Assignment 1

CSL 858

Due date: January 16, 2007 (Tuesday)

Topics: Probability, Entropy

1. In this problem we try to understand why people waste money buying lottery tickets although their chances of winning maybe near-zero. A man pays Rs. 1 to buy a lottery ticket. If he wins the lottery he will receive Rs. 10^6. The probability of his winning is 10^{-7}. What is the expected value of the money earned? Does the man on average gain money or lose money?

   Now consider the satisfaction of the man which we call random variable S. The satisfaction is modeled as $S = X^3$ where X is the *net* money earned from the lottery. Compute the expected value of satisfaction. Note that the man buys a lottery ticket if this expected value is positive.

2. Say that you are given a random variable U which is uniformly distributed between 0 and 1. How would you transform U to get a random variable G which has cumulative distribution function F_G?

3. (Contact the instructor if you have doubts about the rules of the following game)

   In the game of tennis, a player is allowed a maximum of two serves per point. If the first serve goes "in" then he is not allowed to serve a second time; the play resulting from the first serve decides if he wins or loses the point. However, if the first serve goes "out" then he can serve a second time and the play resulting from the second serve decides if he wins or loses the point.

   Mahesh Bhupati has collected the following information about his tennis serve. He can serve at two speeds: 120mph and 90mph. The probability of his 120mph serve going "in" is $p_1 = 0.6$ and that of his 90mph serve going "in" is $p_2 = 0.9$. Conditioned on the serve being "in", the probability that he wins the play resulting from his 120mph serve is $q_1 = 0.9$ and that from his 90mph serve is $q_2 = 0.7$. What strategy for serving must Mahesh choose to maximize the probability that he wins the point? That is, what speed must he choose for his first serve and what speed for his second serve?

   What if Mahesh is out of form and p_1 drops to 0.2? Must he change his strategy now?

4. There are n persons in a room. Assume that each person's birthday is randomly chosen among the 365 days of the year (do not consider leap years for this problem). What is the probability that no two persons have the same birthday for the cases $n = 5$, $n = 15$, $n = 20$, $n = 23$?

5. Aliens in a remote galaxy use an alphabet consisting of the following letters: a,z,x,u,e. Different tribes use the same alphabet to speak different languages. Compute the Shannon Entropy of the following languages in units of bits/letter (or bits/word).

   (a) Language "Z22" : The different letters occur independently and with probability 0.2, 0.1, 0.3, 0.15, 0.25 respectively.

   (b) Language "C–" : The different letters occur independently and with probability 0.2, 0.2, 0.2, 0.2, 0.2.

(c) Languate "J2D" : This language consists of only 3 words: axz, zueae, zeu. These 3 words occur independently with probabilities 0.2, 0.7, 0.1.

Which language has the highest entropy? Why (give a sentence or two with your intuition)?

Which language has the lowest entropy? Why?

6. Shannon Entropy is a measure of uncertainty of an information source. We will first quantify the uncertainty of a source by playing the game of 20-questions.

   Suppose a source has 5 symbols, A, B, C, D, E, with corresponding probabilities of occurrence $P(A) = 1/2, P(B) = 1/4, P(C) = 1/8, P(D) = 1/16, P(E) = 1/16$. Let us say that the source has generated a symbol. You are allowed to ask a few questions to the source to find out what the symbol is. All questions must be of the type "Is the symbol among the set of symbols S?" and the source will answer yes or no (similar to giving you one bit of information). Your task is to choose the set S intelligently to minimize the number of questions you ask. Note that S can change from one question to the next. Design a strategy for choosing S and compute the expected number of questions you will ask the source. (For example, your first question can be: "Is the symbol in '$S = \{B, D\}$'?)

   Compare the expected number of questions to the entropy of the source.