Chris Piech
CS109

Fall 2024
Oct 29th

CS109 Midterm Exam

This is a closed calculator/computer exam. You are, however, allowed to use notes in the exam. You have 2 hours (120 minutes) to take the exam. The exam is 120 points, meant to roughly correspond to one point per minute of the exam. You may want to use the point allocation for each problem as an indicator for pacing yourself on the exam.

In the event of an incorrect answer, any explanation you provide of how you obtained your answer can potentially allow us to give you partial credit for a problem. For example, describe the distributions and parameter values you used, where appropriate. It is fine for your answers to include summations, products, factorials, exponentials, and combinations. You can leave your answer in terms of Φ (the CDF of the standard normal) or Φ^{-1}. For example $\Phi(3/4)$ is an acceptable final answer. Code can be written in pseudo code. In pseudo code, ideas are more important than syntax.

I acknowledge and accept the letter and spirit of the honor code. I pledge to write more neatly than I have in my entire life:

Signature: ______________________________

Family Name (print): ______________________________

Given Name (print): ______________________________

Stanford Email (@stanford.edu): ______________________________

1 Adventure is Out There [20 points]

```
def adventure_is_out_there():
    mystery = bernoulli(0.3)
    if mystery == 1:
        return bernoulli(0.5) == 1
    else:
        return good_times()

def good_times():
    treasure = bernoulli(0.2)
    kindness = bernoulli(0.7)
    return treasure == 1 or kindness == 1

def bernoulli(p):
    # returns 1 with probability p
    # returns 0 with probability (1-p)
    if random() < p: return 1
    else: return 0
```

a. (5 points) You call `good_times()`. What is the probability that `good_times` returns `True`?

Let T be the event `treasure` is 1 and K be the event `kindness` is 1.

We can use the Inclusion-Exclusion Rule to find the probability of OR:

$$\begin{aligned} P(\text{good\_times returns True}) &= P(T \text{ or } K) \\ &= P(T) + P(K) - P(T \text{ and } K) \\ &= P(T) + P(K) - P(T) \cdot P(K) \\ &= 0.2 + 0.7 - 0.2 \cdot 0.7 \end{aligned}$$

Alternatively, we could subtract the complement: $1 - 0.3 \cdot 0.8$

b. (5 points) You call `adventure_is_out_there()`. What is the probability that `adventure_is_out_there` returns `True`? if needed let p_a be your answer to part (a).

Let M be the event that `mystery` was 1 and B be the event that the `bernoulli(0.5)` call on line 4 returned a 1.

There are two cases where this function returns True: if M and B both happened, or if M didn't happen, and then `good_times` returned True. These two possibilities are mutually exclusive, so we can break down this probability of OR into a sum.

$$\begin{aligned} P(\text{adventure\_is\_out\_there returns True}) &= P(M \text{ and } B \text{ or } M^C \text{ and good\_times returns True}) \\ &= P(M \text{ and } B) + P(M^C \text{ and good\_times returns True}) \\ &= P(M)P(B|M) + P(M^C)P(\text{good\_times returns True}|M^C) \\ &= 0.3 \cdot 0.5 + (1 - 0.3)p_a \\ &= 0.15 + 0.7p_a \end{aligned}$$

c. (6 points) You call `adventure_is_out_there()` and it returns `True`. What is the probability that `mystery` was 1? if needed let p_b be your answer to part (b).

Let A be the event that `adventure_is_out_there()` returns `True`. Let M be the event that `mystery` was 1. Here, we are told the A happened, and we want to know if M happened, conditioned on A happening. This conditional is hard to reason about directly, but if it was flipped, it would be easier to calculate a number for; this tells us that we should use Bayes' Theorem.

We know that $P(M) = 0.3$, and $P(A) = p_b$ is from part b. If M happens, then the function returns True if the Bernoulli call on line 4 is 1, and False otherwise. Thus, $P(A|M) = P(B) = 0.5$.

$$\begin{aligned} P(M|A) &= \frac{P(A|M)P(M)}{P(A)} \\ &= \frac{0.5 \cdot 0.3}{p_b} \end{aligned}$$

d. (4 points) If `adventure_is_out_there` returns `True`, you win \$100; otherwise, you win \$0. What are your expected winnings? if needed let p_a and p_b be your answer to part (a) and (b) respectively.

