1 T/F

(4 points each) Circle T for True or F for False. We will only grade the answers, and are unlikely to even look at any justifications or explanations.

(a) T F Given some sample space $\Omega = \{1,2,3\}$, and events $A = \{1,2\}$ and $B = \{1\}$, then $Pr[B|A] = \frac{1}{2}$.

False. We don't know the probability distribution.

(b) T F Given some sample space $\Omega = \{1,2,3\}$, and events $A = \{1,2\}$ and $B = \{1\}$, then $Pr[B|A] = \frac{1}{3}$.

False. We don't know the probability distribution.

(c) T F Linearity of expectation applies if and only if the random variables involved are independent.

False. the "only if" direction is wrong, even when our random variables are dependent, linearity of expectation still applies

(d) T F The variance of a random variable that only attains values in the interval $[-1,1]$ is at most 1.

True. Consider the random variable that takes value 1 w.p. $1/2$ and -1 w.p. $1/2$, which is the "extreme case"

(e) T F If two events are disjoint, they are independent.

False. One of them could have 0 probability.

(f) T F If two events are independent, they are disjoint.

False. For example consider two independent coin tosses, there is no reason why they cannot be both heads.

(g) T F The value v which maximizes the probability density function (pdf) of a continuous random variable X is equal to $E[X]$.

False. consider $X \sim exp(\lambda)$, then the pdf is maximized at 0 but the mean is $1/\lambda$.

(h) T F A Markov Chain always has a stationary distribution.

True. If the MC is irreducible, this is easy to see. If it is not irreducible, we can break it up into smaller irreducible components such that of these components has no edges going out of it. Now this component, being irreducible, has a stationary distribution, so we can just use that distribution (and put zero probability for states outside this component) to get a stationary distribution for the entire chain.

2 To Catch a Magikarp

If your solutions involve computing integrals, **you do not have to do the calculations**.

Supose that when you catch a pokemon, it has probability p of being a Magikarp. (A Magikarp is a type of pokemon.) You may assume that different people catch pokemon independently.

(a) (3 pts) What is the expected number of pokemon you have to catch before you catch a Magikarp?

Let M be the random variable denoting number of pokemon that you catch. Then $M \sim Geom(p)$, so $\mathbb{E}(M) = 1/p$.

(b) (7 pts) Suppose you and your friend are out catching pokemon. Each of you catches one pokemon per minute. Both of you continue to catch pokemon until both of you have found Magikarps (so the person who catches a Magikarp first will continue to catch pokemon until the other person has also found a Magikarp). What is the expected time it takes for you to stop? Express your answer in closed form (i.e. not an infinite sum). You may use the fact that $\sum_{i=0}^{\infty} c^k = 1/(1-c)$ for $0 < c < 1$.

Let M be the random variable denoting number of pokemon that you catch, and N be the counterpart for your friend. M and N are thus independent $Geom(p)$, and we are interested in $\mathbb{E}[max(M,N)]$.

Recall the tail-sum formula for the expectation:

$$\mathbb{E}[max(M,N)] = \sum_{i=0}^{\infty} Pr[max(M,N) > i] = \sum_{i=0}^{\infty} 1 - Pr[max(M,N) \leq i]$$

. Notice that $Pr[max(M,N) \leq i]$ is the same as $Pr[M \leq i] \cdot Pr[N \leq i] = (1-(1-p)^i)^2$. Thus,

$$\begin{aligned}
\mathbb{E}[max(M,N)] &= \sum_{i=0}^{\infty}(1-(1-(1-p)^i)^2) \\
&= \sum_{i=0}^{\infty}(1-(1+(1-p)^{2i}-2(1-p)^i)) \\
&= \sum_{i=0}^{\infty} 2(1-p)^i-(1-p)^{2i} \\
&= \sum_{i=0}^{\infty} 2(1-p)^i-\sum_{i=0}^{\infty}(1-p)^{2i} \\
&= \frac{2}{p}-\frac{1}{1-(1-p)^2} \\
&= \frac{2p-3}{p^2-2p} \quad .
\end{aligned}$$

(c) (3 pts) Suppose that you catch a pokemon with probability q every minute (you never catch more than one pokemon in a minute). What is the distribution (including parameters) of the time before you catch your first Pokemon?

$Geom(q)$. Imagine you toss a coin (with probability of success q) every minute that determines whether you catch a pokemon that minute. The first minute you catch a Pokemon is the first time your coin comes up heads.

