1. (25 pts.) Bit Torrent
Consider the Bit Torrent problem described in Lecture 8, where a movie is broken down into $n$ chunks. Each server has a uniformly and independently selected chunk of the movie. I query the servers one-by-one and download whatever chunk I find on each server to my local hard disk. Suppose each server query takes 1 second and each chunk of the movie plays for $t$ seconds ($t$ integer). Answer each of these questions. None of the final answers should involve summations.

(a) (5 pts.) What is the expected time to get the first chunk of the movie so that I can start watching the movie? What is the distribution of this random variable?

**Answer:** Let $T_1$ be the random variable denoting the time to get the first chunk of the movie then $T_1 \sim \text{Geom}(1/n)$ which we know has mean given by $E[T_1] = n$

(b) (10 pts.) What is the expected time to get the first two chunks of the movie? (Hint: consider the first time to see either the first chunk or the second chunk of the movie.)

**Answer:** Let $\hat{T}_1 \sim \text{Geom}(2/n)$ be the time to get either the first or second chunk of the movie and $\hat{T}_2 \sim \text{Geom}(1/n)$ be the time to get the remaining chunk. The second random variable is geometric by the memoryless property of the geometric distribution. Our total time to collect the first two movies is $\hat{T} = \hat{T}_1 + \hat{T}_2$ so by linearity of expectation and the mean of a geometric random variable,

$$E[\hat{T}] = E[\hat{T}_1] + E[\hat{T}_2]$$

$$= \frac{n}{2} + n$$

$$= 3/2 \cdot n$$

(c) (10 pts.) I start playing the movie once I get the first chunk of the movie. What is the probability that I have the second chunk of the movie on my hard disk before I finish playing the first chunk so that I can immediately continue playing the movie?

**Answer:** Let $T_1$ and $T_2$ be the time until we collect the first and second chunks respectively then we want to calculate the following probability $P(T_2 - T_1 \leq t)$ where $t \in \mathbb{Z}_{++}$ (this notation means $t$ is a positive integer. With just one + means including 0). Using total probability we can write

$$P(T_2 - T_1 \leq t) = P(T_2 - T_1 \leq t, T_2 > T_1) + P(T_2 - T_1 \leq t, T_2 < T_1)$$

$$= P(T_2 - T_1 \leq t | T_2 > T_1)P(T_2 > T_1) + P(T_2 - T_1 \leq t | T_2 < T_1)P(T_2 < T_1)$$

By the symmetry of the problem we know $P(T_2 < T_1) = P(T_2 > T_1) = 1/2$. Furthermore, the quantity $P(T_2 - T_1 \leq t | T_2 < T_1) = 1$ since $t$ is defined to be a positive quantity. Plugging in we can write,

$$P(T_2 - T_1 \leq t) = 1/2 + P(T_2 - T_1 \leq t | T_2 > T_1)1/2$$

$$= 1/2 + 1/2 \cdot P(T_2 \leq t)$$
Now using the PMF of the geometric distribution and the sum of a geometric series we can simplify

\[
P(T_2 - T_1 \leq t) = 1/2 + 1/2 \sum_{i=1}^{t-1} (1 - 1/n)^i \frac{1}{n}
\]

\[
P(T_2 - T_1 \leq t) = 1/2 + 1/2 \cdot \frac{1}{n} \sum_{i=0}^{t-1} (1 - 1/n)^i
\]

\[
P(T_2 - T_1 \leq t) = 1/2 + 1/2 \cdot \frac{1 - (1 - 1/n)^t}{n - (1 - 1/n)}
\]

\[= 1 - 1/2 \cdot (1 - 1/n)^t\]

So our final answer is

\[P(T_2 - T_1 \leq t) = 1 - \frac{1}{2}(1 - 1/n)^t\]

2. (25 pts.) Estimating Variance

You have available \(n\) samples \(X_1, X_2, \ldots, X_n\), drawn independently from a pmf \(p_X\). Suppose you know the mean \(\mu\) but not the variance \(\sigma^2\) of the pmf. Find a way to estimate \(\sigma^2\) based on the data, such that as the number of samples increases, the estimate becomes more and more accurate. Make mathematically precise what you mean by "more and more accurate".