Let W be our winnings. The possible values for W are 0 and 100, and the probability that $W = 100$ is p_b, so $P(W = 0) = 1 - p_b$. Using the general formula for expectation:

$$E[W] = 100 \cdot P(W = 100) + 0 \cdot P(W = 0) = 100p_b$$

2 Chance Encounters [20 points]

For each of the following scenarios, **Step 1:** Declare a random variable of a type introduced in class that best models the scenario, specifying its parameters e.g., $X \sim \text{Bin}(n = 10, p = 0.3)$. **Step 2:** State the question in terms of your random variable e.g., $P(X = 2)$. **Do not** solve the probability statements from Step 2.

Important notes: If the problem cannot be modeled using one of the random variables types introduced in class, write **"N/A"** and move on. **Use an approximation** variable if doing so would provide a substantial computational speedup.

a. (5 points) During lecture, Chris has an probability of 0.05 of giving out one fruit for each slide he presents (independently). What is the probability that he gives out his first fruit on slide 5?

Let S be the first slide when Chris gives out fruit. There are multiple possible solutions:

$S \sim \text{Geo}(p = 0.05)$. We want to find $P(S = 5)$.

$S \sim \text{NegBin}(r = 1, p = 0.05)$. We want to find $P(S = 5)$.

We can also consider the number of fruits in which Chris gives out fruits using $S \sim \text{Bin}(n = 5, p = 0.05)$. We want to find $\frac{P(S=1)}{5}$. We have to divide by 5 because we must only count the outcome where Chris gives out fruit on the last of the 5 slides.

b. (5 points) On a clear night at the Stanford Observatory, astronomers see an average of 5 stars per 2-degree by 2-degree area of the night sky. Assuming stars are randomly scattered across the sky, what is the probability that exactly three stars will be seen in a randomly selected 2x2 degree patch?

Let S be the number of stars in a 2x2 degree patch of the night sky. $S \sim \text{Poi}(\lambda = 5)$. We want to find $P(S = 3)$.

c. (5 points) You are flipping 10 unusual coins. The first coin has a probability of 1/1 of landing heads, the second has a probability of 1/2, the third has a probability of 1/3, and so on. What is the probability of observing exactly 5 heads if you flip each of these 10 coins once?

N/A; because the probabilities of success or heads on each coin flip are not consistent, there is no random variable we've learned in class that fits this scenario.

d. (5 points) A blossoming cherry tree currently has 300,000 beautiful pink petals. Today each petal has a 0.45 probability of falling off the tree, independent of the other petals. What is the probability that more than 135,001 petals will fall today?

Let F be the number of petals that fall from the tree today. $F \sim \text{Bin}(n = 300{,}000, p = 0.45)$. We want to find $P(F > 135{,}001)$.

Because this Binomial has extreme n, we need to approximate. Let X be an approximation of F. $X \sim N(\mu = n \cdot p = 300{,}000 \cdot 0.45,\ \sigma^2 = n \cdot p \cdot (1-p) = 300{,}000 \cdot 0.45 \cdot 0.55)$. Using continuity correction, we want to approximate $P(F > 135{,}001)$ using $P(X > 135{,}001.5)$.

3 Bingo! [18 points]

A Bingo card is constructed as a 5x5 grid, with each cell containing a unique integer. The possible values for each cell depend on the column in which it is located:

Column 1: contains numbers in the range 1 to 15 inclusive
Column 2: contains numbers in the range 16 to 30 inclusive
Column 3: contains numbers in the range 31 to 45 inclusive
Column 4: contains numbers in the range 46 to 60 inclusive
Column 5: contains numbers in the range 61 to 75 inclusive

Each column contains five distinct, randomly chosen numbers from its respective range. The center cell of the grid (in the third column) is an exception: it is always marked as "free" and does not contain a number. Below is an example Bingo card, with each column annotated to indicate the possible values for that column.

