EECS 70 Discrete Mathematics and Probability Theory
Fall 2014 Anant Sahai

# Homework 13

## This homework is due December 3, 2014, at 12:00 noon.

1. **Section Rollcall!**

   In your self-grading for this question, give yourself a 10, and write down what you wrote for parts (a) and (b) below as a comment. Put the answers in your written homework as well.

   (a) What discussion did you attend on Monday last week? If you did not attend section on that day, please tell us why.

   (b) What discussion did you attend on Wednesday last week? If you did not attend section on that day, please tell us why.

2. **Practice Makes Perfect**

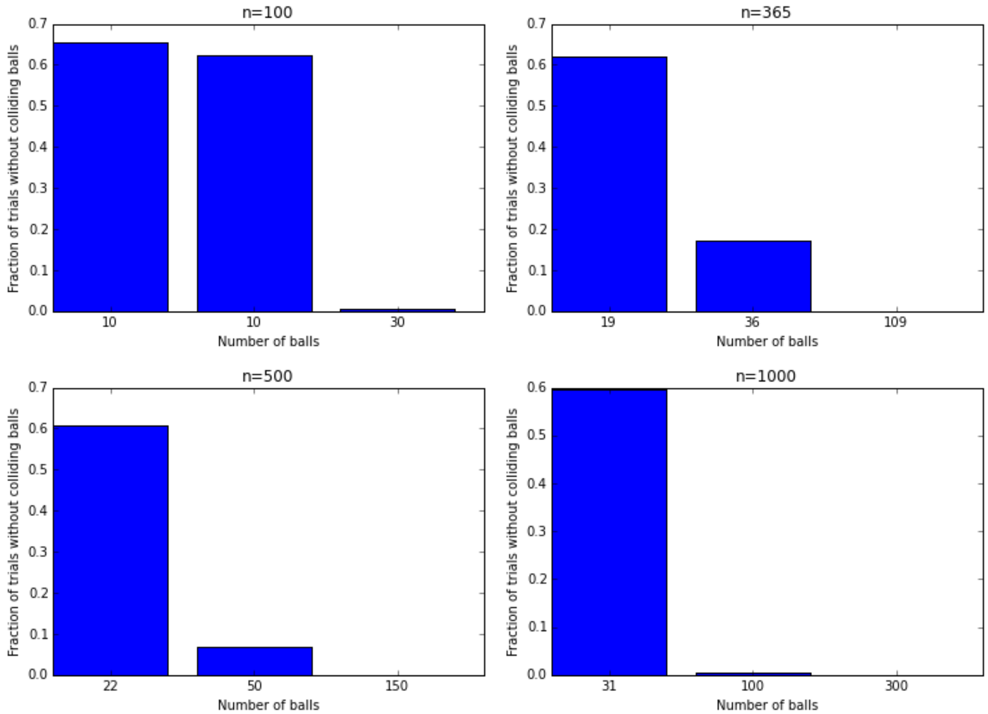
   For this question, do 5 of the online practice problems. For your answer, write down which problems you did (the problem set title and the number of the question).

3. **Hashing & Drunk Man Lab**

   Please complete the Virtual Lab that was released in Homework 12. For consistency, rename the lab file to `lab13.ipynb`, then zip and submit it on the instructional server as `hw13.zip`.

   (a) For $n = 100, 365, 500, 1000$, and for $m = 0.1n, 0.3n, \sqrt{n}$, simulate throwing $m$ balls into $n$ bins. For every $(m,n)$ pair, do this for 1000 trials. For each value of $n$, plot the fraction of trials for which there is no collision vs. $m$. What do you observe as $m$ gets larger?

   **Solutions:** As the number of balls $m$ increases, it is more likely for a collision to happen. This is very similar to the Birthday Paradox, where the chance of at least two people having the same birthday increases as the number of people increases.




(b) Let's take a detour and look at a very useful approximation for $ln(1-x)$. Plot $ln(1-x)$ for 1000 values of $x$ from $0.01-0.1$. What is the approximate slope of this graph and why? (think about Taylor expansion.)

