Universität Bielefeld
Technische Fakultät, AG Genominformatik
WinterSemester 2007/2008

Advanced Algorithmic Techniques for Bioinformatics

Prof. Dr. Ferdinando Cicalese, Dr. Zsuzsanna Lipták
Dipl.-Inform. Roland Wittler

Exercise sheet no. 4 (out: 5 Dec. 2007, in: 12 Dec. 2007)

1. **Markov Chains.**

   Suppose that Cola-drinkers purchase once a week either a Coke or a Pepsi.

   Given that a person's last cola purchase was Coke, there is a 90% chance that her next cola purchase will also be Coke.

   If a person's last cola purchase was Pepsi, there is an 80 % chance that her next cola purchase will be Pepsi.

   Model this situation as a Markov Chain.

   Compute the probability that a person that is currently a pepsi purchaser will buy a coke two weeks from now.

   Assume 60% of all people now drink Coke, and the rest drinks Pepsi. What is the percentage of people that will be drinking Coke 3 weeks from now?

2. **Bayes' Theorem**

   A rare genetic disease is discovered. Although only one in a million people carry it, you consider being screened. You are told that the genetic test is extremely good: it is 100% sensitive (it is always correct if you have the disease) and 99.99% specific (it gives a false positive result only 0.01% of the time). Using Bayes' theorem, explain why you might decide not to take the test.

3. **Random Variables** Let $S = \{1, 2, \ldots, 8\}$. Let X be an element of S chosen randomly in accordance to the following distribution $Pr(X = 1) = 1/2$, $Pr(X = 2) = 1/4$, $Pr(X = 3) = 1/8$, $Pr(X = 4) = 1/16$, $Pr(X = 5) = 1/32$, $Pr(X = 6) = 1/64$, $Pr(X = 7) = 1/128$, $Pr(X = 8) = 1/128$.

   What is the average number of questions performed by someone who tries to find out X by using the classical binary search algorithm?

   Can you find an algorithm with a better average-case performance?

   [Consider the random variable $Y : S \to \mathbb{N}$ that for each outcome x for X counts the number of questions necessary to find out that $X = x$.]