Polytechnic Institute of NYU Prof. Özgür Özkan CS6003—Fall 2013

Final Exam. December 14th, 2013

- You have four hours.
- Use a word processor to write your responses. Clearly label your solutions. Once you are done with the exam, submit the file as a File Response on Blackboard in one of the following formats: .doc, .docx, .pdf, .tex, .txt (if using Piazza)
- You can also use pen and paper and scan.
- If you are using Piazza LaTeX, then also submit a copy to Piazza as a private note.
- Write your name and ID number on the first page.
- Answer **all** questions.
- Explain your answers. If you do not show work, you will not get credit (except for multiple choice questions).
- 200 points total.
- Closed book, NO notes. You are allowed to use a word processor to write your answers, and a pdf reader to view the exam questions. You may use a SIMPLE calculator app that allows arithmetic operations.
- You can use your approved cheat sheets. They are not on Blackboard this time. You can use your local files. **Do not waste too much time browsing your cheat sheet.**

Reference

Recall that if $X_1, X_2, ..., X_n$ **are independent Poisson trials with** $Pr[X_i = 1] = p_i$ **where** $X = \sum_i X_i$ **and** $\mu = E[X]$**, then for any** $\delta \in (0, 1]$**:**

$$Pr[X > (1 + \delta)\mu] < e^{-\mu\delta^2/4}$$

(10pts) 1. Mark all that apply. $f(n) = \max(n^2, n^{1.5}\log^{16} n)$

√ $f(n) = O(n^2)$
○ $f(n) = O(n^{1.5}\log^{16} n)$
√ $f(n) = \Omega(n^2)$
√ $f(n) = \Omega(n^{1.5}\log^{16} n)$

(10pts) 2. Mark all that apply. $f(n) = \max(n\log n, n\log\log\log n)$

√ $f(n) = O(n\log n)$
○ $f(n) = O(n\log\log n)$
√ $f(n) = \Omega(n\log n)$
√ $f(n) = \Omega(n\log\log n)$

(10pts) 3. Use the Master Method to determine the best solution to the recurrence $T(n) = 16T(n/4) + n^2$ where $T(1) = 1$.

Solution: $T(n) = \Theta(n^2\log n)$

(10pts) 4. Use the Master Method to determine the best solution to the recurrence $T(n) = T(n/4) + 1$ where $T(1) = 1$.

Solution: $T(n) = \Theta(\log n)$

(10pts) 5. Use the Master Method to determine the best solution to the recurrence $T(n) = 3T(n/3) + n\log^2 n$ where $T(1) = 1$.

Solution: $T(n) = \Theta(n\log^3 n)$

6. Extra Credit [10 points]: Determine which of the following two functions grows faster as n does to infinity. Justify your answer. (Do not simply plug in values). $f(n) = \log^*(\log n)$, or $g(n) = \log(\log^* n)$.

Solution: $f(n)$ grows faster. Let $x = \log^* n$. Then $f(n) = \log^*(\log n) = \log^* n - 1 = x - 1 = \omega(\log x)$ as $x \to \infty$.

(10pts) 7. You are given a chess board of size n by n. Your goal is to place a pebble on exactly one square in each row and on exactly one square in each column so that none of the pebbles are diagonal from each other. (Note that this implicitly means they are not horizontal or vertical from each other since by definition, you place exactly one per row/column.)

- Is this goal achievable for a 2 by 2 board?
- If not, what is the smallest value of $n > 1$ for which such a placement is possible?
- Call the value you obtain in the last question n_0. Prove by induction that there exists a placement of pebbles satisfying the described criteria for all $n \geq n_0$.

Solution: a) No
b) 4
c) Not included. Extra credit.

(10pts) 8. Which of the following generally yields the best (tightest) bounds?

○ Markov's Inequality ○ Chebyshev's Inequality √ **Chernoff bound**

(10pts) 9. Which of the following generally yields the worst (loosest) bounds?

