Probabilistic models in molecular biology

Exercises 1

1. Define A and B as two sets which exhaust the sample space Ω (i.e. $A \cup B = \Omega$). Let $P(A) = 0.4$ and $P(A \backslash B) = 0.15$. What is $P(B)$?

2. In a survey of people attending a clinic, 80% have disease A and 90% have disease B. Show that at least 70% have both diseases.

3. (*multiple choice*) Events A and B satisfy $P(A) = 0.6$, $P(B) = 0.2$ and $P(A \cap B) = 0.12$. Hence A and B are

   (A) disjoint [i.e. mutually exclusive such that $P(A \cap B) = \emptyset$] and independent

   (B) disjoint but not independent

   (C) not disjoint but are independent

   (D) neither disjoint nor independent

4. A blood test is 99% effective in detecting a certain disease when the disease is present. However, the test also yields a false positive result for 2% of healthy patients tested. Suppose that 0.5% of the population have the disease.

   (a) What proportion of healthy patients are correctly diagnosed?

   (b) What is the *unconditional* (or marginal) probability of a positive test result?

   (c) What is the probability of having the disease given a positive test result?

   (d) Compare your result in part (c) to the probability of detecting the disease when it is present. Why do these probabilities differ?

5. In a basketball tournament, suppose that the probability that the 'better' team wins is 0.6. Given that there are 32 teams and that only the winning team goes through after each game, what is the probability that the best team wins the tournament?

6. Show that $P(A \cap B) - P(A)P(B) = P((A \cup B)^c) - P(A^c)P(B^c)$
7. Show that if A is independent of itself then either $P(A) = 1$ or $P(A) = 0$
8. You deal yourself two cards from a conventional pack and your opponent two cards. Your opponent reveals that the sum of his/her two cards is 21; what is the probability that the sum of your two cards is 21? (treat picture cards as 10 and aces as 11)
9. Consider a genome sequence of length 3×10^9 nucleotides (represented by the letters A,C,G and T). Suppose that 10×10^6 mutations occur in this sequence, entirely at random (a mutation changes a nucleotide to another nucleotide). Estimate the number of positions in the genome sequence at which more than one mutation will have occurred (ideally without using a calculator).
10. Suppose that in a given year a man who has at least one teenage daughter is twice as likely to have a nervous breakdown as a man with no teenage daughters. A man has four teenage children and a nervous breakdown. What is the probability that he has at least one daughter?