CS 70 Spring 2022 — Discrete Mathematics and Probability Theory — Satish Rao and Koushik Sen — HW 12

1 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 α 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.)

Solution:

(a) Note that $\sigma^2 = \text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2]$. Using the inequality presented in the problem, we have:

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

(b) We set the derivative with respect to c of the above expression equal to 0, and solve for c.

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}c}\frac{\sigma^2+c^2}{(\alpha+c)^2} &= 0 \\
\frac{2c(\alpha+c)^2 - 2(\alpha+c)(\sigma^2+c^2)}{(\alpha+c)^4} &= 0 \\
2c(\alpha+c)^2 - 2(\alpha+c)(\sigma^2+c^2) &= 0 \\
\alpha c^2 + (\alpha^2 - \sigma^2)c - \sigma^2\alpha &= 0 \\
c &= \frac{\sigma^2}{\alpha}
\end{aligned}$$

To get the last step we use the quadratic equation and take the positive solution.

(c) It is possible for one of the tails to contain more probability than the other. One example of a random variable which demonstrates this is X, where $\mathbb{P}[X=0] = 0.75$ and $\mathbb{P}[X=10] = 0.25$, with $\alpha = 7$. Here, $\mathbb{E}[X] = 2.5$ and $\text{Var}(X) = 100 \cdot 0.25 \cdot 0.75$, so we have:

$$\mathbb{P}[X \geq \mathbb{E}[X] + 7] = 0.25 > \frac{\text{Var}(X)}{2 \cdot 7^2} \approx 0.19$$

(d) Using Markov's: $\mathbb{P}[X \geq 5] \leq \frac{\mathbb{E}[X]}{5} = \frac{3}{5}$

Using Chebyshev's: $\mathbb{P}[X \geq 5] \leq \mathbb{P}[|X - \mathbb{E}[X]| \geq 2] \leq \frac{\text{Var}(X)}{2^2} = \frac{1}{2}$

Using bound shown above (Cantelli's):
Since we have the condition that this bound applies to zero-mean random variables, let us define $Y = X - \mathbb{E}[X] = X - 3$. Note that $\text{Var}(Y) = \text{Var}(X)$.

Then we get: $\mathbb{P}[X \geq 5] = \mathbb{P}[Y \geq 2] \leq \frac{\text{Var}(Y)}{2^2 + \text{Var}(Y)} = \frac{1}{3}$.

We see that Cantelli's inequality (the bound from part (b)) does better than Chebyshev's, which does better than Markov's (note that having a smaller upper bound is better)! This is a good demonstration on how we might derive better bounds using Markov's inequality, if we know further information about the random variable like its variance.

2 Law of Large Numbers

Recall that the *Law of Large Numbers* holds if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| > \varepsilon\right] = 0.$$

In class, we saw that the Law of Large Numbers holds for $S_n = X_1 + \cdots + X_n$, where the X_i's are i.i.d. random variables. This problem explores if the Law of Large Numbers holds under other circumstances.

Packets are sent from a source to a destination node over the Internet. Each packet is sent on a certain route, and the routes are disjoint. Each route has a failure probability of $p \in (0,1)$ and different routes fail independently. If a route fails, all packets sent along that route are lost. You can assume that the routing protocol has no knowledge of which route fails.

For each of the following routing protocols, determine whether the Law of Large Numbers holds when S_n is defined as the total number of received packets out of n packets sent. Answer **Yes** if the Law of Large Number holds, or **No** if not. Give a justification of your answer. (Whenever convenient, you can assume that n is even.)

(a) **Yes** or **No**: Each packet is sent on a completely different route.

(b) **Yes** or **No**: The packets are split into $n/2$ pairs of packets. Each pair is sent together on its own route (i.e., different pairs are sent on different routes).

(c) **Yes** or **No**: The packets are split into 2 groups of $n/2$ packets. All the packets in each group are sent on the same route, and the two groups are sent on different routes.

(d) **Yes** or **No**: All the packets are sent on one route.

Solution:

(a) **Yes.** Define X_i to be 1 if a packet is sent successfully on route i. Then X_i, $i = 1, \ldots, n$ is 0 with probability p and 1 otherwise. Since we have individual routes for each packet, we have a total of n routes. The total number of successful packets sent is hence $S_n = X_1 + \cdots + X_n$. Since S_n is a sum of i.i.d. Bernoulli random variables, $S_n \sim \text{Binomial}(n, 1-p)$.