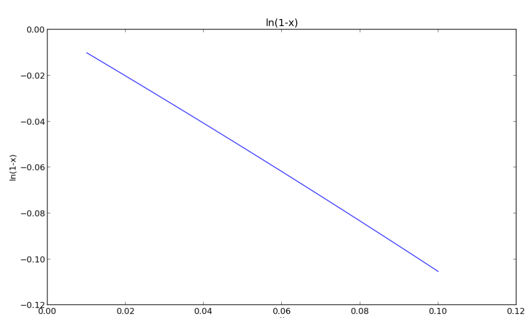
**Solutions:** The slope is $\approx -1$. This is because the Taylor expansion of $ln(1-x)$ is

$$ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} \ldots$$

and when $x$ is small, we expect to see

$$ln(1-x) \approx -x$$

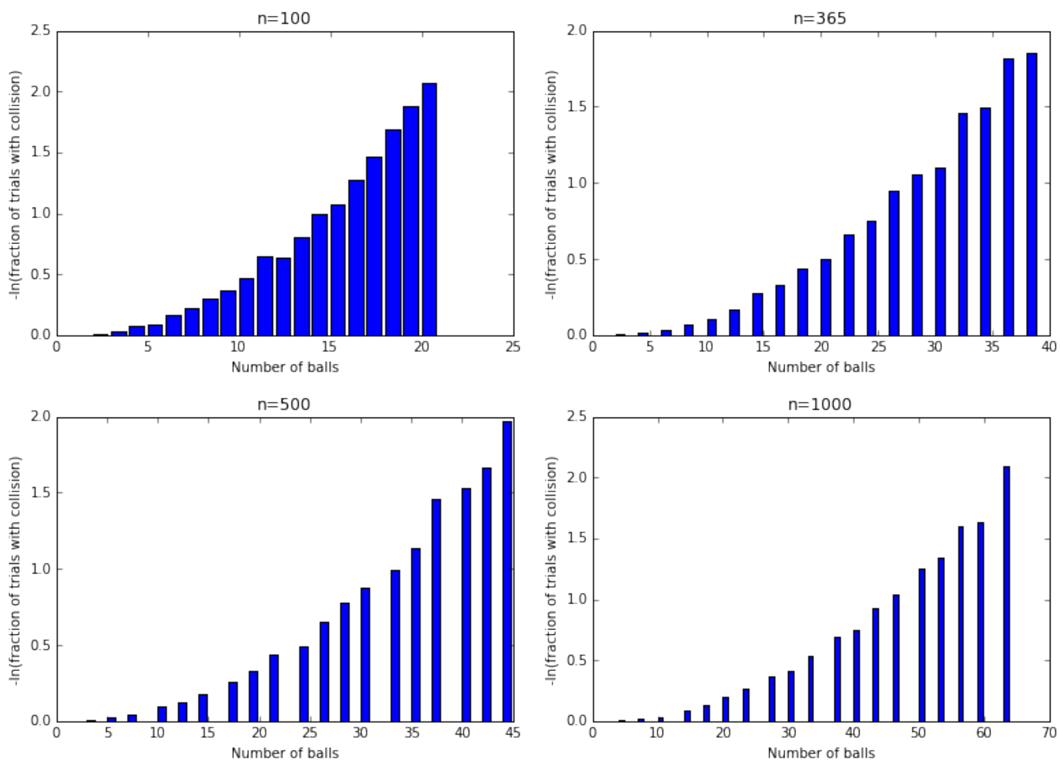
, as seen from the graph below.




(c) Repeat simulating throwing $m$ balls in $n$ bins for $n = 100, 365, 500, 1000$, and 20 values of $m$ between 1 and $2\sqrt{n}$. Now for each $n$, plot the negative logarithm of the fraction of trials for which there were no collisions, for different values of $m$. Does this plot look linear in $m$? Quadratic? Exponential?

