CS 683 — Spring '07 **Learning, Games, and Electronic Markets**

Homework #1

Due February 2, 2007.

This homework assignment will be graded pass-fail. Its main purpose is to assess your mathematical background for the course. You are not expected to solve all of the questions; some of them are quite difficult, while others should be readily solvable if you have mastered the prerequisites for the course. You must work on this assignment by yourself. (On future assignments students will be allowed to work in groups.) Turn in your solutions in class on Friday, February 2, or email them to Professor Kleinberg on that day.

1. Suppose x and y are real numbers and $x > y$. Prove that $e^x > \frac{e^x - e^y}{x-y} > e^y$.

2. A standard 52-card deck[1] is randomly partitioned into four 13-element sets, which are dealt to players named North, South, East, and West.

   (a) Calculate Pr(South gets exactly 2 aces | North gets exactly 1 ace).

   (b) Let H and D denote the number of hearts and diamonds, respectively, dealt to North. Calculate $\mathbf{E}(H \mid D)$ as a function of D.

3. Let $x_1, x_2, \ldots, x_n$ be independent uniformly-distributed random samples from the interval $[0, 1]$. Define the following probabilities:

   - $p(n)$ is the probability that $\min_k x_k > 0.01$.
   - $q(n)$ is the probability that $\min_{i \neq j} |x_i - x_j| < \frac{1}{n^2}$.
   - $r(n)$ is the probability that $\min_{i \neq j} |x_i - x_j| > \frac{1}{100n}$.
   - $s(n)$ is the probability that exactly $\lfloor n/2 \rfloor$ of the numbers $x_1, \ldots, x_n$ lie in $[0, \frac{1}{2}]$.

   Estimate the asymptotic behavior of each of these probabilities as n tends to infinity. Specifically, for each of the sequences $p(n), q(n), r(n), s(n)$, determine whether the sequence

   (A) tends to zero exponentially fast, i.e. is bounded above by c^n for some constant $c < 1$, for all sufficiently large n;

   (B) tends to zero, but not exponentially fast;

   (C) remains bounded away from 0 and 1;

   (D) tends to 1.

   You do not need to justify your answer; simply specify choice A, B, C, or D.

[1] A *standard 52-card deck* is the set
$\{2, 3, 4, 5, 6, 7, 8, 9, 10, \text{jack}, \text{queen}, \text{king}, \text{ace}\} \times \{\text{clubs}, \text{diamonds}, \text{hearts}, \text{spades}\}$.

4. Let x be a real-valued random variable which is exponentially distributed with decay rate 4, i.e. $\Pr(x > r) = e^{-4r}$ for all $r > 0$.

   (a) What is the probability density function of x?

   (b) What is the expected value of x?

   (c) Let $y = \sqrt{x}$. What is the probability density function of y?

5. (a) You are shopping for a house by observing a sequence of houses one by one, in random order. You have decided to stop and buy a house as soon as you see one which is nicer than the first house you observed. What is the expected number of houses you will have to look at, including the first one? Mathematically, let's model this process as follows. Let $z_1, z_2, \ldots$ denote an infinite sequence of independent uniformly-distributed random samples from the interval $[0, 1]$. (Interpretation: z_i is the "quality" of the i-th house you observed.) Let τ be the smallest $i > 1$ such that $z_i > z_1$. What is $\mathbf{E}(\tau)$?

   (b) Now suppose that you modify your stopping rule. For some fixed predetermined number k, you observe the first k houses without purchasing. Let h be the second-best house observed among the first k. Your policy is to buy the next house (after the first k) which is nicer than h. In more precise terms, let $z_1, z_2, \ldots$ be a sequence of independent random variables uniformly distributed in $[0, 1]$ as before, let $z_a > z_b$ be the two largest elements of the set $\{z_1, \ldots, z_k\}$, and let ρ be the smallest $i > k$ such that $z_i > z_b$. What is $\mathbf{E}(\rho)$, as a function of k?

6. (a) Find a non-zero vector $\vec{v} = (x, y, z)$ such that $\vec{v}$ is a linear combination of $(0, 1, -1)$ and $(1, 0, 0)$, and $\vec{v}$ is also a linear combination of $(-1, 2, 0)$ and $(1, 1, 1)$.

   (b) Let
   $$S = \{(x, y, z) \mid x + y + z = 5, x \geq 0, y \geq 0, z \geq 0\}.$$
   Let $\vec{v} = (13, 16, 6)$. Find the point in S which is closest to $\vec{v}$, i.e. the $\vec{w} \in S$ which minimizes $\|\vec{v} - \vec{w}\|_2$.

7. Recall the prediction problem we discussed on the first day of class: there are n experts predicting a binary sequence $B_1, B_2, \ldots$, and one of the experts never makes a mistake. (In other words, if b_{ij} denote the prediction of expert i on the j-th trial, we are assuming that there exists some i ($1 \leq i \leq n$) such that $b_{ij} = B_j$ for all j.) We saw that there is a deterministic prediction algorithm which makes at most $\lfloor \log_2(n) \rfloor$ mistakes, and that this mistake bound is optimal.

   (a) Show that there is a randomized prediction algorithm whose *expected* number of mistakes is at most $\frac{1}{2} \log_2(n)$.

   (b) Show that no randomized prediction algorithm can make fewer than $\lfloor \frac{1}{2} \log_2(n) \rfloor$ mistakes, in expectation.