Now similar to notation in the lecture notes, we define $A_n = S_n/n$ to be the fraction of successful packets sent, out of the n packets. Moreover, for each X_i,

$$\mathbb{E}[X_i] = 1-p$$

and

$$\text{Var}(X_i) = p(1-p).$$

Using Chebyshev's inequality:

$$\mathbb{P}[|A_n - \mathbb{E}[A_n]| > \varepsilon] = \mathbb{P}[|A_n - (1-p)| > \varepsilon] \le \frac{\text{Var}[A_n]}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty.$$

(b) **Yes.** Now we need $n/2$ routes for each pair of packets. Similarly to the previous question, we define X_i, $i = 1, \ldots, n/2$ to be 0 with probability p and 2 (packets) otherwise. Now the total number of packets is $S_n = X_1 + \cdots + X_{n/2}$ and the fraction of received packets is $A_n = S_n/n$.

Now for each $i = 1, \ldots, n/2$,

$$\mathbb{E}[X_i] = 2(1-p)$$

and

$$\text{Var}(X_i) = 4p(1-p).$$

Thus,

$$\mathbb{E}[A_n] = \frac{\mathbb{E}[X_1] + \ldots + \mathbb{E}[X_{n/2}]}{n} = \frac{1}{n} \cdot \frac{n}{2} \cdot 2(1-p) = 1-p$$

and

$$\text{Var}[A_n] = \frac{1}{n^2}\left(\text{Var}[X_1] + \ldots + \text{Var}[X_{n/2}]\right) = \frac{1}{n^2} \cdot \frac{n}{2} 4p(1-p) = \frac{2p(1-p)}{n}.$$

Finally, we get:

$$\mathbb{P}[|A_n - \mathbb{E}[A_n]| > \varepsilon] = \mathbb{P}[|A_n - (1-p)| > \varepsilon] \le \frac{2p(1-p)}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty.$$

(c) **No.** In this situation, we have that no packets get through with probability p^2, half the packets get through with probability $2p(1-p)$, and all the packets get through with probability $(1-p)^2$. This tells us that $\frac{1}{n}S_n$ is 0 with probability p^2, $\frac{1}{2}$ with probability $2p(1-p)$, and 1 with probability $(1-p)^2$. Since $\mathbb{E}\left[\frac{1}{n}S_n\right] = 1-p$, this gives us that

$$\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| = \begin{cases} 1-p & \text{with probability } p^2 \\ \left|p - \frac{1}{2}\right| & \text{with probability } 2p(1-p) \\ p & \text{with probability } (1-p)^2 \end{cases}$$

We now consider two cases: either $p = \frac{1}{2}$ or $p \neq \frac{1}{2}$. In the former case, we can take $\varepsilon = \frac{1}{4}$, and we'll have that

$$\begin{aligned}\mathbb{P}\left[\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| > \varepsilon\right] &= \mathbb{P}\left[\frac{1}{n}S_n = 0 \cup \frac{1}{n}S_n = 1\right] \\ &= \frac{1}{2}\end{aligned}$$

In the latter case, we can take $\varepsilon = \frac{\min\left(1-p, \left|p-\frac{1}{2}\right|, p\right)}{2}$ and we'll have that

$$\mathbb{P}\left[\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| > \varepsilon\right] = 1$$

Since neither of these probabilities converge to zero as $n \to \infty$, we have that the WLLN does not hold in either case.

(d) **No.** In this case, we have that no packets get through with probability p and all the packets get through with probability $(1-p)$. Hence,

$$\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| = \begin{cases} 1-p & \text{with probability } p \\ p & \text{with probability } (1-p) \end{cases}$$

So if we take $\varepsilon = \frac{\min(p, 1-p)}{2}$, we have that

$$\mathbb{P}\left[\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| > \varepsilon\right] = 1$$

As in the previous part, because this does not converge to 0 as $n \to \infty$, we have that the WLLN does not hold.

For problems (c) and (d), you should've had the intuition that since the packets are automatically sent through 1 or 2 routes, increasing n does not really help for LLN.

