CS 574: Randomized Algorithms, Spring 2022 — **Instructor:** Sariel Har-Peled

Final: Tuesday, May 3, 3:35PM–4:35PM, 2022 — SC 1304

Name:	
NetID:	

- **This exam lasts 60 minutes.**
- ***Don't panic!***
- If you brought anything except your writing implements, your double-sided **handwritten** (in the original) $8\frac{1}{2}" \times 11"$ cheat sheet, and your university ID, please put it away for the duration of the exam. In particular, please turn off and put away *all* medically unnecessary electronic devices.
- **Best answer.** Choose best possible choice if multiple options seems correct to you – for algorithms, faster is always better (as long as the answer is correct).
- Please ask for clarification if any question is unclear.
- You should answer all the questions in the exam. There is no penalty for wrong answers, so feel free to guess.
- Fill your answers in the Scantron form using a pencil. We also **strongly** recommend you circle/mark your answer in the exam booklet.
- Do not fill more than one answer on the Scantron form per question - such answers would not be graded. Also, fill your answer once you are sure of your answer – erasing an answer might make the form unscannable.
- You have to submit both this booklet and the Scantron form when you are done with the exam.
- ***Good luck!***

Before doing the exam...

- Fill your name and netid in the back of the Scantron form, and also on the top of this page.
- **Fill in the pattern shown on the right in the Scantron form.**

  This encodes which version of the exam you are taking, so that we can grade it.

92 ● Ⓑ Ⓒ Ⓓ Ⓔ
93 ● Ⓑ Ⓒ Ⓓ Ⓔ
94 ● Ⓑ Ⓒ Ⓓ Ⓔ
95 Ⓐ Ⓑ Ⓒ ● Ⓔ
96 Ⓐ Ⓑ ● Ⓓ Ⓔ

1

1. (2 points) Let P be a set of n coupons. In each round, you randomly, uniformly and independently get a coupon from the set P. In the first n rounds, let T be the number of distinct coupons you picked. We have that

(A) $\mathbf{E}[T] = n(1 - (1 - 1/n)^n)$.
(B) $\mathbf{E}[T] = n/e$.
(C) $|T - \mathbf{E}[T]| \leq c \log n$, with probability $\geq 1 - 1/n^2$ for c sufficiently large constant.
(D) $\mathbf{E}[T] = n(1 - 1/e)$.

---

2. (8 points) Consider an undirected connected graph G with n vertices, and each vertex has degree at most 20. Each vertex $v \in \mathsf{V}(\mathsf{G})$ picks randomly, uniformly and independently a number $\alpha(v) \in [0, 1]$. Starting at any vertex u of G, consider the path that starts at $X_1 = u$, and always continue from the current vertex X_i, to the neighbor of X_i that has the lowest value of α – let this vertex be X_{i+1}. The walk then continues to the next iteration. If $\alpha(X_i)$ is smaller than the α value of all its neighbors, the walk stops. Let M be the length of this walk. We have that

(A) $M = \Theta(\sqrt{n})$ with high probability.
(B) $M = O(\log \log n)$.
(C) M can be infinite.
(D) $\mathbf{E}[M] = O(1)$.
(E) $M = \Theta(\sqrt{n \log n})$ with high probability.

---

3. (4 points) Consider a graph G with n vertices and m edges. Let χ be a random coloring of the vertices where each vertex gets assigned a random color in the range $\{1, \ldots, k\}$ (independently and uniformly). The expected number of edges that are not colored correctly (i.e., edges that both endpoints have the same color) is

(A) $m(1 - 1/k)$.
(B) m/k^2
(C) m/k.
(D) $m/\binom{k}{2}$

---

4. (2 points) Given a connected graph G with n vertices and m edges. Consider the algorithm that randomly assigns weights to the edges from the range $[0, 1]$ uniformly at random. Let T be the minimum spanning tree of G according to these random weights. Let e be the heaviest edge in T. Removing e from T breaks T into to two subtrees T_1, T_2. The probability that $(V(T_1), V(T_2))$ is a minimum cut is (bigger is better):

(A) $\geq 1/n$.
(B) $\geq 1/2^n$.
(C) $\geq 1/n^2$

---

5. (4 points) Let $\mathcal{S} = (s_1, \ldots, s_m)$ be a stream (i.e., sequence) of m elements from $N = \{1, \ldots, n\}$. Let f_i be the number of times the number i appears in $\mathcal{S}$. For $k \geq 0$, let

$$F_k = \sum_{i=1}^{n} f_i^k$$

be the ***k*th frequency moment** of $\mathcal{S}$. One can

(A) $(1 \pm \varepsilon)$-approximate F_2 using $O(\varepsilon^{-2} \log n)$ words, and this estimate is correct with high probability. No better solution is possible.
(B) $(1 \pm \varepsilon)$-approximate F_2 using $O(\varepsilon^{-2} \sqrt{n})$ words, and this estimate is correct with high probability. No better solution is possible.
(C) compute F_2 exactly using $O(1)$ words, and no better solution is possible.

