1 Sundry

Before you start your homework, write down your team. Who else did you work with on this homework? List names and email addresses. (In case of homework party, you can also just describe the group.) How did you work on this homework? Working in groups of 3-5 will earn credit for your "Sundry" grade.

Please copy the following statement and sign next to it:

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up. (Signature here)

2 How Many Kings?

Suppose that you draw 3 cards from a standard deck without replacement. Let X denote the number of kings you draw.

(a) What is $\Pr(X = 0)$?

(b) What is $\Pr(X = 1)$?

(c) What is $\Pr(X = 2)$?

(d) What is $\Pr(X = 3)$?

(e) Do the answers you computed in parts (a) through (d) add up to 1, as expected?

(f) Compute $\mathbf{E}(X)$ from the definition of expectation.

(g) Suppose we define indicators X_i, $1 \leq i \leq 3$, where X_i is the indicator variable that equals 1 if the ith card is a king and 0 otherwise. Compute $\mathbf{E}(X)$.

(h) Are the X_i indicators independent? How does this affect your answer to part (g)?

Solution:

(a) We must draw 3 non-king cards in a row, so the probability is

$$\Pr(X = 0) = \frac{48}{52} \cdot \frac{47}{51} \cdot \frac{46}{50} = \frac{4324}{5525}.$$

Alternatively, every 3-card hand is equally likely, so we can use counting. There are $\binom{52}{3}$ total 3-card hands, and $\binom{48}{3}$ hands with only non-king cards, which gives us the same result.

$$\Pr(X=0) = \frac{\binom{48}{3}}{\binom{52}{3}} = \frac{4324}{5525}$$

(b) We will continue to use counting. The number of hands with exactly one king amounts to the number of ways to choose 1 king out of 4, and 2 non-kings out of 48.

$$\Pr(X=1) = \frac{\binom{4}{1}\binom{48}{2}}{\binom{52}{3}} = \frac{1128}{5525}$$

(c) Choose 2 kings out of 4, and 1 non-king out of 48.

$$\Pr(X=2) = \frac{\binom{4}{2}\binom{48}{1}}{\binom{52}{3}} = \frac{72}{5525}$$

(d) Choose 3 kings out of 4.

$$\Pr(X=3) = \frac{\binom{4}{3}}{\binom{52}{3}} = \frac{1}{5525}$$

(e) We check:

$$\Pr(X=0)+\Pr(X=1)+\Pr(X=2)+\Pr(X=3) = \frac{4324+1128+72+1}{5525} = 1$$

(f) From the definition, $\mathbf{E}(X) = \sum_{k=0}^{3} k \Pr(X=k)$, so

$$\mathbf{E}(X) = 0 \cdot \frac{4324}{5525} + 1 \cdot \frac{1128}{5525} + 2 \cdot \frac{72}{5525} + 3 \cdot \frac{1}{5525} = \frac{3}{13}.$$

(g) We know that $\mathbf{E}(X_i) = \Pr(\text{card } i \text{ is an king}) = 1/13$, so

$$\mathbf{E}(X) = \mathbf{E}(X_1) + \mathbf{E}(X_2) + \mathbf{E}(X_3) = \frac{1}{13} + \frac{1}{13} + \frac{1}{13} = \frac{3}{13}.$$

Notice how much faster it was to compute the expectation using indicators!

(h) No, they are not independent. As an example:

$$\Pr(X_1=1)\Pr(X_2=1) = \frac{1}{13} \cdot \frac{1}{13} = \frac{1}{169}$$

However,

$$\Pr(X_1=1, X_2=1) = \Pr(\text{the first and second cards are both kings}) = \frac{4}{52} \cdot \frac{3}{51} = \frac{1}{221}.$$

Even though the indicators are not independent, this does not change our answer for part (g). Linearity of expectation *always* holds, which makes it an extremely powerful tool.

3 Sisters

Consider a family with n children, each with a 50% chance of being male or female. Let X be the total number of sisters that the male children have, and let Y be the total number of sisters that the female children have (for example, if $n = 3$ and there are two boys and one girl, then $X = 2$ and $Y = 0$). Find expressions for $\mathbf{E}(X)$ and $\mathbf{E}(Y)$ in terms of n. Do we expect that boys have more sisters or girls have more sisters?