√ **Markov's Inequality** ○ Chebyshev's Inequality ○ Chernoff bound

(20pts) 10. What assumptions do you minimally need to use each of the following tools?

- Markov's Inequality:
- Chebyshev's Inequality:
- Chernoff bound:

Solution:

- Markov's Inequality: positive r.v.
- Chebyshev's Inequality: variance.
- Chernoff bound: independence.

(10pts) 11. A new web service decides to use a program called BinCodes to generate coupon codes for its customers. BinCodes generates 10 digit binary strings as coupon codes where each digit is 0 or 1 uniformly at random. If two users receive the same coupon code, one of them cannot use it. How many customers does there need to be so that the expected number of pairs of customers that will receive the same coupon code is roughly 1.

Solution:

$$\binom{m}{2}1/1024 = 1 \to m \approx \sqrt{2048} = 45\text{ish}$$

(10pts) 12. Consider an x by y by z three dimensional grid of routers. A packet starts at router located at $(0, 0, 0)$ on the grid and is traveling through the grid towards the router at (x, y, z). At each step, if the packet is at router (s, t, k) it can move to $(s + 1, t, k)$ or $(s, t + 1, k)$ or $(s, t, k + 1)$, but not to $(s - 1, t, k)$ or $(s, t - 1, k)$ or $(s, t, k - 1)$. How many paths are there in total for the packet to travel?

Solution: First, ignore dimension z. Let $x' = x - 1, y' = y - 1, z' = z - 1$. The total number of paths ignoring this dimension is $\binom{x'+y'}{y'}$. We can consider each such path and mix in the third dimension by collapsing the first two dimensions. For each path, we get $\binom{x'+y'+z'}{z'}$. Total number of paths is therefore $\binom{x'+y'}{y'}\binom{x'+y'+z'}{z'}$.

(10pts) 13. Prove that every planar graph has a vertex of degree at most 5 using the pigeonhole principle. Hint: $|E| \leq 3 \cdot |V| - 6$.

Solution: Proof by contradiction. Assume no such vertex exists. Then the degree of each vertex is at least 6. By the handshaking lemma ($\sum_v \deg(v) = 2|E|$), the number of edges is at least $3|V|$. Since there are at most $3 \cdot |V| - 6$ edges in a planar graph, by the pigeonhole principle there are at least 6 extra vertices. Contradiction.

(10pts) 14. Prove by contradiction that if x is irrational, then $1/x$ is irrational. Use the language of predicate logic. Assume $x \neq 0$. Hint: the definition of *rational* is as follows: A is rational iff $\exists x, y \in Z((A = x/y) \wedge (y \neq 0))$.

Solution: Let $P(A) =$"A is rational". Assume $1/x$ is rational. This implies

$$\begin{aligned} P(1/x) &\to \exists a, b \in Z((1/x = a/b) \wedge (b \neq 0)) \\ &\to \exists a, b \in Z((x = b/a) \wedge (a \neq 0)) \\ &\to P(x) \end{aligned}$$

$P(x)$ is a contradiction.

(10pts) 15. Let R be a relation on the vertices of a simple undirected 2-colorable graph G such that vRw if and only if there is a path in G from v to w containing an even number of vertices (including v and w). Show that R is not an equivalence relation.

(10pts) 16. The fraction of CS majors who know both Python and C++ is 2/7. The conditional probability that a random CS major knows C++, given that she/he knows Python is 3/7. What is the probability that a random CS major knows Python?

Solution:

$$P[E|C] = \frac{P[C \cap E]}{P[C]}.$$

$3/7 = \frac{2/7}{x} \rightarrow x = 2/3.$

(10pts) 17. A fair coin is tossed repeatedly until 3 consecutive heads occurs. What is the expected number of tosses?