(d) (3 pts) Suppose again that you catch a pokemon with probability q every minute (you never catch more than one pokemon in a minute). What is the distribution (including parameters) of the time before you catch your first Magikarp?

$Geom(pq)$. To identify the distribution, again think of "catching a Magikarp" as the same as "getting heads", except the probability of getting a heads is pq now, since each pokemon has probability q of being a Magikarp.

(e) (3 pts) Suppose that people catch pokemon until they catch a Magikarp, at which point they stop. What is the expected number of pokemon you and your 100,000 friends have to catch (in total) before you all catch Magikarp?

Let M be the r.v. $= X_1 + X_2 + \cdots + X_{100,001}$, where each $X_i \sim Geom(p)$. We have $\mathbb{E}(M) = \frac{100,001}{p}$ by linearity of expectation.

Solutions assuming 100,000 people in total instead of 100,001, i.e. $\frac{100,000}{p}$, also receive full credit.

(f) (7 pts) Suppose again that people catch pokemon until they catch a Magikarp, at which point they stop. What is the probability that the total number of pokemon caught by you and your 100,000 friends is greater than x? Compute the best bound you can, assuming the Central Limit Theorem applies.

Clarification note: this is actually an approximation, not a bound. The term "bound" in the question was erroneous.

Let S_n denote sums of n independent and identical $Geom(p)$ and a standard linearity calculation yields $\mathbb{E}[S_n] = \frac{n}{p}$ and $Var[S_n] = \frac{n(1-p)^2}{p^2}$.

Then, using the central limit theorem, when n is considerably large, S_n converges towards

$$S_n \to \mathcal{N}\left(\frac{n}{p}, \frac{n(1-p)^2}{p^2}\right) \ .$$

(In order to see how you get using the CLT form seen in the lecture, notice that

$$\frac{S_n - \frac{n}{p}}{\sqrt{\frac{n(1-p)^2}{p^2}}} \to \mathcal{N}(0,1) \ .$$

From here, just apply shifting and scaling.)

Therefore,

$$P[S_n \geq x] = 1 - \Phi\left(\frac{x - \frac{n}{p}}{\sqrt{\frac{n(1-p)^2}{p^2}}}\right)$$

where Φ is the cumulative distribution function of standard normal distribution.

Finally, we plug in $n = 100,001$ to get

$$1 - \Phi\left(\frac{x - \frac{100,001}{p}}{\sqrt{\frac{100,001(1-p)^2}{p^2}}}\right) \ .$$

3 Good Proof, Bad Proof

For each of the following propositions and proofs, indicate which of the following cases apply:

1. Correct proposition with correct proof. No further explanation is needed for this case.
2. Correct proposition but incorrect proof. In this case, identify what the error in the proof is and provide a correct proof.
3. Incorrect proposition (therefore the proof is clearly incorrect). In this case, identify what the error in the proof is and provide a counterexample to the proposition.

(a) (10 points) **Let X be a random variable with expectation 1 and variance 1. Then $Pr[X \geq 7] \leq \frac{1}{36}$**

Proof. $Pr[X \geq 7] = Pr[X-1 \geq 6] = Pr[|X-1| \geq 6] = Pr[|X-E(X)| \geq 6] \leq \frac{Var(X)}{6^2} = \frac{1}{36}$ □

Correct proposition, incorrect proof: $Pr[X-1 \geq 6] \neq Pr[|X-1| \geq 6]$. To fix the proof, just replace that equals sign with a $\leq$: $Pr[X-1 \geq 6] \leq Pr[|X-1| \geq 6]$

(b) (10 points) **Suppose i is a state in a finite, irreducible, aperiodic Markov chain with transition matrix P and states $\{0,...,n\}$. Suppose that at timestep 1, the Markov chain is distributed according to its stationary distribution π. Then in the next timestep, 2, the probability that we leave i (i.e. that we started in i at timestep 1, and go to something that's not i at timestep 2), is the same as the probability that we enter i (i.e. we started outside i at timestep 1 and are at i at timestep 2).**

Proof. The probability that we start at i is just π_i. The probability that we leave i if we started at i is $\sum_{j \neq i} P_{i,j}$. Therefore, the probability that we leave i is $\pi_i \sum_{j \neq i} P_{i,j}$.