**Solutions:** This looks quadratic, since

$$\begin{aligned}
P(A) &= \left(1-\frac{0}{n}\right)\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{m-1}{n}\right) \\
\Rightarrow ln(P(A)) &= ln\left(1-\frac{0}{n}\right)+ln\left(1-\frac{1}{n}\right)+\cdots+ln\left(1-\frac{m-1}{n}\right) \\
&\approx \left(-\frac{0}{n}\right)+\left(-\frac{1}{n}\right)+\cdots+\left(-\frac{m-1}{n}\right) \\
&= -\frac{m(m-1)}{2n} \\
&\approx -\frac{m^2}{2n}
\end{aligned}$$




(d) Now overlay the plots in part (c) with $\frac{m^2}{2n}$. What do you notice?
From these graphs, and for each value of $n$, find the approximate $m$ value such that $P(A) \approx 0.5$ for both the simulated data and $\frac{m^2}{2n}$. How much does this differ?

**Solutions:** The $\frac{m^2}{2n}$ curve touches each bar nicely near the top. It looks like an upper bound for the negative logarithm of the fraction of collision.

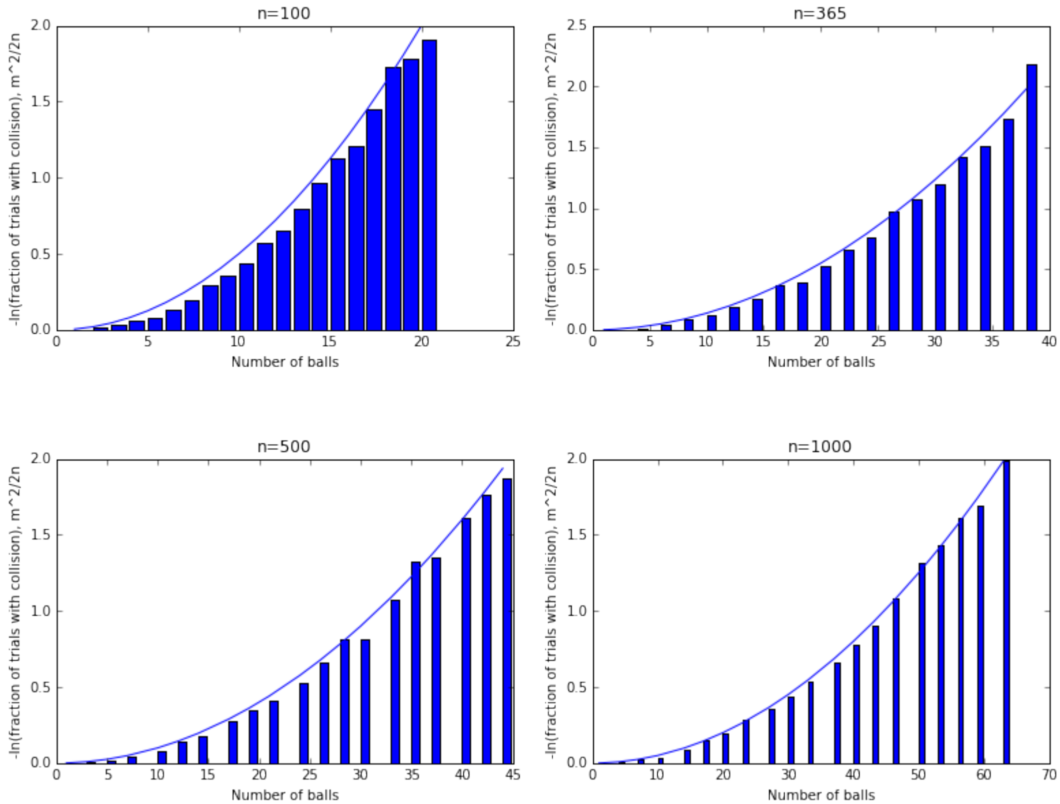
The values for which $P(A) \approx 0.5$ are:

For $n = 100$, $m = 12$ from the trials and $m = 12$ from the calculations.

For $n = 365$, $m = 22$ from the trials and $m = 22$ from the calculations.

For $n = 500$, $m = 26$ from the trials and $m = 26$ from the calculations.

