Sample Final Problems

1. **Short questions**
   a. For a \( m \times n \) matrix \( A \) with rank \( n \), where \( m > n \), is it possible for \( Ax = b \) to have no solution?
   b. For a full-rank \( m \times n \) matrix \( A \) with rank \( m \), where \( n > m \), is it possible for \( Ax = b \) to have no solution?
   c. You are trying to predict the points scored by an NBA team in 50 games. You realize that you can actually fit a linear model that predicts the scores based on the average temperature that day in each of the 50 US states. What is going on? Will your discovery make you rich?
   d. Suppose the average family income in an area is $10,000. Find an upper bound for the percentage of families with income over $50,000. Find a better bound if you know that the standard deviation is $8,000.
   e. In what situations is it desirable to minimize the Type I errors as opposed to the Type II errors?
   f. In what situations is it desirable to minimize the Type II errors as opposed to the Type I errors?
   g. If \( X \) is a uniform random variable between 0 and 1, what is the pdf of \( Y = X^2 \)?

2. **Chad** You find your coworker Chad really annoying. He often works from home, but when he is in the office and you walk by his desk he insists on showing you pictures of his pet iguana and rants endlessly about Fantasy Football. You would like to be able to predict when he is in the office in order to avoid him as much as possible. Another of Chad’s annoying habits is to crank up the AC, so you decide to use the temperature in the office to predict his presence. After a month you have gathered the following data.

<table>
<thead>
<tr>
<th>Chad</th>
<th>61</th>
<th>65</th>
<th>59</th>
<th>61</th>
<th>61</th>
<th>65</th>
<th>61</th>
<th>63</th>
<th>63</th>
<th>59</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Chad</td>
<td>68</td>
<td>70</td>
<td>68</td>
<td>64</td>
<td>64</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Temperature (°F)

a. You model the presence of Chad using a random variable \( C \) which is equal to 1 if he is there and 0 if he is not. Estimate the pmf of \( X \) from the data.

b. You model the temperature using a random variable \( T \). Use a kernel density estimator which is rectangular and has width 2 to estimate the conditional pdf of \( T \) given \( C \). Sketch the distribution.

c. If the temperature is 68° does a maximum-likelihood estimate of \( C \) predict that Chad is in the office?

d. If the temperature is 64°. Does a maximum-a-posteriori estimate of \( C \) predict that Chad is in the office?

e. What happens if the temperature is 57°? Explain how using parametric estimation may alleviate this problem.
3. **3-point shooting** The New York Knicks hire you as a data analyst. Your first task is to come up with a way to determine whether a 3-point shooter is any good. You will use the graph in Figure 1 to help you. It plots the following function.

\[ g(\theta, n) = \theta^n. \]  

a. You decide to make the player shoot 3-pointers until he misses. If we interpret \( \theta \) as the probability that the player makes each shot and \( n \) is a fixed number of shots, what does \( g(\theta, n) \) represent? Under what assumption?

b. The coach tells you: *I want to make sure that the guy has a shooting percentage over 80%.* What is your null hypothesis?

c. What number of shots does a player need to make in a row for you to reject the null hypothesis with a confidence level of 5%?

d. A player makes 9 shots in a row. What is the corresponding \( p \) value? Do you declare him as a good shooter?

e. What is the probability that you do not declare a player who has a shooting percentage of 90% as a *good shooter*? (In other words, that you do not reject the null hypothesis).

f. You apply the test on 100 players that are trying out for the team. You adapt the threshold applying Bonferroni’s method. What is the new threshold?

g. With the correction, what is the probability that you do not declare a player who has a shooting percentage of 90% as a *good shooter*?

h. What is the advantage of adapting the threshold? What is the disadvantage?

