CS 174: Combinatorics and Discrete Probability Fall 2012

Midterm 2 – Solutions

Problem 1. *(Total 20 points)*

For the questions 1-4, read the statement of the problem carefully and circle either TRUE or FALSE. You will get 2 points for a correct answer and 0 points for an incorrect answer.

For questions 5-7, one or more of the choices may be correct. For each choice you will get 1 point if your selection of the choice is correct. Thus, each of the questions is worth 4 points total. There is no negative marking.

1. Let X_n be a sequence of random variables that depend on some object of size n (e.g. number of independent sets in $G_{n,1/2}$). Then, if it is the case that $\lim_{n\to\infty} \mathbb{E}[X_n] = \infty$, it must be the case that $\lim_{n\to\infty} \Pr[X_n = 0] = 0$.

   TRUE FALSE

   Solution. The correct answer is FALSE. It is possible to have a random variable, X, with $\Pr[X = 0] = 0.99$ and yet, $\mathbb{E}[X] = \infty$.

2. A finite Markov Chain may have infinitely many stationary distributions.

   TRUE FALSE

   Solution. The correct answer is TRUE. If the transition graph is disconnected, then the number of stationary distributions may be infinite. For example, in the Markov chain in Problem 1-5.

3. Let ϕ be a 3-SAT formula, in which each clause has *exactly* 3 variables, and any clause shares a variable with at most 2 other clauses. Then ϕ always has a satisfying assignment.

   TRUE FALSE

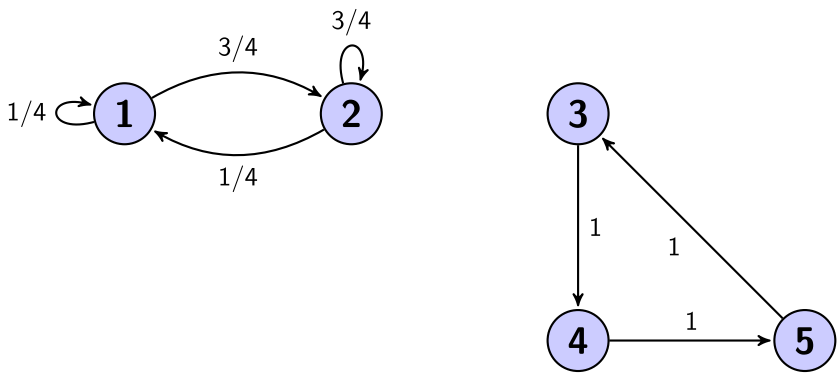
   Solution. The correct answer is TRUE. Let E_i be the event that the i^{th} clause is not satisfied by a random assignment. Then, $p = \Pr(E_i) = 1/8$. The maximum degree of the dependency graph on events E_i is 2 (since each clause shares a variable with at most 2 other clauses). Thus, $4pd \leq 1$ and by Lovasz Local Lemma, it is guaranteed that $\Pr(\cap_i \bar{E}_i) > 0$, and hence there must be a satisfying assignment.

4. If a state, i, of a Markov Chain has period 3, then there must be a cycle containing i of length 3 in the transition graph.

   TRUE FALSE

Solution. The correct answer is FALSE. It is possible, for example, that there is a cycle of length 6 and another of length 9, containing the state i. The period of state i would still be 3.

5. Observe the Markov chain in the figure below. Which of the following are valid stationary distributions? More than one answer may be correct.



   (a) $(1/5, 1/5, 1/5, 1/5, 1/5)$
   (b) $(1/4, 3/4, 0, 0, 0)$
   (c) $(0, 0, 1/3, 1/3, 1/3)$
   (d) $(1/12, 1/4, 2/9, 2/9, 2/9)$

   Solution. The correct answers are (b), (c), and (d). This can be verified easily. Alternatively, if we restrict attention to states 1 and 2, the stationary distributions is $(1/4, 3/4, 0, 0, 0)$ and if we restrict attention to states 3, 4, and 5, the stationary distribution is $(0, 0, 1/3, 1/3, 1/3)$. Then, any convex combination of the above two states is also a stationary distribution.

6. Which of the following statements is true? You are given a graph, G. Then some edges are added to G, by some (not necessarily random) process, to obtain a graph G'.