[1 to 15]	[16 to 30]	[31 to 45]	[46 to 60]	[61 to 75]
8	17	45	55	61
12	16	37	58	66
1	29	Free	47	73
15	20	39	59	71
13	22	31	52	70

a. (6 points) How many unique Bingo cards can be created under these rules? Two cards are unique if they have different values, or if they have the same values, but in different cells.

We can break down the count of the total number of unique Bingo cards using the step rule of counting, where each step is choosing the values for one column.

If there was no free space: for each column, there are 5 cells we need to choose values for. Since the numbers must be distinct, we can't repeat values; so for the first cell in a column, we have 15 choices, and then for the second cell, we have 14, for the third, we have 13, and so on until for the fifth cell, we have 11 choices. Thus for each column there are $15 \cdot 14 \cdot 13 \cdot 12 \cdot 11$ or $\frac{15!}{10!}$ possible ways to choose 5 values.

For the column with only 4 values to pick, we instead have $15 \cdot 14 \cdot 13 \cdot 12$ or $\frac{15!}{11!}$ possibilities.

Our final answer is obtained by multiplying together the possibilities for all five columns together:

$$\left(\frac{15!}{10!}\right)^4 \left(\frac{15!}{11!}\right)$$

Why didn't we use $\binom{15}{5}$ and $\binom{15}{4}$? Because that would treat the values in each column as unordered, but we are told that two bingo cards are unique if a column has the same values in different orders, so we want to count ordered. Writing $\binom{15}{5} \cdot 5!$ is equivalent to $\frac{15!}{10!}$.

b. (6 points) Four distinct integers are randomly selected from the range 1 to 75, where each integer in the range is equally likely. What is the probability that all four selected integers are in the range 31 to 45 inclusive? Note: there are 15 integers in the range [31 to 45].

Since we are in a world where it is doable to count outcomes, we can find this probability by counting the number of outcomes in our event space and sample space and dividing.

The sample space is all the ways we can select 4 distinct integers from the range 1 to 75. Here we do not count about order, so $|S| = \binom{75}{4}$.

The event space is all the ways to select 4 distinct integers from the range 31 to 35, so $|E| = \binom{15}{4}$.

Putting these together, $\frac{|E|}{|S|} = \frac{\binom{15}{4}}{\binom{75}{4}}$.

c. (6 points) Four distinct integers are randomly selected from the range 1 to 75, where each integer in the range is equally likely. What is the probability that exactly:

One integer is from the range 1 to 15 and
One integer is from the range 16 to 30 and
One integer is from the range 46 to 60 and
One integer is from the range 61 to 75?

Note that each range above consists of exactly 15 integers and is inclusive.

We can also find this probability by counting the number of outcomes in our event space and sample space and dividing.

The sample space is the same as the previous part: $|S| = \binom{75}{4}$.

The event space is all the ways to have one value from each of the four ranges of values. Using the step rule of counting, we can multiply the number of ways to choose one value from the first range, times the number of ways to choose one value from the second range, etc. This is $15 \cdot 15 \cdot 15 \cdot 15 = 15^4$.

Putting these together, $\frac{|E|}{|S|} = \frac{15^4}{\binom{75}{4}}$.

Why is the event space not $75 \cdot 60 \cdot 30 \cdot 15$ or similar? That strategy would be counting ordered, but we want to count unordered to be consistent with our sample space count. Alternative solutions could count both the event space and sample space ordered.

4. The Driftwood Random Variable [18 Points]

Here we introduce a new type of **discrete** random variable called the Driftwood. If $X \sim \text{Driftwood}(n)$, then the probability mass function is defined only when x is an integer between 1 and n inclusive:

$$P(X = x) = K \cdot \log(x+1) \quad \text{for } x \in \{1, 2, \cdots n\}$$

K is a normalizing constant. For each subproblem you can assume you have already correctly calculated the answer to previous parts. Eg after subproblem (a) you may refer to K without needing to recalculate it.

a. (5 points) Let $X \sim \text{Driftwood}(n = 5)$. What is K?

