

Instructions and Policy: Each student should write up their own solutions independently. You need to indicate the names of the people you discussed a problem with; ideally you should discuss with no more than two other people.

You need to submit your TYPED answer in PDF via Blackboard. L^AT_EX is typesetting is encouraged but not required. Also submit your R and Python codes. Please write clearly and concisely - clarity and brevity will be rewarded. Refer to known facts as necessary.

Q1 (2 pts):

- (a) Show that the variance of $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$, where cov is the covariance between X and Y .
- (b) State one condition over X and Y that makes $\frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}}$ achieve its maximum value.
- (c) State one condition over X and Y that makes $\frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}}$ achieve its minimum value.

Q2 (2 pts): Probability and inference.

- (a) If X_1, \dots, X_n are independent and identically normally distributed random variables, $X_i \sim N(\mu, \sigma)$. Estimate μ and σ via maximum likelihood.
- (b) Prove the conditional version of Bayes rule:

$$P(B|A, C) = \frac{P(A|B, C)P(B|C)}{P(A|C)}$$

Q3 (2 pts): Probability and conditional probability.

(a) The Internet is a wonderful source of information about symptoms of rare diseases. Are you sneezing? It could be the West Nile virus! The West Nile virus (WNV) infected approximately 2,000 people in the United States last year¹. Sheldon, your hypochondriac friend, is sneezing and heard about the West Nile virus on Twitter. He demands a test for the West Nile virus, why not? The test correctly identifies the presence of WNV in 95% of cases and only gives false positives in 1/10,000 cases. Unfortunately, the test indicates came back positive for West Nile virus and Sheldon is very concerned. Assume that in the population of the United States there are 300 million people susceptible to WNV.

- (i) What is the probability that Sheldon has WNV?
- (ii) The WNV virus is fatal in 5% of the cases. What is the probability that Sheldon will die this year? Assume a fatality rate of any cause (car accident, etc.) of 0.1%.
- (b) Alice and Bob are playing a simple dice game. Each rolls one dice and the one with higher number wins. If the numbers are the same, they roll again. If Alice just won, what is the probability that she rolled a '4'?

¹Source: the “always-reliable” Wikipedia

Q4 (4 pts): Download the Marvel comic book data² <https://goo.gl/TRXC8T> containing the following files:

- `marvel.txt`: characterID (1st column) and comicID (2nd column).
- `marvelCharacters.txt`: characterID and the name of the character.
- `marvelComicBooks.txt`: comicID and the name of the comic book.

(a) Plot the empirical complementary cumulative distribution (ECCDF) of comic characters appearances in comic books. The ECCDF $P[X > x]$ is defined as the fraction of characters with more than x comic book appearances. For instance, if “superman” appears in 1000 comic books and there are only 10 characters with more than 1000 comic book appearances out of 2000 characters, then $P[X > 1000] = 10/2000$.

IMPORTANT: Your plot should be in log-log scale.

Example of a bare-bones python code to plot in log-log scale (CCDF of exponential distribution). Numpy and matplotlib are libraries that you need to install.

```
import numpy as np
import matplotlib.pyplot as plt
import math
x = xrange(0,100)
y = np.zeros(len(x))
cumul = 1
for i in x:
    y[i-1] = math.exp(-i)
plt.xlim([1,max(x)])
X, Y AXIS
plt.loglog(x,y,'ro')
plt.show()
```

(b) Let A be the adjacency matrix connecting characters to comic books, where $A_{i,j}$ has character i appearing on comic book j . Let A^T be the transpose of matrix A .

- What does $W = AA^T$ represent? Give the name of the entity with the largest degree in the graph that has adjacency matrix W ?
- What does $U = A^T A$ represent? Give the name of the entity with the largest degree in the graph that has adjacency matrix U ?
- Choose the correct option: If $U = A^T A$, then (1) $U^T = AA$, (2) $U^T = A^T A$, (3) $U^T = AA^T$, or (4) $U^T = A^T A^T$.
- Let $P = D^{-1}W$, where $W = AA^T$ and D is a diagonal matrix where $D_{i,i} = \sum_j W_{i,j}$. Find the eigenvector x such that $x = Px$ and $\|x\|^2 = 1$, where $\|x\|^2 = \langle x, x \rangle$ is the inner product of x with itself.

²Courtesy of Jay-Yoon Lee, Cesc Rossello, Ricardo Alberich, and Joe Miro. Reference <http://bioinfo.uib.es/~joemiro/marvel.html>