For $n = 1000$, $m = 37$ from the trials and $m = 37$ from the calculations.

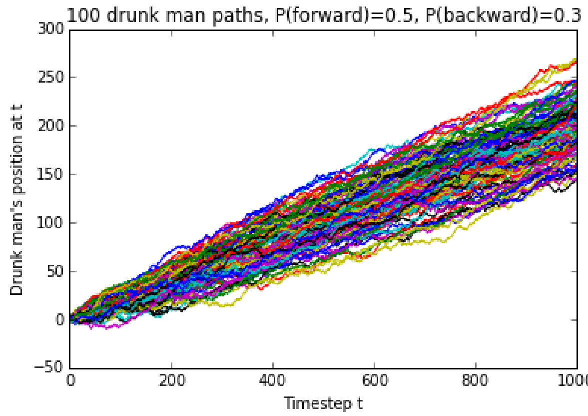



(e) The drunk man is back, and he wants your help to plot 100 sample paths that he could take from time $t = 0$ to $t = 999$ (1000 timesteps). Assume the same probabilities from Homework 11's Question 5, that is, the man moves forward with probability 0.5, backward with probability 0.3, and stays exactly where he is with probability 0.2. What do you observe about his paths? Where do you think he should end up at after 1000 timesteps, on average?

*Hint*: Implement the function `drunk_man`, which returns a list of elements that starts at 0, and every element thereafter is one more, one less, or equal to the previous one, with the correct probability for each possibility.

**Solutions:** All the paths are moving toward the top right corner of the plot. This makes sense because the probability of the drunk man moving forward is higher than the probability of him moving backward.

We can say that at every timestep, the drunk man is expected to increase his distance by $0.5 - 0.3 = 0.2$, so after 1000 timesteps, we expect him to be around 200 (and indeed, the ending locations of most of his paths are around this value).




(f) Redo the previous part, but this time, the man moves forward with probability 0.3, backward with probability 0.5, and stays exactly where he is with probability 0.2. What do you observe about his paths? Where do you think he should end up at after 1000 timesteps, on average?

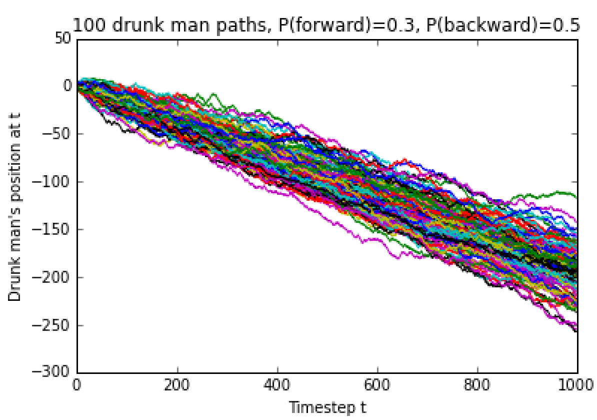
Finally, assume the man moves forward and backward with probability 0.5 each, i.e. he never stays still. Plot his 100 sample paths. What do you observe? Does this graph remind you of something we have done in a previous virtual lab?

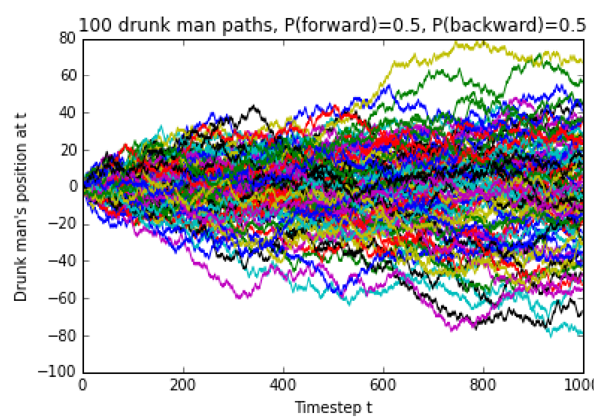
**Solutions:**

For the first case, the plot is the complete opposite of the one in the previous part. All the paths are going toward the lower right corner, and this again makes sense because the probability of moving backward is higher than the probability of moving forward.