   (a) The cover time of G' is larger than the cover time of G.
   (b) The cover time of G' is smaller that the cover time of G.
   (c) The cover time of G' is the same as the cover time of G.
   (d) The cover time of G' may be larger or smaller than G depending on how the edges are added.

   Solution. The only correct answer is (d). It is indeed possible to increase the cover time by adding edges. For example, start with the line graph, (a path of length n), and add edges to make this into a lollipop graph. Then the cover time increases from $\Theta(n^2)$ to $\Theta(n^3)$. On the other hand if the line graph is made into a complete graph on n vertices, the cover time decreases from $\Theta(n^2)$ to $\Theta(n \log(n))$. Thus, (a), (b), and (c) are all incorrect.

7. X is a random variable that takes positive integer values. Let a be an integer. Which of the following statements suffice to prove that $\Pr[X = a] > 0$.

(a) $\Pr[X \geq a] > 0$ and $\Pr[X \leq a] > 0$.

(b) $\mathbb{E}[X] = a$.

(c) $\mathbb{E}[X] = a + (1/4)$ and $\mathsf{VAR}(X) = 1/2$.

(d) $\Pr[X > a] < 1/2$ and $\Pr[X < a] < 1/2$.

Solution. The correct answers in this case are (c) and (d). It is easy to check that (a) and (b) are not sufficient conditions. To see that (c) is sufficient, apply Chebychev's inequality to show that $\Pr[|X - \mathbb{E}[X]| > 0.74] < 1$, and hence $\Pr[|X - a| < 0.74] > 0$. But there is only one integer value in the range $[a - 0.74, a + 0.74]$, which is a. So it must be the case that $\Pr[X = a] > 0$. The fact that (d) is sufficient is obvious.

Problem 2. *(Total 20 points)* For a graph $G = (V, E)$, for any subset of the vertices, $V' \subseteq V$, the *induced subgraph* on V' is $G' = (V', E')$, where $E' = \{e = (i, j) \mid e \in E \text{ and } i \in V', j \in V'\}$. Thus, any edge in the original graph that has both endpoints in V' is included in the induced subgraph. Let G be a random graph generated according to the $G_{n,p}$ model.

(a) Find the expected number of induced 4-cycles in the graph.

(b) Find the value of p for which the expected number of induced 4-cycles is 1, in the limit as $n \to \infty$.

Solution. (a) Let $X_1, \ldots, X_{\binom{n}{4}}$, be the random variables corresponding to all possible subsets of size 4 of the set of vertices. Let $X_i = 1$, if the induced subgraph on the corresponding four vertices is a 4-cycle. Note, that given 4 vertices, say labelled A, B, C, D, there are exactly 3 possible 4-cycle graphs — (A, B, C, D), (A, C, B, D) and (A, B, D, C). In each case, exactly 4 edges must be present and the remaining 2 absent. Thus, $\Pr[X_i = 1] = 3p^4(1 - p)^2$.

Now, if $X = \sum_i X_i$, then $\mathbb{E}[X] = \binom{n}{4} 3p(1 - p)^2$ gives the expected number of induced 4-cycles.

Remark: A lot of you missed the fact that there are three possible 4-cycle graphs, given 4 vertices.

(b) Here there are two possibilities: If p is very small, that is $p \to 0$ as $n \to \infty$, then the graph is sparse, making induced 4-cycles unlikely. Then, we have

$$\lim_{n \to \infty} \mathbb{E}[X] = \frac{n^4 p^4}{8}$$

Solving for p, we get $p = \sqrt[4]{8}/n$.

There is another possibility: If p is large, in the sense that $p \to 1$ as $n \to \infty$. In this case, the graph is very dense, again making induced 4-cycles unlikely. We have:

$$\lim_{n \to \infty} \mathbb{E}[X] = \frac{n^4 (1 - p)^2}{8}$$

Solving for p, we get $p = 1 - (2\sqrt{2}/n^2)$.

Remark: No points were deducted if you only obtained one value of p.

Problem 3. *(Total 20 points)* Consider a random walk (drunkard's walk) in n dimensions. The starting point is $(0, \ldots, 0)$ and at each step you choose one of the n co-ordinates randomly and add either $+1$ or -1 to it. You may assume that when $n = 3$, $(0, 0, 0)$ is not a recurrent state. Using this fact, show that for $n \geq 4$, $(0, \ldots, 0)$ is not a recurrent state. (The proof is expected to be formal, though you will get some partial credit for writing the correct high-level idea.)