3 Practical Confidence Intervals

(a) It's New Year's Eve, and you're re-evaluating your finances for the next year. Based on previous spending patterns, you know that you spend \$1500 per month on average, with a standard deviation of \$500, and each month's expenditure is independently and identically distributed. As a college student, you also don't have any income. How much should you have in your bank account if you don't want to run out of money this year, with probability at least 95%?

(b) As a UC Berkeley CS student, you're always thinking about ways to become the next billionaire in Silicon Valley. After hours of brainstorming, you've finally cut your list of ideas down to 10, all of which you want to implement at the same time. A venture capitalist has agreed to back all 10 ideas, as long as your net return from implementing the ideas is positive with at least 95% probability.

Suppose that implementing an idea requires 50 thousand dollars, and your start-up then succeeds with probability p, generating 150 thousand dollars in revenue (for a net gain of 100 thousand dollars), or fails with probability $1-p$ (for a net loss of 50 thousand dollars). The success of each idea is independent of every other. What is the condition on p that you need to satisfy to secure the venture capitalist's funding?

(c) One of your start-ups uses error-correcting codes, which can recover the original message as long as at least 1000 packets are received (not erased). Each packet gets erased independently with probability 0.8. How many packets should you send such that you can recover the message with probability at least 99%?

Solution:

(a) Let T be the random variable representing the amount of money we spend in the year.

We have $T = \sum_{i=1}^{12} X_i$, where X_i represents the spending in the i-th month. So, $\mathbb{E}[T] = 12 \cdot \mathbb{E}[E_1] = 18000$.

And, since the X_is are independent, $\text{Var}(T) = 12 \cdot \text{Var}(X_1) = 12 \cdot 500^2 = 3,000,000$.

We want to have enough money in our bank account so that we don't finish the year in debt with 95% confidence. So, we want to keep some money ε more than the mean expenditure such that the probability of deviating above the mean by more than ε is less than 0.05.

Let's use Chebyshev's inequality here to express this.

$$\mathbb{P}[|T - \mathbb{E}[T]| \geq \varepsilon] \leq \frac{\text{Var}(T)}{\varepsilon^2} \leq 0.05$$

This gives us $\varepsilon^2 \geq \frac{3,000,000}{0.05}$. So, $\varepsilon \geq 7746$. This means that we want to have a balance of $\geq \mathbb{E}[T] + \varepsilon = 25746$.

Observe that here, while we wanted to estimate $\mathbb{P}[T - \mathbb{E}[T] \geq \varepsilon]$, Chebyshev's inequality only gives us information about $\mathbb{P}[|T - \mathbb{E}[T]| \geq \varepsilon]$. But since

$$\mathbb{P}[|T - \mathbb{E}[T]| \geq \varepsilon] \geq \mathbb{P}[T - \mathbb{E}[T] \geq \varepsilon],$$

this is fine. We just get a more conservative estimate.

(b) For this question, to keep the numbers from exploding, let's work in thousands of dollars. Let X_i be the profit made from idea i, and T be the total profit made. We have $T = \sum_{i=1}^{10} X_i$.

Here, $\mathbb{E}[X_1] = 100p - 50(1-p) = 150p - 50$.

And $\text{Var}(X_1) = 150^2 p(1-p)$ as the distribution of X_1 is a shifted and scaled Bernoulli distribution. Using $\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2$ yields the same answer.

We have, $\mathbb{E}[T] = 10 \cdot \mathbb{E}[X_1]$. Similarly, $\text{Var}(T) = 10 \cdot \text{Var}(X_1)$.

Now, we want to bound the probability of T going below 0 by 0.05. In other words, we want $\mathbb{P}[T < 0] \leq 0.05$.

But, in order to apply Chebyshev's inequality, we need to look at deviation from the mean. We use the assumption that to get our funding we obviously need $\mathbb{E}[T] > 0$. Then:

$$\mathbb{P}[T < 0] \leq \mathbb{P}[T \leq 0 \cup T \geq 2\mathbb{E}[T]] = \mathbb{P}[|T - \mathbb{E}[T]| \geq \mathbb{E}[T]] \leq \frac{\mathrm{Var}(T)}{\mathbb{E}[T]^2} \leq 0.05$$