**Answer:** We can estimate the variance as

\[
\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2.
\]

Let \(Z_i = (X_i - \mu)^2\), then \(\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} Z_i\), and \(\mathbb{E}[Z_i] = \sigma^2\). Note that \(Z_i\) is an i.i.d. sequence. By the law of large numbers we know that for any \(\epsilon > 0\),

\[
P\left(\left|\frac{1}{n} \sum_{i=1}^{n} Z_i - \sigma^2\right| > \epsilon\right) \xrightarrow{n \to \infty} 0.
\]

Hence

\[
P(|\hat{\sigma}^2 - \sigma^2| > \epsilon) \xrightarrow{n \to \infty} 0.
\]

You can also argue that the estimator becomes more accurate by using Chebyshev’s inequality to show that the variance of the estimator goes to 0 as \(n \to \infty\) (it is essentially the same thing, since we used Chebyshev’s inequality to prove the LLN).

3. (25 pts.) Bounding Tail Events

The per bit error rate over a certain binary communication channel is \(10^{-10}\). No other statistics are known about the channel or the data.

(a) (10 pts.) What is the expected number of erroneous bits in a block of 1000 bits? Hint: Define the random variable \(X_i\) to be 1 if the \(i\)-th bit is in error and 0 otherwise. Thus \(P\{X_i = 1\} = 10^{-10}\). Now the number of bits in error in a block of 1000 bits is the random variable \(N = \sum_{i=1}^{1000} X_i\).

**Answer:**

As suggested, let the Bernoulli random variables \(X_1, \ldots, X_{1000}\) have the common pmf

\[
X_i = \begin{cases} 
1 & \text{with probability } 10^{-10} \\
0 & \text{with probability } 1 - 10^{-10}.
\end{cases}
\]
Then, $E(X_i) = 10^{-10}$. If $N$ is the number of errors in 1000 bits, then

$$E(N) = E\left(\sum_{i=1}^{1000} X_i\right) = \sum_{i=1}^{1000} E(X_i) = 1000E(X_i) = 1000 \cdot 10^{-10} = 10^{-7}.$$ 

(b) **(15 pts.)** Find an upper bound on the probability that a block of 1000 bits has 10 or more erroneous bits.

**Answer:**
The upper bound can be obtained from the Markov inequality:

$$P\{N \geq 10\} = P\{N \geq aE(N)\} \leq \frac{1}{a},$$

where $a = 10^8$ and $E(N) = 10^{-7}$. Therefore $P\{N \geq 10\} \leq 10^{-8}$. The maximum probability of $\geq 10$ errors occurs when the errors occur in groups of 10.

4. **(25 pts.) Gambling**

Let $X_n$ be the amount you win on the $n$th round of a game of chance. Assume that $X_1, X_2, \ldots, X_n$ are i.i.d. with finite mean $E(X)$ and variance $\sigma^2$. Make the realistic assumption that $E(X) < 0$. Show that $P\{(X_1 + X_2 + \ldots + X_n)/n < E(X)/2\} \rightarrow 1$. What is the moral of this result?

**Answer:**
We know that for $A_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ by the weak law of large numbers for any $\epsilon > 0$:

$$P(|A_n - E(X)| \geq \epsilon) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Note that $|A_n - E(X)| \geq \epsilon \implies A_n - E(X) \leq -\epsilon$ or $A_n - E(X) \leq -\epsilon$, as the events $\{A_n - E(X) \geq \epsilon\}$ and $\{A_n - E(X) \leq -\epsilon\}$ are non-intersecting, the probability of their union is the sum of probabilities of each event:

$$P(|A_n - E(X)| \geq \epsilon) = P(A_n - E(X) \geq \epsilon) + P(A_n - E(X) \leq -\epsilon) \geq P(A_n - E(X) \geq \epsilon) \geq 0.$$

Note that the LHS converges to 0 and the RHS is 0, so the value in between has to also converge to 0 (by the squeeze theorem), which means that

$$P(A_n - E(X) \geq \epsilon) \rightarrow 0.$$

Now, choosing $\epsilon = -\frac{E(X)}{2} > 0$ we get

$$P(A_n \geq E(X)/2) = P(A_n - E(X) \geq \epsilon) \rightarrow 0$$

$$P\{(X_1 + X_2 + \ldots + X_n)/n < E(X)/2\} = 1 - P(A_n > E(X)/2) \rightarrow 1$$

3