Solution. *High-level Idea*: Consider the random walk in n-dimensions – this can be thought of as a (random) infinite string over an alphabet of size $2n$, say using $(i, +1)$ and $(i, -1)$, for $1 \leq i \leq n$. If we consider the substring corresponding to the movement in the first 3 dimensions, then this is a (random) *infinite* string on an alphabet of size 6, with $(1, +1), (1, -1), (2, +1), (2, -1), (3, +1), (3, -1)$ – and exactly corresponds to a random walk in 3 dimensions. But, if the original random walk in n dimensions returned to $(0, \ldots, 0)$ with probability 1, then so must the restriction to 3 dimensions.

Remark: Some credit was given for stating any reasonable proof idea. The closer your verbal description was to a formal proof, the more points you got.

Now, for the formal proof. Let $p^{t,k}$ denote the probability that a random walk in k dimensions, starting at $(0, \ldots, 0)$ returns to $(0, \ldots, 0)$ after t time-steps. Note that $p^{0,k} = 1$ for all k.

Then consider,

$$p^{t,n} = \sum_{t'=0}^{t} p^{t',3} \cdot p^{t-t',n-3} \cdot \Pr[t' \text{ out of } t \text{ steps were in the first 3 dimensions}]$$

Note that, $\Pr[t' \text{ out of } t \text{ steps were in the first 3 dimensions}] = \binom{t}{t'}(3/n)^{t'}(1 - 3/n)^{t-t'}$. Also, conditioned on t' out of t steps being the first 3-dimensions, the random walks in the first 3 dimensions and the remaining $n - 3$ dimensions are independent. Summing both sides over $t \geq 0$, we get

$$\begin{aligned}
\sum_{t=0}^{\infty} p^{t,n} &= \sum_{t=0}^{\infty} \sum_{t'=0}^{t} p^{t',3} p^{t-t',n-3} \binom{t}{t'} \left(\frac{3}{n}\right)^{t'} \left(1 - \frac{3}{n}\right)^{t-t'} \\
&= \sum_{t'=0}^{\infty} p^{t',3} \left(\sum_{t=t'}^{\infty} p^{t-t',n-3} \binom{t}{t'} \left(\frac{3}{n}\right)^{t'} \left(1 - \frac{3}{n}\right)^{t-t'} \right)
\end{aligned}$$

Since $p^{t-t',n-3} \leq 1$, we get

$$\sum_{t=1}^{\infty} p^{t,n} \leq \sum_{t'=0}^{\infty} p^{t',3} \left(\sum_{t=t'}^{\infty} \binom{t}{t'} \left(\frac{3}{n}\right)^{t'} \left(1 - \frac{3}{n}\right)^{t-t'} \right)$$

Now, we claim that $\sum_{t=t'}^{\infty} \binom{t}{t'} \left(\frac{3}{n}\right)^{t'} \left(1 - \frac{3}{n}\right)^{t-t'} \leq \frac{n}{3}$. To see why: notice that $\binom{t}{t'} \left(\frac{3}{n}\right)^{t'+1} \left(1 - \frac{3}{n}\right)^{t-t'}$ (note the one extra power in the $(3/n)$ term) is the probability that the $t' + 1^{th}$ step of the random walk that is in one of the first three dimensions happens *exactly* after a total of t time-steps (this includes all n dimensions). Thus, $\sum_{t=t'}^{\infty} \binom{t}{t'} \left(\frac{3}{n}\right)^{t'+1} \left(1 - \frac{3}{n}\right)^{t-t'}$ is simply the probability that a total of $t' + 1$ steps are taken in the first 3 dimensions. Since this is a probability this can be at most 1 (in fact it is exactly 1 in this case). Thus, we have

$$\sum_{t=0}^{\infty} p^{t,n} \leq \sum_{t'=0}^{\infty} p^{t',3} \frac{n}{3}$$

Now, the RHS of the last expression is finite, by the assertion of the problem statement that the 3-d walk is not recurrent. But, then the LHS must also be finite, and hence the n-d walk is not recurrent.

Remark: Some of you have said that not being recurrent is implied by the fact that $p^{t,n} \to 0$ as $t \to \infty$. This is not true. This only means that the state is not *positive recurrent*. In particular, even in the 1-d and 2-d case, where the random walk is recurrent, this limit is 0.