[Hint: Define a random variable B to denote the number of boys, find an expression for X as a function of B, and apply linearity of expectation. Use a similar approach for girls.]

Solution:

Let B be the number of boys. Each boy has $n - B$ sisters, so boys have $B(n - B)$ sisters in total. Using linearity of expectation,

$$\begin{aligned}
\mathbf{E}(X) &= \mathbf{E}(B(n-B)) \\
&= \mathbf{E}(Bn - B^2) \\
&= n\mathbf{E}(B) - \mathbf{E}(B^2).
\end{aligned}$$

Note that B follows the distribution $\text{Bin}(n, p)$ where $p = 1/2$. So $\mathbf{E}(B) = np = n/2$ and $\text{Var}(B) = np(1-p) = n/4$. From the definition of variance, we also have

$$\begin{aligned}
\mathbf{E}(B^2) - \mathbf{E}(B)^2 &= \text{Var}(B) \\
\mathbf{E}(B^2) &= \text{Var}(B) + \mathbf{E}(B)^2 \\
&= \frac{n}{4} + \left(\frac{n}{2}\right)^2 \\
&= \frac{n}{4} + \frac{n^2}{4}.
\end{aligned}$$

Substituting into our expression for $\mathbf{E}(X)$, we obtain

$$\mathbf{E}(X) = n\mathbf{E}(B) - \mathbf{E}(B^2) = n \cdot \frac{n}{2} - \frac{n}{4} - \frac{n^2}{4} = \frac{n^2}{4} - \frac{n}{4} = \frac{n(n-1)}{4}.$$

Let G be the number of girls. Each girl has $G - 1$ sisters, so girls have $G(G-1)$ sisters in total. Using linearity of expectation,

$$\begin{aligned}
\mathbf{E}(Y) &= \mathbf{E}(G(G-1)) \\
&= \mathbf{E}(G^2 - G) \\
&= \mathbf{E}(G^2) - \mathbf{E}(G).
\end{aligned}$$

Note that G follows exactly the same distribution as B, so $\mathbf{E}(G^2) = n^2/4 + n/4$ and $\mathbf{E}(G) = n/2$. Thus

$$\mathbf{E}(Y) = \mathbf{E}(G^2) - \mathbf{E}(G) = \frac{n^2}{4} + \frac{n}{4} - \frac{n}{2} = \frac{n^2}{4} - \frac{n}{4} = \frac{n(n-1)}{4}.$$

We expect boys and girls to have the same number of sisters!

4 Unbiased Variance Estimation

We have a random variable X and want to estimate its variance, σ^2 and mean, μ, by sampling from it. In this problem, we will derive an "unbiased estimator" for the variance.

(a) We define a random variable Y that corresponds to drawing n values from the distribution for X and averaging, or $Y = (X_1 + \ldots + X_n)/n$. What is $\mathbf{E}(Y)$? Note that if $\mathbf{E}(Y) = \mathbf{E}(X)$ then Y is an unbiased estimator of $\mu = \mathbf{E}(X)$. *Hint: This should not be difficult.*

(b) Now let's assume the actual mean is 0 as variance doesn't change when one shifts the mean. Before attempting to define an estimator for variance, show that $\mathbf{E}(Y^2) = \sigma^2/n$.

(c) In practice, we don't know the mean of X so following part (a), we estimate it as Y. With this in mind, we consider the random variable $Z = \sum_{i=1}^n (X_i - Y)^2$. What is $\mathbf{E}(Z)$?

(d) What is a good unbiased estimator for the $\mathrm{Var}(X)$?

(e) How does this differ from what you might expect? Why? (Just tell us your intuition here, it is all good!)

Solution:

(a) By linearity of expecation, the value is $(\sum_{i=1}^n \mathbf{E}(X_i))/n = \mathbf{E}(X)$.

(b) The variables X_i are independent, so

$$\mathbf{E}(Y^2) = \mathbf{E}((Y - \mathbf{E}(Y))^2) = \mathrm{var}\left(\frac{1}{n}\left(\sum_{i=1}^n X_i\right)\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{var}(X_i) = \frac{\sigma^2}{n}.$$

The first equality follows the fact that $\mathbf{E}(Y) = \mathbf{E}(X) = 0$, the second from the definition of variance, the third from linearity of variance for independent variables, and the others by substitution.