For the equation above to be a valid probability mass function, we must have $\sum_{i=1}^{n} P(X = i) = 1$. (Note that we sum over all possible values of $x \in \{1, 2, \cdots n\}$.) Using $x = 5$, we can solve for K:

$$\begin{aligned} 1 &= \sum_{i=1}^{5} P(X = i) = \sum_{i=1}^{5} K \cdot \log(i+1) = K \cdot \sum_{i=1}^{5} \log(i+1) \\ K &= \frac{1}{\sum_{i=1}^{5} \log(i+1)} = \frac{1}{\log(2) + \log(3) + \log(4) + \log(5) + \log(6)} \end{aligned}$$

b. (5 points) Let $X \sim \text{Driftwood}(n = 5)$. What is $P(X \leq 2)$?

$$\begin{aligned} P(X \leq 2) &= P(X = 1) + P(X = 2) \\ &= K \cdot \log(2) + K \cdot \log(3) \end{aligned}$$

c. (5 points) Let $X \sim \text{Driftwood}(n = 5)$. What is $E[X]$?

Using the definition of expectation:

$$\begin{aligned} E[X] &= \sum_{i=1}^{5} i \cdot P(X = i) \\ &= K(1 \cdot \log(2) + 2 \cdot \log(3) + 3 \cdot \log(4) + 4 \cdot \log(5) + 5 \cdot \log(6)) \end{aligned}$$

d. (3 points) Let $X \sim \text{Driftwood}(n = 5)$. What is $\text{Var}(X)$?

Using the definition of variance (and re-using our $E[X]$ from the previous part):

$$\begin{aligned} \text{Var}(X) &= E[X^2] - E[X]^2 \\ &= \left(\sum_{i=1}^{5} i^2 \cdot P(X = i) \right) - E[X]^2 \\ &= K\left(\left(1^2 \cdot \log(2)\right) + \left(2^2 \cdot \log(3)\right) + \left(3^2 \cdot \log(4)\right) + \left(4^2 \cdot \log(5)\right) + \left(5^2 \cdot \log(6)\right) \right) - E[X]^2 \end{aligned}$$

5. True Random [20 Points]

Atmospheric Noise is an authentic source of true random values. Atmospheric Noise follows a normal distribution with mean 2 and variance of 16. You have a electronic box that can sample and return a value from Atmospheric Noise when a program calls a function `true_random()`.

a. (5 points) Let B be the value returned by `true_random()`. What is the probability that B is less than 0?

Since $B \sim N(\mu = 2, \sigma^2 = 16)$,

$$P(B < 0) = P(\frac{B-\mu}{\sqrt{\sigma^2}} < \frac{0-\mu}{\sqrt{\sigma^2}}) = P(\frac{B-2}{\sqrt{16}} < \frac{0-2}{\sqrt{16}}) = \Phi(-1/2) = \boxed{1 - \Phi(1/2)} \approx 0.309$$

As an explicit integral, this would be

$$P(B < 0) = \int_{-\infty}^{0} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx = \boxed{\int_{-\infty}^{0} \frac{1}{4\sqrt{2\pi}} e^{-\frac{(x-2)^2}{32}} \, dx}.$$

But this formula can not be integrated, so using phi as above is necessary.

b. (7 points) Let B be the value returned by `true_random()`. What is the value x where $P(B < x) = \frac{1}{5}$?

The definition of the Normal distribution CDF is $P(B < x) = \Phi(\frac{x-\mu}{\sqrt{\sigma^2}}) = \Phi(\frac{x-2}{4})$. We wish to solve for x, where

$$\Phi(\frac{x-2}{4}) = \frac{1}{5}$$

$$\frac{x-2}{4} = \Phi^{-1}(1/5) \qquad \text{(apply } \Phi^{-1} \text{ function to both sides)}$$

$$x = \boxed{2 + 4\Phi^{-1}(1/5)} \qquad \text{(solve for } x \text{ by isolating it on the LHS).}$$

We need to use the inverse phi function above to isolate x, as it effectively cancels out the phi function on the left-hand side and allows us to get x by itself.

c. (8 points) Write code for a function `true_int(n)` that returns an integer in the range $[0, 1, \ldots, n-1]$ such that each value is equally likely. Your function should make a single call to `true_random()` and use the result to decide which value to return. Your code can call `phi(x)` and `inverse_phi(x)`.

