1 Student Life

In an attempt to avoid having to do laundry often, Marcus comes up with a system. Every night, he designates one of his shirts as his dirtiest shirt. In the morning, he randomly picks one of his shirts to wear. If he picked the dirtiest one, he puts it in a dirty pile at the end of the day (a shirt in the dirty pile is not used again until it is cleaned).

When Marcus puts his last shirt into the dirty pile, he finally does his laundry, and again designates one of his shirts as his dirtiest shirt (laundry isn’t perfect) before going to bed. This process then repeats.

(a) If Marcus has $n$ shirts, what is the expected number of days that transpire between laundry events? Your answer should be a function of $n$ involving no summations.

(b) Say he gets even lazier, and instead of organizing his shirts in his dresser every night, he throws his shirts randomly onto one of $n$ different locations in his room (one shirt per location), designates one of his shirts as his dirtiest shirt, and one location as the dirtiest location.

In the morning, if he happens to pick the dirtiest shirt, and the dirtiest shirt was in the dirtiest location, then he puts the shirt into the dirty pile at the end of the day and does not throw any future shirts into that location and also does not consider it as a candidate for future dirtiest locations (it is too dirty).

What is the expected number of days that transpire between laundry events now? Again, your answer should be a function of $n$ involving no summations.

2 Coupon Collector Variance

It’s that time of the year again—Safeway is offering its Monopoly Card promotion. Each time you visit Safeway, you are given one of $n$ different Monopoly Cards with equal probability. You need
to collect them all to redeem the grand prize.

Let $X$ be the number of visits you have to make before you can redeem the grand prize. We’ve shown in discussion that $\text{Var}(X) = n^2 \left( \sum_{i=1}^{n} i^{-2} \right) - \mathbb{E}[X]$.

The series $\sum_{i=1}^{\infty} i^{-2}$ converges to the constant value $\pi^2 / 6$. Using this fact and Chebyshev’s Inequality, find a lower bound on $\beta$ for which the probability you need to make more than $\mathbb{E}[X] + \beta n$ visits is less than $1/100$, for large $n$. [Hint: Use the approximation $\sum_{i=1}^{n} i^{-1} \approx \ln n$ as $n$ grows large.]

3 Probabilistically Buying Probability Books

Chuck will go shopping for probability books for $K$ hours. Here, $K$ is a random variable and is equally likely to be 1, 2, or 3. The number of books $N$ that he buys is random and depends on how long he shops. We are told that

$$
\mathbb{P}[N = n \mid K = k] = \begin{cases} 
\frac{c}{k} & \text{for } n = 1, \ldots, k \\
0 & \text{otherwise}
\end{cases}
$$

for some constant $c$.

(a) Compute $c$.

(b) Find the joint distribution of $K$ and $N$.

(c) Find the marginal distribution of $N$.

(d) Find the conditional distribution of $K$ given that $N = 1$.

(e) We are now told that he bought at least 1 but no more than 2 books. Find the conditional mean and variance of $K$, given this piece of information.

(f) The cost of each book is a random variable with mean 3. What is the expectation of his total expenditure? [Hint: Condition on events $N = 1, N = 2, N = 3$ and use total expectation.]

4 Dice Games

(a) Alice and Bob are playing a game. Alice picks a random integer $X$ between 0 and 100 inclusive, where each value is equally likely to be chosen. Bob then picks a random integer $Y$ between 0 and $X$ inclusive. What is $\mathbb{E}[Y]$?

(b) Alice rolls a 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 $\mathbb{E}[Y]$.

(c) Bob plays a game in which he starts off with one 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 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.

What is the expected number of dice Bob will have after $n$ time steps?

5 Just One Tail, Please

Let $X$ be some random variable with finite mean and variance which is not necessarily non-negative. The extended version of Markov’s Inequality states that for a non-negative function $\phi(x)$ which is monotonically increasing for $x > 0$ and some constant $\alpha > 0$,

$$\mathbb{P}[X \geq \alpha] \leq \frac{\mathbb{E}[\phi(X)]}{\phi(\alpha)}$$

Suppose $\mathbb{E}[X] = 0$, $\text{Var}(X) = \sigma^2 < \infty$, and $\alpha > 0$.

(a) Use the extended version of Markov’s Inequality stated above with $\phi(x) = (x + c)^2$, where $c$ is some positive constant, to show that:

$$\mathbb{P}[X \geq \alpha] \leq \frac{\sigma^2 + c^2}{(\alpha + c)^2}$$

(b) Note that the above bound applies for all positive $c$, so we can choose a value of $c$ to minimize the expression, yielding the best possible bound. Find the value for $c$ which will minimize the RHS expression (you may assume that the expression has a unique minimum).

We can plug in the minimizing value of $c$ you found in part (b) to prove the following bound:

$$\mathbb{P}[X \geq \alpha] \leq \frac{\sigma^2}{\alpha^2 + \sigma^2}.$$ 

This bound is also known as Cantelli’s inequality.

(c) Recall that Chebyshev’s inequality provides a two-sided bound. That is, it provides a bound on $\mathbb{P}[|X - \mathbb{E}[X]| \geq \alpha] = \mathbb{P}[X \geq \mathbb{E}[X] + \alpha] + \mathbb{P}[X \leq \mathbb{E}[X] - \alpha]$. If we only wanted to bound the probability of one of the tails, e.g. if we wanted to bound $\mathbb{P}[X \geq \mathbb{E}[X] + \alpha]$, it is tempting to just divide the bound we get from Chebyshev’s by two.

