Final Exam

5:00-8:00pm, 16 May

Read these instructions carefully

1. This is a **closed book** exam. Calculators **are** permitted.
2. This exam consists of 10 multiple choice questions and 4 longer questions with multiple parts.
3. **Multiple choice questions:** Answer these by **circling** the correct answer or answers. You should be able to answer these from memory, by inspection, or with a very small calculation. Incorrect answers will receive a **negative** score, so if you do not know the answer you should **not** guess.
4. **Other questions:** Write your answers to these in the spaces provided below them. None of these questions requires a long answer, so you should have enough space; if not, continue on the back of the page and state clearly that you have done so. **Show your working**.
5. Approximate point totals for each question part are indicated in the margin. The maximum total number of points is 130.
6. The questions vary in difficulty: if you get stuck on some part of a question, leave it and go on to the next one.

Your Name:

For official use; please do not write below this line!

MC	
Q11	
Q12	
Q13	
Q14	
Total	

[exam starts on next page]

1. A multiple choice exam has six possible answers for each question, only one of which is correct. A correct answer receives 3 points, while an incorrect answer incurs a penalty of b points. If we wish to ensure that the expected score for a student who randomly guesses on every question is zero, we should set b to be *3pts*

$$\frac{1}{6} \qquad \frac{1}{2} \qquad \frac{3}{5} \qquad \frac{4}{5} \qquad 1 \qquad 5$$

---

2. A certain device fails each time it is used with probability $\frac{1}{100}$, and the failures are independent.

(a) The expected number of uses until the device fails twice is *3pts*

$$2 \qquad 50 \qquad 100 \qquad 150 \qquad 200 \qquad 10000$$

(b) Conditional on the device not having failed in the first 100 uses, the expected total number of uses until the device fails for the first time is *3pts*

$$100 \qquad 101 \qquad 150 \qquad 200 \qquad 300 \qquad 10000$$

---

3. **(a)** Let a be a fixed n-vector with entries in $\{0,1,2\}$ such that $a \neq \mathbf{0}$, and let r be an n-vector with entries chosen uniformly at random in $\{0,1,2\}$. The probability that the dot product $r \cdot a \equiv \sum_i a_i r_i$ is equal to zero (mod 3) is *3pts*

$$\frac{1}{2^n} \qquad \frac{1}{3^n} \qquad \frac{1}{n} \qquad \frac{1}{2} \qquad \frac{1}{3} \qquad 1 - \frac{1}{3^n}$$

(b) Now let R be an $n \times n$ matrix with entries chosen u.a.r. from $\{0,1,2\}$. The probability that the vector Ra has all components equal to zero (mod 3) is *3pts*

$$\frac{1}{2^n} \qquad \frac{1}{3^n} \qquad \frac{1}{n} \qquad \frac{1}{2} \qquad \frac{1}{3} \qquad 1 - \frac{1}{3^n}$$

---

4. Each of n people is assigned a random t-digit *decimal* number, independently and u.a.r. The smallest value of t that ensures all numbers are distinct with constant probability is about *3pts*

$$n \ln n \qquad \log_2 n \qquad \sqrt{n} \qquad \log_{10} n \qquad (\log_{10} n)^2 \qquad 2 \log_{10} n$$

---

5. **(a)** A book contains 10^6 characters. A character is mis-typed with probability 10^{-5}, independently of all other characters. The probability that the book contains no mis-typed characters is approximately *3pts*

$$\tfrac{1}{10} \qquad 10^{-6} \qquad \mathrm{e}^{-10} \qquad \mathrm{e}^{-10^5} \qquad \mathrm{e}^{-10^6} \qquad 1 - \mathrm{e}^{-10}$$

(b) In the same book as in part (a), a character is smudged with probability 10^{-5}, independently of all other characters and independently of whether the character is mis-typed. An "error" is a mis-typed or smudged character (with a mis-typed, smudged character counting as two errors). The probability that the book contains exactly 10 errors is approximately *3pts*

$$\mathrm{e}^{-20}\frac{20^{10}}{10!} \qquad \mathrm{e}^{-10}\frac{10^{10}}{10!} \qquad \mathrm{e}^{-20}\frac{20^{10}}{20!} \qquad \mathrm{e}^{-20}\frac{10^{20}}{10!} \qquad 10^{-50} \qquad 10^{-5}$$

[continued overleaf]

6. **(a)** Let G be a random graph in the $\mathcal{G}_{n,p}$ model. The expected number of k-cliques in G is on the order of *3pts*

$(\frac{n}{p})^k$ $\qquad (\frac{n}{p})^{\binom{k}{2}}$ $\qquad n^k p^k$ $\qquad n^k p^{\binom{k}{2}}$ $\qquad n^k p^k (1-p)^{n-k}$

(b) For any fixed k, the threshold value of p for the existence of a k-clique in G is on the order of *3pts*

$n^{-2/(k-1)}$ $\qquad n^{-1/k}$ $\qquad n^{-1}$ $\qquad (k \ln n)^{-1}$ $\qquad k/n$ $\qquad \binom{n}{k}^{-1}$

---

7. A Markov chain on two states $\{0, 1\}$ has transition matrix $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix}$.

