Midterm Solution Sketches

- Don't Panic.
- The midterm contains six problems (and one just for fun). You have 120 minutes to earn 100 points.
- The midterm contains 16 pages, including this one and 3 pages of scratch paper.
- The midterm is closed book. You may bring one double-sided sheet of A4 paper to the midterm. (You may not bring any magnification equipment!) You may not use a calculator, your mobile phone, or any other electronic device.
- Write your solutions in the space provided. If you need more space, please use the scratch paper at the end of the midterm. Do not put part of the answer to one problem on a page for another problem.
- Read through the problems before starting. Do not spend too much time on any one problem.
- Show your work. Partial credit will be given. You will be graded not only on the correctness of your answer, but also on the clarity with which you express it. Be neat.
- Draw pictures and give examples.
- Good luck!

Problem #	**Name**	**Possible Points**	**Achieved Points**
1	True, False, Explain	15	
2	Faulty Servers	10	
3	Minimum Cuts for Fun and Profit	12	
4	Room Allocation	30	
3	Morris Likes to Count	16	
6	The World Congress	17	
Total:		100	

Student Number: ______________________________

Problem 1. True, False, and Explain [15 points]

For each statement, indicate whether it is true or false, and briefly explain why. (No credit will be given for a blank or incorrect explanation.)

For any random variable X, assume that $\mathrm{E}[X] \geq \sqrt{\mathrm{Var}[X]} > 0$. Then $\Pr[X \geq 3\mathrm{E}[X]] \leq 1/4$.

TRUE **FALSE**

Solution: True. $\Pr[X \geq 3\mathrm{E}[X]] \leq \Pr[|X - \mathrm{E}[X]| \leq 2\mathrm{E}[X]] \leq \mathrm{Var}[X]/4\mathrm{E}[X]^2 \leq 1/4$.

For every non-negative random variable X and for every value a, $\Pr[X \geq a] \geq \Pr[X^2 \geq a^2]$.

TRUE **FALSE**

Solution: True. More specifically, $\Pr[X \geq a] = \Pr[X^2 \geq a^2]$ since for all non-negative values, $X \geq a$ if and only if $X^2 \geq a$.

Stochastic domination is transitive: given three independent random variables A, B, and C, if A stochastically dominates B and B stochastically dominates C, then A stochastically dominates C. **Solution:** True. For all values t, $\Pr[(]A \geq t) \geq \Pr[(]B \geq t) \geq \Pr[(]C \geq t)$, as required.	**TRUE**	**FALSE**
The expected search time for Cuckoo Hashing is asymptotically faster than the expected search time for a hash table with chaining (where the table size $m = \Theta(n)$, the number of keys). **Solution:** False. Both are $O(1)$ in expectation. In fact, for Cuckoo Hashing it is deterministically $O(1)$, while for hashing with chaining it is $O(1)$ in expectation.	**TRUE**	**FALSE**
If you throw n balls randomly into n bins, where $n > 2$, then the expected number of empty bins is $< n/9$. **Solution:** False. The probability that a bin is empty is $(1-1/n)^2 \geq 1/e^2 > 1/9$. Therefore the expected number of empty bins is $> n/9$.	**TRUE**	**FALSE**

Problem 2. Faulty Servers [10 points]

You are the system administrator for a collection of n servers. Unfortunately, every so often, a server fails. It's your job to detect when servers fail and identify them. To check if a server is running properly, you can *ping* it, sending it a message and getting a response. Unfortunately, communication is faulty, so sometime your ping fails.

- If the server has failed, then there is no response to the ping.
- If the server is functional, then with probability 1/2 you get a response to your ping; with probability 1/2 you get no response.

Thus, when there is no response to a ping, you cannot be sure whether the server failed or whether the ping failed.

This morning, an alarm goes off, indicating that (exactly) one server has failed! Alas, you do not know which server has failed. (You may assume that a randomly chosen server failed.)

Assume that $n = 4$, and you ping server X and get no reponse.

