Final Take-Home

(This is an open-notes, open-book, and open-internet final. You may not interact with anyone about the material on the exam until after 12/12/13—interaction includes posting questions on piazza, internet forums, etc. The only exception are questions emailed to both the instructor and TAs, and there will be no office hours held until after the exam is due on 12/11/13.)

Electronic submission is strongly encouraged. Email as attachment to psetCS265@gmail.com). If submitting a hard copy, there is a box on the 1st floor of Gates Building, by the East entrance, labelled CS265/CME309. Submissions must be received by 2pm Wednesday 12/11/13.

Please include your full name, and SUID. A small number of points may be subtracted for “poor style”, which includes excessively long/convoluted writeups, or the inclusion of false statements within the context of a correct proof.

1. Recall the “martingale stopping theorem”, which states that for a martingale $\{Z_i\}$ with respect to $\{X_i\}$, then if $T$ is a stopping time for $\{X_i\}$, then $E[Z_T] = E[Z_0]$ provided at least one of the following conditions hold:

   - There exists a constant $c$ s.t. $|Z_i| \leq c$ for all $i$.
   - There exists a constant $c$ s.t. $T < c$.
   - $E[T] < \infty$, and there exists a constant $c$ s.t. $E[|Z_{i+1} - Z_i| |X_1, \ldots, X_i] < c$.

   (a) Give an example of a martingale and stopping time for which $E[Z_T] \neq E[Z_0]$ but for which $E[T] < \infty$. Briefly justify why your example satisfies the desired conditions.

   (b) Give an example of a martingale and stopping time for which $E[Z_T] \neq E[Z_0]$ but for which $E[|Z_{i+1} - Z_i| |X_1, \ldots, X_i] < 1$. Briefly justify why your example satisfies the desired conditions.

2. In class we commented that an unbiased one-dimensional random walk will cross every point infinitely many times (with probability 1). Here, we show that this is not the case for biased random walks. Consider a gambling game where at each time step, you win one dollar with probability $p \geq 3/4$, and lose a dollar with probability $1-p$. Prove that if one starts playing with $1000$, with probability at least $1/2$, one will be able to play the above gambling game forever without ever being bankrupt. Note that the amount $1000$ and probability $1/2$ can be tightened significantly. [Hint: one can either do this problem by solving some recurrence relationship, or one can use Chernoff bounds as follows: first show that after playing for 2000 time steps, with probability at least $3/4$ one will have $\geq 1000 + 500 = 1500$. Conditioned on this, show that after playing for an additional 3000 time steps, with probability at least $7/8$ one will have $\geq 1500 + 3000/4 = 2250$. Given this, show that after playing for an additional 4500 time steps with probability at least $15/16$ one will have $\geq 2250 + 4500/4 = 3375$, etc.]
3. Suppose that we have a class with \( n \) students, and that every student “likes” exactly \( k \) other students, and is “liked” by exactly \( k \) students.

(a) Prove that there exists a way to partition the class into \( \lfloor \frac{k}{3 \log k} \rfloor \) groups, such that each person is in a group that contains at least one person that he/she “likes”. [Hint: Use the Lovasz Local Lemma, in which case make sure you define some probability space to which you apply the lemma....]

(b) Describe a polynomial-time (in expectation) randomized algorithm for finding such a partition (assuming, of course, that you know who likes who). [Just describe the algorithm, no need to justify the run-time.]

4. Suppose you have a flock of sheep, with \( w \) white sheep and \( b \) black sheep, and you randomly select \( n \leq w + b \) sheep to take to market.

(a) Letting the random variable \( X \) denote the number of white sheep that are sent off to market, what is \( E[X] \) as a function of \( n, w, \) and \( b \)?

(b) Prove that \( \Pr[|X - E[X]| \geq \lambda] \leq 2e^{-\frac{\lambda^2}{2}} \). Note that the stronger statement \( \Pr[|X - E[X]| \geq \lambda] \leq 2e^{-\frac{\lambda^2}{2n}} \) is true, though you don’t need to prove it. [Hint: Azuma’s inequality!!! Make sure that if you use Azuma’s inequality, you actually define a martingale, and justify why it is a martingale, etc.]

5. Consider the following stochastic process that, in some sense, captures the “rich get richer” phenomena found in many real world settings: There is a room with an infinite number of lunch tables. At time \( t = 1 \), one person sits down at a randomly chosen table. For all \( i > 1 \), let \( k \) denote the number of non-empty tables before the \( i \)th person is seated and let \( n_1, \ldots, n_k \) denote the number of people sitting at each of these tables; the \( i \)th person will choose a table according to the following probability distribution: choose table \( j \) (that has \( n_j \) people at it) with probability \( \frac{n_j}{1 + \sum_{\ell=1}^{k} n_\ell} \), and with probability \( \frac{1}{1 + \sum_{\ell=1}^{k} n_\ell} \) choose an empty table.

(a) Let \( S_i \) denote the number of non-empty tables after the \( i \)th person has been seated. Show that \( E[S_i] = \sum_{j=1}^{i} \frac{1}{j} \).

(b) Prove that \( \Pr[|S_i - E[S_i]| \geq \lambda] \leq 2\sqrt{\sum_{j=1}^{i} \frac{1}{j}} \leq \frac{1}{4} \).

(c) Prove that there is some positive constant \( p > 0 \) such that with probability at least \( p \), the table chosen by the first person will be the most popular table for all time steps if the table-choosing process were continued indefinitely. [This is a powerful result—the moral is that it is often a good idea to be the first one to a party, whether that party is a new research area, new startup, new electronic currency, or your neighbor’s new year’s eve party.] [Hint: Imagine that the first person is lucky, and the first 2000 people all join his table. This occurs with some (small, but constant) probability. Now argue that given this promising start, with probability at least \( 1/2 \) this table will maintain its dominance for all time by relating this process to that of Problem 2.]

(d) Letting \( T_i \) denote the size of the table chosen by the first person after \( i \) time steps, what is the variance of \( T_i \) as a function of \( i \)? Prove your answer, though you do not worry about the constants, the asymptotic “big-O” dependence is sufficient (e.g. \( Var[T_i] = \Theta(\log(i)) \), or \( Var[T_i] = \Theta(i^2) \), etc.).