(a) In the stationary distribution, the probability of being in state 0 is *3pts*

0 $\qquad 1$ $\qquad \dfrac{1}{2}$ $\qquad \dfrac{1}{3}$ $\qquad \dfrac{2}{3}$ $\qquad$ not well defined

(b) Starting from the uniform distribution on the two states, the probability of being in state 0 after two steps is *3pts*

0 $\qquad 1$ $\qquad \dfrac{1}{2}$ $\qquad \dfrac{1}{8}$ $\qquad \dfrac{3}{4}$ $\qquad \dfrac{5}{8}$

---

8. Let $\{X_i\}$, for $i = 1, \ldots, n$, be a sequence of 0-1 random variables such that $E[X_i] = \frac{1}{2}$ for all i and $\mathrm{Cov}[X_i, X_j] \le \epsilon$ for all $i \ne j$. Let $X = \sum_{i=1}^n X_i$. Circle those **three** of the following statements that **must** be true about X: *6pts*

$E[X] = \frac{n}{2}$ $\qquad \Pr[X \ge \frac{3}{4}n] \le \mathrm{e}^{-n/16}$ $\qquad \mathrm{Var}[X] \le \frac{n}{4} + n(n-1)\epsilon$

$E[X^2] = \frac{n^2}{4}$ $\qquad \Pr[X \ge \frac{2}{3}n] \le \frac{3}{4}$ $\qquad \Pr[X = \frac{n}{2}] < 1$

---

9. Let π be a uniform random permutation on $\{1, 2, \ldots, n\}$.

(a) The probability that $\pi(1) < \pi(2) < \pi(3)$ is *3pts*

$\dfrac{1}{\binom{n}{3}}$ $\qquad \dfrac{1}{n}$ $\qquad \dfrac{1}{8}$ $\qquad \dfrac{1}{6}$ $\qquad \dfrac{1}{3}$ $\qquad \dfrac{1}{2}$

(b) *Conditional on* the event $\pi(1) < \pi(3)$, the probability that $\pi(1) < \pi(2) < \pi(3)$ is *3pts*

$\dfrac{1}{\binom{n}{2}}$ $\qquad \dfrac{1}{n}$ $\qquad \dfrac{1}{6}$ $\qquad \dfrac{1}{3}$ $\qquad \dfrac{1}{2}$ $\qquad$ none of these

---

10. We have n bins and we throw balls into them in two phases as follows. In the first phase, we toss $2n$ balls into the n bins, independently and uniformly at random. In the second phase, the numbers of balls tossed into each bin are drawn from independent Poisson distributions each with parameter 2.

(a) After both phases, the expected (total) number of balls in the first three bins is *3pts*

6 $\qquad 6 + 3e^{-2}$ $\qquad 12$ $\qquad 2n$ $\qquad$ none of these

(b) After both phases, the probability that there are a total of $3n$ balls in all n bins combined is *3pts*

$\frac{1}{\binom{3n}{n}}$ $\qquad 2^{-n}$ $\qquad \mathrm{e}^{-2n}\frac{(2n)^n}{n!}$ $\qquad \mathrm{e}^{-4n}\frac{(4n)^{3n}}{(3n)!}$ $\qquad$ none of these

---

[continued overleaf]

11. Stock Market (16 pts)

A simple model of the stock market suggests that, each day, a stock with price q will increase by a factor $r > 1$ to qr with probability p, and will fall to q/r with probability $1-p$. Suppose that the initial price of the stock is \$100, and for $t \geq 0$ let the r.v. Z_t denote the stock price in dollars after t days (so that $Z_0 = 100$).

(a) Write down the expression for $E[Z_t \mid Z_{t-1}]$ for $t > 0$ (as a function of Z_{t-1}). *3pts*

(b) Use part (a) to show carefully that $E[Z_t] = 100\left(pr + \frac{1-p}{r}\right)^t$. [NOTE: Hand-waving arguments will be penalized!] *4pts*

(c) Show that $\text{Var}[Z_t] = 10000\left(\left(pr^2 + \frac{1-p}{r^2}\right)^t - \left(pr + \frac{1-p}{r}\right)^{2t}\right)$. [HINT: Begin by computing the conditional expectation of a relevant expression, as in part (a).] *5pts*

(d) A new company seems to be doing very well with $p = \frac{1}{3}$ and $r = 2$. Show that in this case $E[Z_{100}] = 100$ and $\text{Var}[Z_{100}] = 10000((\frac{3}{2})^{100} - 1)$. *1pt*

(e) An unscrupulous analyst claims that the stock price after 100 days is likely to be at least \$1000. Show that the probability of this event is at most $\frac{1}{10}$. *3pts*

[continued overleaf]

12. Testing Software (20 pts)

A new start-up company is developing an online course for improving students' performance on the SAT. The company has a database of $2n$ students, exactly half of whom are classified as "strong" and the other half as "weak" (though this information is not available to the company). The company would like to select a small random sample of students to participate in a trial to find out if the course does indeed improve SAT performance for weak students. Ideally, a *good* sample will have at least $\delta n/2$ weak students and at most $2\delta n$ strong students, where $\delta < \frac{1}{2}$ is a (small) constant. As consultants for the start-up, the CS174 teaching staff recommends that the company select a sample by independently picking each student to participate in the trial with probability δ.

