EECS 70 Discrete Mathematics and Probability Theory
Spring 2013 Anant Sahai MT 2 Solution

PRINT your student ID: _____

PRINT AND SIGN your name: _____, _____ _____
(last) (first) (signature)

PRINT your Unix account login: cs70-_____
PRINT where you are taking this exam: _____
PRINT your discussion section and GSI: (where you want it back) _____

Mark Here	Section	Time	Location	GSI
	1	9-10am	6 Evans	Ramtin
	2	10-11am	71 Evans	Ramtin
	3	11-12pm	71 Evans	Nima
	4	12-1pm	2 Evans	Nima
	5	1-2pm	87 Evans	Sridhar
	6	2-3pm	2070 VLSB	Sridhar
	7	3-4pm	85 Evans	Chung-Wei
	8	4-5pm	9 Evans	Chung-Wei
	9	5-6pm	9 Evans	Richard
	10	1-2pm	3105 Etch.	Chenyu
	11	2-3pm	151 Barr.	Kate
	12	4-5pm	B51 Hilde.	Richard
	13	6-7pm	70 Evans	Sibi
	14	12-1pm	101 Wheel.	Chenyu
	15	4-5pm	156 Dwin.	Sibi

Prob. 1	
Prob. 2	
Prob. 3	
Prob. 4	
Total	

Names of the people sitting next to you: _____

You may consult your two handwritten note sheets. **(You must turn them in with your exam, along with any scratch paper you might have used. Your name and SID should be on each sheet of paper.)** Phones, calculators, tablets, and computers are not permitted. No collaboration is allowed at all and you are not allowed to look at another's work.

Please write your answers in the spaces provided in the test; in particular, we will not grade anything on the back of an exam page unless we are clearly told on the front of the page to look there.

You have 120 minutes. There are 4 questions, of varying numbers of points. The questions are of varying difficulty, so avoid spending too long on any one question.

PRINT your name and student ID: _____

SOME APPROXIMATIONS AND OTHER USEFUL TRICKS THAT MAY OR MAY NOT COME IN HANDY:

$$n! \approx \sqrt{2\pi n}(\frac{n}{e})^n$$

$$\binom{n}{k} \leq (\frac{ne}{k})^k$$

When x is small, $\ln(1+x) \approx x$

When x is small, $(1+x)^n \approx 1+nx$

$$\lim_{n\to\infty}(1+\frac{x}{n})^n = e^x$$

You probably have others on your note sheet and in your minds. Good for you.

The Golden Rule of 70 (and Engineering generally) applies: if you can't solve the problem in front of you, state and solve a simpler one that captures at least some of its essence. You'll get partial credit for doing so, and maybe you'll find yourself on a path to the solution.

Do not turn this page until your instructor tells you to do so.

PRINT your name and student ID: _____

Problem 1. Secret Sharing (15 points)

Suppose that N is the secret number, and is restricted to be one of $0,1,2,3,4$.

A person decides to divide this secret into four shares by using the following K,L secret sharing scheme. (He has kept the identity of K and L hidden from the people he is sharing the secret with.)

To person i, where i is either $1,2,3$, or 4, he gives the share $S_i = (N + K*i + L*i*i) \bmod 5$.

a. 5 points Suppose that three individuals had the following shares: $S_1 = 2, S_2 = 0, S_3 = 2$. **Calculate the secret N and show your work.**

Answer:

$$\left\{\begin{array}{lllllll} S_1 & = & N+K+L & = & N+K+L & = & 2 \pmod 5 \quad \ldots\langle A\rangle \\ S_2 & = & N+2K+4L & = & N+2K+4L & = & 0 \pmod 5 \quad \ldots\langle B\rangle \\ S_3 & = & N+3K+9L & = & N+3K+4L & = & 2 \pmod 5 \quad \ldots\langle C\rangle \end{array}\right.$$

$$\Longrightarrow \left\{\begin{array}{lllll} \langle B\rangle - \langle A\rangle: & K+3L & = & 3 \pmod 5 & \ldots\langle D\rangle \\ \langle C\rangle - \langle B\rangle: & K & = & 2 \pmod 5 & \ldots\langle E\rangle \end{array}\right.$$

$\Longrightarrow$ $K = 2$ to $\langle D\rangle$: $2+3L = 3 \pmod 5$ and then $L = 2 \pmod 5$

$\Longrightarrow$ $K = 2, L = 2$ to $\langle A\rangle$: $N+4 = 2 \pmod 5$ and then $N = 3 \pmod 5$

$\Longrightarrow$ $\boxed{N = 3.}$

PRINT your name and student ID: _____

b. 10 points Now suppose that the people learn the method by which the secret N and the keys K, L were generated:

First, he rolls a 5-sided fair die (labeled 0,1,2,3,4) and calls the first number K.

