CS 174: Combinatorics and Discrete Probability	Spring 2015

Final

Tuesday, May 12, 2015

Notes: Please read all the questions carefully and ask for any clarifications. Some questions are harder than others, so you may want to solve the easier questions first. If you are stuck anywhere, move on to a different question and come back to where you were stuck later.

Please write clearly and legibly, and manage your space wisely!

Good Luck!

Name:

SID:

Score (for instructor use)

1.1		1.2		1.3	
2.1		2.2		2.3	
3.1		3.2		3.3	
4.1		4.2		4.3	

Some Useful Definitions and Results

Markov's Inequality. Let X be a random variable that assumes only nonnegative values. Then, for all $a > 0$,

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Chernoff Bound for Sums of Bernoullis. Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with mean p. Let $X = \sum_{i=1}^{n} X_i$. Then for any $0 < \delta \leq 1$,

$$P(X \geq (1+\delta)np) \leq e^{-np\delta^2/3}.$$

Azuma-Hoeffding. Let $X_0, \ldots, X_n$ be a martingale such that

$$B_k \leq X_k - X_{k-1} \leq B_k + d_k$$

for some constants d_k and for some random variables B_k that may be functions of $X_0, \ldots, X_{k-1}$. Then, for all $t \geq 0$ and for any $\lambda > 0$,

$$P(|X_t - X_0| \geq \lambda) \leq 2e^{-2\lambda^2/(\sum_{k=1}^{t} d_k^2)}.$$

Martingales. A sequence of random variables $Z_0, Z_1, \ldots$ is a martingale with respect to the sequence $X_0, X_1, \ldots$ if, for all $n \geq 0$,

- Z_n is a function of $X_0, X_1, \ldots, X_n$,
- $\mathbb{E}[|Z_n|] < \infty$, and
- $\mathbb{E}[Z_{n+1} | X_0, \ldots, X_n] = Z_n$.

Stopping Times. A nonnegative, integer-valued random variable is a stopping time for the sequence $\{Z_n, n \geq 0\}$ if the event $T = n$ depends only on the value of the random variables $Z_0, Z_1, \ldots, Z_n$.

Markov Chains. The transition matrix P of a Markov chain $\{X_t\}$ is defined so that $P(X_{t+1} = j \mid X_t = i) = P_{ij}$. The row vector π is a stationary distribution if it is a probability distribution and if $\pi P = \pi$.

Problem 1.
Let $G = (E, V)$ be an undirected graph and suppose that $|V| = 2n$. Let S be a subset of V chosen uniformly at random. Let E_S be the subset of edges in E that have exactly one vertex in S.

(a) Compute $\mathbb{E}[|E_S|]$, the expected size of E_S, in terms of $|E|$.

(b) Using the probabilistic method, prove that there exists some subgraph $H = (E', V')$ such that H is bipartite and $|E'| \geq |E|/2$. By subgraph, we mean that $E' \subseteq E$, $V' \subseteq V$, and $u, v \in V'$ whenever $(u, v) \in E'$.

Now let T be a subset of V chosen uniformly at random from all subsets of size n. Let E_T be the subset of edges in E that have exactly one vertex in T.

(c) Compute $\mathbb{E}[|E_T|]$

(d) Argue that there exists some subgraph $F = (E'', V'')$ such that F is bipartite and $|E''| \geq |E| f(n)$ for some function f, where $f(n) > 1/2$.

(e) Compare the results from parts (b) and (d). Which one is better?

Answer here:

Use this page if you need more space

Problem 2.
Below, P and Q denote the transition matrices of two different Markov chains, and π and ϕ denote distributions over states.

(a) Suppose that π and ϕ are both stationary distributions of P. Assuming $\lambda \in [0, 1]$, is $\lambda\pi + (1 - \lambda)\phi$ also a stationary distribution of P?

(b) Suppose that π is a stationary distribution for both P and Q. Assuming $\lambda \in [0, 1]$, is π a stationary distribution of $\lambda P + (1 - \lambda)Q$?

(c) Is π a stationary distribution of PQ (the product of the matrices P and Q)?

Consider a Markov chain $X_0, X_1, \ldots$ with transition matrix P and unique stationary distribution π. Suppose that the distribution of X_0 is π.

(d) If n is a positive integer, what is the distribution of X_n.

(e) If T is a stopping time for the sequence $X_0, X_1, \ldots$, is it necessarily the case that the distribution of X_T is π?

Answer here:

Use this page if you need more space

Problem 3.

In the 0th generation, there is 1 amoeba. At every time step, each amoeba independently either splits into two amoeba (with probability p) or dies (with probability $1 - p$). Let X_n be the number of amoeba alive at time n.

(a) Compute $\mathbb{E}[X_n | X_{n-1}]$.

(b) Compute $\mathbb{E}[X_n]$.

(c) We say that the population goes extinct if $X_n = 0$ for any n. Suppose that $p < \frac{1}{2}$. Show that the population goes extinct with probability 1. **Hint:** use Markov's inequality.

Answer here:

Use this page if you need more space

Problem 4.
In this problem, let $X_0, X_1, \ldots$ and $Z_0, Z_1, \ldots$ be sequences of random variables.

