Name________________________________

IE 4521 – Midterm #2

Prof. John Gunnar Carlsson

April 20, 2010

Before you begin: This exam has 11 numbered pages and a total of 8 problems. Make sure that all pages are present. To obtain credit for a problem, you must show all your work. if you use a formula to answer a problem, write the formula down. Do not open this exam until instructed to do so.

1. (12 points) Consider the probability density function

$$f(x) = \begin{cases} c(1+\theta x) & -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Compute the value of the constant c (since $f(x)$ must define a valid p.d.f.)

Solution We must have

$$\begin{aligned} \int_{-1}^{1} c(1+\theta x)\,dx &= 1 \\ c\left[x+\theta\frac{x^2}{2}\right]_{-1}^{1} &= 1 \\ c &= 1/2 \end{aligned}$$

(b) Use the method of moments to estimate θ.

Solution We have

$$\begin{aligned} E(X) &= \int_{-1}^{1} \frac{1}{2}(1+\theta x)\,x\,dx \\ &= \theta/3 \end{aligned}$$

and therefore $\theta = 3E(X)$, so $\hat{\theta} = 3\bar{x}$ is the method-of-moments estimate.

(c) Show that $\hat{\theta} = 3\bar{x}$ is an unbiased estimator for θ.

Solution The bias of $\hat{\theta}$ is

$$E\left(\hat{\theta}\right) - \theta = E\left(3\bar{X}\right) - \theta = 3E\left(\bar{X}\right) - \theta = 3E(X) - \theta$$

and since $\bar{x}$ is always an unbiased estimator, the bias is zero.

(d) Suppose we have two samples x_1 and x_2. Find the maximum likelihood estimator for θ. For the time being, do not worry about any additional constraints on θ that may be necessary.

Solution We have

$$\begin{aligned} \mathcal{L}(x_1,x_2;\theta) &= \frac{1}{2}(1+\theta x_1)\cdot\frac{1}{2}(1+\theta x_2) \\ \ell(x_1,x_2;\theta) = \log\mathcal{L}(x_1,x_2;\theta) &= \log(1/4) + \log(1+\theta x_1) + \log(1+\theta x_2) \\ \frac{\partial \ell}{\partial \theta} &= \frac{x_1}{1+\theta x_1} + \frac{x_2}{1+\theta x_2} = 0 \end{aligned}$$

solving this we find that $\hat{\theta} = \frac{-x_1-x_2}{2x_1x_2}$.

(e) (Bonus; 2 points) Suppose we sample $x_1 = 0.2$, $x_2 = 0.8$, $x_3 = 0.3$. What is wrong with the moment estimator in (c)?

(f) (Bonus; 2 points) Other than having $x_1 = 0$ or $x_2 = 0$, what else could go wrong in (d), and how might you resolve it?

2. (12 points) Consider the probability density function

$$f(x) = \begin{cases} \frac{1}{\theta^2} x e^{-x/\theta} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Find the maximum likelihood estimator of θ, given a collection of samples $x_1, \ldots, x_n$.

Solution We have

$$\begin{aligned} \mathcal{L}(x_1, \ldots, x_n; \theta) &= \left(\frac{1}{\theta^2} x_1 e^{-x_1/\theta}\right) \cdots \left(\frac{1}{\theta^2} x_n e^{-x_n/\theta}\right) \\ \ell(x_1, \ldots, x_n; \theta) &= -2n \log \theta + (\log x_1 + \cdots + \log x_n) - \frac{x_1 + \cdots + x_n}{\theta} \\ \frac{\partial \ell}{\partial \theta} &= \frac{-2n}{\theta} + \frac{x_1 + \cdots + x_n}{\theta^2} = 0 \end{aligned}$$

Solving for θ we find that $\hat{\theta} = \frac{x_1 + \cdots + x_n}{2n}$.

3. (12 points) The *Rayleigh distribution* has a probability density function given by

$$f(x) = \begin{cases} \frac{x}{\theta} e^{-x^2/(2\theta)} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the maximum likelihood estimator of θ, given a collection of samples $x_1, \ldots, x_n$.

Solution We have

$$\begin{aligned} \mathcal{L}(x_1, \ldots, x_n; \theta) &= \left[\frac{x_1}{\theta} e^{-x_1^2/(2\theta)}\right] \cdots \left[\frac{x_n}{\theta} e^{-x_n^2/(2\theta)}\right] \\ \ell(x_1, \ldots, x_n; \theta) &= (\log x_1 + \cdots + \log x_n) - n \log \theta - \frac{x_1^2 + \cdots + x_n^2}{2\theta} \\ \frac{\partial \ell}{\partial \theta} &= \left(\frac{x_1^2 + \cdots + x_n^2}{2\theta^2}\right) - \frac{n}{\theta} = 0 \end{aligned}$$

and therefore $\hat{\theta} = \frac{x_1^2 + \cdots + x_n^2}{2n}$.

(b) It can be shown that $E(X^2) = 2\theta$. Use this information to construct an unbiased estimator for θ.