For the second case, we see that the drunk man's paths are somewhat random, because he has the same probability of moving forward and moving backward. This is exactly the same plot that we did in part (a) of VL 9, where we plot 100 random fair coin tosses sequences.

The drunk man's position at each timestep is in fact just a biased coin toss with three outcomes: head (forward), tail (backward), and landing on the edge (staying still).




(g) One simple way of solving counting problems is to enumerate all possibilities, and then count the ones that we are looking for. In this question, let's check your answer to Question 5 from the last homework. Implement the function `count_paths(t)`, which generates all the paths the Drunk Man can take in $t$ timesteps, then counts the number of paths in which he returns to 0 at time $t$ and it is his first return. Remember that we no longer care about probabilities when counting paths.

*Hint*: One way to generate all possible paths is to use `itertools.product`. You might also want to implement the function `catalan(n)`, which computes the $n^{th}$ Catalan number.

**Solutions:** See *lab13sol.pdf*. We simply generate all possible paths in $t$ timesteps, and then for each path, check whether the Drunk Man returns to 0 at time $t$ and it is his first return.

(h) The Bernoulli distribution is the probability distribution of a random variable which takes value 1 with success probability $p$ and value 0 with failure probability $1-p$. It can be used, for example, to represent a coin toss, where 1 is defined to mean "heads" and 0 is defined to mean "tails".

In other words:

$$f(k;p) = \begin{cases} p & \text{if } k=1 \\ 1-p & \text{if } k=0 \end{cases}$$

Related values:

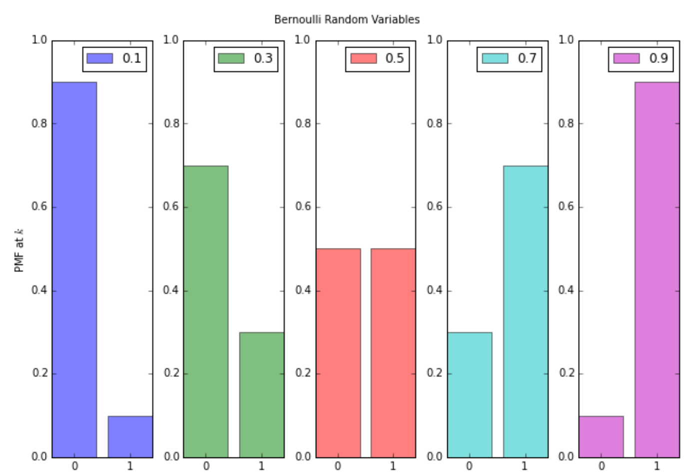
$$E[X] = p$$

$$Var(X) = p(1-p)$$

Plot the probability mass functions (pmf) for Bernoulli random variables with success probabilities of $0.1, 0.3, 0.5, 0.7$, and $0.9$, respectively.

*Hint*: `scipy.stats.bernoulli.pmf` will be useful for this question.

**Solutions:**




(i) The binomial distribution gives us the probability of observing $k$ successes, each with a probability $p$, out of $N$ attempts.

$$f(k;N;p) = \binom{N}{k} p^k (1-p)^{N-k}$$

where

$$\binom{N}{k} = \frac{N!}{k!(N-k)!}$$

with $k = 0, 1, 2, \ldots, N$

Related values:

$$E[X] = Np$$

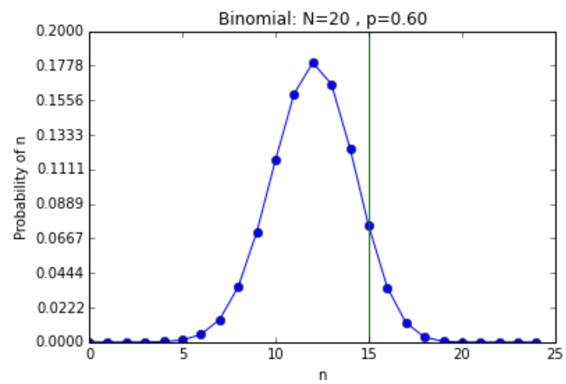
$$Var(X) = Np(1-p)$$

If I toss a 0.6-biased coin 20 times, what is the approximate probability of getting exactly 15 heads? Answer this question graphically by plotting the PMF of a binomial random variable with parameters $(20, 0.6)$.