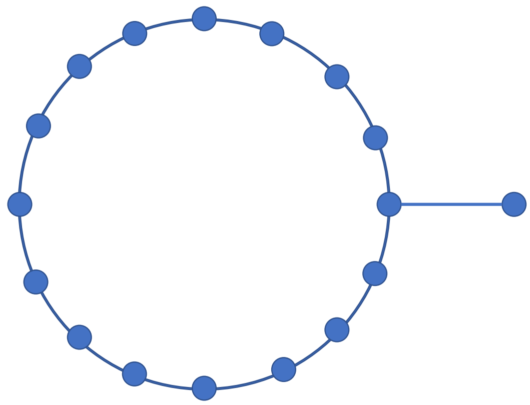
What is the probability that server X has failed? | 2/5 |

Show that your answer is correct:

Solution: Let f be the event that the server X has failed, and p be the event there is no response to a ping. By the rule of conditional probability, we know that $\Pr[f|p] = \frac{\Pr[f]}{\Pr[p]}$. In this case, $\Pr[f] = 1/4$, since $n = 4$ and one server has failed. The probability that a ping gets a response is $(3/4)(1/2)$, since the failure of the ping is independent of which server failed. So $\Pr[p] = 5/8$. We conclude that $\Pr[f|p] = (1/4)/(5/8) = 2/5$.

Problem 3. Minimum Cuts for Fun and Profit [12 points]

Consider the following graph G containing n nodes and n edges, where $n-1$ nodes are arranged in a ring and one node is attached outside:



Problem 3.a. What is the minimum cut of G?

Solution: The minimum cut is of size 1, separating the outside node from the ring.

Problem 3.b. Prove (carefully, in detail) that the probability that one execution of the contraction algorithm (i.e., executing `Collapse`$(G, n, 2)$, reducing the number of nodes to two) in Karger's Min-Cut algorithm has probability exactly $2/n$ of successfully finding the minimum cut.

Solution: We need to show that for $n-1$ iterations, we never choose to contract the edge outside the ring, i.e., the only edge in the minimum cut. We refer to that edge as the "critical edge."

In the first step, there are n edges total and $n-1$ other edges to choose; in the second step, there are $n-1$ edges total and $n-2$ other edges to choose; in the third step, there are $n-2$ edges total and $n-3$ other edges to choose, etc. Notice that this is an exact calculation because whenever an edge on the ring is collapsed, only one edge is removed (i.e., there are no parallel edges until the very last step).

Let E_j be the event that we choose the critical edge in step j, and $\bar{E}_j$ be the event that we do not choose the critical edge in step j. Thus the probability of *not* choosing the critical edge is:

$$
\begin{aligned}
\Pr\left[\bigcap_{j=1}^{n-1} \bar{E}_j\right] &= \prod_{j=1}^{n-2} \Pr\left[\bar{E}_j | \bigcap_{i=1}^{j-1} \bar{E}_j\right] \\
&= \prod_{j=1}^{n-1} \frac{n-j}{n-j+1} \\
&= \left(\frac{n-1}{n}\right)\left(\frac{n-2}{n-1}\right)\left(\frac{n-3}{n-2}\right)\cdots\left(\frac{2}{3}\right) \\
&= \frac{2}{n}
\end{aligned}
$$

Problem 4. Room Allocation [30 points]

Professor Rogammer has n study rooms and n high school students, each of whom needs to be assigned to a room. Her goal is to assign students to study rooms in a uniform fashion, so that no study room has too many students. To do this, she wants to use a random load balancing strategy: when each student arrives, she chooses a room uniformly at random and assigns the student to the room.

It turns out that all the students arrive in pairs. (They were each assigned partners in class. Assume n is even.) Some of the pairs want to work together, while others do not. When a pair of students arrives:

- With probability 1/2, the pair want to study together. Professor Stubbins chooses a random room and assigns them both to the same room.
- With probability 1/2, the pair does not want to study together. Professor Stubbins assigns each of the two students to randomly chosen rooms, independently of each other. (They may, by chance, still end up in the same room, of course.)

Choose one room that we want to analyze (maybe, room number seven). Let X be the number of students assigned to that room, after all n students have arrived.

Problem 4.a. What is the expected value of X? | 1 |

Show that your answer is correct:

Solution: Let y_i be the indicator random variable equal to 1 if student i is assigned to the room. We know that $\mathrm{E}[y_i] = \Pr[y_i = 1] = 1/n$. (The fact that some choices are correlated does not change that.) By linearity of expectation, we know that $\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[y_i] = \sum_{i=1}^{n} 1/n = 1$.

Problem 4.b. What is the variance of X?