Let B be the result of a single call to `true_random()`.

Solution 1: Our strategy is to divide the area under the curve of the PDF of B into n equal-probability chunks, then figure out which chunk B falls into.

To formalize that mathematically: we want to find chunk boundaries $c_0, c_1, c_2, \ldots, c_n$ such that

$$P(c_i < B < c_{i+1}) = 1/n \qquad \text{for } i = 0, 1, 2, \ldots, n-1$$

These boundaries c_i need to satisfy $P(B < c_i) = i/n$.

Generalizing from Part (b), where we use i/n instead of $1/5$, we can solve for each c_i by plugging in i and n to: $c_i = 4\Phi^{-1}(i/n) + 2$. (Technically, we can say that $c_0 = -\infty$ and $c_n = \infty$.)

Then, `true_int(n)` should return the value i when B lies between chunk i's boundaries of c_i and c_{i+1}.

In code:

```
def true_int(n):
→B = true_random()
→for i in range(n):
→→if (B < 4 * inverse_phi((i+1)/n) + 2):
→→→return i
```

Solution 2:

To sample from a normal distribution, most computers first generate a sample from a standard uniform distribution, then apply the inverse phi function to get a random number from $-\infty$ to ∞ that looks like it was sampled from a normal distribution.

Since this process is reversible, we can do the same thing in reverse to convert a Normal distribution sample into a standard uniform sample between 0 to 1. From there, we can obtain an integer between 0 and $n-1$ by dividing the range from 0 to 1 into equally-sized buckets to ensure equal likelihood of each integer.

In code:

```
def true_int(n):
→B = true_random()
→B_to_standard_normal = (B - 2) / 4
→prob = phi(B_standardized)
→return int(n * prob)
```

6. DNA Mutation Clock [24 Points]

A Mitochondrial DNA base pair is represented by a single letter drawn from {A, T, G, C}. A mutation is when a base pair changes from one of these letters to a different one.

Each Mitochondrial DNA base pair mutates at a rate of 3×10^{-8} mutations per year.

Aside: Assume that the rate of mutation is constant and that the occurrence of one mutation doesn't change the probability of another. Assume that the mutation process is the only way that Mitochondrial DNA changes.

a. (5 points) Consider one base pair. If 10^6 years have passed, what is the probability that the base pair has mutated at least once?

Since the rate of mutation is constant, and the occurrence of one mutation doesn't change the probability of another, mutation is is a poisson process.

Using the Poisson
Let X represent the number of mutations in 10^6 years. Since we are interested in the number of mutations in 10^6 years, we can convert our given rate to be in terms of 10^6 years. Since we observe an average of 3×10^{-8} mutations in one year, we will observe an average of $(3 \times 10^{-8}) \times (10^6) = 0.03$ mutations in 10^6 years.

$$X \sim \text{Poi}(\lambda = 0.03)$$

To find the probability of at least one mutation in 10^6 years, we can solve for $P(X \geq 1)$:

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - \frac{(0.03)^0 e^{-0.03}}{0!} \\ &= 1 - e^{-0.03} \end{aligned}$$

Using the Exponential
Alternatively, let Y be the amount of years till the first mutation. We know that $Y \sim Exp(3 \times 10^{-8})$ and we need to find $P(Y < 10^6)$. After plugging into the CDF, we get $(1 - e^{-10^6 \times 3 \times 10^{-8}}) = 1 - e^{-0.03} \approx 0.0296$. Other choices of λ also work, as long as $\lambda \cdot x = 0.03$.

b. (5 points) A strand of Mitochondrial DNA has 10,000 base pairs. What is the probability that exactly 10 distinct base pairs on the strand will mutate at least once in 10^6 years? Let p_a be your answer to part (a).