Solution: $x = \sum_{i=1}^{3}(1/2^i)(x+i) + 3/8 \rightarrow x = 14$

(10pts) 18. Below are the vertices of a tree listed using in-order and pre-order traversal. What is the right child of the root? (Hint: ask yourself where the root would be in these two traversals and where the rest of the nodes relative to the root would be.)

```
 In-order :  ABDCVLKJPZ
Pre-order :  LCBADVZJKP
```

Solution:

```
                    L
          C                   Z
   B           V         J
A     D               K     P
```

(10pts) 19. Give a context free grammar that generates the language $L = \{0^{5n}1^k2^n |\ n, k \geq 0\}$.

Solution: Here is the grammar:

$S \rightarrow 00000S1|G$
$G \rightarrow 1G|\epsilon$

(10pts) 20. Randomized QuickSort algorithm (R-QUICK) works as follows. Given a list of n comparable elements, it picks one of them uniformly at random as a pivot. It then compares all elements to this pivot and creates two lists where each element goes either to the first list if it has a value less than or equal to the pivot, or to the second list otherwise. Then, in recurses on the two lists to sort them. Finally it returns the first sorted list followed by the pivot followed by the second sorted list. What is the expected number of comparisons performed by this algorithm? Hint: use indicator random variables and linearity of expectation. Consider when two elements are ever compared.

Solution: Note that two items are compared exactly once if and only if one of them is picked as the pivot within the same subproblem. Let $X_{i,j}$ be an indicator random variable where $X_{i,j} = 1$ if i and j are compared, and 0 otherwise. Then, let $X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{i,j}$. Expected number of comparisons of the R-QUICK algorithm is then $E[X]$. Lastly, note that $Pr(X_{i,j} = 1) = 2/(j - i + 1)$.

$$\begin{aligned}
E[X] &= E\left[\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{i,j}\right] \\
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E[X_{i,j}] \\
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Pr(X_{i,j} = 1) \\
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2/(j - i + 1) \\
&= \Theta(n \log n)
\end{aligned}$$

21. Extra Credit [20 points]: Consider the R-QUICK algorithm. Given an upper bound on the probability that the algorithm will perform more than $50 \cdot \alpha$ comparisons where α is the expected number of comparisons performed by R-QUICK computed in the last question.

Solution:

$$Pr[X > 50\alpha] \leq 1/50$$

22. Extra Credit [10 points]: Show an upper bound for the number of codewords in a code where codewords are bit strings of length 10 and the minimum distance between codewords is 5.

Solution:

$$\frac{2^{10}}{\binom{10}{0} \cdot \binom{10}{1} \cdot \binom{10}{2}}$$

23. Extra Credit [10 points]: If $\mathbf{T} = [p_{ij}]$ is the matrix of transition probabilities of some Markov chain, explain why the sum of the entries in any of the rows of $\mathbf{T}$ is 1.

Solution: Because the sum of probabilities in any probability distribution adds up to 1.

24. Extra Credit [30 points]: In Seattle, it has been observed that if the weather was sunny today, it will be sunny tomorrow with probability 0.5, and if it was cloudy today (for our purposes cloudy = NOT sunny) it will be cloudy tomorrow with probability 0.9. Model this situation as a Markov chain with two states, describe the state space, and find the matrix of transition probabilities. Then, find the equilibrium distribution.

Solution: The state space $S = \{s_1, s_2\}$, where s_1 is the state that it is sunny s_2 is the state that it is cloudy. Let X_k be the weather on the k-th day. This will be a Markov chain with matrix of transition probabilities $T = \begin{bmatrix} 0.5 & 0.5 \\ 0.1 & 0.9 \end{bmatrix}$.

$(pq)\begin{bmatrix} 0.5 & 0.5 \\ 0.1 & 0.9 \end{bmatrix} = (pq)$ is the system of equations $0.5p + 0.1q = p, 0.5p + 0.9q = q$. We also have $p + q = 1$. Solving the system of three equations for the unknowns p and q yields $p = 0.167$ and $q = 0.833$.