(1/2)(3 + 5/n)

Show that your answer is correct:

Solution: The variance $\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$. We know that $\mathrm{E}[X]^2 = 1$. We need to compute the $\mathrm{E}[X^2] = \mathrm{E}[(\sum_{i=1}^{n} y_i)^2]$. By linearity of expectation, we know that $\mathrm{E}[X^2] = \sum_{k,\ell} \mathrm{E}[y_k y_\ell]$. There are three cases here to consider:

- k is odd and $\ell = k + 1$ (or vice versa): In this case, k and ℓ represent two students in a pair. Thus with probability $1/(2n)$ they are both assigned to the room because they are inseparable partners; with probability $(1/2(1/n)(1/n)$ they are both assigned to the room independently. Thus the total probability that k and ℓ are both assigned to the room is $(1/2n)(1 + 1/n)$. Therefore $\mathrm{E}[y_k y_\ell] = (1/2n)(1 + 1/n)$ (as the only case where this is 1 is when both are assigned to the room). This case occurs n times in total (i.e., $n/2$ different pairs, where the pairs can appear in both orders).

- $k = \ell$: In this case, k and ℓ represent the same student. In this case, $\mathrm{E}[y_k y_\ell] = 1/n$. The fact that this may depend on some other student is irrelevant. This case occurs n times.

- None of the above cases hold: In this case, k and ℓ represent independent students. Therefore the probability that they are both assigned to the room is $(1/n)(1/n) = 1/n^2$. Thus $\mathrm{E}[y_k y_\ell] = 1/n^2$. This case occurs $n^2 - 2n$ times.

It remains only to compute the total:

$$n(1/2n)(1 + 1/n) + n(1/n) + (n^2 - 2n)(1/n^2) = 1/2 + 1/2n + 1 + 1 - 2/n = (5/2)(1 + 1/n) \ .$$

Therefore, the variance is $(5/2)(1 + 1/n) - 1 = (1/2)(3 + 5/n)$. If you observe that $1/n \leq 1$, then this is bounded by 4.

Problem 4.c. Use Chebychev's Inequality to upper bound the probability that the room has more than 9 students. For this approximation, you may assume that $n \geq 5$, which should simplify your variance calculation from the previous part. (If you were not able to solve the previous part, then for partial credit you may use V to represent the variance of X.)

Solution: From the previous part, we know that $\mathrm{Var}[X] = (1/2)(3 + 5/n) \leq (1/2)(4) \leq 2$.

$$\Pr[|X - 1| \geq 8] \leq \frac{\mathrm{Var}[X]}{64} \leq 2/64 \leq 1/32$$

Problem 4.d. Now prove that *every* room has $O(\log n)$ students, with high probability. (Remember to show that this holds for *all* rooms, not just for one room.) For this part, you may again assume that $n \geq 5$.

Hint: You might want to define some new random variables that are independent.

Solution: Fix one room. Consider students in pairs: let $z_k = (y_{2k-1} + y_{2k})/2$ where k ranges from 1 to $n/2$. Each $z_k \in [0, 1]$ depending on whether neither, one, or both students are allocated to the room. Notice that the actual number of student assigned to the room is $X_j = 2\sum_{k=1}^{n/2} z_k$. So if we can show that $\sum_{k=1}^{n/2} z_k = O(\log n)$ with high probability, then we have successfully proved the desired property.

The $\mathrm{E}[z_k] = (1/2)(2/n) = 1/n$ (by linearity of expectation). Each of the z_k are independent. Thus we can apply a Chernoff Bound to show that:

$$\Pr\left[\sum_{k=1}^{n/2} z_k \geq (1 + 6\ln n)(n/2)(1/n)\right] \leq e^{-(1/2)36\ln^2(n)/(2+6\ln n)} \leq e^{-18\ln^2(n)/8\ln n} \leq e^{-2\ln n} \leq 1/n^2$$

Taking a union bound over all $n/2$ pairs, we see that with probability at least $1 - 1/n$, we know that $\sum_{k=1}^{n/2} z_k \leq 6\ln n$, and hence $X \leq 12\ln n$.

Problem 5. Morris Likes To Count [16 points]

Morris invented the following clever algorithm for (approximate) counting:

Algorithm 1: Morris Approximate Counter

```
c = 0
INCREMENT()
    With probability 1/2^c: set c = c + 1.

ANSWER()
    return 2^c − 1
```

Start with the counter c equal to 0. On each increment, increment the counter c with probability $1/2^c$. Morris showed that after n increment operations:

$$\mathrm{E}\left[2^c\right] = n + 1$$

$$\mathrm{Var}\left[2^c\right] \leq n^2$$

(You can assume this is true for today. As a fun exercise, you can prove it by induction.) This means that, luckily, it is an unbiased estimator, giving the propert expected answer! Unfortunately, the variance is quite high.

Problem 5.a. Design an algorithm using the Morris Counter as a block box that returns a $(1 \pm \epsilon)$ approximation of the correct count with probability at least 3/4. Do this by running α copies of the Morris Counter in parallel and combining the answers in some way. Explain how your algorithm works. (Give your analysis/proof on the next page.)