With the Binomial (Exact)
Let X be the distribution of the number of distinct base pairs that have mutated at least once in 10^6 years with a population of 10,000. Since each base pair is an independent trial, let $X \sim \text{Bin}(n = 10000, p = p_a)$. We want to determine $P(X = 10)$.

$$P(X = 10) = \binom{10000}{10} p_a^{10} (1 - p_a)^{10000-10}$$

With the Poisson (Approximate)
Alternatively, since n is large and p_a is small (< 0.05), a Poisson can approximate the binomial. Let Y be the approximating Poisson distribution of X. Therefore, $Y \sim \text{Poi}(\lambda = 10000 p_a)$. We want to determine $P(Y = 10)$.

$$P(Y = 10) = \frac{(10000 p_a)^{10} e^{-10000 p_a}}{10!}$$

c. (8 points) Suppose you observe that exactly 10 distinct base pairs have mutated at least once across 10,000 base pairs. What is the probability that exactly k years have passed? Assume a prior belief that the number of years passed is equally likely to have been any value from 0 to 10^{100} years.

This is in inference problem. We have a prior belief in how many years have passed, and we want to update that belief given information.

Let Y be the number of years that have passed (the random variable we want to update our belief in). Let be B be the number of distinct pairs that have mutated at least once (our observation). We want to find $P(Y = k|B = 10)$.

We are told that our prior is $P(Y = k) = \frac{1}{10^{100}}$, since any year in the past 10^{100} is equally likely.

Our likelihood is represented by $P(B = 10|Y = k)$. As we saw in parts a and b, the probability of a certain number of basepairs mutating changes depending on how much time has passed. Specifically, the amount of time passed is incorporated into the calculation in part a, which tells us the probability of a single basepair mutating. If we generalize the calculation in part a for any k years, then part b would generalize to the following binomial:

$$B|Y = k \sim Bin(n = 10^4, p = 1 - e^{-(3\times10^{-8})k})$$

From here, we plug the PMF of the binomial at $B = 10$, along with our prior probability, into Bayes' Theorem:

$$\begin{aligned} P(Y = k|B = 10) &= \frac{P(B = 10|Y = k)P(Y = k)}{P(B = 10)} \\ &= \frac{\binom{10^4}{10}(1 - e^{-(3\times10^{-8})k})^{10} \cdot e^{-(3\times10^{-8})k\cdot(10^4-10)} \cdot \frac{1}{10^{100}}}{\sum_{k=1}^{10100} \binom{10^4}{10}(1 - e^{-(3\times10^{-8})k})^{10} \cdot e^{-(3\times10^{-8})k\cdot(10^4-10)} \cdot \frac{1}{10^{100}}} \\ &= \frac{(1 - e^{-(3\times10^{-8})k})^{10} \cdot e^{-(3\times10^{-8})k\cdot(10^4-10)}}{\sum_{k=1}^{10100}(1 - e^{-(3\times10^{-8})k})^{10} \cdot e^{-(3\times10^{-8})k\cdot(10^4-10)}} \end{aligned}$$

Since the prior was equally likely for all years, it cancels between the numerator and denominator.

The Law of Total Probability in the denominator has one term for each possible year, and we are given that any of 10^{100} years as possible, so the sum must loop over all of those years. In practice, this would be treated as some constant, where rather than calculating it's exact value out by hand, we would just computationally normalize the distribution to sum to 1.

Using the exact value or the Poisson approximation for this problem were both accepted answers.

d. (6 points) Two base pairs start with the same letter: "A". What is the probability that the two base pairs end up with the same letter after 10^6 years? Assume when a base pair mutates it is equally likely to change to any of the other three letters. *Note: the course staff won't take clarification questions on this problem. If anything about the problem is unclear to you, write down your assumption and solve from there.*

This is a super challenging problem. It's ok if you didn't get it!