Then, he rolls a 5-sided fair die (labeled 0,1,2,3,4) and calls the second number L.

Finally, he rolls a 4-sided fair die (labeled 0,1,2,3) and calls the third number N. (So they already know it can't be 4.)

Prove that the event $\{N = 2\}$ is independent of the event $\{S_1 = 2, S_2 = 0\}$.

Answer: We will use the triples (N, K, L) as the sample space. Clearly, all 100 outcomes are equally likely.

First, given $S_1 = 2$ and $S_2 = 0$, we have

$$\begin{aligned}
&\begin{cases} S_1 &= N+K+L &= 2 \pmod 5 & \dots\langle A\rangle \\ S_2 &= N+2K+4L &= 0 \pmod 5 & \dots\langle B\rangle \end{cases} \\
\iff &\begin{cases} \langle A\rangle + \langle B\rangle: & 2N+3K &= 2 \pmod 5 \\ \langle A\rangle \times 3 + \langle B\rangle: & 4N+2L &= 1 \pmod 5 \end{cases} \\
\iff &\begin{cases} K &= N+4 \pmod 5 \\ L &= 3N+3 \pmod 5 \end{cases}
\end{aligned}$$

so

- If $N = 0$, then $K = 4, L = 3$.
- If $N = 1$, then $K = 0, L = 1$.
- If $N = 2$, then $K = 1, L = 4$.
- If $N = 3$, then $K = 2, L = 2$.

Next, we can compute the following probabilities:

- $\Pr[N = 2] = \frac{1\times5\times5}{4\times5\times5} = \frac{1}{4}$ (obviously since this was given in the problem statement).
- $\Pr[S_1 = 2, S_2 = 0] = \Pr[(N,K,L) = (0,4,3) \text{ or } (1,0,1) \text{ or } (2,1,4) \text{ or } (3,2,2)] = \frac{4}{4\times5\times5} = \frac{1}{25}$.
- $\Pr[N = 2, S_1 = 2, S_2 = 0] = \Pr[(N,K,L) = (2,1,4)] = \frac{1}{4\times5\times5} = \frac{1}{100}$.

Since $\Pr[N = 2] \times \Pr[S_1 = 2, S_2 = 0] = \frac{1}{100} = \Pr[N = 2, S_1 = 2, S_2 = 0]$, the two events are independent.

Alternatively, the probability of $N = 2$ given $S_1 = 2, S_2 = 0$ is clearly $\frac{1}{4}$ since, of the 4 possible choices above, exactly one has $N = 2$. Since conditioning on the event $S_1 = 2, S_2 = 0$ did not change the probability of the event $N = 2$, the two events are independent.

PRINT your name and student ID: _____

Problem 2. Counting your ABCs (20 points)

a. (8 points) **How many strings of length $4n$ are there that have exactly n letter 'a's, n letter 'b's, and $2n$ letter 'c's?**

Answer: Imagine one had $4n$ distinct tiles, n of which were marked a (but still considered distinct), n of which were marked b, and $2n$ of which were marked c. The number of ways to order these tiles can be computed in two different ways: the simpler way is to note that it is $(4n)!$, since it is the number of orderings of $4n$ distinct items, and the more complicated way is note that it is equal to the number of ways to first choose where in the tile-ordering each of the three letters will show up (the quantity we are looking for), times the number of ways to order the n 'a' tiles among themselves ($n!$), times the number of ways to order the n 'b' tiles among themselves ($n!$), times the number of ways to order the $2n$ 'c' tiles among themselves ($(2n)!$). As these two quantities are equal, we see that the value we are looking for must equal $\frac{(4n)!}{n!n!(2n)!}$.