The probability that we start at some state j (that's not i) is π_j, and the probability that we go from j to i during the timestep is $P_{j,i}$. Therefore, the probability that we enter i is $\sum_{j \neq i} \pi_j P_{j,i}$.

Since π is a stationary distribution, $\pi_i = \sum_{j=0}^{n} \pi_j P_{j,i}$. We also know that, by definition of a Markov chain, $\sum_{j=0}^{n} P_{i,j} = 1$. Therefore, $\pi_i = \pi_i * 1 = \pi_i \sum_{j=0}^{n} P_{i,j}$.

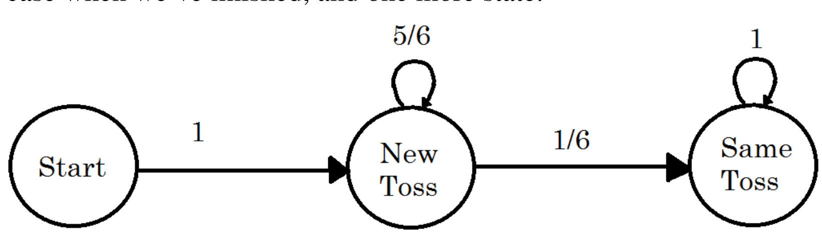
Therefore, $\pi_i \sum_{j=0}^{n} P_{i,j} = \sum_{j=0}^{n} \pi_j P_{j,i}$. Subtracting $\pi_i P_{i,i}$ from both sides gives us $\sum_{j \neq i} \pi_j P_{j,i} = \sum_{j \neq i} \pi_i P_{i,j}$, so the probability that we leave i and the probability that we enter i are the same. □

Correct statement, correct proof. The intuition is that if I have a huge amount of people distributed among the states of a Markov chain according to a stationary distribution, and all the people follow the timsteps, I expect the number of people in a particular state to remain constant. In order for that to hold, the number of people I expect to leave a state and the number of people I expect to enter the state must be the same at each timestep.

4 Markov Chains

Suppose you play the following game: You toss a fair six-sided die repeatedly until the same number comes up twice in a row, whereupon you stop.

(a) (4 points) Draw a Markov chain corresponding to this process with three states: a single state corresponding to the start of the process (before we've tossed any dice), a single state corresponding to the case when we've finished, and one more state.



(b) (4 points) Is this Markov chain aperiodic? Why or why not?

Periodic. Notice that the initial state is never returned to once you touch it. Therefore, there exists some positive integer greater than 1 (actually, this holds for all positive integers) Δ such that for all s not a multiple of Δ, $P_{1,1}^{s} = 0$ (where state 1 is the initial state) so the state is periodic. Therefore, the Markov chain is periodic.

(c) (4 points) Is this Markov chain irreducible? Why or why not?

The chain is not irreducible; we cannot get back to “New Toss” from “Same Toss”.

(d) (4 points) Does this Markov chain have a stationary distribution? If it does, tell us what it is. If no stationary exists, why not?

Yes, it does: $\begin{bmatrix}0 & 0 & 1\end{bmatrix}$. If you're in the "Same Toss" state, you'll never get out of that state!

(e) (4 points) Write down the transition matrix for this Markov chain.

Assuming right-stochastic matrix: $\begin{bmatrix}0 & 1 & 0\\ 0 & \frac{5}{6} & \frac{1}{6}\\ 0 & 0 & 1\end{bmatrix}$

(f) (5 points) How many tosses should we expect to do before we stop?

Let's number the states: let state 0 be the beginning state, state 1 = after we toss one coin (the middle state), and state 2 be the state where we have the same number in two consecutive tosses.

Let $\beta(i)$ represent the time it takes to get to state 2 from state i. Then we can set up the following equation:

$$\beta(1) = 1 + \frac{5}{6}\beta(1)) + \frac{1}{6}(0) \ .$$

Solving, we get $\beta(1) = 6$. Therefore, $\beta(0)$, the expected time to hit state 2 from state 0, is $1 + \beta(1) = 7$.

5 Money bags

I have a bag containing either a \$1 or a \$5 bill (with equal probability assigned to both possibilities).

(a) (4 points) How much money would you be willing to pay for this bag? In other words, what amount of money and this bag would you be indifferent between? This amount should make your expected profit zero, and your expected loss zero.

Suppose we pay a dollars for the bag. Let X be the net profit we obtain. Then we have $E[X] = \frac{1}{2}(-a+1) + \frac{1}{2}(-1+5) = 0$. Solve the equation to get $a = 3$.