(a) Let the random variables X, Y denote the numbers of weak students and strong students respectively in the random sample. Write down $E[X]$ and $E[Y]$. *2pts*

(b) Show that $\Pr[X < \frac{\delta n}{2}] < e^{-\Omega(\delta n)}$. *3pts*

(c) Show that $\Pr[Y > 2\delta n] < e^{-\Omega(\delta n)}$. *3pts*

(d) Deduce from parts (b) and (c) that the random sample is good with probability at least $1 - e^{-\Omega(\delta n)}$. *2pts*

[continued overleaf]

(e) Now suppose that the students in the sample are picked with probability δ but only *pairwise independently*. Again, let X, Y denote the numbers of weak and strong students respectively. Clearly, $E[X]$ and $E[Y]$ are as before. Compute $\mathrm{Var}[X]$ and $\mathrm{Var}[Y]$ and deduce that both variances are at most δn. *2pts*

(f) In the pairwise independent scenario of part (e), show that the sample is good with probability at least $1 - \frac{5}{\delta n}$. *6pts*

(g) Explain briefly one advantage of using pairwise independent sampling. *2pts*

[continued overleaf]

13. Consumer Choice (20 pts)

A large supermarket chain supplies its stores with toothpaste using trucks sent out from various distribution centers. Each of the n stores is serviced by t different trucks, and each truck may service multiple stores on its route. For reasons of efficiency, however, each truck can carry only one brand of toothpaste, out of the b brands available to the company. In order to provide at least a minimum level of choice for its customers, the company wants to ensure that every store receives a delivery of at least two different brands of toothpaste. The company's goal is to assign brands of toothpaste to trucks (one brand per truck) so as to satisfy this requirement.

(a) Suppose that the company assigns brands of toothpaste to the trucks u.a.r. Show that the expected number of stores that do not receive at least two different brands is nb^{1-t}. *3pts*

(b) Show that if $n < b^{t-1}$ then there is an assignment of brands to trucks so that all stores receive at least two different brands. *3pts*

(c) Assuming $n < b^{t-1}$, show that under the randomized algorithm of part (a), with probability at least $\frac{1}{2}$ all but at most two of the stores receive at least two different brands. *3pts*

[continued overleaf]

(d) Assuming again that $n < b^{t-1}$, explain how to apply the Method of Conditional Probabilities to the randomized algorithm of part (a) and obtain a deterministic algorithm that produces an assignment that *guarantees* that all stores receive at least two different brands. [NOTE: You should explain the following clearly (i) what your algorithm is; (ii) how to compute any conditional expectations that it uses; and (iii) why your algorithm yields an assignment with the required guarantee.] *11pts*

[continued overleaf]

14. Random Walk on a Cycle (20 points)

This question concerns random walk on the simple cycle with n vertices, where we assume throughout that n is even. You will find it useful to recall the following fact from class: for random walk on the line $[0, 1, \ldots, n]$, the expected hitting time from vertex i to vertex n is exactly $n^2 - i^2$.

(a) By quoting a standard upper bound on cover time, show that the cover time of the cycle is at most $O(n^2)$. *2pts*

(b) Show that the cover time of the cycle is at least $\Omega(n^2)$. [HINT: Consider the hitting time from a vertex to the diametrically opposite vertex, and relate this to hitting time on the line, as in the introduction.] *3pts*

(c) Suppose now that we add a self-loop probability of $1/2$ to every vertex of the cycle. (I.e., at each step, with probability $1/2$ the walk does nothing, else it moves to a random neighbor.) Explain briefly why this modification ensures that the random walk converges to a unique stationary distribution, and write down this stationary distribution. *3pts*

(d) Consider the following coupling (X_t, Y_t) for the random walk with self-loops as in part (c). If $X_t = Y_t$ then X_t, Y_t both make the same move. Otherwise, with probability $1/2$ X_t does nothing while Y_t moves to a random neighbor; and with probability $1/2$ Y_t does nothing while X_t moves to a random neighbor. Explain briefly why this is a valid coupling. *2pts*

[continued overleaf]

(e) For the coupling in part (d), show that the expected time until $X_t = Y_t$ (for any pair of initial states (X_0, Y_0)) is at most $n^2/4$. [HINT: Relate this problem to hitting time for random walk on the line, as in the introduction.] *4pts*

(f) Deduce from part (e) that the time until the variation distance from the stationary distribution reaches ϵ is at most $n^2/4\epsilon$. [HINT: Use Markov's inequality and the Coupling Lemma.] *4pts*

(g) Show that the bound on mixing time in part (f) can be improved to $O(n^2 \log \epsilon^{-1})$. *2pts*

[The End]