Looking at just the last inequality, we have:

$$\frac{\mathrm{Var}(T)}{\mathbb{E}[T]^2} = \frac{10 \cdot \mathrm{Var}(X_1)}{100 \cdot \mathbb{E}[X_1]^2} = \frac{\mathrm{Var}(X_1)}{10 \cdot \mathbb{E}[X_1]^2} \leq 0.05$$

$$\therefore \frac{\mathrm{Var}(X_1)}{\mathbb{E}[X_1]^2} \leq 0.5$$

Now, substituting what we have for variance and expectation, we get the following:

$$-22500p^2 + 22500p \leq 0.5(150p - 50)^2$$

which gives us the quadratic:

$$33750p^2 - 30000p + 1250 \geq 0$$

The solutions for p are $p \geq \frac{1}{9}(4 + \sqrt{13})$ and $p \leq \frac{1}{9}(4 - \sqrt{13})$. So $p \geq 0.845$ or ≤ 0.0438.

The relevant solution here is to pick $p \geq 0.845$, since the other solution yields negative expectation (contradicting the earlier assumption of positive expectation).

(c) We want $k = 1000$ packets to get across without being erased. Say we send n packets. Let X_i be the indicator random variable representing whether the ith packet got across or not.

Let the total number of unerased packets sent across be T. We have $T = \sum_{i=1}^{n} X_i$ and we want $T \geq 1000$.

We want $\mathbb{P}[T < 1000] \leq 0.01$. Now, let's try to get this in a form so that we can use Chebyshev's inequality. We know that $\mathbb{E}[T] > 1000$, so we can say that

$$\begin{aligned}\mathbb{P}[T < 1000] &\leq \mathbb{P}[T \leq 1000 \cup T \geq \mathbb{E}[T] + (\mathbb{E}[T] - 1000)] \\ &= \mathbb{P}[|T - \mathbb{E}[T]| \geq (\mathbb{E}[T] - 1000)] \leq \frac{\mathrm{Var}(T)}{(\mathbb{E}[T] - 1000)^2} \leq 0.01.\end{aligned}$$

What is $\mathbb{E}[T]$? $\mathbb{E}[T] = n\mathbb{E}[X_1] = n(1-p) = 0.2n$.

Next, what is $\mathrm{Var}(T)$? $\mathrm{Var}(T) = n\,\mathrm{Var}(X_1) = np(1-p) = 0.16n$.

Now, $\frac{\mathrm{Var}(T)}{(\mathbb{E}[T] - k)^2} \leq 0.01 \implies 16n \leq (0.2n - 1000)^2$. This gives us the quadratic:

$$0.04n^2 - 416n + 1000000 \geq 0$$

Solving the last quadratic, we get $n \geq 6629$ or $n \leq 3774$. Since the second inequality doesn't make sense for our situation, our answer is $n \geq 6629$.

4 Estimating π

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

Technique 1: Buffon's needle is a method that can be used to estimate the value of π. 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 π.

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

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

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

where the $o(\varepsilon^2)$ represents terms containing powers of ε 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 ε 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

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

(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 ε of π with $1-\delta$ confidence. Your answer should be in terms of ε and δ.

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

Solution:

(a) $\hat{p}_1$ is an estimate of $\frac{2}{\pi}$, so $X_1 = \frac{2}{\hat{p}_1}$ would be an estimate of π. Similarly, $X_2 = 4\hat{p}_2$ and $X_3 = \sqrt{\frac{6}{\hat{p}_3}}$ are estimates of π.

(b) We have

$$\mathbb{P}\left[|X_2 - \pi| \geq \varepsilon\right] = \mathbb{P}\left[\left|\hat{p}_2 - \frac{\pi}{4}\right| \geq \frac{1}{4}\varepsilon\right]$$
$$\geq \frac{\text{Var}\left(\hat{p}_2\right)}{\left(\frac{1}{4}\varepsilon\right)^2}$$

by Chebyshev's Inequality and using the fact that $X_2 = 4\hat{p}_2$. We want our estimate to have confidence $1-\delta$, so we want $\frac{\text{Var}(\hat{p}_2)}{\left(\frac{1}{4}\varepsilon\right)^2} < \delta$. Since $\hat{p}_2$ is a Binomial(N, p_2) variable, it has variance $\frac{p_2(1-p_2)}{N}$. Combining everything gives

$$\frac{\frac{p_2(1-p_2)}{N}}{\left(\frac{1}{4}\varepsilon\right)^2} < \delta \implies N > \frac{16p_2(1-p_2)}{\delta\varepsilon^2} = \frac{\pi(4-\pi)}{\delta\varepsilon^2}.$$