(c)

$$\begin{aligned}
\mathbf{E}(Z) &= \sum_{i=1}^n (\mathbf{E}(X_i^2) - \mathbf{E}(2YX_i) + \mathbf{E}(Y^2)) \\
&= (n+1)\sigma^2 - 2\sum_{i=1}^n \mathbf{E}(X_i Y) \\
&= (n+1)\sigma^2 - \frac{2}{n}\left(\sum_{i=1}^n \mathbf{E}(X_i^2) + \sum_{j \neq i} \mathbf{E}(X_i X_j)\right) \\
&= (n-1)\sigma^2 - \frac{2}{n}\left(\sum_{j \neq i} \mathbf{E}(X_i X_j)\right) \\
&= (n-1)\sigma^2
\end{aligned}$$

The first equality is plugging in definition of Z and uses linearity of expectation. The second line uses $\mathbf{E}(X_i^2) = \sigma^2$ and $\mathbf{E}(Y^2) = \sigma^2/n$. The third line plugs in the definition of Y and uses linearity of expectation. The fourth again uses $\mathbf{E}(X_i^2) = \sigma^2$. The final line follows from $\mathbf{E}(X_iX_j) = \mathbf{E}(X_i)\mathbf{E}(X_j) = 0$ since the X_i are chosen independently and have expectation 0.

(d) $Z/(n-1)$, since $\mathbf{E}(Z/(n-1)) = \sigma^2$.

(e) Maybe one could guess Z/n since there are n terms in Z. But in fact, each term is a bit smaller than expected as Y contains a bit of X_i/n in it. So a term, $(X_i - Y)^2$ is actually

$$\left(\frac{n-1}{n}X_i - \frac{1}{n}\sum_{i\neq j} X_j\right)^2,$$

so it is a bit smaller than the variance of X_i.

5 Markov Bound for Coupon Collectors

Suppose you are trying to collect a set of n different baseball cards. You get the cards by buying boxes of cereal: each box contains exactly one card, and it is equally likely to be any of the n cards. You are interested in finding m, a lower bound on the number of boxes you should buy to ensure that the probability of you collecting all n cards is at least $\frac{1}{2}$. In class, we used the Union Bound to show that it suffices to have $m \geq n\ln(2n)$.

Use Markov's Inequality to find a different (weaker) lower bound on m.

Solution:

Let X be the random variable denoting the number of cereal boxes you need to buy before completing your set of n cards. We desire m, the number we need to buy to guarantee success with probability at least $1/2$. Then we want

$$\Pr(X > m) \leq \frac{1}{2}.$$

Note that $\Pr(X > m) \leq \Pr(X \geq m)$ and

$$\Pr(X \geq m) \leq \frac{\mathbf{E}(X)}{m},$$

by Markov's Inequality. Therefore it would suffice to satisfy

$$\begin{aligned} \frac{\mathbf{E}(X)}{m} &\leq \frac{1}{2}, \\ \frac{n(\ln n + \gamma)}{m} &\leq \frac{1}{2}, \\ m &\geq 2n(\ln n + \gamma), \end{aligned}$$

where we used $\mathbf{E}(X) \approx n(\ln n + \gamma)$ for the expectation of the coupon collector's problem from Note 19 and $\gamma \approx 0.5772$ is the Euler constant.

6 Safeway Monopoly Cards

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.