Put another way: to pick such a string, one can first pick any n of the $4n$ possible positions to contain the n 'a's, then pick any n of the remaining $4n-n=3n$ remaining positions to contain the 'b's, then finally place the 'c's in all $3n-n=2n$ remaining positions. This means the number of such strings is $\binom{4n}{n}\binom{3n}{n} = \frac{(4n)!}{n!(3n)!}\frac{(3n)!}{n!(2n)!} = \frac{(4n)!}{n!n!(2n)!}$.

b. (12 points) **Use Stirling's approximation to estimate how fast the answer to the previous part grows as a function of n. Is it essentially exponential in n? If so, it is like what to the nth power?**

Answer: Stirling's approximation tells us that $n!$ is approximately $\sqrt{2\pi n}(\frac{n}{e})^n$, for large n.

Simply mechanically plugging our answer to part a) into this approximation for the factorial yields $\frac{(4n)!}{n!n!(2n)!} \approx \frac{\sqrt{2\pi(4n)}(\frac{4n}{e})^{4n}}{\sqrt{2\pi n}(\frac{n}{e})^n\sqrt{2\pi n}(\frac{n}{e})^n\sqrt{2\pi(2n)}(\frac{2n}{e})^{2n}}$.

The rest is just algebra to simplify this: we can rewrite this as

$$\frac{\sqrt{2\pi(4n)}}{\sqrt{2\pi n}\sqrt{2\pi n}\sqrt{2\pi(2n)}}\frac{(\frac{4n}{e})^{4n}}{(\frac{n}{e})^n(\frac{n}{e})^n(\frac{2n}{e})^{2n}}$$

$$= \frac{1}{\sqrt{2}\pi n}\frac{4^{4n}(\frac{n}{e})^{4n}}{(\frac{n}{e})^n(\frac{n}{e})^n 2^{2n}(\frac{n}{e})^{2n}} = \frac{1}{\sqrt{2}\pi n}\frac{4^{4n}}{2^{2n}} = \frac{(\frac{4^4}{2^2})^n}{\sqrt{2}\pi n}$$

The multiplicative effect of the denominator here is negligible in comparison to the numerator; this function is essentially exponential with base $\frac{4^4}{2^2} = 64$.

PRINT your name and student ID: _____

Problem 3. [True or false] (24 points)

Circle TRUE or FALSE.

Prove all statements that you think are true and disprove (e.g. by showing a counterexample) all statements that you think are false.

Reminder: $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ represents the set of non-negative integers.

(a) TRUE or FALSE: If $0 < P(A) < P(B) < 1$, then $P(A|B) < P(B)$.

Answer: False

The simplest counterexample would be one in which the event A is completely contained in the event B (though this is not sufficient—be careful!). For example, we can consider the outcome of two fair coin tosses. Let A be the event that both tosses come up heads and B be the event that the first toss comes up heads. Then $P(A) = 1/4$ and $P(B) = 1/2$ so we have satisfied the condition that $0 < P(A) < P(B) < 1$.

Observe that

$$P(A \cap B) = 1/4 = P(A|B) \cdot P(B) = P(A|B) \cdot 1/2 \Rightarrow P(A|B) = 1/2$$

so we have shown that $P(A|B) = P(B)$ which means that the statement above ($P(A|B) < P(B)$) is *not* true.

Another counterexample is one in which event A is completely contained in B and just slightly smaller than it (e.g. only one outcome out of very many is in B but not A). To make this concrete, flip 100 fair coins and let B be the event that we flip 99 heads in a row (starting with the first toss) and A be the event that we flip 100 heads in a row. Clearly $P(A) < P(B)$ (and both are much smaller than 1—and much smaller than 1/2) but $P(A|B)$ is $1/2 > P(B)$.

Common mistakes:

1. Some students claimed to have constructed an example in which $P(A \cap B)$ was greater than both $P(A)$ and $P(B)$, which cannot happen (draw a picture to see why this is true).
2. Some students correctly identified that this statement is true when A and B are independent. However, the problem statement does not state that they are independent so you cannot assume this is true.
3. Some students falsely concluded that $P(A|B) = 1$ for their example. With $P(A) < P(B)$, this cannot be true (again, draw a picture to see why this is true). (Also note that $P(B|A) = 1$ *is* possible, e.g. if the event A is completely contained within event B as in our example above.)
4. Some students did not enforce the condition $0 < P(A) < P(B) < 1$, often using $P(A) > P(B)$.

(b) TRUE or FALSE: If $P(A) = \frac{1}{4}$ and $P(B) = \frac{1}{3}$ and $P(A \cup B) = \frac{1}{2}$, then A and B are independent.

