Problem 1: 30 points
Intrusion Detection System

A computer intrusion detection system (IDS) is designed to provide an alarm whenever an intrusion (e.g., unauthorized access) into a computer system is being attempted. A probabilistic evaluation of a system with two independent operating intrusion detection systems (a double IDS) was published in the Journal of Research of the National Institute of Standards and Technology (Nov/Dec 2003).

Consider a double IDS with System A and System B. If there is an intruder, System A sounds an alarm with probability 0.9, and System B sounds an alarm with probability 0.95. If there is no intruder, System A sounds an alarm with probability 0.2, and System B sounds an alarm with probability 0.1. Assume that Systems A and B operate independently.

Let A be the event that System A sounds an alarm. Let B be the event that System B sounds an alarm. Let I be the event that there is in fact an intruder.

(a)
Define the relevant events, and formally express the four probabilities given in the example.
(e.g., $P(A|I) = x.xx$)

(b)
If there is an intruder, what is the probability that both systems sound an alarm?

(c)
If there is no intruder, what is the probability that both systems sound an alarm?

(d)
Given an intruder, what is the probability that at least one of the systems sound an alarm?

(e)
Assume that the probability of an intruder is 0.4. Also continue to assume that both systems operate independently. If both systems sound an alarm, what is the probability that an intruder is detected?

Problem 2: 25 points
Lie Detector Test

A new type of lie detector, called the Computerized Voice Stress Analyzer (CVSA) has been developed. The manufacturer claims that the CVSA is 98% accurate — that is, it correctly determines if a suspect is lying or not 98% of the time — and unlike a polygraph machine, it will not be thrown off by drugs and medical factors. However, laboratory studies by the DoD found that the CVSA had an accuracy rate of 49.8% — slightly less than pure chance. Suppose the CVSA is used to test the veracity of four suspects. Assume the suspects' responses are independent.

(a)
If the manufacturer's claim is true, what is the probability that the CVSA will correctly determine the veracity of all four suspects?

(b)
If the manufacturer's claim is true, what is the probability that the CVSA will yield an incorrect result for at least one of the four suspects?

(c)
Suppose that in a laboratory experiment conducted by the DoD on four suspects, the CVSA yielded incorrect results for two of the suspects. Use this result to make an inference about the true accuracy rate of the new lie detector.

Problem 3: 20 points
Auditing an Accounting System

In auditing a firm's financial statements, an auditor will (1) assess the capability of the firm's accounting system to accumulate, measure, and synthesize transactional data properly and (2) assess the operational effectiveness of the accounting system. In performing the second assessment, the auditor frequently relies on a random sample of actual transactions (Stickney and Weil, Financial Accounting: An Introduction to Concepts, Methods, and Uses, 2002). A particular firm has 5,382 customer accounts that are numbers from 0001 to 5382.

(a)
One account is to be selected at random for audit. What is the probability that account number 3,241 is selected?

(b)
Draw a random sample of 10 accounts, and explain in detail the procedure you used. (Hint: Python can do this)

(c)
Referring to part b, is one sample of size 10 more likely to be chosen than any other? What is the probability that the sample you drew in part b was selected?

Problem 4: 25 points
Fish Contamination

A U.S. Army Corps of Engineers (USACE) study focused on DDT contamination of fish in the Tennessee River in Alabama. Part of that investigation studied how far upstream contaminated fish have migrated. A fish is considered to be contaminated if its measured DDT concentration is greater than 5.0 parts per million (ppm).

(a)
Considering only contaminated fish captured from the Tennessee River, the data reveal that 52% of the fish are found 275–300 miles upstream, 39% are found 305–325 miles upstream, and 9% are found 330–350 miles upstream. Use the percentages to estimate the probabilities $P(275\text{–}300)$, $P(305\text{–}325)$, and $P(330\text{–}350)$.

(b)
Given that a contaminated fish is found a certain distance upstream, the probability that it is a channel catfish (CC) is determined from the data as $P(CC|275\text{–}300) = 0.775$, $P(CC|305\text{–}325) = 0.77$, and $P(CC|330\text{–}350) = 0.86$. If a contaminated channel catfish is captured from the Tennessee River, what is the probability that it was captured 275–300 miles upstream?