USF Math 371, Spring 2015 – Midterm Exam #1

Instructor: Mario Micheli
Date/time: Tuesday, March 10, 2015, 12:45-2:30 p.m.
Place: Lo Schiavo 209, University of San Francisco

- *You have 105 minutes to complete this midterm. This exam has 9 pages, including this cover sheet.*
- *There are 7 problems. I suggest that you start with the problems that you find the easiest. Problem 7 is a set of true/false questions, where each answer must be justified.*
  If you are stuck, move on to the next problem!
- *No books, no printed material, and no solved homework assignments are allowed. Besides pencil, paper, and a regular calculator, only a one-sided hand-written, letter-sized, sheet of paper with your own choice of definitions, results and formulas is allowed.*
- *Please do your work on these sheets; you may use the back of each sheet if necessary.*
- **Show your work.** *Ambiguous or otherwise unreadable answers will be marked incorrect. So:* **write clearly** *and provide only one answer to each question.*

Your Name (LAST, First) & USF ID#: ____

I certify that the work appearing on this exam is completely my own. __________________________
(Your signature)

Problem #	Points	Your score
1	12	
2	12	
3	8	
4	12	
5	16	
6	24	
7	16	
Total	100	

Problem 1. (12 points total) Consider a population consisting of $N = 5$ values: 1, 2, 4, 4, 7.

(a) (6 points) Find the exact probability distribution of the sample mean $\overline{X}$ for a simple random sample (without replacement) of size $n = 3$ from this population.

(b) (6 points) What are the expectation and the standard deviation of $\overline{X}$?

Problem 2. (12 points total) Two independent samples of the same population are taken with the goal of estimating the population mean μ. The two samples are of sizes n_1 and n_2 respectively, and the two sample means are $\overline{X}_1$ and $\overline{X}_2$ (note that these are independent random variables, because the samples are independent). Consider combining the two in the form $\hat{\mu} = a\overline{X}_1 + b\overline{X}_2$.

(a) (4 points) Derive a condition on a and b so that $\hat{\mu}$ is an unbiased estimate of μ.

(b) (8 points) For a and b such that $\hat{\mu}$ is unbiased, as in part **(a)**, what choice minimizes $\text{Var}(\hat{\mu})$?

Problem 3. (8 points) In surveying a *large* population in order to estimate the proportion p of voters who support a proposition, what sample size n is necessary to guarantee that the standard deviation of the estimate (i.e. the standard error) is less than 2%? (Note: this value of n must be large enough for any value of p).

Problem 4. (12 points total) Consider a Geometric(p) random variable X, i.e. with probability distribution:

$$P(X = k) = p(1-p)^{k-1}, \qquad k = 1, 2, 3, \ldots.$$

(a) (8 points) Compute its moment generating function, $M_X(t)$. For which values of $t \in \mathbb{R}$ is it defined?

(b) (4 points) Compute $E(X)$ by calculating the appropriate derivative of the moment generating function.

Hint: you will need the formula $\displaystyle\sum_{\ell=0}^{\infty} x^{\ell} = \frac{1}{1-x}$, valid for $|x| < 1$. Be careful with the indices.

Problem 5. (16 points total) Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. discrete random variables, all with probability distribution $P(X = -1|\theta) = \theta^2$, $P(X = 0|\theta) = 2\theta(1-\theta)$, and $P(X = 1|\theta) = (1-\theta)^2$. Here, $\theta \in [0, 1]$ is an unknown parameter.

(a) (4 pts.) What is the Method of Moments estimate $\hat{\theta}_{\mathrm{MM}}$ of the parameter θ (as a function of $X_1, X_2, \ldots, X_n$)?

(b) (8 pts.) What is the Maximum Likelihood estimate $\hat{\theta}_{\mathrm{ML}}$ of the parameter θ (as a function of $X_1, X_2, \ldots, X_n$)?

(c) (4 pts.) Assume that the independent observations $X_1 = 1$, $X_2 = 1$, and $X_3 = 0$ are made. What are $\hat{\theta}_{\mathrm{MM}}$ and $\hat{\theta}_{\mathrm{ML}}$ in this case?

Problem 6. (24 points total) Suppose that X_1, X_2, $\ldots$, X_n, are i.i.d. continuous random variables with probability density function, or PDF, given by

$$f(x|\theta) = \begin{cases} \theta x^{\theta-1} & \text{for } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Here, $\theta \in [0, 1]$ is an unknown parameter.

Equivalently, the cumulative distribution function, or CDF, is $F(x|\theta) = P(X_i \le x) = \begin{cases} x^\theta & \text{for } 0 \le x \le 1, \\ 0 & \text{for } x < 0, \\ 1 & \text{for } x > 0. \end{cases}$.

- **(a)** (4 pts.) What is the Method of Moments estimate $\hat{\theta}_{\text{MM}}$ of the parameter θ (as a function of $X_1, X_2, \ldots, X_n$)?
- **(b)** (8 pts.) What is the Maximum Likelihood estimate $\hat{\theta}_{\text{ML}}$ of the parameter θ (as a function of $X_1, X_2, \ldots, X_n$)?
- **(c)** (4 pts.) What is the approximate variance of the Maximum Likelihood estimate $\hat{\theta}_{\text{ML}}$ for large n?
- **(d)** (4 pts.) Suppose that $n = 100$ and $\hat{\theta}_{\text{ML}} = 1.5$. Give an approximate 95% confidence interval for θ.
- **(e)** (4 pts.) The *median* of the distribution of a continuous random variable X is a number $m \in \mathbb{R}$ for which $P(X \le m) = 1/2$. Using $\hat{\theta}_{\text{ML}}$ from part **(d)**, what is the estimate of the *median* of the distribution of each of the random variables X_i? *Hint:* use the expression given above for the CDF.

Note: part **(d)** needs the answer from part **(c)**, otherwise the questions can be answered independently. Also, more space is provided on the next page.

(Extra page for your solution to Problem 6.)

Problem 7. (2 points each, 16 points total.) **True or False? JUSTIFY your answer precisely.**

7.1. The sample mean of the values of the observations in simple random sample (SRS) is a random variable.

7.2. The population variance is a random variable.

7.3. The estimated standard error of the sample mean from a simple random sample is a random variable.

7.4. A histogram of the values in a simple random sample $X_1, X_2, \ldots, X_n$ of size $n = 1000$ should look approximately like a Normal distribution.

7.5. A 90% confidence interval for the average number of children per household based on a simple random sample is found to be $I_{90\%} = [.7, 2.1]$. Under these assumptions, we conclude that 90% of households have between .7 and 2.1 children.

7.6. Suppose X_1, X_2, $\ldots$, X_n are the values of the observations of a simple random sample from a population with mean μ. Under these assumptions, $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ is an unbiased estimate of μ^2 (note: "μ square", not μ_2).

7.7. Under the assumptions of **7.6**, X_1 and $3X_1 - 2X_2$ are two unbiased estimates of μ.

7.8. Under the assumptions of **7.6**, the estimate X_1 has a smaller variance than the estimate $3X_1 - 2X_2$.