Final - Winter 2015 - Math 462

NAME: ______________________________

INSTRUCTIONS: You must show all your work to obtain credit for any answer. If your handwriting is not legible, your work may not be possible to grade. No calculators, smartphones, or other electronic devices may be used during this exam, nor are written notes of any kind.

Problem	Points Possible	Points Earned
1	15	
2	15	
3	10	
4	10	
5	10	
6	10	
TOTAL	70	

Problem 1 (15 points). Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. from the Normal$(0, \theta)$ pdf

$$f(x;\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2\theta}x^2}$$

Note: $E[X_1^4] = 3\theta^2$.

(a) Find a sufficient statistic for θ.

(b) Find the MLE of θ.

(c) Find the method of moments estimator of θ.

(d) Is the MLE unbiased? (Justify.) If not, modify it to obtain an unbiased estimator.

(e) Compute the Cramer-Rao Lower bound for the variance of unbiased estimators of θ. Calculate the variance for the unbiased estimator found above and compare to this bound.

Problem 2 (15 points)**.** Let $X_1, X_2, \ldots, X_n$ be i.i.d. with pdf

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \le x \le \theta, \\ 0 & \text{otherwise}. \end{cases}$$

(a) Find a sufficient statistic for θ. Justify your answer.

(b) Find the pdf for $T_n = \max\{X_1, \ldots, X_n\}$.

(c) Calculate $P(\theta - \epsilon < T_n \le \theta)$, and then show that T_n is a consistent estimator of θ

(d) Suppose that the prior distribution on θ (with known parameters α and b) has pdf

$$f(\theta) = \begin{cases} \frac{\alpha b^\alpha}{\theta^{\alpha+1}} & \text{if } \theta \ge b \\ 0 & \text{otherwise}. \end{cases}$$

Note: This pdf has mean $\frac{\alpha b}{\alpha - 1}$.

Find the Bayes Estimator for θ assuming squared-error loss.

Hint: Use indicator functions to keep track of where the pdf's are positive. The posterior distribution for θ has the same form as the prior distribution (but with different values replacing parameters α and b).

(e) Is the Bayes estimator consistent? Justify your answer.

Problem 3 (10 points)**.** Suppose that you have two samples from Normal distributions, $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$, respectively. The first sample is of size 4, and the second is of size 8. The sample variances are 2 and $\frac{4}{7}$, respectively.

(a) If the sample means are 3.1 and 4.3, test at level 0.05 the hypothesis that $\mu_X = \mu_Y$ against the two-sided alternative. Assume that the variances of the distributions are the same when carrying out this test.

Use the approximation $\sqrt{\frac{3}{8}} \approx 0.6$.

(b) Test at level 0.01 the hypothesis that $\sigma_X^2 = \sigma_Y^2$ against the alternative that $\sigma_X^2 > \sigma_Y^2$.

(c) Prior knowledge leads you to believe that $\sigma_X^2 = 3$. Test this hypothesis, at level 0.01, against the alternative that $\sigma_X^2 \neq 3$.

Problem 4 (10 points)**.** Suppose that a stock can either go up, stay the same, or go down each day. Denote the probabilities of these three events as p_u, p_s and p_d, respectively.

You have a theory that the probabilities of these three events are $0.4, 0.2, 0.4$, respectively.

You observe the stock on 100 days and find that it goes up, stays the same, and goes down $45, 15$ and 40 days, respectively.

(a) Find the expected number of days out of 100 for which the stock goes up, stays the same, and goes down. Write down the sum, but no need to complete the arithmetic, which equals the chi-squared statistic for testing the hypothesis $H_0 : (p_u, p_s, p_d) = (0.4, 0.2, 0.4)$.

(b) The value of the chi-squared statistic is 1.875. Carry out, at level 0.025, the hypothesis test of $H_0 : (p_u, p_s, p_d) = (0.4, 0.2, 0.4)$.

Problem 5 (10 points)**.** Suppose that the random variable X has the pdf

$$f(x;\theta) = \frac{1}{2}e^{-|x-\theta|}$$

(a) Consider the test $H_0 : \theta = 0$ vs $H_0 : \theta \neq 0$, based on the single observation X, which rejects H_0 if and only if $|X| > 2$. Find the significance level of this test.

