

Final

Problem 1. [8 points] Prove that for all $n \in \mathbb{N}$, the following identity holds

$$\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}.$$

Solution. By induction on $n \geq 1$, with induction hypothesis

$$P(n) : \sum_{i=0}^n i^2 = \frac{n(n+1)(2n+1)}{6}$$

for all $n \in \mathbb{N}$

Base case ($n = 1$):

$$\frac{1(1+1)(2+1)}{6} = \frac{6}{6} = 1 = 1^2$$

Inductive step: Assume $P(n)$, we need to show that $P(n+1)$ holds.

$$\begin{aligned} \sum_{i=0}^{n+1} i^2 &= \left(\sum_{i=0}^n i^2 \right) + (n+1)^2 \\ &= \frac{n(n+1)(2n+1)}{6} + (n+1)^2 \\ &= \frac{n(n+1)(2n+1) + 6(n+1)^2}{6} \\ &= \frac{(n+1)(2n^2 + n + 6n + 6)}{6} \\ &= \frac{(n+1)(2n^2 + 7n + 6)}{6} \\ &= \frac{(n+1)(n+2)(2n+3)}{6} \\ &= \frac{(n+1)((n+1)+1)(2(n+1)+1)}{6} \\ &\Rightarrow P(n+1) \end{aligned}$$

as required.

We have shown that $P(n) \Rightarrow P(n+1)$. Thus, $P(n)$ is true for all $n \in \mathbb{N}$.

Problem 2. [20 points] Coin-Flip is a 2 player game. Each player wins with probability exactly 0.5. There are no ties.

n people are playing a Coin-Flip tournament. Every person plays a Coin-Flip game with every other person exactly once. Thus everybody plays $n - 1$ games. The outcomes of all the games are mutually independent of one another.

We say that the tournament is a *success* if for every $i \in \{0, 1, \dots, n - 1\}$, there is exactly one player, which we will refer to as p_i , with exactly i wins.

(a) [10 points] Prove that if the tournament is a success, then for any integers j, k with $0 \leq k < j \leq n - 1$, p_j defeats p_k .

Solution. We prove it by induction on k . The inductive hypothesis $P(k)$ is that for all $0 \leq \ell \leq k$ and all $\ell < j < n$, p_j defeats p_ℓ .

The base case is $k = 0$. Now p_0 loses all $n - 1$ games. Thus, for every $j > 0$, p_j defeats p_0 . Suppose $P(k)$ and let us show $P(k + 1)$. Since the tournament is a success, p_{k+1} wins exactly $k + 1$ games. Because $P(k)$ holds, p_{k+1} defeats the $k + 1$ players p_0, p_1, \dots, p_k . Thus, for all $k + 1 < j < n$, p_j defeats p_{k+1} . So $P(k + 1)$ holds.

(b) [6 points] What is the probability that the tournament will be a success?

Solution. Let π be a permutation of $\{0, 1, \dots, n - 1\}$ and define the event \mathcal{E}_π to be that the tournament is a success with players p_i winning exactly i games, where p_i is the $\pi(i)$ th player. Then the events \mathcal{E}_π and \mathcal{E}_σ are disjoint for $\pi \neq \sigma$. Moreover, by symmetry, $\Pr[\mathcal{E}_\pi] = \Pr[\mathcal{E}_\sigma]$ for all permutations σ, π . Let \mathcal{E} be the event the tournament is a success. Since the events are disjoint,

$$\Pr[\mathcal{E}] = \sum_{\pi} \Pr[\mathcal{E}_\pi] = n! \Pr[\mathcal{E}_\sigma],$$

where σ is some fixed permutation. Now \mathcal{E}_σ determines all the outcomes of the games. As there are $\binom{n}{2}$ games, we have

$$\Pr[\mathcal{E}_\sigma] = \left(\frac{1}{2}\right)^{\binom{n}{2}}.$$

In total,

$$\Pr[\mathcal{E}] = n! \left(\frac{1}{2}\right)^{\binom{n}{2}}.$$

(c) [4 points] Show that your answer to part (b) is $o(1)$. **Solution.** We have,

$$\Pr[\mathcal{E}] \leq n^n \left(\frac{1}{2}\right)^{\binom{n}{2}} = 2^{n \log n} \left(\frac{1}{2}\right)^{\binom{n}{2}} = \left(\frac{1}{2}\right)^{\binom{n}{2} - n \log n}.$$

Now, $\binom{n}{2} - n \log n = \Omega(n^2)$, so there is a positive constant c such that for sufficiently large n , this probability is at most

$$\left(\frac{1}{2}\right)^{cn^2},$$

which for sufficiently large n , is clearly less than any positive constant, and thus is $o(1)$.

