NEW YORK INSTITUTE OF TECHNOLOGY
College of Engineering and Computing Sciences

DTSC 620: Statistics for Data Science (Fall 2022)

Base Rate Fallacy Worksheet

1. Suppose that your doctor performs a test that is 99% accurate; that is, when the test was administered to a test population all of whom had the disease, 99% of the tests indicated disease, and likewise, when the test population was known to be 100% free of the disease, 99% of the test results were negative. Upon visiting your doctor to learn the results, he tells you he has good news and bad news. The bad news is that indeed you tested positive for the disease. The good news however, is that, out of the entire population, the rate of incidence is only 1/10000; that is, only 1 in 10000 people have this ailment. What, given this information, is the probability of your having the disease?

2. An attack detection system "*X*" is 99.5 percent accurate in detecting an attack. The false alarm rate of "*X*" is 1 percent. On any given day, "*X*" monitors 10^8 traffic sessions. In a single day, one could expect 1 attack that generates 100 attack sessions. Alice is a security expert and notices that "*X*" identified 2 attacks (200 sessions as belonging to attack). Calculate the following:
    a. *Bayesian detection accuracy* and
    b. *Bayes probability that the absence of an alarm signifies absence of an attack*

Note: *Detection accuracy* is defined as the number of attack sessions correctly detected as attacks divided by total number of attack sessions presented to the detection system. *False alarm rate* is defined as the number of normal sessions erroneously detected as attacks divided by total number of normal sessions presented to the detection system.