Answer: True

By the inclusion-exclusion principle, we can *always* write (regardless of independence!) that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

thus we see that

$$1/2 = 1/4 + 1/3 - P(A \cap B)$$

so $P(A \cap B) = 1/12$. We then notice that $P(A) \cdot P(B) = 1/12 = P(A \cap B)$. We know that A and B are independent if and only if $P(A) \cdot P(B) = P(A \cap B)$ so A and B are independent.

Common mistakes:

1. There was a lot of confusion about disjoint vs. independent. $P(A \cap B)$ is zero for disjoint events but it need not be zero for independent events.
2. Similarly, many students falsely concluded that since $P(A \cup B) \neq P(A) + P(B)$, the events were dependent. This equation holds for disjoint events only.
3. Some students asserted that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ is only true when the events are independent. This statement is always true whereas $P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B)$ is only true if they're independent (since it's using the definition of independence).
4. In many cases, proofs could have been greatly improved if the author had simply written out all of the steps, but this is admittedly hard on an exam when you're heavily constrained in time.

(c) TRUE or FALSE: Let $f_k(x)$ be a real polynomial of degree k that takes the value 1 at $x = k$ and takes the value 0 for $x = 0, 1, \ldots, k-1$. Then, $f_k(x)$ must take non-negative integer values at all non-negative integers x.

Answer: True

Since $f_k(x)$ has roots at $0, 1, ..., k-1$, we know that it must be divisible by ("include") the polynomial $q_k(x) = (x-0)(x-1)\cdots(x-(k-1))$. We also know that $f_k(k) = 1$. Since $q_k(k) = k \cdot (k-1)\cdots 1 = k!$ and $f_k(x)$ cannot have any more roots (by its degree), we know that

$$f_k(x) = q_k(x)/k! = (x-0)(x-1)\cdots(x-(k-1))/k!$$

Next, we notice that we can write

$$q_k(x) = \frac{x!}{(x-k)!}$$

(check this yourself by fully writing out $q_k(x)$) and so

$$f_k(x) = \frac{x!}{(x-k)! \cdot k!} = \binom{x}{k}$$

Note that $f_k(x)$ for $x = 0, 1, ..., k-1$ is already guaranteed to be a non-negative integer by the problem statement. Thus we now restrict our attention to the case where $x \geq k$.

We know that $\binom{x}{k}$ (in general) will always result in positive integer (since it is counting the number of ways to choose k things from x things which can only be an integer) when $x \geq k$ and both x and k are integers, which is given in the problem statement. Thus $f_k(x)$ will be non-negative and integer-valued for integer values of $x \geq k$ and we are done.

Common mistakes:

1. By far the biggest mistake was forgetting to show that $f_k(x)$ will have *integer* values when x is a non-negative integer.
2. Many students gave the following argument: $f_k(x)$ has already "used up" its k roots, thus it can no longer cross the x axis. The function is positive ($f_k(k) = 1$) at $x = k$, so it must remain above the x-axis for $x \geq k$ since crossing the x-axis would require another root. This argument is correct (and most students made it rigorously—yay!) but it fails to show that $f_k(x)$ must take on integer values.
3. Many students attempted to prove this statement via induction. There is a correct inductive proof for this statement[1] but very few used this technique. A key observation is that the functions $f_k(x)$ (for various k) are not related, so induction on k is meaningless.
4. A few students argued that $f_3(1)$ (for example) was equal to zero which was not a non-negative integer and therefore the statement was false. However, 0 is a non-negative integer so this reasoning is faulty.

[1] This essentially involves noticing that $f_k(m+1) = d(m) \cdot f_k(m)$ for some $d(m)$ and then using the fact that $f_k(m)$ is positive (by the inductive hypothesis) and that $d(m) > 0$ by its form. Finding $d(m)$ is left as an exercise for the reader. The reader must also carefully argue that $f_k(m+1)$ will have an integer value if m is an integer.