---

6. (4 points)
Consider a sequence of variables $X_0, X_1, \ldots, X_h$. Here, $X_0 = 1$, and X_i is picked uniformly (and independently) from the range $[0, X_{i-1}]$. We have that

(A) $\mathbf{E}[X_{h-1}/X_h] = 1/2$.
(B) $X_0, X_1, \ldots, X_h$ is a martingale.
(C) $\mathbf{E}[X_h] = 1/2^h$.
(D) $\mathbf{E}[X_{h-1} \mid X_h] = X_h/2$.

---

7. (8 points)
Consider an algorithm that in the ith iteration picks a number r_i uniformly and independently from $[0, 1]$. The algorithm continues to the next iteration if $r_i = \min(r_1, \ldots, r_i)$. What is the expected number of iterations the algorithm performs?

(A) ∞.
(B) $\pi^2/6$.
(C) 2.
(D) e.

---

8. (2 points) Consider any two random variables X and Y. The statement $\mathbf{E}\big[X + Y\big] = \mathbf{E}[X] + \mathbf{E}[Y]$ is

(A) correct only if X and Y are independent.
(B) correct.
(C) not well defined.
(D) incorrect.

---

9. (2 points) Consider any random variable X. Let $\alpha, \beta > 0$ be two real numbers. We always have that

(A) $\mathbb{V}[\alpha + \beta X] = \beta^2\, \mathbb{V}[X]$.
(B) $\mathbb{V}[\alpha + \beta X] = \alpha^2 + \beta^2\, \mathbb{V}[X]$.
(C) $\mathbb{V}[\alpha + \beta X] = \alpha + \beta\, \mathbb{V}[X]$.
(D) $\mathbb{V}[\alpha + \beta X] = \alpha + \beta^2\, \mathbb{V}[X]$.

10. (2 points) Let G be an undirected, non-bipartite and connected graph with n vertices, and let u and v be any two vertices in the graph. The *hitting time* h_{uv} is the expected number of steps in a random walk that starts at u and ends upon first reaching v. Which of the following statements is always correct?

(A) $h_{uv} \neq 2m$.
(B) $h_{uv} \leq 2mn$.
(C) $h_{uv} \neq n^2$.

11. (4 points) Given a set P of n points in $\mathbb{R}^n$, and a parameter $\varepsilon \in (0,1)$, one can compute a transformation $T : \mathbb{R}^n \to \mathbb{R}^D$, such that for any $p, q \in \mathbb{R}^n$, we have

$$||p - q|| \in \Big[(1-\varepsilon)||T(p) - T(q)||,\ (1+\varepsilon)||T(p) - T(q)||\Big].$$

This statement

(A) is correct, with $D = O(\varepsilon^{-2} \log n)$.
(B) is correct with $D = O(\varepsilon^{-2} \log \varepsilon^{-1})$.
(C) incorrect.

12. (2 points)
Deciding if a 2SAT formula F with n variables and m clauses, has an assignment that satisfies at least 3/4 of the clauses, can be done by a program in (faster is better):

(A) None of the other answers.
(B) $O(1)$ time.
(C) $O(2^{n+m})$ time.
(D) $O((n+m)^2)$ time.

13. (4 points) Consider any random variable X and a value $t > 0$. The quantity $\mathbb{P}[(X - \mathbf{E}[X])^2 \geq t\,\mathbb{V}[X]]$ is

(A) none of the other answers.
(B) $\leq \frac{1}{t}$.
(C) not well defined.
(D) $\leq \frac{1}{t^2}$.
(E) $\leq \exp(-t^2/2)$.

14. (2 points) Consider quick sort. With the input being $A[1 \ldots n]$ (all of them being distinct), and the output being the array $B[1 \ldots n]$. The probability that during the algorithm execution the numbers $A[i]$ and $A[j]$ are being compared, for $i < j$, is

(A) $1/H_n$, where $H_n = \sum_{i-1}^{n} 1/i$.
(B) 1.
(C) in the range $[2/n, 1]$.
(D) $\frac{2}{j-i+1}$.
(E) 0.

---

15. (4 points) Consider a kSAT formula F with n variables and m clauses (each clauses has exactly k literals), where $k = O(1)$ is a constant. One can compute an assignment to F that satisfies $\geq (1-1/2^k)m$ of the clauses. This can be done in

(A) Can be done in randomized polynomial time (but not deterministic polynomial time).
(B) Deterministic polynomial time.
(C) The problem is NP-Hard.