(c) For $i = 1$, we have

$$\begin{aligned}
|X_1 - \pi| < \varepsilon &\implies \left|\frac{2}{\hat{p}_1} - \pi\right| < \varepsilon \\
&\implies \pi - \varepsilon < \frac{2}{\hat{p}_1} < \pi + \varepsilon \\
&\implies \frac{2}{\pi + \varepsilon} < \hat{p}_1 < \frac{2}{\pi - \varepsilon} \\
&\implies \frac{2}{\pi + \varepsilon} - \frac{2}{\pi} < \hat{p}_1 - \frac{2}{\pi} < \frac{2}{\pi - \varepsilon} - \frac{2}{\pi} \\
&\implies -\frac{2\varepsilon}{\pi(\pi + \varepsilon)} < \hat{p}_1 - \frac{2}{\pi} < \frac{2\varepsilon}{\pi(\pi - \varepsilon)}.
\end{aligned}$$

We apply the Taylor series expansion $\frac{1}{a-x} = \frac{1}{a} + o(x)$ to get

$$\frac{2\varepsilon}{\pi(\pi - \varepsilon)} = \frac{2\varepsilon}{\pi}\left(\frac{1}{\pi} + o(\varepsilon)\right) = \frac{2\varepsilon}{\pi^2} + o(\varepsilon^2)$$

and

$$-\frac{2\varepsilon}{\pi(\pi + \varepsilon)} = -\frac{2\varepsilon}{\pi}\left(\frac{1}{\pi} + o(\varepsilon)\right) = -\frac{2\varepsilon}{\pi^2} + o(\varepsilon^2).$$

We conclude that

$$|X_1 - \pi| < \varepsilon \implies -\frac{2\varepsilon}{\pi^2} + o(\varepsilon^2) < \hat{p}_1 - \frac{2}{\pi} < -\frac{2\varepsilon}{\pi^2} + o(\varepsilon^2) \implies |\hat{p}_1 - \frac{2}{\pi}| < \frac{2\varepsilon}{\pi^2} + o(\varepsilon^2),$$

so $c_1 = \frac{2}{\pi^2}$.

Similarly, for $i = 3$, we have

$$\begin{aligned}
|X_3 - \pi| < \varepsilon &\implies \left|\sqrt{\frac{6}{\hat{p}_3}} - \pi\right| < \varepsilon \\
&\implies \pi - \varepsilon < \sqrt{\frac{6}{\hat{p}_3}} < \pi + \varepsilon \\
&\implies \frac{6}{(\pi + \varepsilon)^2} < \hat{p}_3 < \frac{6}{(\pi - \varepsilon)^2} \\
&\implies \frac{6}{(\pi + \varepsilon)^2} - \frac{6}{\pi^2} < \hat{p}_3 - \frac{6}{\pi^2} < \frac{6}{(\pi - \varepsilon)^2} - \frac{6}{\pi^2} \\
&\implies -\frac{12\pi\varepsilon + 6\varepsilon^2}{\pi^2(\pi + \varepsilon)^2} < \hat{p}_3 - \frac{6}{\pi^2} < \frac{12\pi\varepsilon - 6\varepsilon^2}{\pi^2(\pi - \varepsilon)^2}.
\end{aligned}$$

We apply the Taylor series expansion $\frac{1}{(a-x)^2} = \frac{1}{a^2} + o(x)$ to get

$$\frac{12\pi\varepsilon - 6\varepsilon^2}{\pi^2(\pi - \varepsilon)^2} = \frac{12\pi\varepsilon - 6\varepsilon^2}{\pi^2}\left(\frac{1}{\pi^2} + o(\varepsilon)\right) = \frac{12\pi\varepsilon}{\pi^4} + o(\varepsilon^2) = \frac{12\varepsilon}{\pi^3} + o(\varepsilon^2)$$

and

$$-\frac{12\pi\varepsilon+6\varepsilon^2}{\pi^2(\pi+\varepsilon)^2} = -\frac{12\pi\varepsilon+6\varepsilon^2}{\pi^2}\left(\frac{1}{\pi^2}+o(\varepsilon)\right) = -\frac{12\pi\varepsilon}{\pi^4}+o(\varepsilon^2) = -\frac{12\varepsilon}{\pi^3}+o(\varepsilon^2).$$