Solution: Run α independent copies of the Morris Counter and return the average value.

Problem 5.b. Prove that your algorithm is correct, i.e., after n increment operations it returns a value $n(1 \pm \epsilon)$ with probability at least 3/4.

Solution: Let A be the average value of α copies of the Morris Counter. Since each copy of the Morris Counter returns a value with expectation n, we know that $\mathrm{E}[A] = n$. Similarly, since the variance of the Morris Counter is $\leq n^2$, we know that $\mathrm{Var}[A] = \alpha \cdot \mathrm{Var}[2^c]/\alpha^2 \leq n^2/\alpha$. By Chebychev's Inequality, we know that:

$$\begin{aligned}
\Pr\left[|A - n| \geq \tfrac{2n}{\sqrt{\alpha}}\right] &\leq \frac{\mathrm{Var}[A]}{(2n/\sqrt{\alpha})^2} \\
&\leq \frac{n^2}{\alpha}\frac{\alpha}{4n^2} \\
&\leq 1/4
\end{aligned}$$

Therefore, we conclude that $A = n(1 \pm 2/\sqrt{\alpha})$ with probability at least 3/4. By choosing $\alpha = 4/\epsilon^2$, we conclude that $A = n(1 \pm \epsilon)$ with probability at least 3/4.

Problem 5.c. What value of α did you choose, as a function of ϵ? $O(4/\epsilon^2)$

Briefly explain your choice of α:

Solution: See the above analysis. We showed that $A = n(1 \pm 2/\sqrt{\alpha})$ with probability at least 3/4, so by choosing $\alpha = 4/\epsilon^2$, we conclude that $A = n(1 \pm \epsilon)$ with probability at least 3/4.

Problem 6. The World Congress[1] [17 points]

In the future, we will all be governed by The World Congress, which has representatives from n countries. The World Congress has many subcommittees, each of which governs an important aspect of daily life (e.g., the Committee on Clean Air, the Committee on the Prevention of War, the Committee on Superhero Management, etc.). Your job is to help design a randomized algorithm for selecting membership of the various committees. There are $k \leq n$ committees in total. Each committee makes decisions based on majority vote.

Unfortunately, some of the representatives represent evil countries that want to overthrow The World Congress and destroy the world.[2] If evil representatives take control of a committee, who knows what harm they may do! It is imperative that we ensure that each committee has a majority of representatives that are good (i.e., *not evil*).

Luckily, we know that at most $1/4$ of the representatives are evil, and at least $3/4$ of the representatives are good. So it shouldn't be too hard to design some good committees, right? Each committee is assigned α members, chosen uniformly and independently at random from the n representatives. (A representative may therefore be on more than one committee, and for simplicity, a representative may be chosen more than once for a single committee.)

Your goal is to choose a value of α so that with probability at least $1-1/n$, all the committees have a good majority.

What value of α do you choose: $24(\ln n + \ln k)$

Prove that all the committees have a good majority with probability at least $1 - 1/n$:

[1] This problem may seem apocryphal, but the basic idea in fact underlies recently popular sharding ideas in cyptocurrencies, where transactions are balanced across smaller committees that can process transactions more quickly.

[2] Their plot is currently unknown, but rumor goes that it involves Thanos, Harvestors, Ice Nine, and burning a lot of oil.

Solution: Fix one committee. Let x_i be an indicator random variable that is equal to 1 if committee member i is good and equal to 0 if committee member i is evil. Therefore we know that $\mathrm{E}[x_i] = \Pr[x_i = 1] \geq 3/4$. Let $X = \sum_{i=1}^{\alpha}(x_i)$ be the number of honest members of the committee. We know that $\mathrm{E}[X] \geq \alpha n/4$. Since the x_i are 0/1 random variables, and are independent, we can use a Chernoff Bound to show that:

$$\begin{aligned} \Pr[X \leq (1 - 1/3)\mathrm{E}[X]] &\leq e^{-\mathrm{E}[X](1/3)^2/2} \\ &\leq e^{-(3/4)(1/9)(1/2)\alpha} \\ &\leq e^{-(1/24)\alpha} \end{aligned}$$

Therefore, if we choose $\alpha = 24(\ln n + \ln k)$, we find that $\Pr[X \leq (1/2)\alpha] \leq 1/nk$. Taking a union bound over the k committees, we conclude that the probability that *any* committees has a majority of evil representatives is at most $1/n$, as desired.

(Extra space on the next page.)

Optional extra space for the previous question:

Scratch Paper

Scratch Paper

Scratch Paper