5. Many students seemed to forget that $f_k(k) = 1$ and thus forgot to divide $q_k(x)$ by $k!$ when finding $f_k(x)$.
6. Some students incorrectly found the formula for $f_k(x)$ via Lagrange interpolation (the right method but they messed up somewhere). This led them to non-integer results for their test cases. To prevent yourself from making this mistake or the previous one, it is always a good idea to check some test cases that you know *must* be right by the problem formulation, e.g. $f_2(1) = 0, f_2(2) = 1$.
7. Some students made minor arithmetic/off-by-one errors such as omitting the $(x-0)$ term from $q_k(x)$ (and the corresponding k from the $k!$) or simplifying $(x-(k-1))$ as $(x-k-1)$. These led to difficulties in finding the simpler form of $f_k(x)$.

PRINT your name and student ID: _____

Problem 4. Orpheus' Adventures in the Halls of Time (55 points)

You're designing a new role-playing game for a mathematically themed production house. Your eccentric colleague comes to you with an idea for a key scene and he wants you to think about it.

The backstory is that the mortal Orpheus wants to gain knowledge of the dates of certain key events in the year to come: call these the prophecies of interest. He has heard that in the Halls of Time, these things are already known so he quests through the underworld till he comes upon them.

In the Halls of Time, he encounters the Guardians. They have access to the knowledge of the Fates.

a. (20 points) On the medium difficulty setting, the game behaves as follows. There are 12 guardians (corresponding to the 12 constellations of the Zodiac or the 12 months) and each knows all the prophecies, but they have a peculiar property. Half of them are honest and answer questions posed to them exactly. One quarter of them consider mortals to be beneath them and will simply say "Begone mortal!" And one quarter despise mortals and will answer maliciously.

But mortals do not know the secret forms of the guardians and so Orpheus doesn't know who he is talking to.

On this setting, Orpheus can only ask questions (he can invoke arithmetic operations in $GF(367)$ if he wants) whose answer is a number from $\{0,1,2,\ldots,366\}$.

(Hint: You can ask them to encode a prophecy of interest as follows: 1, ..., 365 for the days in the coming year. 0 for the past. 366 to represent the future beyond this coming year. Fortunately for Orpheus, 367 happens to be prime.)

(The prophecies he wants are answers to questions like: "When will my wife die?" Using the above hint, these can be viewed as numbers.)

He can only ask any individual guardian one question. After that, that particular guardian will magically leave the room. He gets to question all 12 guardians.

How many prophecies can Orpheus reliably extract from the 12 guardians? How can he do it? (Be explicit) Why will this work?

Answer: We map the date of prophecy i to a number $p_i \in \{0,1,\ldots,366\}$ as stated in the hint. Orpheus asks each guardian j the following question: "What is the value of the polynomial $P(x) = p_1x^2 + p_2x + p_3$ at $x = j$?".

From the 12 questions he asks, he receives 9 points. At least 6 of these points lie on the polynomial P. This problem is now exactly the same as the general error correcting problem when using Reed-Solomon codes ($n = 3$,$k = 3$). We apply the Reed-Solomon algorithm to reconstruct the polynomial P. From the coefficients of P, Orpheus can determine the dates of prophecy $1,2,3$.

b. (5 points) Your friend comes to you with a further twist for the hard-mode of the game. He says that one-quarter (3) of the guardians are lazy instead of being honest. Each lazy guardian just randomly chooses one of its sibling guardians and telepathically asks them how they would answer this particular question. (If it happens to choose another lazy guardian, that guardian will again telepathically ask a random sibling until someone answers or would say "begone.") Once it gets some answer telepathically, the lazy guardian being questioned simply repeats that answer out loud and leaves the room. (Telepathic conversations don't cause guardians to leave the room, only speaking an answer out loud.)

If we choose one of the 12 guardians uniformly at random to query, what is the probability that we get a malicious answer?

Answer: The last guardian to provide an answer must be a non-lazy guardian, each of which are equally likely to be chosen. There are 3 malicious and 9 total non-lazy. Thus the probability of getting a malicious answer is $\frac{3}{9} = \frac{1}{3}$.

Notice that this answer is independent of whether or not a lazy guardian can ask another lazy guardian or if there are cycles.

c. (10 points) Using the setup from part (b) but asking questions of all 12 guardians, **if Orpheus only wants to extract a single prophecy, how should Orpheus proceed and what is his probability of being successful?**

Answer: There are an equal number of truthful and malicious guardians so by symmetry, Orpheus can do no better than 50% (since the lazy ones are not more likely to choose truthful vs malicious telepathically). One way to achieve this is to ask all 12 guardians and take the answer that comes up most often (breaking ties by choosing one answer randomly).

d. (20 points) On the easy difficulty setting, the game is quite different. There are just 400 guards and they don't know much. They can carry prophecy books, each containing one prophecy. But the Fates have scattered 200 distinct books of prophecy across them by assigning each prophecy at random (by rolling a fair 400-sided die) to one of the guards. The guards carry their books and because Orpheus can only get the prophecies from one guard, he finds the one who is carrying the most and asks her for everything she is carrying.