We conclude that

$$|X_3-\pi|<\varepsilon \implies -\frac{12\varepsilon}{\pi^3}+o(\varepsilon^2) < \hat{p}_3-\frac{6}{\pi^2} < \frac{12\varepsilon}{\pi^3}+o(\varepsilon^2) \implies \left|\hat{p}_3-\frac{6}{\pi^2}\right| < \frac{12\varepsilon}{\pi^3}+o(\varepsilon^2),$$

so $c_3 = \frac{12}{\pi^3}$.

(d) By Chebyshev's Inequality,

$$\begin{aligned}\mathbb{P}\left[|X_i-\pi|\geq\varepsilon\right] &= \mathbb{P}\left[|\hat{p}_i-p_i|\geq c_i\varepsilon\right]\\ &\geq \frac{\operatorname{Var}(\hat{p}_i)}{(c_i\varepsilon)^2}.\end{aligned}$$

We want our estimate to have confidence $1-\delta$, so we want $\frac{\operatorname{Var}(\hat{p}_i)}{(c_i\varepsilon)^2}<\delta$. Since $\hat{p}_i$ is a Binomial(N,p_i) variable, it has variance $\frac{p_i(1-p_i)}{N}$. Combining everything gives

$$\frac{\frac{p_i(1-p_i)}{N}}{(c_i\varepsilon)^2}<\delta \implies N>\frac{p_i(1-p_i)}{c_i^2\varepsilon^2\delta}.$$

Plugging in $i=1$ gives

$$N>\frac{\frac{2}{\pi}\left(1-\frac{2}{\pi}\right)}{\left(\frac{2}{\pi^2}\right)^2\varepsilon^2\delta} = \frac{\pi^2(\pi-2)}{2\varepsilon^2\delta}.$$

and plugging in $i=3$ gives

$$N>\frac{\frac{6}{\pi^2}\left(1-\frac{6}{\pi^2}\right)}{\left(\frac{12}{\pi^3}\right)^2\varepsilon^2\delta} = \frac{\pi^2(\pi^2-6)}{24\varepsilon^2\delta}.$$

(e) Looking at our values from parts (b) and (d),

$$\frac{\pi^2(\pi^2-6)}{24\varepsilon^2\delta} < \frac{\pi(4-\pi)}{\delta\varepsilon^2} < \frac{\pi^2(\pi-2)}{2\varepsilon^2\delta},$$

so technique 3 required the lowest value for N, while technique 1 required the highest value for N.

5 Balls in Bins Estimation

We throw $n > 0$ balls into $m \geq 2$ bins. Let X and Y represent the number of balls that land in bin 1 and 2 respectively.

(a) Calculate $\mathbb{E}[Y \mid X]$. [*Hint*: Your intuition may be more useful than formal calculations.]

(b) What is $L[Y \mid X]$ (where $L[Y \mid X]$ is the best linear estimator of Y given X)? [*Hint*: Your justification should be no more than two or three sentences, no calculations necessary! Think carefully about the meaning of the conditional expectation.]

(c) Unfortunately, your friend is not convinced by your answer to the previous part. Compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

(d) Compute $\text{Var}(X)$.

(e) Compute $\text{cov}(X,Y)$.

(f) Compute $L[Y \mid X]$ using the formula. Ensure that your answer is the same as your answer to part (b).

Solution:

(a) $\mathbb{E}[Y \mid X = x] = (n-x)/(m-1)$, because once we condition on x balls landing in bin 1, the remaining $n-x$ balls are distributed uniformly among the other $m-1$ bins. Therefore,

$$\mathbb{E}[Y \mid X] = \frac{n-X}{m-1}.$$

(b) We showed that $\mathbb{E}[Y \mid X]$ is a linear function of X. Since $\mathbb{E}[Y \mid X]$ is the best *general* estimator of Y given X, it must also be the best *linear* estimator of Y given X, i.e. $\mathbb{E}[Y \mid X]$ and $L[Y \mid X]$ coincide.

(c) Let X_i be the indicator that the ith ball falls in bin 1. Then, $X = \sum_{i=1}^n X_i$, and by linearity of expectation, $\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = n/m$, since there are n indicators and each ball has a probability $1/m$ of landing in bin 1. By symmetry, $\mathbb{E}[Y] = n/m$ as well.