4. **Linear regression with an intercept** Consider the model

\[ y_i \approx \alpha x_i + \beta, \]  

where \( \beta \) and \( \alpha \) are real-valued parameters. Our aim is to estimate \( \alpha \) and \( \beta \) from some training data: \((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\).

a. We decide to fit the parameters solving a least-squares problem. Write the cost function for the least-squares problem in terms of \( \alpha, \beta \) and the vectors

\[
x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad 1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.
\]  

b. Compute the projection of \( x \) onto the orthogonal complement of span (1), which we denote by \( \tilde{x} \). Interpret the result.

c. Prove that span (1, x) = span (1, \( \tilde{x} \)).

d. Consider the model

\[ y_i \approx \tilde{\alpha} \tilde{x}_i + \tilde{\beta}. \]  

Write the cost function to fit \( \tilde{\alpha} \) and \( \tilde{\beta} \) using least-squares in terms of \( \tilde{\alpha}, \tilde{\beta}, x, 1 \) and \( y \).
Figure 1: Graph for Problem 3.

e. Compute the least-squares fit of $\tilde{\alpha}$ and $\tilde{\beta}$: $\tilde{\alpha}_{LS}$ and $\tilde{\beta}_{LS}$. Write them as a function of the centered $\tilde{x}$, $y$ and the average of $y \frac{1}{n} \sum_{i=1}^{n} y_i$.

f. Show that

$$\tilde{\alpha}_{LS} \tilde{x}_i + \tilde{\beta}_{LS} = \alpha_{LS} x_i + \beta_{LS}$$

for $1 \leq i \leq n$, where $\alpha_{LS}$ and $\beta_{LS}$ are the solution of the least-squares problem in part (a).

5. Foxes and rabbits A biologist studies the populations of foxes and rabbits in Central Park. She determines that the population of foxes $f_n$ and of rabbits $r_n$ in year $n$ are given by the equations

$$r_n = 1.06 r_{n-1} - 0.16 f_{n-1}$$

(6)

$$f_n = 0.06 r_{n-1} + 0.84 f_{n-1}.$$  (7)

The eigenvalues of the matrix

$$A := \begin{bmatrix} 1.06 & -0.16 \\ 0.06 & 0.84 \end{bmatrix}$$  (8)

are $\lambda_1 = 1$ and $\lambda_2 = 0.9$. The corresponding eigenvectors are

$$u_1 = \begin{bmatrix} 0.8 \\ 0.3 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \quad (9)$$

a. Describe the equations in words.
b. If there are exactly the same number of foxes as rabbits, what will the population of foxes and rabbits tend to eventually?

c. Represent the vectors

\[ x_1 = \begin{bmatrix} 200 \\ 100 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 100 \\ 200 \end{bmatrix} \quad (10) \]

in terms of \( u_1 \) and \( u_2 \) using the formula for the inverse of a \( 2 \times 2 \) matrix

\[
\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.
\quad (11)
\]

d. If there are 200 rabbits and 100 foxes, what will the population of each animal tend to eventually?

e. How about if there are 100 rabbits and 200 foxes?

f. Prove that the populations are only sustainable if there are more rabbits than foxes.

6. **Defective pixels.** A company makes HD TVs. Each television typically has some defective pixels. From past observations, the company has determined that the average number of defective pixels in a single TV is 20. An engineer is hired to make TVs that have fewer defective pixels. The engineer claims that she can improve the current method. You are hired to test whether this is the case.

a. You test 100 new televisions. The average number of defective pixels turns out to be 19.1. What are the chances that this happens if the mean is actually 20? Use an exact bound applying Chebyshev’s inequality, assuming that the standard deviation is equal to 4.

b. If you want to test whether the new method is better, what is your null hypothesis?

c. Using the previous result, is the new method significantly better? The required significance level is \( \alpha = 0.05 \) level.

d. Use the Central Limit Theorem to obtain an approximate bound. Is the result statistically significant at a level of \( \alpha = 0.05 \)?

e. Explain why Bonferroni’s method is not relevant for this problem, even though we are testing many TVs.

7. **Camera measurement.** The measurement from a camera can be expressed as \( Y = AX + Z \), where \( X \) is the object position with mean \( \mu \) and variance \( \sigma_X^2 \), \( A \) is the **occlusion** indicator function and is equal to 1 (if the camera can see the object) with probability \( p \), and 0 (if the camera cannot see the object) with probability \( (1 - p) \), and \( Z \) is the measurement error with mean 0 and variance \( \sigma_Z^2 \). Assume that \( X, A, \) and \( Z \) are independent.

a. Compute the mean and variance of \( Y \).

b. Compute the covariance between \( X \) and \( Y \).

c. Find the best linear MSE estimate of \( X \) given the camera measurement \( Y \). Your answer should be in terms only of \( \mu, \sigma_X^2, \sigma_Z^2, \) and \( p \).

d. What does is your estimate if \( p = 0 \)? Argue that in this case your estimate is the best MSE estimate.