---

16. (2 points) Consider a binomial random variable $X \sim \mathrm{Bin}(n, p)$. The variance of X is

(A) npq.
(B) None of the other answers.
(C) np.
(D) $\sqrt{np}$.

---

17. (6 points) Consider the random walk on the integer grid that starts at $(0, 0)$. If at time t the walk is located at (x, y), then it moves randomly (independently, and uniformly) to one of the locations

$$(y-1, x), \qquad (y+1, x), \qquad (y, x-1), \qquad (y, x+1).$$

In expectation, this infinite random walk would visit the origin $(0, 0)$ how many times?

(A) ∞.
(B) $O(1)$.

---

18. (4 points) Let P be a set of n points in the plane, and let $k < n$ be a parameter. Using the algorithm seen in class, computing the kth smallest distance in P can be done in (faster is better):

(A) $O(k^2)$.
(B) $O(nk)$.
(C) $O(n + k)$.
(D) $O(n^2)$.

---

19. (2 points) Ohm's law states that

(A) all the electric flow arriving into a node, leaves it.
(B) All electric current entering a circuit is distributed uniformly in the circuit.
(C) voltage = resistance $*$ current.

---

20. (4 points) Let P be a set of n coupons. In each iteration, you randomly and uniformly get a coupon from the set P. Assume that to pick all coupons at least once, with probability at least $p = 0.99$, one needs T rounds. The time required to pick two copies of all coupon, with probability a least p, is

(A) $< 2T$ rounds.
(B) $2T$ rounds.
(C) $> 2T$ rounds.

---

21. (8 points) Let P be a set of n coupons. In each round, you randomly, uniformly and independently get a coupon from the set P. Alternatively, you can replace any ten coupons you have with (for free) for any coupon you want. Which of the following strategies requires the fewest number of steps in expectation till you pick all coupons (assume $n > 10000$):

(A) Pick $3n$ coupons, and then replace any unneeded ten coupons with a missing coupon till you have all coupons. If this fails, repeat this process till success.
(B) Pick random coupons till you pick all the coupons.
(C) Pick $10n$ coupons, and then replace any unneeded ten coupons with a missing coupon till you have all coupons. This strategy always succeeds.

---

22. (4 points)
For a string $s \in \{0,1\}^d$, let $f_i(s)$ be the function that returns the ith bit of s. Let $\mathcal{F} = \{f_1, \ldots, f_d\}$. For a parameter k, consider a random function $F : \{0,1\}^d \to \{0,1\}^k$, where $F(s) = (g_1(s), g_2(s), \ldots, g_k(s))$, where g_i is picked uniformly and independently at random from $\mathcal{F}$ (with replacement). Consider two points $p, q \in \{0,1\}^d$, with Hamming distance at most r. For $k = \lfloor d/r \rfloor$, we have that

(A) $\mathbb{P}[F(p) = F(q)] \geq (1 - r/d)^k \approx 1/e$.
(B) $\mathbb{P}[F(p) \neq F(q)] \geq (1 - r/k)^d \approx \exp(-r^2)$.
(C) $\mathbb{P}[F(p) \neq F(q)] \leq \exp(-r)$.

---

23. (4 points) Consider a group of n people, and assume their birthday dates have uniform (independent) distribution over $N = 365$. The expected number of triples of people with the same birthday date is:

(A) $\binom{n}{3}/N^2$
(B) n^2/N^3
(C) $\binom{n}{2}/N$
(D) $1/N$
(E) n/N^3

24. (4 points) You are given a polynomial $f(x, y, z)$ of degree d that is defined over $\mathbb{Z}_p$. Importantly, f is not the zero polynomial. Here p is a prime number and the calculations are done modulo p. Assume that $4d \leq p \leq 8d$. For a random numbers $x', y', z' \in \{0, 1, \dots, p-1\}$ picked uniformly and independently, let $\alpha = \mathbb{P}[f(x', y', z') = 0]$. Which of the following statements is correct.

(A) $\alpha = 1/d$.
(B) $\alpha \leq 1/p$.
(C) $\alpha = d/p$.
(D) $\alpha \leq d/p$.

25. (4 points) Consider a random variables X and Y. The statement $\mathbf{E}\big[\mathbb{P}[X = x \mid Y]\big] = \mathbb{P}[X = x]$ is

(A) incorrect.
(B) correct.
(C) correct only if X and Y are independent.
(D) not well defined.

26. (4 points) Consider a graph G with n vertices and m edges. We randomly assign every vertex a weight between $[0, 1]$, and let I be the set of vertices in G, such that their weight is smaller than the weight of all their neighbors. The set I form an independent set. The size of I is

(A) in expectation $\geq n/(d_{\mathrm{avg}} + 1)$, where d_{avg} is the average degree in G.
(B) strongly concentrated since this can be represented as a martingale.