(b) Find the power of the above test when $\theta = 1$.

Problem 6 (10 points)**.** Suppose you want to evaluate a new drug for the flu. The manufacturer claims that the drug increases the probability p that a patient recovers in under three days. You give the drug to 100 flu patients and record the number X of patients who recover in under three days. Note that, if you give no treatment, the chance a person will recover from the flu in under three days is 0.5.

(a) Write down, in terms of X, an approximate level 0.05 test of the hypothesis that the treatment is no better than doing nothing (i.e. that $p = 1/2$) versus the alternative that it is better ($p > 1/2$).

(b) Carry out this test if you observe $X = 55$.

(c) Find the approximate power of the test above when $p = 0.6$.

(You can use the approximation $\sqrt{24} \approx 5$.)

Final

Important: Show all work. A correct answer without work and explanation may not receive credit.

NAME: ______________________________

Problem	Points
1	15
2	10
3	10
4	10
5	20
6	10
TOTAL	75

I want the weight of the FINAL exam to equal 50% of my total grade and the following midterm to weigh 0% of my grade (check ONLY one):

MIDTERM I	
MIDTERM II	

Problem 1 (15 points)**.** Suppose that $X_1, \dots, X_n$ are i.i.d. with mass function

$$p_X(x;\theta) = \frac{1}{\theta}\left(1 - \frac{1}{\theta}\right)^{x-1}, \quad x = 1, 2, \dots$$

The parameter θ satisfies $\theta \geq 1$. Note that $E(X_1) = \theta$ and $\text{Var}(X_1) = \theta(\theta - 1)$.

(a) Find a sufficient statistic for θ.

(b) Find the MLE of θ and show that it is an unbiased estimator of θ.

(c) Show the MLE above has minimum variance among all unbiased estimates.

Note: Take care when differentiating. For example,

$$\frac{d}{d\theta}\left[\log\left(1 - \frac{1}{\theta}\right)\right] = \frac{1}{\theta(\theta-1)}, \qquad \frac{d}{d\theta}\left[\frac{1}{\theta(\theta-1)}\right] = -\frac{2\theta-1}{\theta^2(\theta-1)^2}$$

Solution. Because the likelihood function is

$$L(\theta; x_1, \dots, x_n) = \theta^{-n}(1-\theta^{-1})^{\sum x_i - n},$$

and only depends on $(x_1, \dots, x_n)$ through $\sum x_i$, the statistic $\sum x_i$ is sufficient for θ.

To find the MLE:

$$\ell(\theta) = -n\log\theta + (s-n)\log(1-\theta^{-1})$$

$$\ell'(\theta) = -\frac{n}{\theta} + \frac{s-n}{(1-\theta^{-1})\theta^2} = \frac{(s-n) - n(\theta-1)}{\theta(\theta-1)} = \frac{s - \theta n}{\theta(\theta-1)}$$

Setting $\ell'(\theta) = 0$ and solving for θ shows $\hat{\theta} = s/n = \bar{X}$.

The Fisher Information is found by calculating as follows:

$$\ell''(\theta; X_1) = \frac{1}{\theta^2} - (X_1 - 1)\frac{1}{(\theta(\theta-1))^2}[2\theta - 1]$$

$$E[\ell''(\theta; X_1)] = \frac{1}{\theta^2} - \frac{(\theta-1)(2\theta-1)}{\theta^2(\theta-1)^2} = \frac{(\theta-1)-(2\theta-1)}{\theta^2(\theta-1)} = -\frac{1}{\theta(\theta-1)}$$

$$I(\theta) = \frac{1}{\theta(\theta-1)}$$

Since

$$\text{Var}(\bar{X}) = \frac{\text{Var}(X_1)}{n} = \frac{\theta(\theta-1)}{n} = \frac{1}{nI(\theta)},$$

we conclude that $\text{Var}(\bar{X})$ attains the Cramer-Rao lower bound, whence $\bar{X}$ is minimum variance among all unbiased estimators. □

Problem 2 (10 points)**.** Suppose that there are four colors of tulips planted in a field. Below are the totals in a random sample of size 60

color	red	yellow	orange	pink
count	16	11	14	19

Test, at level 0.05, the hypothesis that the four colors occur with equal proportions in the entire field.