(b) (4 points) I add a \$1 bill to the bag, so it now contains two bills. The bag is shaken. I draw out a random bill out of the bag, and it is a \$1 bill. How much money are you willing to pay now for this bag?

Define the following events:

- A: the bag contains \$1 after drawing.
- B: the bag contains \$5 after drawing.
- C: I draw \$1 bill from the bag.
- D: the bag originally contained \$1
- E: the bag originally contained \$5

Then:

$$\begin{aligned}
Pr[A \mid C] &= \frac{Pr[A \cap C]}{Pr[C]} \\
&= \frac{Pr[D]}{Pr[C \cap D] + Pr[C \cap E]} \\
&= \frac{Pr[D]}{Pr[C \mid D] \times Pr[D] + Pr[C \mid E] \times Pr[E]} \\
&= \frac{\frac{1}{2}}{1 \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2}} \\
&= \frac{2}{3}\ .
\end{aligned}$$

Therefore, $Pr[B \mid C] = 1 - Pr[A \mid C] = \frac{1}{3}$.

Again, suppose we pay a dollars for the bag. Then we have: $E[X] = (-a+1)\frac{2}{3} + (-a+5)\frac{1}{3}$. Setting this to get the value of a at which we expect to break even, we get $a = \frac{7}{3}$

(c) (5 points) Consider the original bag (before part (b)). Your friend, who lies with probability 0.7, takes a look inside the bag, and tells you that it has the \$5 bill. How much money are you willing to pay now?

Define the following events:

- A: the bag contains \$1.
- B: the bag contains \$5.
- C: your friend says it has \$5.
- D: your friend lies.

Then:

$$\begin{aligned}
Pr[A \mid C] &= \frac{Pr[A \cap C]}{Pr[C]} \\
&= \frac{Pr[C \mid A] \times Pr[A]}{Pr[C \cap A] + Pr[C \cap B]} \\
&= \frac{Pr[D] \times Pr[A]}{Pr[D] \times Pr[A] + (1 - Pr[D]) \times Pr[B]} \\
&= \frac{\frac{1}{2} \times \frac{7}{10}}{\frac{1}{2} \times 0.7 + \frac{1}{2} \times 0.3} \\
&= \frac{7}{10}\ .
\end{aligned}$$

Therefore, $Pr[B \mid C] = 1 - Pr[A \mid C] = \frac{3}{10}$. Again supposing that we pay \$$a$ for the bag, we have: $E[X] = (-a+1)\frac{7}{10} + (-a+5)\frac{3}{10}$. Setting this to zero to get the value of a at which we expect to break even, we get $a = \frac{22}{10} = \frac{11}{5}$.

(d) (7 points) A company is selling these bags (same as the original one - each bag contains \$1 or \$5 with equal probability, and the amount of money in each bag is mutually independent of the amount in the others) for \$2. You decide to buy $1,000,000$ of these bags, hoping to make a profit. Find an upper bound on the probability that you lose money. Find a linear bound for 2 points, a quadratic bound for 4 points, or an exponential bound for full credit.

As the "exponential" part in the problem hints, we will bound our probability using Chernoff bounds.

There are two possible ways to use Chernoff here: the first (and simpler) one is to consider each bag as an indicator random variable X_i that is 0 when there is a 1-dollar bill in the bag (so we lose a dollar) and 1 when the there is a 5-dollar bill in the bag (so we gain four dollars). The second approach is to bound the total profit directly; this will only differ from the first one by a constant factor so we will only present the former.

Full Credit: Chernoff Bounds

Consider the indicator random variables X_i defined as in the previous paragraph. then in order to get 2×10^6 dollars in the end, we need at least $\frac{2\times 10^6}{5} = 4\times 10^5$ of them to take up value 1. Now the expectation of $\sum_{i=1}^{10^6} X_i$ is $\mathbb{E}[\sum_i^{10^6} X_i] = \sum_{i=1}^{10^6}\mathbb{E}[X_i] = \frac{1}{2}\times 10^6 = 5\times 10^5$. Then if we use the following form of Chernoff as given in the formula sheet: $P[X \le (1-\delta)\mu] \le e^{\frac{-\mu\delta^2}{2}}$, we can solve for δ by $(1-\delta)5\times 10^5 = 4\times 10^5 \implies 1-\delta = \frac{4}{5} \implies \delta = \frac{1}{5}$. Now plug in δ, μ on the RHS gives us an expoential bound: $e^{\frac{-5\times 10^5(\frac{1}{5})^2}{2}} = e^{-10^4}$

