_0 \)

14. An online search using Orbitz.com has been performed in order to estimate the price (in U.S. Dollars) difference when choosing between renting a car from Budget and Hertz. 

The sample mean rental prices and the sample variances for Budget and Hertz, respectively, are:

\( n_1 = 10, \overline{X}_1 = 82.90, s_1 = 30.36, S_1^2 = 921.73 \)

\( n_2 = 10, \overline{X}_2 = 93.60, s_2 = 29.99, S_2^2 = 899.40 \)

a) Test the claim that the variance of rental prices are different between the two rental companies at \( \alpha = 0.05 \) level of significance. Use both p-value and critical value to test the hypothesis[10]

\( H_0: \sigma_B^2 = \sigma_H^2 \)

\( H_a: \sigma_B^2 \neq \sigma_H^2 \)

\[
f^* = \frac{S_B^2}{S_H^2} = \frac{921.73}{899.40} = 1.0248
\]

Since \( f^* = 1.0248 > 4.03 \) and \( f^* = 1.0248 < 0.248 \)

we fail to reject \( H_0 \).

\[
P\text{-value} = 1 - P(f(1.0248, 9, 9)) = 0.486
\]

Since \( P\text{-value} > 0.486 \), we fail to reject \( H_0 \).
b) Test the claim that the mean rental prices are different between the two rental companies at \( \alpha = 0.05 \) level of significance. Use both p-value and critical value to test the hypothesis [10]

\[
\begin{align*}
H_0: \mu_B - \mu_H &= 0 \\
H_a: \mu_B - \mu_H &\neq 0
\end{align*}
\]

Since \( S_1^2 = S_2^2 \) we use pooled variance = 910.565

\[
T^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{82.9 - 93.6}{\sqrt{910.565 \left( \frac{1}{10} + \frac{1}{10} \right)}} = -0.7929
\]

\[
d.f = n_1 + n_2 - 2 = 18
\]

Reject \( H_0 \) if \( T^* \geq t_{d/2, d.f} \) or \( T^* \leq -t_{d/2, d.f} \)

\[
t_{0.025, 18} = 2.101
\]

\[
|T^*| = | -0.7929 | \leq 2.101
\]

\[
p-value = 2 \times P(T \geq |T^*|)
\]

15. Now take into account that each data point reflects quote prices in Boston, Los Angeles, New York, Las Vegas, Washington, Salt Lake City, Pittsburgh, Miami, Akron and Orlando, respectively. Set up a two-sided paired test, and compute a 95% confidence levels for the parameter of interest. [5+5]

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Budget</td>
<td>116</td>
<td>100</td>
<td>114</td>
<td>55</td>
<td>101</td>
<td>45</td>
<td>117</td>
<td>82</td>
<td>50</td>
<td>49</td>
</tr>
<tr>
<td>Hertz</td>
<td>122</td>
<td>113</td>
<td>126</td>
<td>93</td>
<td>113</td>
<td>52</td>
<td>118</td>
<td>91</td>
<td>51</td>
<td>57</td>
</tr>
</tbody>
</table>

\[
d = \text{Budget} - \text{Hertz}
\]

\[
T^* = \frac{-10.7}{10.48/\sqrt{10}} = -3.23
\]

Reject \( H_0 \) if \( |T^*| \geq t_{d/2, d.f} \)

\[
t_{0.025, 9} = 2.262
\]

\[
|T^*| = 3.23 > 2.262, \text{ we reject } H_0
\]

95% C.I for \( \mu_d \)

\[
\bar{d} \pm t_{d/2, n-1} \sqrt{\frac{S_d^2}{n}} = -10.7 \pm 7.496
\]

\[
-18.1964 \leq \mu_d \leq -3.21
\]

\[
S.E = \frac{S_d}{\sqrt{n}} = \frac{3.314}{\sqrt{10}} = 2.262
\]

\[
ME = (2.262)(5.56) = 12.696
\]
16. On June 25, 1995, The Associated Press reported the results of a national survey conducted by the Center for Social and Religious Research at the Hartford Seminary. The study was on the divorce rate of a group of 5000 Protestant clergymen and 5000 Protestant clergywomen. It was found that 25% out of 2458 clergywomen responding had been divorced at least once and 20% out of 2086 clergymen responding had been divorced at least once. If true \( \pi_1 = .20 \) and \( \pi_2 = .18 \), what sample sizes (m = n) would be necessary to detect such a difference with probability 0.90? [5]

\[
\begin{align*}
\pi &= 0.2 \quad \pi_2 = 0.18, \quad \alpha = 0.1 \quad (\text{two-sided}), \quad \beta = 0.1 \quad Z^{1.645} \\
Z &= \left[ \frac{Z_{\alpha/2} \sqrt{(\pi_1 + \pi_2)(1 - \pi_1 - \pi_2)}}{\sqrt{\pi_1 \pi_2}} \right] \sqrt{n} \quad Z_{\beta} = 1.28 \\
&= \left[ 1.645 \sqrt{(0.2 + 0.18)(0.8 + 0.82)} \right] \sqrt{\frac{1}{0.02} + 1.28 \sqrt{(0.2)(0.8) + (0.18)(0.82)}} \\
&= 6582
\end{align*}
\]

17. The following partial ANOVA table is taken from a study in which the abilities of three different groups to identify a perceptual incongruity were assessed and compared. All individuals in the experiment had been hospitalized to undergo psychiatric treatment. There were 21 individuals in the depressive group, 32 individuals in the functional “other” group, and 21 individuals in the brain-damaged group. Complete the ANOVA table and carry out the F test at level \( \alpha = 0.1 \) [10]

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups</td>
<td>2</td>
<td>150</td>
<td>75</td>
<td>5.357</td>
</tr>
<tr>
<td>Error</td>
<td>71</td>
<td>994</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>73</td>
<td>1144</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\text{P-value} &= P(F \geq 5.357) \\
&= 1 - Pf(5.357, 2, 71) \\
&= 0.0068 \\
\text{Since p-value < 0.05, we reject } H_0
\end{align*}
\]