Assignment #6

Due on Friday, March 13, 2020

Introduction to Chapter 4, *Stochastic Models,* in the class lecture notes at `http://pages.pomona.edu/~ajr04747/`.

Read Section 4.1 on *Random Variables* in the class lecture notes at `http://pages.pomona.edu/~ajr04747/`.

Read Section 4.1.1 on *A Brief Excursion into Probability* in the class lecture notes at `http://pages.pomona.edu/~ajr04747/`.

Do the following problems

1. Deoxyribonucleic acid (DNA) is a molecule that encodes genetic information in living organisms. It is made up of two strands of nucleotides that coil around one another in a helical structure. There are four types of nucleotides that make up DNA: Adenine (A), Thymine (T), Cytosine (C) and Guanine (G). The nucleotides in each strand of DNA form a sequence that parirs up with the nucleotides of the other stand accoring to the pairings: A and T, and C and G.

   Suppose a stand of DNA has the following the 20–nucleotide sequence[1]

   $$AGGGATACATGACCCATACA.$$

   Let $P(A)$ denote the probability that a nucleotide picked at random from the sequence will be an A nucleotide; $P(T)$ denote the probability that a nucleotide picked at random from the sequence will be an T nucleotide; and similarly for $P(C)$ and $P(G)$.

   (a) Use the first five nucleotides in the sequence to estimate the probabilities $P(A)$, $P(G)$, $P(C)$ and $P(T)$.

   (b) Repeat part (a) using the first 10 nucleotides.

   (c) Repeat part (a) using all the nucleotides in the sequence.

   (d) Is there a pattern to the way the probabilities you computed in parts (a)–(c) changed? If so, what features of the original sequence does this pattern reflect?

---

[1]This problem is based on Problem 4.2.3 on page 128 in *Mathematical Models in Biology* by E. S. Allman and J. A. Rhodes, Cambridge University Press, 2004.

2. A simple model for human offsprings[2] is that each child is equally likely to be male or female.

   With this model, a three–child family can be thought of as an ordered triple of F's for females and M's for males, in which each of the triples is determined at random.

   (a) What are the 8 possible outcomes? What are their probabilities?

   (b) What outcomes make up the event "the oldest child is a daughter" and what is the probability of this event?

   (c) What outcomes make up the event "the family has one daughter and two sons" and what is its probability?

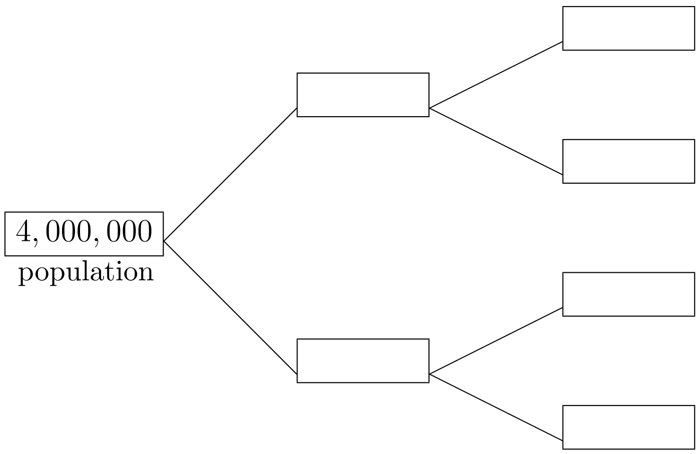
   (d) What is the complement of the event in part (c)? List the outcomes in it and describe it in words. What is its probability?

   (e) What outcomes make up the event "the family has at least one daughter"? What is its probability?

3. Suppose a clinic in a large city uses a new test to determine if a patient has hepatitis. If a person tests *positive*, then the test is telling us that he/she does have hepatitis (the test could be wrong!) Similarly, if a person tests negative, then the test's conclusion is that he/she does not have hepatitis. Assume that out of every 100 people who do have hepatitis, 95 test positive (that is, the test result is that they do have hepatitis) and 5 test negative. Out of every 100 people who don't have hepatitis, 90 test negative and 10 test positive. Suppose that among the 4 million people who live in the city, 0.05% do have hepatitis.

   (a) Complete the chart below. First decide on what the branches represent (there are two ways to do the branching but the information provided is suitable for only one of them). Then, for each of the boxes, find the corresponding number of people. Be careful not to make any assumptions other than the facts presented above.

[2] This is Problem 4.2.5 on page 128 in *Mathematical Models in Biology* by E. S. Allman and J. A. Rhodes, Cambridge University Press, 2004.



(b) A person is selected at random and given the test. If the test is positive, what is the probability that the person actually has hepatitis.

Hint: Recall than in calculating a probability you write a fraction. Make sure that you choose the numerator and the denominator thoughtfully.

(c) Is your answer to the previous part surprising? In what way?

4. Medical tests, such as those for diseases, are sometimes characterized by their *sensitivity* and *specificity.* [3] The sensitivity of a test is the probability that a person with the disease will show a positive test result (a correct positive). The specificity of a test is the probability that a person who does not have the disease will show a negative test result (a correct negative).

(a) Both sensitivity and specificity are conditional probabilities. Which of the following are they?

$$P(-\text{result} \mid \text{disease}), \quad P(-\text{result} \mid \text{no disease})$$
$$P(+\text{result} \mid \text{disease}), \quad P(+\text{result} \mid \text{no disease})$$

(b) The other conditional probabilities listed in (a) can be interpreted as probability of false positives and false negatives. Which is which?

[3] This is Problem 4.3.3 on pages 134 and 135 in *Mathematical Models in Biology* by E. S. Allman and J. A. Rhodes, Cambridge University Press, 2004.

(c) A study by Yerushalmy *et al.*[4] investigated the use of X–ray readings to diagnose tuberculosis. Diagnosis of $1,820$ individuals produces the data in Table 1. Compute both the sensitivity and specificity for this method of diagnosis.

Table 1: Data from Tuberculosis (TB) Diagnostic Study

	Persons without TB	Persons with TB
Negative X–ray	$1,739$	8
Positive X–ray	51	22

5. Ideally, the specificity and sensitivity of medical tests should be high (close to 1). [5] However, even with a highly sensitive and specific test, screening a large population for a disease that is rare can produce surprising results.

   (a) Suppose the specificity and sensitivity of a test for the disease are both 0.99. The test is applied to a population of $100,000$ individuals, only 100 of whom have the disease. Compute how many individual with or without the disease would be expected to test positive or negative.

   (b) Compute the conditional probability that a person who tests positive actually has the disease.

[4] Yerushalmy, J., Harkness, J. T., Cope, J. H., and Kennedy, B. R. (1950). *The role of dual readings in mass radiography.* Am. Rev. Tuber., **61**, 443–464

[5] This is Problem 4.3.4 on page 135 in *Mathematical Models in Biology* by E. S. Allman and J. A. Rhodes, Cambridge University Press, 2004.