CS 361 Sample Midterm 2

NAME: SOLUTIONS

NETID:

CIRCLE YOUR DISCUSSION:

Thu 2-3 **Thu 4-5** **Fri 10-11** **Fri 11-12**

- Be sure that your exam booklet has 6 pages including this cover page
- Make sure to write your name exactly as it appears on your i-card
- Write your netid and circle your discussion section on this page
- **Show your work**
- This is a closed book exam
- You are allowed one handwritten 8.5 x 11-inch sheet of notes (both sides)
- You may **not** use a calculator or any other electronic device
- Turn off your phone and store it in your backpack
- Store away any other electronic devices including earphones and smartwatches
- Absolutely no interaction between students is allowed
- Use backs of pages for scratch work if needed
- Show your i-card when handing in your exam

Problem	1	2	3	4	5	Total
Possible	30	30	30	30	30	150
Score						

Problem 1 (30 pts)

1. (15 points) The Career Center wants to estimate the mean monthly salary of U of I students who go out on summer internships. The staff surveys 100 students and finds a sample mean of \$5000 and a sample (unbiased) standard deviation of \$1200. Provide a 95% confidence interval for the population mean monthly salary. Draw a box around your answer.

$$\text{stderr} = \frac{1200}{\sqrt{100}} = 120$$

95% confidence interval

$$= [5000 - 2 \times 120,\ 5000 + 2 \times 120]$$

$$= \boxed{[\$4760, \$5240]}$$

2. (15 points) You are trying to estimate the mean time that students wait in the CS 225 office hours queue. After 50 samples, you have calculated a standard error of 3 minutes. How many **additional** samples will you need to reduce the standard error to 1 minute? Draw a box around your answer.

Since $\text{stderr} \propto \frac{1}{\sqrt{N}}$,

we need a total of $9 \times 50 = 450$ samples

$$\boxed{\text{Number of additional samples} = 400}$$

Problem 2 (30 pts)

1. (15 points) You ask your CS 361 classmates whether they have started the next homework. Out of 10 students in the sample, 9 have not started yet. Calculate the sample (unbiased) standard deviation of the fraction of students who have started the homework. Draw a box around your answer.

$$\text{std. dev.} = \sqrt{\frac{9(0-0.1)^2 + 1(1-0.1)^2}{10-1}}$$

$$= \sqrt{\frac{0.09 + 0.81}{9}}$$

$$= \boxed{\sqrt{0.1}}$$

2. (15 points) Your instructor grades 5 midterms out of a much larger stack in order to estimate the median score. The grades of the exams in the sample are 125, 135, 140, 140 and 150. For a bootstrap replicate of this sample, what is the probability that the median score is 150? You may use choose and summation notation in your answer. Draw a box around your answer.

A bootstrap replicate is a sample of $N=5$ drawn uniformly with replacement

For the median to be 150, at least 3 samples in the replicate must be 150.

$$\boxed{\text{So probability} = \sum_{k=3}^{5} \binom{5}{k} \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{N-k}}$$

Problem 3 (30 pts)

1. (15 points) You hypothesize that the average mass of apples in the cafeteria is 8 ounces. You weigh an apple every day for 9 days and find that the sample mean is 7.5 ounces and the standard error is 0.25 ounces. Complete the following formula for the two-tailed p-value of this experiment by providing values for a, b and n, where $f_n(x)$ is the probability density function of Student's t-distribution with n degrees of freedom. Draw a box around your answer.

$$\text{p-value} = 1 - \int_a^b f_n(x)dx$$

test statistic $g = \frac{7.5-8}{0.25} = -2$

So $\boxed{\begin{aligned} a &= -2 \\ b &= 2 \\ n &= 9-1 = 8 \end{aligned}}$

2. (15 points) A roulette wheel has 36 nonzero slots and an unknown number of zero slots. You observe that in N spins of the wheel, the ball lands in a zero slot k times. Write down the likelihood function $L(\theta)$ for the number of zero slots on the wheel. Draw a box around your answer.

Let θ be number of zero slots

$$\boxed{L(\theta) = \binom{N}{k}\left(\frac{\theta}{\theta+36}\right)^k \left(\frac{36}{\theta+36}\right)^{N-k}}$$

Problem 4 (30 pts)

1. (15 points) You flip a coin 10 times and observe 3 heads. Use the prior distribution $P(\theta)$ below to calculate the maximum a posteriori (MAP) estimate $\hat{\theta}$ of the probability of heads. Draw a box around your answer.

N k

$$P(\theta) = \binom{5}{2}\theta^2(1-\theta)^3$$

$$P(\theta) \sim \text{Beta}(\alpha = 3, \beta = 4)$$

$$\hat{\theta} = \frac{\alpha - 1 + k}{\alpha + \beta - 2 + N} = \frac{5}{15} = \boxed{\frac{1}{3}}$$

2. (15 points) You begin with a belief that your upstairs neighbor has gone away for the weekend with probability 0.1. When she is home, she stomps on your ceiling according to a Poisson process at an intensity of 1 stomp per hour. Between 9 am and noon on Saturday morning, you don't hear any stomps. What is the posterior probability that your neighbor is away? Assume that she is either at home or away. Draw a box around your answer.

Let A be the event that she is away

Let B be the number of stomps in a 3 hour period

$$P(A|B=0) = \frac{P(B=0|A)P(A)}{P(B=0|A)P(A) + P(B=0|A^c)P(A^c)}$$

$$= \frac{1 \times 0.1}{1 \times 0.1 + e^{-3}\frac{3^0}{0!} \times 0.9} = \frac{0.1}{0.1 + 0.9e^{-3}}$$

$$= \boxed{\frac{e^3}{e^3 + 9}}$$

Poisson with $\lambda = 3$

Problem 5 (30 pts)

1. (15 points) Can the matrix shown below be a covariance matrix? Justify your answer in one sentence.

$$\begin{bmatrix} 7 & 1 & -0.5 \\ 0.5 & 3 & 0.1 \\ -0.7 & 0.2 & 0.9 \end{bmatrix}$$

No, because a covariance matrix must be symmetric

2. (15 points) Suppose dataset $\{\mathbf{x}\}$ has the covariance matrix shown below. What is the mean square error incurred by projecting $\{\mathbf{x}\}$ on to its first principal component? Draw a box around your answer.

$$\text{Covmat}(\{\mathbf{x}\}) = \begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix}$$

Find eigenvalues:

$(9-\lambda)(3-\lambda) - 4^2 = \lambda^2 - 12\lambda + 11 = (\lambda-11)(\lambda-1) = 0$

So, $\lambda_1 = 11$ and $\lambda_2 = 1$

Since the PCA is reducing from $d=2$ to $s=1$ dimensions,

mean square error $= \sum_{j=2}^{2} \lambda_j = \lambda_2 = \boxed{1}$