(d) The number of balls that falls into the first bin is binomially distributed with parameters n and $1/m$. Hence the variance is $n(1/m)(1-1/m)$.

(e) Let X_i be as before, and let Y_i be the indicator that the ith ball falls into bin 2.

$$\text{cov}(X,Y) = \sum_{i=1}^n \sum_{j=1}^n \text{cov}(X_i, Y_j)$$

We can compute $\text{cov}(X_i,Y_i) = \mathbb{E}[X_i Y_i] - \mathbb{E}[X_i]\,\mathbb{E}[Y_i] = 0 - (1/m)(1/m) = -1/m^2$ (note that $\mathbb{E}[X_i Y_i] = 0$ because it is impossible for a ball to land in both bins 1 and 2). Also, we have $\text{cov}(X_i,Y_j) = 0$ because the indicator for the ith ball is independent of the indicator for the jth ball when $i \neq j$. Hence, $\text{cov}(X,Y) = n(-1/m^2) = -n/m^2$.

(f)

$$\begin{aligned}
L[Y \mid X] &= \mathbb{E}[Y] + \frac{\operatorname{cov}(X,Y)}{\operatorname{var}(X)}(X - \mathbb{E}[X]) \\
&= \frac{n}{m} + \frac{-n/m^2}{n(1/m)(1-1/m)}\left(X - \frac{n}{m}\right) \\
&= \frac{n}{m} - \frac{1}{m-1}\left(X - \frac{n}{m}\right) \\
&= \frac{mn - n - mX + n}{m(m-1)} = \frac{n-X}{m-1}
\end{aligned}$$

6 Gambling Woes

Forest proposes a gambling game to you (uh oh!). Every day, you flip two independent fair coins. If both of the coins come up heads, then your fortune triples on that day. If one coin comes up heads and the other coin comes up tails, then your fortune is cut in half. If both of the coins comes up tails, then game over: you lose all of your money! Forest claims that you can get rich quickly with this scheme, but you decide to calculate some probabilities first.

(a) Let M_0 denote your money at the start of the game, and let M_n denote the amount of money you have at the end of the nth day. Compute $\mathbb{E}[M_{n+1} \mid M_n]$.

(b) Use the law of iterated expectation to calculate $\mathbb{E}[M_{n+1}]$ in terms of $\mathbb{E}[M_n]$. Solve your recurrence to obtain an expression for $\mathbb{E}[M_{n+1}]$. Do you think this is a fair game?

(c) Calculate $\mathbb{P}(M_n > 0)$. What is the behavior as $n \to \infty$? Would you still play this game?

Solution:

(a) Suppose that you have $M_n = m$ dollars at the end of day n. At the end of day $n+1$: with probability $1/4$, your fortune is $3m$; with probability $1/2$, your fortune is $m/2$; with probability $1/4$, your fortune is 0. Therefore,

$$\mathbb{E}[M_{n+1} \mid M_n = m] = \frac{1}{4} \cdot 3m + \frac{1}{2} \cdot \frac{m}{2} = m$$

and $\mathbb{E}[M_{n+1} \mid M_n] = M_n$.

(b) $\mathbb{E}[M_{n+1}] = \mathbb{E}[\mathbb{E}[M_{n+1} \mid M_n]] = \mathbb{E}[M_n]$. Therefore, we can see that

$$\mathbb{E}[M_{n+1}] = \mathbb{E}[M_n] = \mathbb{E}[M_{n-1}] = \cdots = \mathbb{E}[M_0] = M_0.$$

Your expected fortune never changes! Technically, we would be justified in calling this a fair game.

(c) The probability that your fortune is non-zero on day n is the probability that you never flipped two tails on a day, which has probability $(3/4)^n$. As $n \to \infty$, $(3/4)^n \to 0$, so after many days of playing the game, it is highly probable that we are broke!

Remark: In part (b), we have shown that $\mathbb{P}(M_n > 0) \to 0$ as $n \to \infty$, which means that M_n converges to 0 in probability. In part (a), we have shown that $\mathbb{E}[M_n] = M_0$ for all n. That is, $M_n \to 0$ but $\mathbb{E}[M_n] \not\to 0$.