Solution. The expected counts are

color	red	yellow	orange	pink
count	16	11	14	19
expected	15	15	15	15

Thus

$$\chi^2 = \frac{(16-15)^2}{15} + \frac{(11-15)^2}{15} + \frac{(14-15)^2}{15} + \frac{(19-15)^2}{15} = 2.267$$

The critical value for a chi-squared with $4-1=3$ degrees of freedom is 7.82, so we do not reject H_0. □

Problem 3 (10 points)**.** Suppose that $X_1,\dots,X_{10}$ are i.i.d. $N(\mu_x,\sigma_x^2)$ and $Y_1,\dots,Y_8$ are i.i.d. $N(\mu_y,\sigma_y^2)$. Suppose that $S_x^2 = 11.5$ and $S_y^2 = 8.4$, and $\bar{X} = 20.5$ and $\bar{Y} = 16.2$.

(a) Test $H_0 : \sigma_x^2 = \sigma_y^2$ vs. $H_1 : \sigma_x^2 \neq \sigma_y^2$ at level 0.01.

(b) If the result of the previous test is not significant, assume that $\sigma_x = \sigma_y$; otherwise assume that $\sigma_x \neq \sigma_y$. Now test $H_0 : \mu_x = \mu_y$ at level 0.01.

Solution. The F statistic (with degrees of freedom $n = 9$ and $m = 7$) equals

$$F = \frac{S_x^2}{S_y^2} = \frac{11.5}{8.4} = 1.369$$

The 99.5-th percentile of the $F_{9,7}$ distribution equals 8.51. The 99.5-th percentile of the $F_{7,9}$ distribution equals 6.88, so the 0.5-th percentile of the $F_{9,7}$ distributions equals $1/6.88 = 0.15$.

Since $0.15 < 1.369 < 8.51$, we do not reject H_0.

We use the pooled estimate of σ^2 :

$$\begin{aligned} s_p^2 &= \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2} = \frac{9 \cdot 11.5 + 7 \cdot 8.4}{16} = 10.14 \\ s_p &= 3.185 \end{aligned}$$

Then

$$T = \frac{20.5 - 16.2}{3.185\sqrt{\frac{1}{10} + \frac{1}{8}}} = 2.846$$

The critical value of the T-distribution with 16 degrees of freedom is $t^\star = 2.92$, whence we do not reject H_0.

$\square$

Problem 4 (10 points)**.** Suppose that $X_1, \ldots, X_{16}$ are i.i.d. $N(\mu, 1)$ random variables.

(a) (3 points) Consider testing at level 0.01 the hypothesis $H_0 : \mu = 0$ vs $H_1 : \mu > 0$. The usual test rejects if $\bar{X}$ is large. Explicitly specify the rejection region.

(b) (7 points) Compute the power of this test when $\mu = 0.1$.

Solution. The test rejects if and only if

$$Z = \frac{\bar{X} - 0}{1/\sqrt{16}} = 4\bar{X} > 2.326\,.$$

That is, the test rejects iff

$$\bar{X} > \frac{2.326}{4} = 0.5816\,.$$

Thus

$$\begin{aligned} P(\bar{X} > 0.5816 \mid \mu = 0.1) &= P\left(4(\bar{X} - 0.1) > 4 \cdot (0.5816 - 0.1)\right) \\ &= P(Z > 1.9264) = 0.027\,. \end{aligned}$$

$\square$

Problem 5 (20 points)**.** Let $X_1, \ldots, X_n$ be i.i.d. with pdf

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}, \quad x \geq 0\,.$$

Note that $E(X_1) = \theta$.

(a) (4 points) Find the MLE of θ, and show that is it is consistent.

(b) (4 points) Suppose $n = 5$. The distribution of $S = \sum_{i=1}^{5} X_i$ has density

$$f(s;\theta) = \frac{1}{24\theta^5} s^4 e^{-s/\theta}, \quad s \geq 0\,.$$

Find the density of S/θ; your answer should not depend on θ.