(a) Show that if $Z_0, Z_1, \ldots$ is a martingale with respect to $X_0, X_1, \ldots$, then it is also a martingale with respect to itself.

(b) Suppose that $Z_n = f(X_n)$ for all n and for some function f. If T is a stopping time with respect to the sequence $X_0, X_1, \ldots$, is it necessarily a stopping time with respect to the sequence $Z_0, Z_1, \ldots$?

Answer here:

Problem 5.
In the Bin Packing problem, we are given n real numbers $x_1, \cdots, x_n \in [0,1]$ and asked to partition the set $\{1, \cdots, n\}$ into k (non overlapping) sets $S_1, \cdots, S_k$ such that $\sum_{i \in S_j} x_i \leq 1$ for all $j = 1, \cdots, k$, and such that k is as small as possible. We denote by $B(X)$ the smallest such k, with $X = (x_i)_{i=1}^n$. The elements $i \in S_j$ are thought of as being placed in bin j which has capacity 1. Throughout the problem, we assume that $x_1, \cdots, x_n$ are generated independently and uniformly at random on the interval $[0,1]$.

(a) Let $\mu = \mathbb{E}[B(X)]$. Prove that $\mu \geq n/2$.

(b) Use the bounded differences inequality (Azuma-Hoeffding) to prove that

$$\Pr\left(B(X) - \mu \geq t\right) \leq e^{-2t^2/n}.$$

By taking $t = \sqrt{n \log n}$, this implies that

$$B(X) \leq n/2 + \sqrt{n \log n}$$

with probability at least $1 - 1/n^2$.

Now we describe a bin packing algorithm `FOLD` for which the expected number of bins used is at most $n/2 + c\sqrt{n \log n}$: let $\alpha = 1 - 6\sqrt{\frac{\log n}{n}}$.

1. Place each element $x_i \geq \alpha$ into a bin on its own. Suppose there are B_1 such elements.
2. Let $N = n - B_1$ be the number of the items remaining to be placed.
3. Order the items so that $x_1 \leq x_2 \leq \cdots \leq x_N \leq \alpha$.
4. For $i = 1, 2, \cdots, \lfloor N/2 \rfloor$
   - Put x_i and x_{N-i+1} into one bin if $x_i + x_{N-i+1} \leq 1$.
   - Put x_i and x_{N-i+1} into separate bins otherwise.
5. Put the item $x_{\lfloor N/2 \rfloor}$ into a separate bin if N is odd.

(c) Prove that the number of bins used by this algorithm can be written as

$$B = B_1 + N - \sum_{i=1}^{N/2} \mathbf{1}\{x_i + x_{N-i+1} \leq 1\}.$$

(d) Compute $\mathbb{E}[B_1]$.

(e) By using a tail bound on the binomial distribution, show that

$$\Pr\left(x_i \geq \frac{i + \sqrt{\frac{n}{2}\log n}}{n}\right) \leq 1/n.$$

(f) Assume that

$$\Pr\left(x_{N-i+1} \geq \frac{n - i - \sqrt{\frac{n}{2}\log n}}{n}\right) \leq 1/n$$

which can be proved in the same way. Show now that

$$\Pr(x_i + x_{N-i+1} > 1) \leq 2/n.$$

(g) Assume that N is even (i.e. ignore step 5 of the algorithm). Show that

$$\mathbb{E}[B] \leq n/2 + 3\sqrt{n \log n} + 1.$$

Answer here:

Use this page if you need more space

Problem 6.
Consider a Markov chain on n points $S = \{0, 1, \ldots, n-1\}$ lying in order on a circle. At each step, the chain stays at the current point with probability $1/2$ or moves to the next point in the clockwise direction with probability $1/2$.

(a) Find the stationary distribution π of this Markov chain.

(b) Define a Markov chain $Z_t = (X_t, Y_t)$ on $S \times S$. To update Z_t, we consider two cases. If $X_t \neq Y_t$, then both X_t and Y_t independently either stay the same or move to the next point (each with probability $1/2$). If $X_t = Y_t$, then either they both stay the same (with probability $1/2$) or they both move to the next point (with probability $1/2$). Prove that this is a coupling of the original Markov chain.

(c) Let $D_t = X_t - Y_t \bmod n$. Describe the transition probabilities $P(D_{t+1}|D_t)$. You will need to handle the cases $D_t = 0$ and $D_t \neq 0$ separately.

(d) Define the stopping time T to be the first t such that $D_t = 0$. You may use the fact that $\mathbb{E}[T] \leq n^2/2$ (this was essentially shown in the Gambler's ruin problem in the homework). Prove that

$$P(T \geq n^2) \leq \frac{1}{2}.$$

(e) Prove that for positive integer k,

$$P(T \geq kn^2) \leq \left(\frac{1}{2}\right)^k.$$

(f) Show that for any $x, y \in S$ and $t \geq n^2 \log_2(1/\epsilon)$ we have

$$P(X_t \neq Y_t \,|\, X_0 = x, Y_0 = y) \leq \epsilon.$$

This implies (by the "coupling lemma"), that $\tau(\epsilon) \leq n^2 \log(1/\epsilon)$, where $\tau(\epsilon)$ is the mixing time.

Answer here:

Use this page if you need more space

Use this page if you need more space