Yeshiva University

AIM 5009 Bayesian Statistics
Problem Set 1.2

Instructions

This problem set reviews some of the key concepts from the Computational Statistics and Probability course that are used in Bayesian Statistics. This assignment focuses on the key definitions and concepts of probability you should have learned previously.

Complete the following problems. Your full solutions should be typed or neatly handwritten and should be submitted as a PDF file by the due date given in Canvas.

Problems

1. You have accepted a position as an artificial intelligence analyst at an international crime-fighting organization. For security purposes, all employees are assigned a unique identifier as well as a random three-character codename using the 26 letters of the alphabet and the ten digits 0-9. There are no constraints on these codenames – the three characters can be the same or different, all letters or all numbers or a mix of each, etc. (So, for example, PRS, 9X3, UUU, and 007 are all legitimate codenames.)

    a. How many employees need to be in the same room for there to be at least a 25% chance that two employees have the same codename?
    b. How would this answer change if we used four characters instead of three?
    c. How many employees need to be in the same room for you to be “pretty sure” that two people will have the same codename? (Specify what you mean by “pretty sure” and use the three-character codename case.)

2. (Classic Monte Hall Problem) Everyone who studies Bayesian Statistics should be familiar with the Monte Hall problem. You are a contestant on a game show, and you are presented with three doors. Behind one of the doors is a new car, while behind each of the other two doors is a goat. You are instructed to choose a door, and if the car is behind that door, you get to keep the car. If your chosen door reveals a goat, you get nothing (not even the goat).

    You choose door number two. After you make your choice, the game show host opens door number three and reveals a goat. He then offers you the opportunity to switch to door number one instead. Should you stay with door two or switch to door one? Or does it even matter? Use an appropriate analysis to answer this question.

    Note: You can easily find solutions to the Monte Hall problem online. Don’t do that. Work it out for yourself.

3. Consider the following two problems:

    - Mr. Martin has two children. The older child is a girl. What is the probability that both children are girls?
    - Mrs. Gardner (unrelated to Mr. Martin) has two children. At least one of them is a girl. What is the probability that both children are girls?

    Answer the following questions:

    a. Calculate the probability that both children are girls in the first case (Mr. Martin).
    b. Calculate the probability that both children are girls in the second case (Mrs. Gardner). Clearly identify any assumptions you need to make.
    c. After you have completed your calculation for the second case, can you identify a different set of assumptions (or perhaps a different approach to the problem) that might result in a different answer? Explain.

4. The following problem was used in a test of the ability of physicians to make judgments about probabilities in fairly realistic cases.

    > Suppose you are a physician who has just examined a woman for breast cancer. The woman has a lump in her breast, but based on many years of experience, you estimate the odds of a malignancy as one in a hundred. Just to be safe, though, you order a mammogram. A mammogram is an X-ray test that accurately classifies roughly 80 percent of malignant tumors and 90 percent of benign tumors. The test report comes back, and much to your surprise, the consulting radiologist believes that the breast mass is malignant.
    >
    > Question: Given your prior view that the chances of a malignancy were only 1 percent and given the test results that are 80 or 90 percent reliable, as above, what would you say the overall chances of a malignancy are now?

    Try to do better than the physicians did. (They did not do well...)

5. Complete exercise 2.8 (Flight delays) from the textbook.

Grading Rubric

For this assignment, the following rubric will be used:

Correctness of Answers	0 – 6
Clarity and Organization of Presentation	0 – 4