(c) (4 points) Below is a table for certain values of the cdf of S/θ:

$p = P(S/\theta \leq x)$	0.005	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990	0.995
x	0.6722	0.8232	1.0899	1.3663	1.7448	6.6808	7.7537	8.7673	10.0451	10.9775

Use this table to find a 99% confidence interval for θ based on S.

(d) (4 points) Consider the hypotheses $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$. Give an explicit test at level 0.01 using the test statistic S/θ_0.

(e) (4 points) Use the table to give upper and lower bounds on the power of the test in (d) when $\theta = 1.5$.

Solution. If $s = \sum_{i=1}^{n} x_i$, then

$$\begin{aligned} L(\theta; x_1, \ldots, x_n) &= \theta^{-n} e^{-s/\theta} \\ \ell(\theta) &= -n \log\theta - \frac{s}{\theta} \\ \ell'(\theta) &= -\frac{n}{\theta} + \frac{s}{\theta^2}\,. \end{aligned}$$

Setting the derivative to zero and solving shows that $\hat{\theta} = \frac{s}{n} = \bar{x}$.

Thus, the MLE is $\bar{X}$. Since the Law of Large Numbers says that

$$\bar{X} \xrightarrow{Pr} E(X_1) = \theta,$$

the estimator $\bar{X}$ is consistent.

If $T = S/\theta$, then $s = t\theta$, so

$$f_T(t) = f_S(s)\frac{ds}{dt} = \frac{1}{24\theta^5}(t\theta)^4 e^{-t}\theta = \frac{1}{24}t^4 e^{-t}.$$

We have

$$\begin{aligned} 0.01 &= P\left(0.6722 \le \frac{S}{\theta} \le 10.9775\right) \\ &= P\left(\frac{S}{10.9775} \le \theta \le \frac{S}{0.6722}\right), \end{aligned}$$

so $[S/10.9775,\ S/0.6722]$ is a 99% confidence interval.

We would reject if $S > 10.0451$. The power equals

$$\begin{aligned} 1-\beta &= P(S > 10.0451 \mid \theta = 1.5) \\ &= P\left(\frac{S}{1.5} > \frac{10.0451}{1.5} \,\middle|\, \theta = 1.5\right) \\ &= P(S/\theta > 6.70). \end{aligned}$$

From the table,

$$\begin{aligned} P(S/\theta > 6.68) > P(S/\theta > 6.70) > P(S/\theta > 7.75) \\ 1 - 0.9 > 1-\beta > 1 - 0.05 \end{aligned}$$

□

Problem 6 (10 points)**.** Suppose that $X_1, \dots, X_n$ are i.i.d. Normal with mean 0 and variance σ^2. Find the GLRT of $H_0 : \sigma = \sigma_0$ against $H_1 : \sigma \neq \sigma_0$. If $n = 8$ and $\sigma_0 = 2$, use a table to give an explicit critical region for the level 0.01 test.

Solution. Let $s = \sum x_i^2$. Then

$$\log L(\sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{s}{2\sigma^2} - \frac{n\log(2\pi)}{2}$$
$$\frac{d}{d\sigma^2}\log L(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{s}{2(\sigma^2)^2}$$

Thus there is a critical point at $\sigma^2 = s/n$, and in fact a global maximum, and $\hat{\sigma^2} = s/n$.

The likelihood ratio is

$$\Lambda = \frac{f(x_1, \dots, x_n; \sigma_0^2)}{f(x_1, \dots, x_n; \hat{\sigma^2})} = \frac{(\sigma_0^2)^{-n/2} e^{-\frac{s}{2\sigma_0^2}}}{(s/n)^{-n/2} e^{-\frac{n}{2s}s}} = s^{n/2} e^{-\frac{s}{2\sigma_0^2}} n^{-n/2} e^{-n/2} \sigma_0^{-n}$$

Thus Λ is small if s is large or small.

Observe that S/σ_0^2 has a chi-squared distribution with 8 degrees of freedom. The 0.5-th and 99.5-th percentiles of the chi-squared with 8 degrees of freedom are 1.34 and 21.95, respectively. So we reject if $S < 4(1.34) = 5.36$ or if $S > 4(21.95) = 87.8$.

□