Partial Credit: Chebyshev's Inequality For the partial credit, it's easier to use random variables corresponding to the amount of money in the bag, i.e. one that takes value 1 w.p. $\frac{1}{2}$ and 5 otherwises. Then if we denote the sum as S_{10^6}, then $E[S_{10^6}] = 3\times 10^6$, $Var[S_{10^6}] = 4\times 10^6$ and probability we are interested in bounding is $P[S_{10^6} \le 2\times 10^6]$ and the probability is $\le P[|S_{10^6} - 3\times 10^6| \le 1\times 10^6] \le \frac{4\times 10^6}{(1\times 10^6)^2} = \frac{4}{10^6} = \frac{1}{250000}$.

Partial Credit: Markov's Inequality For Markov, if we follow the notation in last item, since we are interested in $P[S_{10^6} \le 2\times 10^6]$, we need to consider a new random variable $5 - X_i$ for each X_i. Obviously, $5 - X_i$ is non-negative, so their sum is thus non-negative and $P[S_{10^6} \le 2\times 10^6] = P[5\times 10^6 - S_{10^6} \ge 3\times 10^6] \le \frac{5\times 10^6 - 3\times 10^6}{3\times 10^6} = \frac{2}{3}$.

6 Squared Exponential

Suppose x is distributed exponentially with parameter λ.

(a) (5 points) What is the probability that x is in the interval $[t, t+\varepsilon]$ for infinitesimally small ε? Express your answer in terms of t and ε.

Recall from note 16 that for very small ε,we have

$$P[t \le x \le t+\varepsilon] = \int_t^{t+\varepsilon} f_X(x)dx \approx f_X(t)\varepsilon = \varepsilon\lambda e^{-\lambda t} \quad .$$

Another way to approach this problem is to use the CDF of the exponential distribution:

$$\begin{aligned} P[t \le x \le t+\varepsilon] &= F_x(t+\varepsilon) - F_x(t) \\ &= 1 - e^{-\lambda(t+\varepsilon)} - (1 - e^{-\lambda t}) \\ &= e^{-\lambda t} - e^{-\lambda(t+\varepsilon)} \\ &= e^{-\lambda t}(1 - e^{-\lambda\varepsilon}) \quad . \end{aligned}$$

Now recall the Taylor series expansion for e^x:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \quad .$$

For x with infinitesimally small absolute value (such as $x = -\lambda\varepsilon$), we can use a first-order approximation (as in the hint for the next part) to approximate this as $1+x$. Therefore,

$$P[t \le x \le t+\varepsilon] = e^{-\lambda t}(1-(1-\lambda\varepsilon)) = \varepsilon\lambda e^{-\lambda t} \quad ,$$

the same as the answer we got using the pdf-based approach.

(b) (8 points) If $x^2 \in [q, q+\delta]$ for infinitesimally small δ, what interval must x fall into? Express your answer in terms of q and δ. *Hint: If you are dealing with an infinitesimally small number s and encounter a polynomial in s, you can drop all but the constant and first-order terms. In other words, if s is infinitesimally small, you can approximate* $a_0 + a_1 s + a_2 s^2 + a_3 s^3 + \dots$ *as* $a_0 + a_1 s$.

Denote the interval that x has to into as $[p', p'+\varepsilon';]$ and note that we can do this since $x \sim exp(\lambda)$ guarantees x is non-negative. Now in order for the $x^2 \in [q, q+\delta]$ to hold, we have $(p')^2 = q$ and $(p'+\varepsilon')^2 = q+\delta$ and we solve for $p' = \sqrt{q}$.

Additionally, since δ is small, ε' would also be small and if we expand the LHS of the second equation, we get $(p'+\varepsilon')^2 = (p')^2 + (\varepsilon')^2 + 2p'\varepsilon'$. Now according to the hint, we can approximate the polynomial by dropping the $(\varepsilon')^2$ term and that gives us $(p')^2 + 2p'\varepsilon' = q + 2\sqrt{q}\varepsilon' = q+\delta$. Therefore we should set ε' to $\frac{\delta}{2\sqrt{q}}$.