Solution Let $Y = X^2$, so that we have samples $y_1 = x_1^2, \ldots, y_n = x_n^2$. We know that $E(Y) = 2\theta$, so $\theta = E(Y)/2$. The estimator $\bar{y} = \frac{y_1 + \cdots + y_n}{n}$ is an unbiased estimator of $E(Y)$, so an unbiased estimator of θ is $\hat{\theta} = \bar{y}/2 = \frac{x_1^2 + \cdots + x_n^2}{2n}$.

4. (12 points) Consider a sequence of random samples $x_1, \ldots, x_n$, with each $X_i \sim \mathcal{U}(0, \theta)$. In a previous problem set we showed that the estimator $\hat{\theta}_1 := \frac{n+1}{n} \max_i \{x_1, \ldots, x_n\}$ is unbiased.

   (a) Argue intuitively why $\hat{\theta}_{BAD} := \max_i \{x_1, \ldots, x_n\}$ must be a biased estimator.

   Solution For each measurement x_i we have $x_i \leq \theta$, and hence $\max_i \{x_1, \ldots, x_n\} \leq \theta$.

   (b) Show that the method-of-moments estimator $\hat{\theta}_2$ of θ is unbiased.

   Solution For uniformly distributed random variables we have $E(X) = \theta/2$, so the method-of-moments estimator is $\hat{\theta}_2 = 2\bar{x}$. The bias is

$$\begin{aligned} E\left(\hat{\theta}_2 - \theta\right) &= E(2\bar{x}) - \theta \\ &= 2E(\bar{x}) - \theta \\ &= 2(\theta/2) - \theta = 0 \end{aligned}$$

   (c) It can be shown that $\operatorname{Var}\left(\hat{\theta}_1\right) = \frac{\theta^2}{n(n+2)}$ and $\operatorname{Var}\left(\hat{\theta}_2\right) = \frac{\theta^2}{3n}$. Show that if $n > 1$, then $\hat{\theta}_1$ is a better estimator than $\hat{\theta}_2$. In what sense is it better?

   Solution Since both estimators are unbiased, the mean squared errors of each estimator is its variance. If $n > 1$, then clearly $\operatorname{Var}\left(\hat{\theta}_1\right) < \operatorname{Var}\left(\hat{\theta}_2\right)$, so $\hat{\theta}_1$ is a better estimator in that it has a smaller mean squared error.

5. (12 points) The compressive strength of concrete is being tested by a civil engineer. He tests 12 specimens and obtains the following data:

2216	2237	2249	2204
2225	2301	2281	2263
2318	2255	2275	2295

(a) Construct a 95% two-sided confidence interval for the mean strength, and use this to test the hypothesis $H_0 : \mu = 2280$.

Solution The interval is $\mu \in (2237, 2282)$, so we accept the null hypothesis.

(b) Construct an appropriate 95% one-sided confidence interval and use this to test the hypothesis $H_0 : \mu \geq 2280$.

Solution The interval is $\mu \in (-\infty, 2278)$. We reject the null hypothesis.

(c) Perform the corresponding t-tests from (a) and (b). You do not need to compute the p-value explicitly; giving bounds will suffice.

Solution The t-statistic in both cases is $t_0 = -1.96$; we find that $|t_0| < 2.201 = t_{0.025,11}$ and $-t_0 > 1.796 = t_{0.05,11}$, so we accept and reject the null hypotheses of (a) and (b) respectively.

6. (13 points) In semiconductor manufacturing, wet chemical etching is often used to remove silicon from the backs of wafers prior to metallization. The etch rate is an important characteristic in this process. Two different etching solutions have been compared, using two random samples of 10 wafers for each solution. The observed etch rates are as follows:

Solution 1		Solution 2	
9.9	10.6	10.2	10.0
9.4	10.3	10.6	10.2
9.3	10.0	10.7	10.7
9.6	10.3	10.4	10.4
10.2	10.1	10.5	10.3

(a) Use a t-test to test the claim that the mean etch rate is the same for both solutions. Use significance level $\alpha = 0.05$.

Solution We can use the method of paired samples. We find that $\bar{z} = \bar{x} - \bar{y} = -0.43$ and consequently the t-statistic is $t_0 = -2.25$. Since $|t_0| < 2.262 = t_{0.025,9}$, we should accept the null hypothesis.

(b) Use a confidence interval to test the claim from (a).

Solution The interval is $(-0.8619, 0.0019)$, so we accept the null hypothesis.

7. (13 points) An article in *IEEE International Symposium on Electromagnetic Compatibility* quantified the absorption of electromagnetic energy and the resulting thermal effect from cellular phones. the experimental results were obtained from *in vivo* experiments conducted on rats. The arterial blood pressure values (mmHg) for the control group (8 rats) during the experiment were $\bar{x} = 90$, $S_x^2 = 25$ and for the test group (9 rats) were $\bar{y} = 115$, $S_y^2 = 100$.

(a) Test the hypothesis that the test group has a higher blood pressure than the control group with $\alpha = 0.05$. Use a t-test.

Solution We use the method of unpaired samples. The t-statistic is

$$t_0 = \frac{\bar{x} - \bar{y} - 0}{\sqrt{S_x^2/n + S_y^2/m}} = -6.62$$

So we accept the null hypothesis, since $t_0 < 1.895 = t_{0.05,7}$.