Looking at the plot, at which value of $n$ is the probability of getting $n$ heads maximized?

*Hint*: As before, you will find `scipy.stats.binom.pmf` useful.

**Solutions:** Looking at the plot, the approximate probability of getting exactly 15 heads is about 0.07. We also see that the probability is maximized at $n = 12$, which happens to be the expected number of heads ($20 \times 0.6 = 12$). This is no coincidence, as we will see later in the course.




**4. Law of Large Numbers**

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

$$\lim_{n\to\infty} \Pr\left[\left|\tfrac{1}{n}S_n - \mathbb{E}(\tfrac{1}{n}S_n)\right| > \varepsilon\right] = 0.$$

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 $p$ 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, and give a brief justification of your answer. (Whenever convenient, you can assume that $n$ is even.)

Intuitively, what is this problem asking? In the context of LLN, this is asking as I increase my $n$, does the fraction of successful packets sent approach $1-p$, the success probability.

(a) YES or NO: Each packet is sent on a completely different route.

**Solutions:** 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 $A_n = \frac{S_n}{n}$ to be the fraction of successful packets sent, out of the $n$ packets. Moreover, for each $X_i$,

$$E[X_i] = 1 - p$$

and

$$Var[X_i] = p(1-p).$$

Using Chebyshev's inequality,

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

(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).

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

Now for each $i = 1, \ldots, \frac{n}{2}$

$$E[X_i] = 2(1-p)$$

and

$$Var[X_i] = 4p(1-p).$$

Thus,

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

and

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

Finally, we get

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

(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.

**Solutions:** No. In this situation, we have

$$X_i = \begin{cases} 0 & \text{with probability } p \\ \frac{n}{2} & \text{with probability } (1-p) \end{cases}$$

for $i = 1, 2$. Now $S_n = X_1 + X_2$ and $A_n = \frac{X_1 + X_2}{2}$.

We have

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

and

$$Var[X_i] = \frac{n^2}{4}p(1-p).$$

Thus,

$$E[A_n] = \frac{E[X_1]+E[X_2]}{n} = \frac{1}{n}n(1-p) = 1-p$$

and

$$\text{Var}[A_n] = \frac{1}{n^2}\left(\text{Var}[X_1]+\text{Var}[X_2]\right) = \frac{1}{n^2}\cdot\frac{n^2}{2}p(1-p) = \frac{p(1-p)}{2}.$$

Finally, we get

$$\begin{aligned}&Pr[|A_n - E[A_n]| > \varepsilon]\\ =&Pr[|A_n-(1-p)| > \varepsilon] \leq \frac{p(1-p)}{2\varepsilon^2}\end{aligned}$$

that does not converge to 0 as $n \to \infty$, so the Law of Large Numbers does not hold.

(d) YES or NO: All the packets are sent on one route.

**Solutions:** No. $S_n = X_1$, where $X_1 = n$ with probability $1-p$ and $X_1 = 0$ with probability $p$. $A_n = \frac{X_1}{n}$. Thus,

$$E[A_n] = \frac{E[X_1]}{n} = \frac{n(1-p)}{n} = 1-p$$

and

$$\text{Var}[A_n] = \frac{1}{n^2}\text{Var}[X_1] = \frac{1}{n^2}\cdot n^2 p(1-p) = p(1-p).$$

The inequality results in

$$\begin{aligned}&Pr[|A_n - E[A_n]| > \varepsilon]\\ =&Pr[|A_n-(1-p)| > \varepsilon] \leq \frac{p(1-p)}{\varepsilon^2}.\end{aligned}$$

Same as before, this does not converge to 0 as $n \to \infty$, and the LLN 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.