7 Iterated Expectation

In this question, we will try to achieve more familiarity with the law of iterated expectation.

(a) You lost your phone charger! It will take D days for the new phone charger you ordered to arrive at your house (here, D is a random variable). Suppose that on day i, the amount of battery you lose is B_i, where $\mathbb{E}[B_i] = \beta$. Let $B = \sum_{i=1}^{D} B_i$ be the total amount of battery drained between now and when your new phone charger arrives. Apply the law of iterated expectation to show that $\mathbb{E}[B] = \beta\,\mathbb{E}[D]$. (Here, the law of iterated expectation has a very clear interpretation: the amount of battery you expect to drain is the average number of days it takes for your phone charger to arrive, multiplied by the average amount of battery drained per day.)

(b) Consider now the setting of independent Bernoulli trials, each with probability of success p. Let S_i be the number of successes in the first i trials. Compute $\mathbb{E}[S_m \mid S_n]$. (You will need to consider three cases based on whether $m > n$, $m = n$, or $m < n$. Try using your intuition rather than proceeding by calculations.)

Solution:

(a) This is simply Wald's Identity from lecture. Condition on $D = d$; then $B = \sum_{i=1}^{d} B_i$ and

$$\mathbb{E}[B \mid D = d] = \sum_{i=1}^{d} \mathbb{E}[B_i] = \beta d.$$

Therefore, $\mathbb{E}[B \mid D] = \beta D$ and by the law of iterated expectation, $\mathbb{E}[B] = \mathbb{E}[\mathbb{E}[B \mid D]] = \beta\,\mathbb{E}[D]$.

(b) Suppose $m > n$. Then we already know that the first n trials resulted in S_n successes, and there are $m-n$ trials for which we do not know the outcome. Each of these $m-n$ trials has probability of success p, so we expect $(m-n)p$ additional successes. Hence, $\mathbb{E}[S_m \mid S_n] = S_n + (m-n)p$.

Next, consider when $m = n$. Here, $\mathbb{E}[S_m \mid S_n] = S_n$.

Finally, suppose that $m < n$. In n trials, we have S_n successes, and due to symmetry, we expect the S_n successes to be distributed uniformly among the n trials. In particular, if we look at the first m trials only, then we expect a proportion m/n of the total successes to be distributed among the first m successes. Therefore, $\mathbb{E}[S_m \mid S_n] = mS_n/n$.

8 In the Moments

Suppose a random variable X satisfies the following conditions:

- $\mathbb{P}[X < 0] = \mathbb{P}[X > 1] = 0$
- $\mathbb{E}[X] \geq 2c$
- $\mathbb{E}\left[X^2\right] \leq \frac{19c^2}{4}$

Prove that

$$\mathbb{P}[X \geq c] \geq \max\left(c, \frac{1}{4}\right)$$

(Hint: Break the inequality into cases; conditional expectation may be helpful in one case)

Solution:

We wish to show that

1. $\mathbb{P}[X \geq c] \geq c$
2. $\mathbb{P}[X \geq c] \geq \frac{1}{4}$

1) Using conditional expectation,

$$\begin{aligned}
2c \leq \mathbb{E}[X] = \mathbb{E}[X \mid X < c]\,\mathbb{P}[X < c] + \mathbb{E}[X \mid X \geq c]\,\mathbb{P}[X \geq c] \\
\mathbb{P}[X \geq c] \geq \frac{2c - \mathbb{E}[X \mid X < c]\,\mathbb{P}[X < c]}{\mathbb{E}[X \mid X \geq c]} \\
\geq \frac{2c - c \cdot 1}{1} \\
= c
\end{aligned}$$

as desired.

2) Using Chebyshev,

$$\begin{aligned}
\mathbb{P}[X \geq c] &= 1 - \mathbb{P}[X < c] \\
&\geq 1 - \mathbb{P}[X < \mathbb{E}[X] - c] \\
&\geq 1 - \mathbb{P}[|X - \mathbb{E}[X]| \geq c] \\
&\geq 1 - \frac{\mathrm{Var}(X)}{c^2} \\
&= 1 - \frac{\mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2}{c^2} \\
&\geq 1 - \frac{\frac{19c^2}{4} - 4c^2}{c^2} \\
&= \frac{1}{4}
\end{aligned}$$

as desired.