In summary, the interval x must fall into is $[\sqrt{q}, \sqrt{q} + \frac{\delta}{2\sqrt{q}}]$

Alternatively, we would award full credits to $[\sqrt{q}, \sqrt{q+\delta}]$ or a Taylor expansion of that.

(c) (7 points) What is the pdf of the distribution of x^2?

For the sake of clarity, let's denote $y = x^2$ and we want to find $f_Y(y)$. Now since δ is small, we know from part (a) that analogously, $P[y \in [q, q+\delta]] \approx f_Y(q)\delta$.

Now we know from part(b) that $P(y \in [q, q+\delta]) = P(x \in [\sqrt{q}, \sqrt{q} + \frac{\delta}{2\sqrt{q}}])$ and using prior results from part (a), the latter is $\frac{\delta}{2\sqrt{q}}\lambda e^{-\lambda\sqrt{q}}$.

Combining the previous two results, we have $P(y \in [q, q+\delta]) = f_Y(q)\delta = \frac{\delta}{2\sqrt{q}}\lambda e^{-\lambda\sqrt{q}}$. Now if we cancel the δ on both sides and since q is arbitrary, we get $f_Y(y) = \frac{\lambda}{2\sqrt{y}} e^{-\lambda\sqrt{y}}$ and this is the pdf of $y = x^2$.

An alternate approach (which is helpful if you did not use the first order approximation in the last part) is to use the CDF: $P(y \le k) = P(x^2 \le k) = P(x \le \sqrt{k}) = 1 - e^{-\lambda\sqrt{k}}$. Now talking the derivative of our cdf gives us $\frac{\lambda}{2\sqrt{k}} e^{-\lambda\sqrt{k}}$ But k is just a dummy variable so we obtained the same pdf expression.

7 Bounds

(a) (7 pts) Show by example that Markov's inequality is tight; that is, show that given $k > 0$, there exists a discrete nonnegative random variable X such that $\Pr[X \ge k] = \mathbf{E}[X]/k$.Bernoulli with 1 w.p. $1/k$ and 0 otherwise. Common alternate solutions: X takes the value 0 w.p. 1, or X takes the value k w.p. 1.

(b) (7 pts) Show by example that Chebyshev's inequality is tight; that is, show that given $k > 0$, here exists a random variable X such that $\Pr[|X - \mathbf{E}[X]| \geq k\sigma(X)] = 1/k^2$.$X$ takes value k w.p. $\frac{1}{2k^2}$, $-k$ w.p. $\frac{1}{2k^2}$, and 0 w.p. $1 - \frac{1}{k^2}$. Then, $E[X] = k\frac{1}{2k^2} - k\frac{1}{2k^2} = 0$, $E[X^2] = 2k^2\frac{1}{2k^2} = 1$, $Var[X] = 1$ and $\sigma(X) = 1$.

$$\Pr[|X - \mathbf{E}[X]| \geq k\sigma(X)] = \Pr[|X| \geq k] = \frac{1}{2k^2} + \frac{1}{2k^2} = \frac{1}{k^2}$$

.

(c) (7 pts) Show that there is no random variable X, that takes values in some finite set $\{v_1, \ldots, v_N\}$, such that for all $k > 0$, Markov's inequality is tight; that is, $\Pr[X \geq k] = \mathbf{E}[X]/k$. For k big enough (more precisely larger than $max_i(v_i)$), the bound is positive, but the actual probability is zero.

8 Boolean dot product

Let $C = (A_1 \wedge B_1) \vee (A_2 \wedge B_2) \vee \cdots \vee (A_N \wedge B_N)$. A_i's and B_i's are i.i.d Bernoulli r.v. with parameter $\frac{1}{2}$. **Express your answers as an algebraic expression of numbers and N.**

(a) (13 points) What is the probability that A_1 is *true* provided that C is *true*?

We use Bayes' Theorem.

We know the prior probability (that A_1 is true) is $1/2$ since it is a Bernoulli random variable with parameter $1/2$.

What is the probability that C is true? The probability that any clause is true is $1/4$. Therefore, the probability that C is false is the probability that all N clauses are false, which is $(3/4)^N$. So the probability that C is true is $1 - (3/4)^N$.

What is the probability that C is true given that A_1 is true? We split into cases. Either B_1 is true (in which case C is guaranteed to be true) or B_1 is false, in which case we know that $(A_2 \wedge B_2) \vee \cdots \vee (A_N \wedge B_N)$ has to be true (since the first term is false). Therefore, the probability that C is true given thaht A_1 is true is

$$\Pr[C = 1|A = 1] = \frac{1}{2}\left(1 + 1 - (\frac{3}{4})^{N-1}\right) = 1 - \frac{1}{2}(\frac{3}{4})^{N-1}$$

.