Orpheus knows that the guard will demand 1 chicken for every prophecy book that she carries. **How many chickens should Orpheus carry with him so that he has at least about a fifty percent chance of being able to get all the prophecy books carried by that guard?** (Try not to have Orpheus carry too many chickens if you can.)

Answer:

This is a load balancing problem. It was the hardest question on the midterm. Because of this, we give a very thorough explanation here, but we didn't expect anyone to give the full explanation on the exam or in the homework. If you even got any reasonable approximation that gives something in the range 4-8, you got most of the credit. The first couple of approximations (assuming $m = n$) below were worth most of the points in the question.

We can solve this problem by the following 5 steps: formulation, reduction, probability derivation, probability approximation, and solution estimation. We first define $m = 200$, $n = 400$, X as the number of books carried by the guard who has the most books, X_i as the number of books carried by guard i, and k is the number of chickens.

THE FIRST APPROXIMATION

This is a little different from the lecture note because Orpheus is still fine if $X = k$. If we formulate it as $\Pr[X \geq k] \leq \frac{1}{2}$, we will miss the smallest (optimal) solution. Besides, $m \neq n$ in this case. However, we can still use the results in the lecture note. The main observations are:

- As mentioned above, if we formulate it as $\Pr[X \geq k] \leq \frac{1}{2}$ as the lecture note, we will miss the smallest (optimal) solution. However, we will find a larger solution, which is still a feasible solution, *e.g.*, if Orpheus carries 7 chickens, we can guarantee that he has at least 50% probability to get all the books carried by the guard.
- If we assume $m = n = 200$ or $m = n = 400$, again, we may miss the smallest (optimal) solution, but we will find a larger and feasible solution. This is because, when $m = n = 200$ (200 books are distributed to 200 guards) or $m = n = 400$ (400 books are distributed to 400 guards), a guard has a higher probability to get a book, compared with $m = 200, n = 400$ (200 books are distributed to 400 guards). As a result, we can claim

$$\begin{aligned} \Pr[X \geq k | (m = 200, n = 400)] &< \Pr[X \geq k | (m = 200, n = 200)]; \\ \Pr[X \geq k | (m = 200, n = 400)] &< \Pr[X \geq k | (m = 400, n = 400)]. \end{aligned}$$

Applying the results in the lecture note for $\Pr[X \geq k | (m = 200, n = 200)]$ or $\Pr[X \geq k | (m = 400, n = 400)]$, we can find $k = \lceil \ln(4 \times 200) \rceil = 7$, $k = \lceil \ln(4 \times 400) \rceil = 8$, $k = \left\lceil \frac{2\ln(200)}{\ln(\ln 200)} \right\rceil = 7$, or $k = \left\lceil \frac{2\ln(400)}{\ln(\ln 400)} \right\rceil = 7$. We will see that they are all feasible but not optimal solutions.

A BETTER APPROXIMATION

1. **Formulation**: this problem is to find the smallest integer k such that

$$\Pr[X > k] \leq \frac{1}{2}. \tag{1}$$

2. **Reduction**: it is much easier to analyze the number of books carried by one guard, so we reduce the problem to find the smallest integer k such that

$$\Pr[X_i > k] \leq \frac{1}{2n}. \tag{2}$$

Comments: if k satisfies Equation (2), then it is guaranteed that k satisfies Equation (1). We can proved this by

$$\begin{aligned}
\Pr[X > k] &= \Pr\left[\bigcup_{i=1}^{n}(X_i > k)\right] \\
&\leq \sum_{i=1}^{n}\Pr[X_i > k] \\
&\leq n \times \frac{1}{2n} \\
&= \frac{1}{2}.
\end{aligned}$$