(b) Calculate a one-sided confidence interval (with $-\infty$ on the left) for $\mu_A - \mu_B$ at the 5% significance level.

Solution The confidence interval is

$$(-\infty, -17.85)$$

(c) Do the data support the claim that the mean blood pressure from the test group is at least 15 mmHg higher than the control group?

Solution Yes. The above confidence interval can be used to test the hypothesis $H_0 : \mu_A - \mu_B \geq -15$. Since -15 lies outside this interval, the data suggest that it is implausible to say that $H_0 : \mu_A - \mu_B \geq -15$.

8. (14 points) Multiple choice. Circle the correct answer. You may justify an answer if you like, but it is not required.

   (a) Which of the following statements is correct about maximum likelihood estimation?

      i. The maximum likelihood estimates of the parameters μ and σ^2 for random variables following a log-normal distribution are $\hat{\mu} = \bar{x}$ and $\hat{\sigma}^2 = S^2$.
      ii. Maximum likelihood estimation should only be performed if are only estimating one parameter.
      iii. The maximum likelihood estimator $\hat{\sigma}^2$ of a set of normally distributed random variables is biased.
      iv. The reason that we typically take a logarithm of the likelihood function $L(x_1, \ldots, x_n; \theta)$ in making a maximum likelihood estimate to ensure that $L(x_1, \ldots, x_n; \theta)$ is positive.
      v. It makes no sense to take a maximum likelihood estimate of a parameter for a discrete random variable, since a discrete p.m.f. $f(x)$ is not differentiable with respect to x.

      Solution: iii

   (b) The Gallup Organization is conducting a poll to see the average number of computers in an American household. The organization decides to increase the size of its random sample of voters from 1500 people to 4000 people. The effect of this increase is to:

      i. reduce the bias of the estimator $\bar{x}$ of μ.
      ii. increase the standard error of the estimator.
      iii. reduce the variability of the estimator.
      iv. increase the width of the confidence interval for the parameter.
      v. have no effect because the population size is the same.

      Solution iii

   (c) Which of the following statements is *incorrect*?

      i. The standard error of the sample mean $\bar{x}$ will decrease as the number of samples increases.
      ii. The standard error of the sample mean $\bar{x}$ is a measure of the variability of the sample mean among repeated samples.
      iii. The sample mean $\bar{x}$ is always an unbiased estimator of the true mean μ.
      iv. If $X \sim \mathcal{B}(n, p)$, then for any sample X we have $p = X/n$.
      v. If $X \sim \mathcal{B}(n, p)$, then the estimator $\hat{p} = X/n$ is unbiased.

      Solution iv

   (d) Which of the following statements regarding the sample mean of a collection of samples is *correct*?

      i. The sample mean $\bar{x}$ is always equal to the true mean μ.
      ii. The average sample mean, over all possible collections of samples, equals the population mean.
      iii. The sample mean is always very close to the population mean.
      iv. The sample mean will only vary a little from the population mean.
      v. For certain probability distributions, the sample mean $\bar{x}$ is a biased estimator of μ.

      Solution ii

   (e) In a test of $H_0 : \mu = 100$ against $H_A : \mu \neq 100$, a sample of size 10 produces a sample mean of 103 and a p-value of 0.08. Thus, at the 0.05 level of significance:

      i. there is sufficient evidence to conclude that $\mu \neq 100$.
      ii. there is sufficient evidence to conclude that $\mu = 100$.
      iii. there is insufficient evidence to conclude that $\mu = 100$.
      iv. there is insufficient evidence to conclude that $\mu \neq 100$.
      v. there is sufficient evidence to conclude that $\mu = 103$.

      Solution iv

   (f) We want to test $H_0 : \mu = 1.5$ against $H_A : \mu \neq 1.5$ at $\alpha = 0.05$ with $n = 15$. A 95% confidence interval for μ calculated from a given random sample is $(1.4, 3.6)$. Based on this finding we:

      i. Accept H_0 (at the 95% significance level).

ii. Reject H_0 (at the 95% significance level).
iii. Accept or reject H_0 depending on whether or not $t_{\alpha/2,15}$ lies in the interval.
iv. Accept or reject H_0 depending on whether or not $t_{\alpha,15}$ lies in the interval.
v. Need to know the underlying distribution of the samples to make any decision regarding H_0.

Solution i

(g) A study was carried out to investigate the effectiveness of a cough syrup. 1000 subjects with a cough participated in the study, with 500 being given the cough syrup and the other 500 given a placebo (i.e. a sugar pill with no medicinal properties). A researcher records the number of days that each subject takes to recover from the cough. The p-value corresponding to the hypothesis $H_0 : \mu_A = \mu_B$ is 0.0008. Thus,

i. There is a large difference between the effects of the cough syrup and the placebo.
ii. There is strong evidence that the cough syrup is effective in reducing the number of days for recovery.
iii. There is strong evidence that there is a difference in effect between the cough syrup and the placebo.
iv. There is little evidence that the cough syrup has any effect.

Solution iii