Problem 3. [8 points] A person is passing time by advancing a token on the set of natural numbers. In the beginning, a token is placed on 0.

The person keeps playing *moves* forever. Each move proceeds as follows:

1. First the person tosses a fair coin (with heads/tails equally likely).
2. Suppose the token is currently placed on n . If heads came up, then the person moves the token to $n + 3$, otherwise he moves the token to $n + 4$.

For each $n \in \mathbb{N}$, let E_n be the event "There was a move on which the token landed on n ". Let $p_n = \Pr[E_n]$.

Find a recurrence relation for p_n . *You do not need to solve the recurrence, but you should specify the boundary conditions that would be necessary to find a solution to the recurrence.*

Solution. For all $n \geq 4$,

$$p_n = \frac{1}{2}p_{n-3} + \frac{1}{2}p_{n-4},$$

with boundary conditions $p_0 = 1, p_1 = 0, p_2 = 0, p_3 = 1/2$.

Problem 4. [10 points] Exactly $1/5$ th of the people in a town have Beaver Fever[©].

There are two tests for Beaver Fever, TEST1 and TEST2. When a person goes to a doctor to test for Beaver Fever, with probability $2/3$ the doctor conducts TEST1 on him and with probability $1/3$ the doctor conducts TEST2 on him.

When TEST1 is done on a person, the outcome is as follows:

- If the person has the disease, the result is positive with probability $3/4$.
- If the person does not have the disease, the result is positive with probability $1/4$.

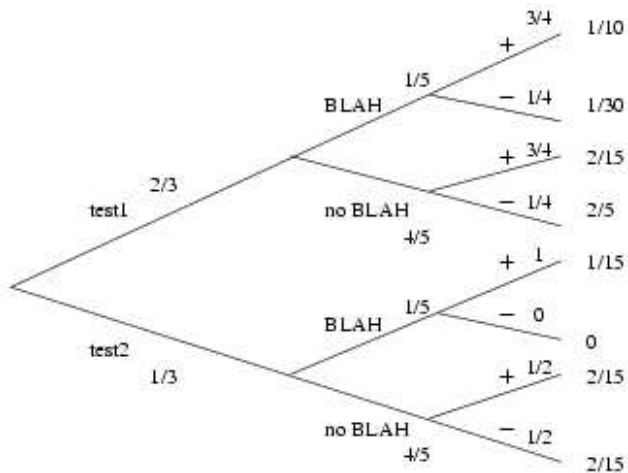
When TEST2 is done on a person, the outcome is as follows:

- If the person has the disease, the result is positive with probability 1.
- If the person does not have the disease, the result is positive with probability $1/2$.

A person is picked uniformly at random from the town and is sent to a doctor to test for Beaver Fever. The result comes out positive. What is the probability that the person has the disease?

Solution. Let B be the event that the person has BLAH. Let $T1$ be the event that the person is tested with test1. Let $T2$ be the event that the person is tested with test2. Let P be the event that the test comes out positive.

A tree diagram is worked out below with the given information:



The probability that a person has BLAH, given that the test comes out positive is:

$$\begin{aligned}
\Pr\{B \mid P\} &= \frac{\Pr(B \cap P)}{\Pr(P)} \\
&= \frac{\Pr(B \cap T1 \cap P) + \Pr(B \cap T2 \cap P)}{\Pr(T1 \cap P) + \Pr(T2 \cap P)} \\
&= \frac{\Pr(B \cap T1 \cap P) + \Pr(B \cap T2 \cap P)}{\Pr(B \cap T1 \cap P) + \Pr(\bar{B} \cap T1 \cap P) + \Pr(B \cap T2 \cap P) + \Pr(\bar{B} \cap T2 \cap P)} \\
&= \frac{(1/5)(2/3)(3/4) + (1/5)(1/3)(1)}{(1/5)(2/3)(3/4) + (4/5)(2/3)(1/4) + (1/5)(1/3)(1) + (4/5)(1/3)(1/2)} \\
&= \frac{5}{13}
\end{aligned}$$

Problem 5. [10 points] Two identical complete decks of cards, each with 52 cards, have been mixed together. A hand of 5 cards is picked uniformly at random from amongst all subsets of exactly 5 cards.

