CS 70 Discrete Mathematics and Probability Theory
Summer 2026 Tate

HW 05

Due: Saturday, 08/01, 4:00 PM
Grace period until Saturday, 08/01, 6:00 PM
Remember to show your work for all problems!

Sundry

Before you start writing your final homework submission, state briefly how you worked on it. Who else did you work with? List names and email addresses. (In case of homework party, you can just describe the group.)

1 Balls and Bins, All Day Every Day

Note 14
Note 15

Suppose n balls are thrown into n labeled bins one at a time, where n is a positive *even* integer.

(a) What is the probability that exactly k balls land in the first bin, where k is an integer $0 \leq k \leq n$?

(b) What is the probability p that at least half of the balls land in the first bin? (You may leave your answer as a summation.)

(c) Using the union bound, give a simple upper bound, in terms of p, on the probability that some bin contains at least half of the balls.

(d) What is the probability, in terms of p, that at least half of the balls land in the first bin, or at least half of the balls land in the second bin?

(e) After you throw the balls into the bins, you walk over to the bin which contains the first ball you threw, and you randomly pick a ball from this bin. What is the probability that you pick up the first ball you threw? (Again, leave your answer as a summation.)

2 Max/Min Dice Rolls

Note 16

Yining rolls three fair six-sided dice.

(a) Let X denote the maximum of the three values rolled. What is the distribution of X? That is, what is $\mathbb{P}[X = x]$ for each possible value of x? Leave your final answer in terms of x.

Hint: Try to first compute $\mathbb{P}[X \leq x]$ for each possible value of x. To check your answer, you can solve this problem using counting and ensure you get the same probabilities.

(b) Let Y denote the minimum of the three values rolled. What is the distribution of Y?

3 Swaps and Cycles

Note 16

A permutation of n objects is a bijection from $(1,\dots,n)$ to itself. For example, the permutation $\pi = (2,1,4,3)$ of 4 objects is the mapping $\pi(1) = 2$, $\pi(2) = 1$, $\pi(3) = 4$, and $\pi(4) = 3$. We'll say that a permutation $\pi = (\pi(1),\dots,\pi(n))$ contains a *swap* if there exist $i, j \in \{1,\dots,n\}$ so that $\pi(i) = j$ and $\pi(j) = i$, where $i \neq j$. The example above contains two swaps: $(1,2)$ and $(3,4)$.

(a) In terms of n, what is the expected number of swaps in a random permutation?

(b) In the same spirit as above, π contains a *k-cycle* if there exist $i_1,\dots,i_k \in \{1,\dots,n\}$ with $\pi(i_1) = i_2, \pi(i_2) = i_3, \dots, \pi(i_k) = i_1$. What is the expected number of k-cycles?

4 Double-Check Your Intuition Again

Note 17

(a) You roll a fair six-sided die and record the result X. You roll the die again and record the result Y.

(i) What is $\text{cov}(X+Y, X-Y)$?

(ii) Prove that $X+Y$ and $X-Y$ are not independent.

The problems below are not related to the scenario above. For each of them, if you think the answer is "yes" then provide a proof. If you think the answer is "no", then provide a counterexample.

(b) If X is a random variable and $\text{Var}(X) = 0$, then must X be a constant?

(c) If X is a random variable and c is a constant, then is $\text{Var}(cX) = c\,\text{Var}(X)$?

(d) If A and B are random variables with nonzero standard deviations and $\text{Corr}(A,B) = 0$, then are A and B independent?

(e) If X and Y are not necessarily independent random variables, but $\text{Corr}(X,Y) = 0$, and X and Y have nonzero standard deviations, then is $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$?

The two subparts below are **optional** and will not be graded but are recommended for practice.

(f) If X and Y are random variables then is $\mathbb{E}[\max(X,Y)\min(X,Y)] = \mathbb{E}[XY]$?

(g) If X and Y are independent random variables with nonzero standard deviations, then is

$$\text{Corr}(\max(X,Y), \min(X,Y)) = \text{Corr}(X,Y)?$$

5 Dice Games

Note 17

(a) Alice rolls a fair six-sided die until she gets a 1. Let X be the number of total rolls she makes (including the last one), and let Y be the number of rolls on which she gets an even number. Compute $\mathbb{E}[Y \mid X = x]$, and use it to calculate $\mathbb{E}[Y]$.

(*Hint*: after applying the law of total expectation, do not substitute the formula for $\mathbb{P}[X = x]$; instead, look for the sums $\sum_x x \cdot \mathbb{P}[X = x]$ and $\sum_x \mathbb{P}[X = x]$, whose values you already know.)

(b) Bob plays a game in which he starts off with one fair six-sided die. At each time step, he rolls all the dice he has. Then, for each die, if it comes up as an odd number, he puts that die back, and adds a number of fair six-sided dice equal to the number displayed to his collection. (For example, if he rolls a one on the first time step, he puts that die back along with an extra die.) However, if it comes up as an even number, he removes that die from his collection.

Compute the expected number of dice Bob will have after n time steps. (Hint: compute the value of $\mathbb{E}[X_k \mid X_{k-1} = m]$ to derive a recursive expression for X_k, where X_i is the random variable representing the number of dice after i time steps. As your base case, $E[X_0] = 1$.)

6 Geometric and Poisson

Note 18

Let $X \sim \text{Geometric}(p)$ and $Y \sim \text{Poisson}(\lambda)$ be independent random variables. Compute $\mathbb{P}[X > Y]$. Your final answer should not have summations.

Hint: Use the total probability rule. You may also find the Taylor series of e^x useful: $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.

7 Self-Consistency

Note 16
Note 17

A common way to make a language model more accurate is *self-consistency*: instead of trusting a single answer, you sample n independent answers to the same question and return the *majority* answer. Suppose each sampled answer is independently correct with probability p, and model the returned answer as correct if and only if a strict majority of the n samples are correct. Throughout, take n to be odd so there are no ties.

(a) Let C be the number of correct answers among the n samples. Name the distribution of C (with its parameters), and compute $\mathbb{E}[C]$ and $\text{Var}(C)$.

(b) Take $n = 3$. Compute the probability q that the majority answer is correct, as a function of p.

(c) Show that majority voting with 3 samples beats a single sample — that is, $q > p$ — if and only if $p > \frac{1}{2}$. For which values of p does majority voting have a *lower* probability of success than a single sample?

(d) Now let $n = 2m+1$ be a general odd number, so the majority answer is correct exactly when $C > \frac{n}{2}$, and suppose $p > \frac{1}{2}$. Let $\sigma(C) = \sqrt{\text{Var}(C)}$ denote the standard deviation of C. Show that the ratio

$$\frac{\mathbb{E}[C] - n/2}{\sigma(C)}$$

grows without bound as $n \to \infty$. Then explain, in one or two sentences, why this implies that the probability the majority answer is correct approaches 1. (You may use the fact that a random variable is very unlikely to be many standard deviations away from its mean.)