3. **Probability Derivation**: we are trying to find the smallest integer k such that $\Pr[X_i > k] \leq \frac{1}{2n}$. $\Pr[X_i > k]$ is the probability that guard i has more than k books, and the probability that guard i has exactly j books is $\binom{m}{j}\left(\frac{1}{n}\right)^j\left(1-\frac{1}{n}\right)^{m-j}$, so

$$\Pr[X_i > k] = \sum_{j=k+1}^{m}\binom{m}{j}\left(\frac{1}{n}\right)^j\left(1-\frac{1}{n}\right)^{m-j}.$$

4. **Probability Approximation**:

$$\begin{aligned}
\Pr[X_i > k] &= \sum_{j=k+1}^{m} \binom{m}{j} \left(\frac{1}{n}\right)^j \left(1-\frac{1}{n}\right)^{m-j} \\
&\quad \lozenge \left(1-\tfrac{1}{n}\right) \le 1 \\
&\le \sum_{j=k+1}^{m} \binom{m}{j} \left(\frac{1}{n}\right)^j \\
&\quad \lozenge \text{ standard approximation} \\
&\le \sum_{j=k+1}^{m} \left(\frac{me}{j}\right)^j \left(\frac{1}{n}\right)^j \\
&= \sum_{j=k+1}^{m} \left(\frac{me}{nj}\right)^j \\
&= \sum_{j=k+1}^{200} \left(\frac{200e}{400j}\right)^j \\
&= \sum_{j=k+1}^{200} \left(\frac{e}{2j}\right)^j \\
&\quad \lozenge\ j \ge k+1 \Rightarrow \tfrac{1}{j} \le \tfrac{1}{k+1} \\
&\le \sum_{j=k+1}^{200} \left(\frac{e}{2(k+1)}\right)^j \\
&= \left(\frac{e}{2(k+1)}\right)^{k+1} \left(1+\frac{e}{2(k+1)}+\left(\frac{e}{2(k+1)}\right)^2+\ldots+\left(\frac{e}{2(k+1)}\right)^{199-k}\right) \\
&\quad \lozenge\ \tfrac{e}{(k+1)} \le 1 \text{ (need to check this after finding a solution)} \\
&\le \left(\frac{e}{2(k+1)}\right)^{k+1} \left(1+\frac{1}{2}+\left(\frac{1}{2}\right)^2+\ldots+\left(\frac{1}{2}\right)^{199-k}\right) \\
&\quad \lozenge \text{ geometric series} \\
&\le \left(\frac{e}{2(k+1)}\right)^{k+1} (2) \\
&= 2\left(\frac{e}{2(k+1)}\right)^{k+1}
\end{aligned}$$

Comments: a good approximation here should use "$\le$" only (do not use "$\ge$") so that $2\left(\frac{e}{2(k+1)}\right)^{k+1} \le \frac{1}{2n}$ can guarantee $\Pr[X_i > k] \le \frac{1}{2n}$.

5. **Solution Estimation**: we are trying to find the smallest integer k such that $2\left(\frac{e}{2(k+1)}\right)^{k+1} \le \frac{1}{2n} = \frac{1}{800}$. For easy calculation (no calculator in the midterm), we approximate e as 3, and we are trying to find the smallest integer k such that

$$2\left(\frac{3}{2(k+1)}\right)^{k+1} \le \frac{1}{800}. \tag{3}$$

- If $k = 4$, the LHS is $2\left(\frac{3}{10}\right)^5 = \frac{486}{10000} > \frac{1}{800}$.
- If $k = 5$, the LHS is $2\left(\frac{3}{12}\right)^6 = \frac{1}{2048} \leq \frac{1}{800}$.

Therefore, we find $\boxed{5}$ as the solution (check $\frac{e}{(k+1)} \leq 1$, yes) — if Orpheus carries 5 chickens, we can guarantee that he has at least 50% probability to get all the books carried by the guard.

Comments: you can still get full credit if you estimate k as 6, 7, or 8 for Equation (3) in this step.