Therefore, applying Bayes' theorem:

$$\begin{aligned}\Pr[A_1 = 1|C = 1] &= \frac{\Pr[A_1 = 1]\Pr[C = 1|A_1 = 1]}{\Pr[C = 1]} \\ &= \frac{\left(\frac{1}{2}\right)\left(1 - \frac{1}{2}(\frac{3}{4})^{N-1}\right)}{1 - \left(\frac{3}{4}\right)^N}\end{aligned}$$

Alternative Solution 1: Compute $\Pr(C = 1) = 1 - (3/4)^N$ as in the official solutions.

In order to calculate $\Pr(A_1 = 1|C = 1)$, observe that it is easier to calculate the complement of the probability:

$$\begin{aligned}
\Pr(A_1 = 1|C = 1) &= 1 - \Pr(A_1 = 0|C = 1) \\
&= 1 - \frac{\Pr(A_1 = 0, C = 1)}{\Pr(C = 1)} \\
&= 1 - \frac{\Pr(A_1 = 0)\Pr(C = 1|A_1 = 0)}{\Pr(C = 1)}
\end{aligned}$$

If $A_1 = 0$, then the first clause is immediately false; therefore, the conditional probability $\Pr(C = 1|A_1 = 0)$ is the probability that one of the other $N-1$ clauses is true (incidentally, this why it was easier to consider the complement above.). Using the same method used to calculate $\Pr(C = 1)$, one calculates that $\Pr(C = 1|A_1 = 0) = 1-(3/4)^{N-1}$. Since $\Pr(A_1 = 1) = 1/2$, the solution is:

$$\Pr(A_1 = 1|C = 1) = 1 - \frac{1}{2}\frac{1-(3/4)^{N-1}}{1-(3/4)^N}$$

Alternative Solution 2: Since the sample space is uniform (because each unique assigment of A_i and B_i has equal probability of occuring), we can count the number of samples in each event directly. We get

$$|C| = 4^N - 3^N$$

There are 4 different ways to assign each clauses, and 3 out of 4 would make the clause evaluate to false. So there are 4^N different ways to assign all the literals, and 3^N of them result in C being false. So the size of the event "C is true" is $4^N - 3^N$. Now we calculate the size of the event "C is true and A_1 is true".

$$|C \cap A_1 = true| = |C \cap A_1 = true \cap B_1 = true| + |C \cap A_1 = true \cap B_1 = false| = 4^{N-1} + (4^{N-1} - 3^{N-1})$$

We break this event into two parts. When A_1 and B_1 are both true, the first clause evaluates to true. In this case, C is true regardless of the assigment of the other $N-1$ clauses. Therefore all the 4^{N-1} assigments to the rest of the clauses are contained in the event. When B_1 is false, then the first clause evaluates to false. To make C true, there need to be true clauses in the other $N-1$ clauses. Similar to how we calculate $|C|$, there are in total 4^{N-1} ways to assign the $N-1$ clauses, and 3^{N-1} assignments make C false, and $4^{N-1} - 3^{N-1}$ assigments make C evaluate to true. Sum the two parts together to get $|C \cap A_1 = true|$. Eventually we get,

$$Pr[A_1 = true|C] = \frac{|C \cap A_1 = true|}{|C|} = \frac{4^{N-1} + (4^{N-1} - 3^{N-1})}{4^N - 3^N}$$

Notice that all three solutions give the same answer.

(b) (13 points) Suppose we pick a variable uniformly at random from the set $\{A_1, \ldots A_N, B_1, \ldots, B_N\}$ and set it to *false*. The remaining variables are i.i.d. Bernoulli with parameter $\frac{1}{2}$. What is the probability that A_1 is *true*, provided that C is true? First, we notice that we pick a variable uniformly at random to set to false. There are three possibilites:

First, with probability $\frac{1}{2N}$, we pick A_1 to be the variable set to false. In this case, it is obvious that the probability that A_1 is true (whether or not we condition on C being true) is zero.

Second, with probability $\frac{1}{2N}$, we pick B_1 to be the variable set to false. Notice that in this case, our choice for the value of A_1 doesn't affect the value of C, since the clause $(A_1 \wedge B_1)$ is going to be false anyway no matter what we set A_1 to. As a result, the probability that A_1 is true, given that C is true, is simply $1/2$.