(i) Why is this not always correct in general?

(ii) Provide an example of a random variable $X$ (does not have to be zero-mean) and a constant $\alpha$ such that using this method (dividing by two to bound one tail) is not correct, that is, $\mathbb{P}[X \geq \mathbb{E}[X] + \alpha] > \frac{\text{Var}(X)}{2\alpha^2}$ or $\mathbb{P}[X \leq \mathbb{E}[X] - \alpha] > \frac{\text{Var}(X)}{2\alpha^2}$.

Now we see the use of the bound proven in part (b) - it allows us to bound just one tail while still taking variance into account, and does not require us to assume any property of the random variable. Note that the bound is also always guaranteed to be less than 1 (and therefore at least somewhat useful), unlike Markov’s and Chebyshev’s inequality!
(d) Let’s try out our new bound on a simple example. Suppose $X$ is a positively-valued random variable with $\mathbb{E}[X] = 3$ and $\text{Var}(X) = 2$.

(i) What bound would Markov’s inequality give for $\mathbb{P}[X \geq 5]$?

(ii) What bound would Chebyshev’s inequality give for $\mathbb{P}[X \geq 5]$?

(iii) What bound would Cantelli’s Inequality give for $\mathbb{P}[X \geq 5]$? (*Note: Recall that Cantelli’s Inequality only applies for zero-mean random variables.*)

6 Estimating $\pi$

In this problem, we discuss some interesting ways that you could probabilistically estimate $\pi$, and see how good these techniques are at estimating $\pi$.

**Technique 1:** Buffon’s needle is a method that can be used to estimate the value of $\pi$. There is a table with infinitely many parallel lines spaced a distance 1 apart, and a needle of length 1. It turns out that if the needle is dropped uniformly at random onto the table, the probability of the needle intersecting a line is $\frac{2}{\pi}$. We have seen a proof of this in the notes.

**Technique 2:** Consider a square dartboard, and a circular target drawn inscribed in the square dartboard. A dart is thrown uniformly at random in the square. The probability the dart lies in the circle is $\frac{\pi}{4}$.

**Technique 3:** Pick two integers $x$ and $y$ independently and uniformly at random from 1 to $M$, inclusive. Let $p_M$ be the probability that $x$ and $y$ are relatively prime. Then

$$\lim_{M \to \infty} p_M = \frac{6}{\pi^2}.$$ 

Let $p_1 = \frac{2}{\pi}, p_2 = \frac{\pi}{4}$, and $p_3 = \frac{6}{\pi^2}$ be the probabilities of the desired events of **Technique 1**, **Technique 2**, and **Technique 3**, respectively. For each technique, we apply each technique $N$ times, then compute the proportion of the times each technique occurred, getting estimates $\hat{p}_1, \hat{p}_2,$ and $\hat{p}_3$, respectively.

(a) For each $\hat{p}_i$, compute an expression $X_i$ in terms of $\hat{p}_i$ that would be an estimate of $\pi$.

(b) Using Chebyshev’s Inequality, compute the minimum value of $N$ such that $X_2$ is within $\varepsilon$ of $\pi$ with $1 - \delta$ confidence. Your answer should be in terms of $\varepsilon$ and $\delta$.

For $X_1$ and $X_3$, computing the minimum value of $N$ will be more tricky, as the expressions for $X_1$ and $X_3$ are not as nice as $X_2$.

(c) For $i = 1$ and 3, compute a constant $c_i$ such that

$$|X_i - \pi| < \varepsilon \iff |\hat{p}_i - p_i| < c_i \varepsilon + o(\varepsilon^2),$$

where the $o(\varepsilon^2)$ represents terms containing powers of $\varepsilon$ that are 2 or higher (i.e. $\varepsilon^2, \varepsilon^3$, etc.).
(Hint: You may find the following Taylor series helpful: For $x$ close to 0,

$$\frac{1}{a-x} = \frac{1}{a} + \frac{x}{a^2} + o(x^2)$$

$$\frac{1}{(a-x)^2} = \frac{1}{a^2} + \frac{2x}{a^3} + o(x^2).$$

The $o(x^2)$ represents terms that have $x^2$ powers or higher. )

In this problem, we assume $\varepsilon$ is close enough to 0 such that $o(\varepsilon^2)$ is 0. In other words,

$$\mathbb{P}\left[|\hat{p}_i - p_i| < c_i \varepsilon + o(\varepsilon^2)\right] = \mathbb{P}\left[|\hat{p}_i - p_i| < c_i \varepsilon\right].$$

Combining with part (c) then gives

$$\mathbb{P}[|X_i - \pi| < \varepsilon] = \mathbb{P}[|\hat{p}_i - p_i| < c_i \varepsilon].$$

(d) For $i = 1$ and 3, use Chebyshev’s Inequality and the above work to compute the minimum value of $N$ such that $X_i$ is within $\varepsilon$ of $\pi$ with $1 - \delta$ confidence. Your answer should be in terms of $\varepsilon$ and $\delta$.

(e) Which technique required the lowest value for $N$? Which technique required the highest?