(a) Let X be the number of visits you have to make before you can redeem the grand prize. Show that $\text{Var}(X) = n^2 \left(\sum_{i=1}^{n} i^{-2}\right) - \mathbf{E}(X)$. *[Hint: Does this remind you of a particular problem? What is the expectation for this problem?]*

(b) 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 β for which the probability you need to make more than $\mathbf{E}(X) + \beta n$ 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.]*

Solution:

(a) Let X_i be the number of visits we need to make before we have collected the ith unique Monopoly card actually obtained, given that we have already collected $i-1$ unique Monopoly cards. Then $X = \sum_{i=1}^{n} X_i$ and each X_i is geometrically distributed with $p = (n-i+1)/n$. Then

$$
\begin{aligned}
\text{Var}(X) &= \sum_{i=1}^{n} \text{Var}(X_i) && \text{(as the } X_i \text{ are independent)} \\
&= \sum_{i=1}^{n} \frac{1-(n-i+1)/n}{[(n-i+1)/n]^2} && \text{(variance of a geometric r.v. is } (1-p)/p^2\text{)} \\
&= \sum_{j=1}^{n} \frac{1-j/n}{(j/n)^2} && \text{(by noticing that } n-i+1 \text{ takes on all values from 1 to } n\text{)} \\
&= \sum_{j=1}^{n} \frac{n(n-j)}{j^2} \\
&= \sum_{j=1}^{n} \frac{n^2}{j^2} - \sum_{j=1}^{n} \frac{n}{j} \\
&= n^2 \left(\sum_{j=1}^{n} \frac{1}{j^2} \right) - \mathbf{E}(X) && \text{(using the coupon collector problem expected value).}
\end{aligned}
$$

(b) We are looking for the smallest value of β for which we can say that $\Pr[X \geq \mathbf{E}(X) + \beta n] < 1/100$ for all values of n. We have:

$$\begin{aligned}
\Pr[X \geq \mathbf{E}(X) + \beta n] &= \Pr[X - \mathbf{E}(X) \geq \beta n] \\
&\leq \Pr[|X - \mathbf{E}(X)| \geq \beta n] \\
&\leq \frac{\text{Var}(X)}{(\beta n)^2} \qquad \text{(by Chebyshev's inequality)} \\
&= \frac{n^2 \sum_{i=1}^{n} i^{-2} - \mathbf{E}(X)}{(\beta n)^2} \\
&\sim \frac{n^2 \sum_{i=1}^{n} i^{-2} - n \ln n}{(\beta n)^2} \\
&= \frac{\sum_{i=1}^{n} i^{-2}}{\beta^2} - \frac{\ln n}{n \beta^2}
\end{aligned}$$

Therefore, we desire a lower bound on β such that the following is satisfied:

$$\frac{\sum_{i=1}^{n} i^{-2}}{\beta^2} - \frac{\ln n}{n\beta^2} < \frac{1}{100}$$

But as $n \to \infty$, the second term approaches zero, since n grows faster than $\ln n$, and the first term approaches $\pi^2/6$. Therefore, we simply need to satisfy

$$\frac{\pi^2}{6\beta^2} < \frac{1}{100}s.$$

This requires $\beta > 10\pi/\sqrt{6} \approx 12.825$.

7 Dice

In this problem, let $X_1, X_2, \ldots X_n$ each denote the outcomes of standard six-sided dice rolls. Let A denote the average of the outcomes $(\sum_{i=1}^{n} X_i)/n$.

(a) For $n = 100$, find some a and b such that A is in the interval $[a, b]$ with probability at least 90% (Don't use trivial intervals like [1, 6]).

(b) For $n = 30$, find a lower bound on $\Pr[3 \leq A \leq 4]$.

(c) Find the minimum n for which you can guarantee that A is within the range [3, 4] with at least 99% probability.

Solution:

(a) A is the average of a 100 dice rolls. We want to find an interval $[a, b]$ such that A falls outside this with probability ≤ 0.1.

In order to use Chebyshev's inequality here, we will find an interval centered at the expectation of A and bound its end points as needed, i.e., we will find a deviation ε to set $a = \mathbf{E}[A] - \varepsilon$ and $b = \mathbf{E}[A] + \varepsilon$ such that $\Pr(A \notin [a,b]) \le 0.1$.