(a) [5 points] What is the probability that the hand has no identical cards (i.e., cards with the same suit and value. For example, the hand $\langle Q\heartsuit, 5\spadesuit, 6\spadesuit, 8\clubsuit, Q\heartsuit \rangle$ has identical cards.)? **Solution.** We can calculate this probability by computing

$$\frac{\text{hands with no identical cards}}{\text{total possible hands}}$$

There are 104 cards. There are 5 cards in a hand. Order does not matter. The total number of possible hands is:

$$\binom{104}{5}$$

There are 52 possible card faces, and we can choose 5 of them if no identical cards are allowed. Additionally, each card can be from either deck 1 or deck 2. Therefore the number of hands with no identical cards, chosen from 2 decks is:

$$\binom{52}{5} \cdot 2^5$$

Therefore the probability of drawing a hand with no identical cards is:

$$\frac{\binom{52}{5} \cdot 2^5}{\binom{104}{5}}$$

(b) [5 points] What is the probability that the hand has exactly one pair of identical cards? **Solution.** This can be solved by a similar approach. A hand of this type can be distinguished by the face (suit and value) of the repeated card, and by the faces

of the 3 non-repeated cards. There are 52 possible values for the face of the repeated card. There are $\binom{51}{3}$ possible faces for the non-repeated cards, since none of these can be repeated. Each of these could come from either the 1st deck or the 2nd deck. There are $\binom{104}{5}$ possible hands, as before. So the probability of getting a hand with exactly one pair of identical cards is:

$$\frac{52 \cdot \binom{51}{3} \cdot 2^3}{\binom{104}{5}}$$

Problem 6. [28 points] Scores for a final exam are given by picking an integer uniformly at random from the set $\{50, 51, \dots, 97, 98\}$. The scores of all 128 students in the class are assigned in this manner. For parts (a), (b), (c) and (d) you may NOT assume that these scores are assigned independently. For parts (e), (f), (g) and (h) you MAY assume that these scores are assigned independently.

Let S_1, \dots, S_{128} be their scores. Let $S = \frac{1}{128}(\sum_{i=1}^{128} S_i)$ be the average score of the class.

(a) [3 points] For $i \in \{1, \dots, 128\}$, what is $\mathbb{E}[S_i]$?

Solution. We simply take the average of the numbers from 50 to 98. Thus, $\mathbb{E}[S_i] = \frac{50+98}{2} = 74$.

(b) [2 points] Show that $\mathbb{E}[S] = 74$. Make no independence assumptions.

Solution. By linearity of expectation,

$$\mathbb{E}[S] = \mathbb{E}\left[\frac{1}{128}\left(\sum_{i=1}^{128} S_i\right)\right] = \frac{1}{128}(128 * \mathbb{E}[S_1]) = \mathbb{E}[S_1] = 74$$

(c) [4 points] Prove that

$$\Pr[S \geq 88] \leq \frac{37}{44}.$$

Make no independence assumptions.

Solution. By Markov's inequality,

$$\Pr[S \geq 88] \leq \frac{\mathbb{E}[S]}{88} = \frac{74}{88} = \frac{37}{44}.$$

(d) [5 points] Improve your previous bound by using the fact that the minimum possible score is 50. Prove that

$$\Pr[S \geq 88] \leq \frac{12}{19}.$$

Make no independence assumptions.

Solution. We define a random variable $T = S - 50$, and thus $\mathbb{E}[T] = \mathbb{E}[S] - 50 = 24$. Now we just apply Markov's inequality:

$$\Pr[S \geq 88] = \Pr[T \geq 38] \leq \frac{\mathbb{E}[T]}{38} = \frac{24}{38} = \frac{12}{19}.$$

(e) [4 points] For the remaining problems, assume that all the scores are assigned mutually independently. Use Problem 1 of this final to find $\text{Var}[S_i]$.