AN EVEN BETTER APPROXIMATION

4. **Probability Approximation**:

$$
\begin{aligned}
\Pr[X_i > k] &= \sum_{j=k+1}^{m} \binom{m}{j} \left(\frac{1}{n}\right)^j \left(1-\frac{1}{n}\right)^{m-j} \\
&\quad \Diamond \left(1-\tfrac{1}{n}\right) \leq 1 \\
&\leq \sum_{j=k+1}^{m} \binom{m}{j} \left(\frac{1}{n}\right)^j \\
&= \sum_{j=k+1}^{m} \left(\frac{m(m-1)(m-2)\ldots(m-(j-1))}{j!}\right) \left(\frac{1}{n}\right)^j \\
&\quad \Diamond\, m-1 \leq m,\, m-2 \leq m,\, \ldots,\, m-(j-1) \leq m \\
&\leq \sum_{j=k+1}^{m} \left(\frac{m^j}{j!}\right) \left(\frac{1}{n}\right)^j \\
&= \sum_{j=k+1}^{m} \left(\frac{1}{j!}\right) \left(\frac{m}{n}\right)^j \\
&= \sum_{j=k+1}^{200} \left(\frac{1}{j!}\right) \left(\frac{200}{400}\right)^j \\
&= \sum_{j=k+1}^{200} \left(\frac{1}{j!}\right) \left(\frac{1}{2}\right)^j \\
&\quad \Diamond\, j \geq k+1 \Rightarrow \tfrac{1}{j} \leq \tfrac{1}{k+1} \\
&\leq \sum_{j=k+1}^{200} \left(\frac{1}{(k+1)!}\right) \left(\frac{1}{2}\right)^j \\
&= \frac{1}{(k+1)!} \left(\frac{1}{2}\right)^{k+1} \left(1+\frac{1}{2}+\left(\frac{1}{2}\right)^2+\ldots+\left(\frac{1}{2}\right)^{199-k}\right) \\
&\quad \Diamond \text{ geometric series} \\
&= \frac{1}{(k+1)!} \left(\frac{1}{2}\right)^{k+1} (2) \\
&= \frac{1}{(k+1)!} \left(\frac{1}{2}\right)^{k}
\end{aligned}
$$

5. **Solution Estimation**: we are trying to find the smallest integer k such that

$$\frac{1}{(k+1)!}\left(\frac{1}{2}\right)^k \leq \frac{1}{800}. \tag{4}$$

- If $k=0$, the LHS is $\left(\frac{1}{1}\right)\left(\frac{1}{2}\right)^0 = 1 > \frac{1}{800}$.
- If $k=1$, the LHS is $\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^1 = \frac{1}{4} > \frac{1}{800}$.
- If $k=2$, the LHS is $\left(\frac{1}{6}\right)\left(\frac{1}{2}\right)^2 = \frac{1}{24} > \frac{1}{800}$.
- If $k=3$, the LHS is $\left(\frac{1}{24}\right)\left(\frac{1}{2}\right)^3 = \frac{1}{192} > \frac{1}{800}$.
- If $k=4$, the LHS is $\left(\frac{1}{120}\right)\left(\frac{1}{2}\right)^4 = \frac{1}{1920} \leq \frac{1}{800}$.

Therefore, we find $\boxed{4}$ as the solution — if Orpheus carries 4 chickens, we can guarantee that he has at least 50% probability to get all the books carried by the guard.

SIMULATION RESULT

The following table is the simulation result with 1,000,000 runs. In each run, each of 200 books is randomly assigned to a guard, and the value of X (the number of books carried by the guard who has the most books) is recorded.

X	2	3	4	5	6	7	8	9
Times (Numbers of Runs)	1,634	494,540	440,045	58,555	4,876	325	23	2

This table shows that there are 494,540 runs (among 1,000,000 runs) with $X = 3$. We can see, if Orpheus carries 4 chickens, he can deal with $\frac{1{,}634+494{,}540+440{,}045}{1{,}000{,}000} > 93\%$ cases, which is much higher than 50%. One may ask: is 3 also a solution? Since $\sqrt{1{,}000{,}000} = 1{,}000$ and the numerical gap from 50% is about $4{,}000$, we are actually kind of sure that 3 is probably not enough. This will be justified later in the course.

EXACT SOLUTION

In fact, $\Pr[X \leq 3]$ can be exactly computed by the following algorithm:

```
1  p = 0;
2  for i = 0 to (200/3) {
3     for j = 0 to ((200-3i)/2) {
4        k = 200 - 3i - 2j;
5        p = p + C(400,i) C(400-i,j) C(400-i-j,k) (200!/((3!)^i (2!)^j (1!)^k)) (1/400^200);
6     }
7  }
8  return p;
```

The algorithm returns $\Pr[X \leq 3] = 49.6\%$, implying that carrying 3 chickens is not sufficient, and 4 is the optimal solution here.