So, we're just trying to find an ε such that $\Pr(A \notin [\mathbf{E}[A] - \varepsilon, \mathbf{E}[A] + \varepsilon]) \le 0.1$. Applying Chebyshev's inequality, we get:

$$\Pr(|A - \mathbf{E}[A]| > \varepsilon) \le \Pr(|A - \mathbf{E}[A]| \ge \varepsilon) \le \frac{\operatorname{var}(A)}{\varepsilon^2} \le 0.1$$

Now, we need $\operatorname{var}(A)$. We have:

$$\begin{aligned}
\operatorname{var}(A) &= \operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\operatorname{var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{var}(X_i) \\
&= \frac{1}{n^2} n \cdot \operatorname{var}(X_1) \\
&= \frac{1}{n}\left(\mathbf{E}\left[X_1^2\right] - \mathbf{E}[X_1]^2\right) \\
&= \frac{1}{n}\left(\frac{1}{6}[1+4+9+16+25+36] - \left(\frac{1}{6}[1+2+3+4+5+6]\right)^2\right) \\
&= \frac{1}{100}\cdot\frac{35}{12} = \frac{7}{240}.
\end{aligned}$$

Thus, for $\operatorname{var}(A)/\varepsilon^2 \le 0.1$, we want $\varepsilon^2 \ge \operatorname{var}(A)/0.1 = 7/24$. This gives $\varepsilon \approx 0.54$.

Now, for calculating our final interval, we need to find $\mathbf{E}[A]$. Using linearity of expectation, we have $\mathbf{E}[A] = \mathbf{E}[X_i] = 3.5$.

So, our interval would be $[\mathbf{E}[A] - \varepsilon, \mathbf{E}[A] + \varepsilon]$ which is $[2.96, 4.04]$.

(b) Now, we're given $\varepsilon = 0.5$ (the deviation from the mean 3.5) and we want to calculate a lower bound on the probability that A lies in $[3,4]$. Let's do this by calculating a upper bound on the probability that A lies outside the interval.

Applying Chebyshev's inequality here, we get:

$$\Pr(|A - \mathbf{E}[A]| \ge 0.5) \le \frac{\operatorname{var}(A)}{0.5^2}$$

This gives the upper bound of $\frac{\operatorname{var}(A)}{0.5^2} = \frac{1}{30}\cdot\frac{35}{12}\cdot\frac{1}{0.25} \approx 0.39$ where we use the same variance calculation as part (a).

So, the probablity $\Pr[A \in [3,4]] = 1 - [A \notin [3,4]] \ge 1 - 0.39 = 0.61$.

(c) This is similar to the previous parts, but here we want to find n so that we are 99% confident of A falling in $[3,4]$. Applying Chebyshev's inequality again, we get:

$$\Pr(|A - \mathbf{E}[A]| \geq 0.5) \leq \frac{\operatorname{var}(A)}{0.5^2} \leq 1 - 0.99 = 0.01$$

From the previous parts, we know that $\operatorname{var}(A) = \dfrac{35}{12n}$. So, we want $\dfrac{35}{12n} \cdot \dfrac{1}{0.5^2} \leq 0.01$.

This gives us $n \geq \dfrac{3500}{3}$. So, we want n to be 1167 or greater.

8 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 poor college student, you also don't have any income. How much should you have in your bank account if you don't want to go broke 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 1000 packets are 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, $\mathbf{E}[T] = 12 \cdot \mathbf{E}[E_1] = 18000$.

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

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.

$$\Pr\left(|T - \mathbf{E}[T]| \geq \varepsilon\right) \leq \frac{\operatorname{var}(T)}{\varepsilon^2} \leq 0.05$$

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

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

$$\Pr\left(|T - \mathbf{E}[T]| \geq \varepsilon\right) \geq \Pr\left(T - \mathbf{E}[T] \geq \varepsilon\right),$$

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

And $\operatorname{var}(X_1) = 150^2 p(1-p)$ as this is just a shifted and scaled binomial distribution. Using $\mathbf{E}\left[X_1^2\right] - \mathbf{E}[X_1]^2$ yields the same answer.

We have, $\mathbf{E}[T] = 10 \cdot \mathbf{E}[X_1]$. Similarly, $\operatorname{var}(T) = 10 \cdot \operatorname{var}(X_1)$.

Now, we want to bound the probability of T going below 0 by 0.05. In other words, we want $\Pr(T < 0) \leq 0.05$.