Solution. We define $T_i = S_i - 50$.

$$\text{Var}[S_i] = \text{Var}[T_i] = \mathbb{E}[T_i^2] - \mathbb{E}^2[T_i] = \left(\frac{1}{49} \sum_{i=0}^{48} i^2\right) - \mathbb{E}^2[T_i] = \frac{1}{49} \frac{(48)(49)(97)}{6} - (24)^2 = 776 - 576 = 200.$$

(f) [3 points] What is $\text{Var}[S]$?

Solution.

$$\text{Var}[S] = \text{Var}\left[\frac{1}{128} \left(\sum_{i=1}^{128} S_i\right)\right] = \left(\frac{1}{128}\right)^2 (128 * \text{Var}[S_1]) = \frac{\text{Var}[S_1]}{128} = \frac{200}{128} = \frac{25}{16}.$$

(g) [2 points] What is the standard deviation of S ?

Solution. The standard deviation of S is simply the square root of the variance of S :

$$\sigma_S = \sqrt{\frac{25}{16}} = \frac{5}{4}.$$

(h) [5 points] Prove, using the Chebyshev Inequality, that

$$\Pr[S \leq 69] \leq \frac{1}{16}.$$

Solution. Using Chebyshev's inequality,

$$\Pr[S \leq 69] \leq \Pr[|S - 74| \leq 5] = \Pr[|S - \mathbb{E}[S]| \leq 4 * \sigma_S] \leq \frac{1}{4^2} = \frac{1}{16}.$$

Problem 7. [16 points] 1000 files $F_1, F_2, \dots, F_{1000}$ have just reached a disk manager for writing onto disk. Each file's size is between $0MB$ and $1MB$. The sum of all files' sizes is $400MB$.

The disk manager has 4 disks under its control. For each file F_i , the disk manager chooses a disk uniformly at random from amongst the 4 disks, and F_i is written to that disk. The choices of disk for the different files are mutually independent.

(a) [2 points] What is the expected number of files that will be written to the first disk?

Solution. We can use indicator variables. For each file, $P_i = 1$ if F_i is written to the first disk. The chance of an individual file being written to the first disk is $1/4$. By linearity of expectation, the expected number of files written to the first disk is the sum of the expected values of P_i 's. The expected value of each indicator variable is $1/4$, and $\sum_{i=1}^{1000} 1/4 = 250$, so the expected number of files to be written to the first disk is 250.

(b) [2 points] What is the expected number of bytes written on the first disk?

Solution. We can say that each file F_i has bit size S_i . Each file has a $1/4$ chance of being written to the first disk. Therefore, by linearity of expectation, the expected number of bytes written to the first disk is the sum of the expected number of bytes per file written to the first disk, which is:

$$\sum_{i=1}^{1000} 1/4 \cdot S_i = 1/4 \sum_{i=1}^{1000} S_i = 1/4 \cdot 400 = 100$$

(c) [8 points] Find the best upper bound you can on the probability that 200MB or more are written on the first disk?

Solution. For this we can use the first Chernoff bound, which is:

$$\Pr(X \geq c \text{Ex}(X)) \leq e^{-(c \ln c - c + 1) \text{Ex}(X)}$$

The Chernoff bound only works if X is the sum of random variables that each take on a value between 0 and 1. The file size of each file in the first disk is between 0 and 1Mb. So we can define X to be the total number of bytes in disk 1. The expected value of X is 100, so we take c to be 2. We get:

$$\Pr(X \geq 2 \cdot 100) \leq e^{-(2 \ln 2 - 2 + 1)100}$$

(d) [4 points] Find the best upper bound you can on the probability that there is some disk with 200MB or more written on it?

Solution. For this we can use the Union Bound along with our result from above. The probability of this event happening in one or more disks is upper bounded by the sum of the probabilities of the event happening in each disk. This gives us an upper bound of

$$4 \cdot e^{-(2 \ln 2 - 1)100}$$