BIO 222 Homework #1
Due Tuesday, September 14, at 8 AM through online submission

1. Of recent major interest is the development of rapid antigen tests to predict SARS-CoV-2. The first two Emergency Use Authorizations (EUAs) from the Food and Drug Administration (FDA) had sensitivities of 84% and 97% compared to the reverse transcription polymerase chain reaction (RT-PCR) test, the current gold standard. The specificities of both of these tests were 100%.

(i) Determine the positive predictive value (PPV) for each of these two tests when the prevalence of SARS-CoV-2 is equal to 2%, 10%, or 20% in the population being tested.

(ii) Using class notation, the negative predictive value (NPV) is given by $P(\bar{D}|-)$. Develop a general formula for NPV as a function of sensitivity, specificity, and prevalence. What is the NPV for each of these two tests when the prevalence of SARS-CoV-2 is equal to 1%, 5%, or 20%.

(iii) Suppose you are a nurse using the first test (sensitivity 84%, specificity 100%, plus your calculations above) on someone who is coming in for testing. Write a few sentences of explanation of what would be most helpful and important for this person to understand.

2. The World Health Organization (WHO) suggests that the preferred and minimally acceptable profile for a human vaccine for SARS-CoV-2 be 70% for one year (https://www.who.int/publications/m/item/who-target-product-profiles-for-covid-19-vaccines).

Assume that a vaccine is 70% successful in preventing an inoculated person from acquiring the disease, if exposed, and that a person not vaccinated has a 50% chance of acquiring the disease, if exposed. Three people, two vaccinated and one not, are not at the same location, are not in contact with the same people, and cannot expose one another (hence are independent). What is the probability that at least one will get the disease if all were exposed? Also, write out the probability mass function and distribution function for the total number of people (out of the three) who will develop the disease.

3. Randomized response is a survey sampling technique employed in the context of asking sensitive questions, e.g., use of illicit drugs, where the idea is for respondents to answer truthfully or not based on a known chance mechanism. Warner (1965) proposed that the respondent draw from a deck of cards with specified proportions p of color A and $1 - p$ of color B: the respondent is to answer the question truthfully if s/he draws color A and lie if s/he draws color B. While the color of the card is unknown to the interviewer, it is possible to use the known proportion p to identify the prevalence of illicit drug use, even though some unknown numbers of respondents have lied.

In this example we consider a survey sampling problem where two questions are asked. In this case, two decks of cards are used, with each deck following the Warner model. For example, each respondent is asked to draw two cards with replacement from two decks of cards, and then answer: “Do you smoke?” and “Do you drink?” Whether or not to answer a question truthfully depends on the color of the corresponding card. Denote p_1 as the known probability of drawing a card with the color corresponding to answer truthfully in deck 1, and p_2 as the known probability for deck 2. Denote parameters π_{11}, π_{10}, and π_{01} as the respective prevalence of smoking and drinking, smoking but not drinking, and drinking but not smoking. Finally, denote the parameters θ_{11}, θ_{10}, and θ_{01} as the proportions of answering ‘yes and yes’, ‘yes and no’ and ‘no and yes’, respectively. How to estimate π_{11}, π_{10}, and π_{01} using θ_{11}, θ_{10}, θ_{01}, and p_1, and p_2?

4. Consider the function $f(x) = kx^2(1 - x)$.

(i) Find the value of k that allows $f(x)$ to integrate to 1 over the range [0, 1].

(ii) For this value of k, the non-negative function $f(x)$ serves as a valid density function over [0, 1]. Calculate the mean of this distribution, given by

$$\int_0^1 x \cdot f(x)dx.$$

5. Exercise 1.52 in C&B.

6. Exercise 2.2 in C&B.