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 $\mathbf{E}[T] > 0$. Then:

$$\begin{aligned}
\Pr(T < 0) &\leq \Pr(T \leq 0 \cup T \geq 2\mathbf{E}[T]) \\
&= \Pr(|T - \mathbf{E}[T]| \geq \mathbf{E}[T]) \\
&\leq \frac{\operatorname{var}(T)}{\mathbf{E}[T]^2} \leq 0.05
\end{aligned}$$

Looking at just the last inequality, we have:

$$\frac{\operatorname{var}(T)}{\mathbf{E}[T]^2} = \frac{10 \cdot \operatorname{var}(X_1)}{100 \cdot \mathbf{E}[X_1]^2} = \frac{\operatorname{var}(X_1)}{10 \cdot \mathbf{E}[X_1]^2} \leq 0.05$$

$$\therefore \frac{\operatorname{var}(X_1)}{\mathbf{E}[X_1]^2} \leq 0.5$$

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

$$\begin{aligned}
\Pr(T < 1000) &\leq \Pr(T \leq 1000 \cup T \geq \mathbf{E}[T] + (\mathbf{E}[T] - 1000)) \\
&= \Pr(|T - \mathbf{E}[T]| \geq (\mathbf{E}[T] - 1000)) \\
&\leq \frac{\operatorname{var}(T)}{(\mathbf{E}[T] - 1000)^2} \leq 0.01.
\end{aligned}$$

What is $\mathbf{E}[T]$? $\mathbf{E}[T] = n\mathbf{E}[X_1] = n(1-p) = 0.2n$.
Next, what is $\operatorname{var}(T)$? $\operatorname{var}(T) = n\operatorname{var}(X_1) = np(1-p) = 0.16n$.

Now, $\frac{\operatorname{var}(T)}{(\mathbf{E}[T] - k)^2} \leq 0.01 \implies 16n \leq (0.2n - 1000)^2$. 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 $n \geq 6629$.

9 Entropy

Let X_i, $1 \leq i \leq n$, be a sequence of i.i.d. Bernoulli random variables with parameter p, i.e. $\Pr(X_i = 1) = p$ and $\Pr(X_i = 0) = 1 - p$.

(a) Express $\Pr(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$ in terms of p and n_0, where $x_i \in \{0, 1\}$ for all $1 \leq i \leq n$ and n_0 depends on $(x_1 \ldots x_n)$ and represents the number of 1's in $(x_1 \ldots x_n)$. Call this result $f(x_1, x_2, \ldots, x_n)$.

(b) Define the random variable Y_n as $Y_n = -n^{-1}\log f(X_1, X_2, \dots, X_n)$. What does Y_n tend to as n grows large?

Solution:

(a) Since the X_i are i.i.d. Bernoulli random variables,

$$f(x_1, x_2, \dots x_n) = \Pr(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

$$= \Pr(X_1 = x_1) \cdot \Pr(X_2 = x_2) \cdots \Pr(X_n = x_n).$$

We know that $\Pr(X_i = x_i) = p$ if $x_i = 1$ and $(1-p)$ otherwise.

$$\therefore f(x_1, x_2, \dots x_n) = p^{\text{\# of 1's in } (x_1, x_2, \dots, x_n)} \times (1-p)^{\text{\# of 0's in } (x_1, x_2, \dots, x_n)} = p^{n_0}(1-p)^{n-n_0}$$

(b) We have

$$\begin{aligned} Y_n &= -\frac{1}{n}\log f(X_1, X_2, \dots, X_n) \\ &= -\frac{1}{n}\log(p^{n_0} \cdot (1-p)^{n-n_0}) \\ &= -\frac{1}{n}(n_0 \log p + (n-n_0)\log(1-p)). \end{aligned}$$

By the Weak Law of Large Numbers, as n becomes very large, the number of ones n_0 moves closer and closer to $\mathbf{E}(n_0) = np$. Substituting that above,

$$\begin{aligned} Y_n &= -\frac{1}{n}(n_o \log p + (n-n_o)\log(1-p)) \\ &\approx -\frac{1}{n}(np\log p + (n-np)\log(1-p)) \\ &= -\frac{1}{n}(np\log p + n(1-p)\log(1-p)) \\ &= -(p\log p + (1-p)\log(1-p)) = -p\log p - (1-p)\log(1-p). \end{aligned}$$

This is called the binary entropy function $H(p)$.