Let $M_1, M_2 \sim \text{Poi}(0.03)$ represent the number of mutations on each base pair after 10^6 years. We can write the final probability generally as:

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} P(M_1 = i)P(M_2 = j)P(\text{bases are same}|M_1 = i, M_2 = j)$$

We can try to solve this by breaking down the summation into a number of cases. Here's some common cases that received partial credit:

- $M_1 = 0, M_2 = 0$. This is the probability that neither mutated, i.e. they're both still A. $P(\text{bases are same}|M_1 = 0, M_2 = 0) = 1$, so we're left with:

  $$P(M_1 = 0)P(M_2 = 0) = (1 - p_a)^2 = e^{-0.06},$$

  where p_a is the answer to part a, i.e. the probability that at least one mutation occurred on a single base pair.
- $M_1 = 1, M_2 = 1$. This is the probability that both bases mutated once, i.e. they're both now one of [C, T, G].

  $$P(M_1 = 1)P(M_2 = 1)P(\text{bases are same}|M_1 = 1, M_2 = 1) = (0.03e^{-0.03})(0.03e^{-0.03}) \cdot \frac{1}{3}$$

  Note that since each of C, T, and G is equally likely for each base pair, we can think of $P(\text{bases are same}|M_1 = 1, M_2 = 1)$ as the probability that two randomly generated base pairs (picking uniformly from [C, T, G]) are the same.

Any other case requires a more complex computation for $P(\text{bases are same}|M_1 = i, M_2 = j)$, since the probability of eventually ending up on any given letter is unequal! For instance, after just 2 mutations, the probability that a base pair returns to A is $\frac{1}{3}$, while the probability that it's C, T, or G is $\frac{2}{9}$ each.

Let $B_1, B_2 \in [A, C, T, G]$ be random variables that represent the final letter value of each base pair after 10^6 years. We can rewrite $P(\text{bases are same}|M_1 = i, M_2 = j)$ as:

$$P(B_1 = B_2|M_1 = i, M_2 = j) = \sum_{b \in [A,C,T,G]} P(B_1 = b|M_1 = i)P(B_2 = b|M_2 = j)$$

We can divide these probabilities of the final letter value given the number of mutations further into two cases: where the final letter is A, and where the final letter is one of [C, T, G]. Because they're equally likely to be mutated into, $P(B = \text{'C'}|M = i) = P(B = \text{'G'}|M = i) = P(B = \text{'G'}|M = i)$, and we have:

$$P(B = b|M = i) = \begin{cases} P(B = \text{'A'}|M = i) & \text{if } b = \text{'A'} \\ \frac{1 - P(B = \text{'A'}|M = i)}{3} & \text{else} \end{cases}$$

Finally, to compute $P(B = \text{'A'}|M = i)$, note there's an inductive relationship: in $\frac{1}{3}$ of cases where the previous mutation was from [C, T, G], the next mutation will be A. So $P(B = \text{'A'}|M = i) = \frac{1}{3}(1 - P(B = \text{'A'}|M = i - 1))$. This does have a closed form using $p_0 = P(B = \text{'A'}|M = 0)) = 1$:

$$P(B = \text{'A'}|M = i) = (-\frac{1}{3})^i p_0 + \sum_{k=1}^{i} \frac{(-1)^{k-1}}{3^k}$$

That's all folks! We hope you had fun. Here are some optional notes for further curiosity.

i. The way Bingo cards are constructed, players are much more likely to win with a horizontal row than a vertical row. A recent paper proved this, and used arguments that started with the math in problem 3.

ii. Chris has never been to the Stanford observatory, but would like to go one day.

iii. Computers traditionally use pseudo random numbers. There are some incredibly sensitive algorithms where that is not good enough. Services like random.org serve APIs where you can get samples from Atmospheric Noise if you need a more authentic source of randomness.

iv. Since Mitochondrial DNA is passed from mother to offspring without undergoing recombination, Mitochondrial DNA serves as a useful (and accurate) molecular clock. By counting the differences between two strands of mitochondrial DNA, scientists can estimate how many years have passed since the two strands shared a common ancestor.