Problem 4. *(Total 40 points)* Here we will consider the problem of finding sum-free subsets. Let $A = \{a_1, a_2, \ldots, a_n\}$ be a set of positive-integers. Let $S \subseteq A$. We say that S is sum-free if there do not exist integers, $x, y, z \in S$, such that $x + y = z$. (Note that the condition does not require the integers to be distinct, thus in particular $0 \notin S$ and there do not exist, x, z such that $2x = z$.)

(a) When $A = \{1, 2, \ldots, n\}$, show that there is a sum-free subset of size at least $n/2$.

Solution. Consider $A_1 = \{i \in A \mid i \text{ is odd}\}$ or $A_2 = \{\lfloor n/2 \rfloor + 1, \ldots, n\}$. It is easy to see that $|A_1| = |A_2| \geq n/2$. A_1 is sum-free because the sum of two odd numbers has to be even. A_2 is sum-free because $2(\lfloor n/2 \rfloor + 1) \geq n + 1$.

Remark: Points were deducted if you were not careful with floors/ceilings.

(b) In the rest of the problem, we will show that any set A, always has a sum-free subset of size at least $n/3$. Let $A = \{a_1, \ldots, a_n\}$ and suppose that a_n is the largest element.

Let $p > 2a_n$ be a prime such that $p \equiv 2 \pmod 3$ (*i.e.* p is of the form $3k+2$. You may assume that there are infinitely many primes of this form). Consider the set $F_p = \{0, 1, \ldots, p-1\}$. We say that a set $S \subseteq F_p$ is a sum-free *modulo* p, if there do not exist, $x, y, z \in S$, such that $x + y = z \pmod p$. Note that this is a stronger condition of *sum-free*ness. Show that F_p contains a sum-free *modulo* p subset of size at least $p/3$. (*Hint*: Consider the middle third elements of F_p.)

Solution. As stated above, let $p = 3k + 2$ be a prime. Let $S = \{k + 1, \ldots, 2k + 1\}$. Then $|S| = k + 1 > p/3$. If we add any two elements in S, as integers (not modulo p), the range for possible sums is $[2k + 2, 4k + 2]$. When we consider these modulo p, $2k + 2 > 2k + 1$, and $4k + 2 = (3k + 2) + k = k < k + 1$. Thus, none of the sums modulo p can fall in S. Hence, S is sum-free modulo p.

Remark 1: Again, if you were not careful with the exact range points have been deducted. In particular, writing $p/3$, $2p/3$ as if they were integers is not acceptable.

Remark 2: Some of you have ignored the fact that you were told that x and y (in the definition of sum-freeness) could be the same element. Hence, your answers were not exactly accurate.

(c) Use the observation in part (b) to show that $A = \{a_1, \ldots, a_n\}$ always has a sum-free subset of size $n/3$ (in integers). (*Hint*: Consider choosing a random element, r, from F_p. Then consider, $A_{r,p} = \{ra_1 \pmod p), ra_2 \pmod p), \ldots, ra_n \pmod p)\} \subseteq F_p$. Let $S \subseteq F_p$ be the subset obtained in part (b). Consider the set $A_{r,p} \cap S$. Then, use the method of expectations.)

Solution. Let $0 < a < p$, then if $r \in F_p$ is chosen uniformly at random, then $ar \pmod p$ is distributed uniformly at random in F_p. To see this, observe that for $r_1 \neq r_2$, $r_1 a \neq r_2 a \pmod p$. This is because, if $p | a(r_1 - r_2)$, then p must divide either a or $r_1 - r_2$, since p is prime; but both a and $r_1 - r_2$ are non-zero and strictly smaller than p. This implies that every distinct element in F_p when multiplied to a modulo p, gives a distinct element in F_p. And hence, $ar \pmod p$ is uniformly distributed in F_p, when r is chosen uniformly at random.

Now, notice that for any $a_i \in A$ and randomly chosen $r \in F_p$, the probability that $ra_i \pmod p$ lands in S (the middle-third) is at least $1/3$. Hence, $\mathbb{E}[|A_{r,p} \cap S|] \geq |A|/3$. Thus, by the method of expectations, there exists r for which $|A_{r,p} \cap S| \geq |A|/3$. For this value of r, let $A_s = \{a \mid ar \pmod p \in S\}$. Then, A_s must be sum-free since S is sum-free modulo p.