Third, with probability $1-\frac{1}{N}$, we pick something else to be false. What happens in this case? Some other clause consisting of the AND of two variables is turned into a guaranteed false, and now we are left with an expression involving $N-1$ clauses. This is essentially just the same expression we had originally, with one less clause, so we can just reuse our answer from part A in this case, except with N reduced by 1: $\frac{\left(\frac{1}{2}\right)\left(1-\frac{1}{2}\left(\frac{3}{4}\right)^{N-2}\right)}{1-\left(\frac{3}{4}\right)^{N-1}}$.

Now, summing up, we get that the desired probability is:

$$\begin{aligned}
\Pr[A_1 = 1 | C = 1] &= \frac{1}{2N}0 + \frac{1}{2N}\frac{1}{2} + \left(1-\frac{1}{N}\right)\left(\frac{\left(\frac{1}{2}\right)\left(1-\frac{1}{2}\left(\frac{3}{4}\right)^{N-2}\right)}{1-\left(\frac{3}{4}\right)^{N-1}}\right) \\
&= \frac{1}{4N} + \left(1-\frac{1}{N}\right)\left(\frac{\left(\frac{1}{2}\right)\left(1-\frac{1}{2}\left(\frac{3}{4}\right)^{N-2}\right)}{1-\left(\frac{3}{4}\right)^{N-1}}\right)
\end{aligned}$$

.

9 Extra Pages

If you use this page as extra space for answers to problems, please indicate clearly which problem(s) you are answering here, and indicate **in the original space for the problem** that you are continuing your work on an extra sheet. You can also use this page to give us feedback or suggestions, report cheating or other suspicious activity, or to draw doodles.

More extra paper. If you fill this sheet up you can request extra sheets from a proctor (just make sure to write your SID on each one, and to staple the extra sheets to your exam when you submit it).

Reference Sheet for Distributions and Bounds

Discrete Distributions

Bernoulli Distribution

- 1 with probability p, 0 with probability $1-p$
- Expectation: p
- Variance: $p(1-p)$

Binomial Distribution with parameters n, p

- $\Pr[X=k] = \binom{n}{k} p^k (1-p)^{n-k}$
- Expectation: np
- Variance: $np(1-p)$

Geometric Distribution with parameters p

- $\Pr[X = k] = (1-p)^{k-1}p$
- Expectation: $1/p$
- Variance: $\frac{1-p}{p^2}$

Uniform Distribution with parameters a, b ($a \leq b$)

- $\Pr[X = k] = \frac{1}{b-a+1}$ for $k \in [a, b]$, 0 otherwise.
- Expectation: $(a+b)/2$
- Variance: $\frac{(b-a+1)^2-1}{12}$

Poisson Distribution with parameter λ

- $\Pr[X = k] = \frac{\lambda^k e^{-\lambda}}{k!}$
- Expectation: λ
- Variance: λ

Continuous Distributions

Uniform Distribution with parameters a, b ($a < b$).

- PDF: $\frac{1}{b-a}$ for $x \in [a, b]$, 0 otherwise.
- Expectation: $(a+b)/2$
- Variance: $\frac{(b-a)^2}{12}$

Exponential Distribution with parameter λ

- PDF: $\lambda e^{-\lambda x}$ for $x > 0$, 0 otherwise
- Expectation: $1/\lambda$
- Variance: $1/\lambda^2$

Normal Distribution with parameters μ, σ^2

- PDF: $\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Expectation: μ
- Variance: σ^2

Chernoff Bounds

Theorem: Let $X_1, \dots, X_n$ be independent indicator random variables such that $Pr[X_i = 1] = p_i$, and $Pr[X_i = 0] = 1 - p_i$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X]$. Then the following Chernoff bounds hold:

- For any $\delta > 0$:
$$Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$$
- For any $1 > \delta > 0$:
$$Pr[X \leq (1-\delta)\mu] \leq \left(\frac{e^{\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$$
- For any $1 > \delta > 0$:
$$Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{\mu\delta^2}{3}}$$
- For any $1 > \delta > 0$:
$$Pr[X \leq (1-\delta)\mu] \leq e^{-\frac{\mu\delta^2}{2}}$$
- For $R > 6\mu$:
$$Pr[X \geq R] \leq 2^{-R}$$