**5. Simplified Self-Grading**

There are about $n = 500$ self-graded question parts in this iteration of EECS 70. For this simplified version of self-grading, we use a scale from 0 to 4 instead of the $0,2,5,8,10$ scale currently being used. On each of them, a student assigns a grade $S_i$. For each homework, readers randomly grade a subset of the problems. Assume that $n/5$ of the question parts are graded by the readers (chosen uniformly over all the problem parts) and the readers assign grades $R_i$. Assume that $R_i$ may deviate from an honest self-grade $S_i$ according to the conditional probabilities given in table 1.

We do the following check: we add up all of the $S_i - R_i$ for a particular student (for the subset of problems graded by readers only). If the result is too high, we suspect that a student might be inflating their grades.

| $R_i \backslash S_i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 0 | 3/4 | 1/4 | 0 | 0 | 0 |
| 1 | 1/4 | 1/2 | 1/4 | 0 | 0 |
| 2 | 0 | 1/4 | 1/2 | 1/4 | 0 |
| 3 | 0 | 0 | 1/4 | 1/2 | 1/4 |
| 4 | 0 | 0 | 0 | 1/4 | 3/4 |

Table 1: $\mathbf{P}(R_i|S_i)$.

a) Suppose that a student is honest. Let $p_0 = \mathbf{P}(S_i = 0)$ and $p_4 = \mathbf{P}(S_i = 4)$. Let $X_i = S_i - R_i$. Express the distribution of $X_i$ as a function of $p_0$ and $p_4$.

**Solutions:** Clearly, $-1 \leq X_i \leq 1$. We can also express the distribution of $X_i$ as a function of the distribution of $S_i$:

$$\begin{aligned}
\mathbf{P}(X_i = 0) &= \mathbf{P}(R_i = S_i) && (1)\\
&= \mathbf{P}(R_i = 0 \cap S_i = 0) + \mathbf{P}(R_i = 1 \cap S_i = 1) + \cdots + \mathbf{P}(R_i = 4 \cap S_i = 4) && (2)\\
&= \mathbf{P}(R_i = 0|S_i = 0)\mathbf{P}(S_i = 0) + \mathbf{P}(R_i = 1|S_i = 1)\mathbf{P}(S_i = 1) + \cdots + \mathbf{P}(R_i = 4|S_i = 4)\mathbf{P}(S_i = 4) && (3)\\
&= \frac{3}{4}\mathbf{P}(S_i = 0) + \frac{1}{2}\mathbf{P}(S_i = 1) + \cdots + \frac{3}{4}\mathbf{P}(S_i = 4) && (4)\\
&= \frac{3}{4}(\mathbf{P}(S_i = 0) + \mathbf{P}(S_i = 4)) + \frac{1}{2}(\mathbf{P}(S_i = 1) + \mathbf{P}(S_i = 2) + \mathbf{P}(S_i = 3)) && (5)
\end{aligned}$$

A similar argument gives:

$$\mathbf{P}(X_i = -1) = \frac{1}{4}(\mathbf{P}(S_i = 0) + \mathbf{P}(S_i = 1) + \mathbf{P}(S_i = 2) + \mathbf{P}(S_i = 3))$$

and

$$\mathbf{P}(X_i = 1) = \frac{1}{4}(\mathbf{P}(S_i = 1) + \mathbf{P}(S_i = 2) + \mathbf{P}(S_i = 3) + \mathbf{P}(S_i = 4))$$

The distribution of $S_i$ is unknown, but we have one more equation from the rule of total probability. Namely $\sum_{k=0}^{n} \mathbf{P}(S_i = k) = 1$. Using it, we can express everything in terms of $p_0$ and $p_4$. We get:

$$\begin{aligned}
\mathbf{P}(X_i = 0) &= \frac{1}{4}(p_0 + p_4) + \frac{1}{2} && (6)\\
\mathbf{P}(X_i = -1) &= \frac{1}{4}(1 - p_4) && (7)\\
\mathbf{P}(X_i = 1) &= \frac{1}{4}(1 - p_0) && (8)
\end{aligned}$$

b) Give the best upper-bounds you can on both $\mathbf{E}[X_i]$ and $\mathrm{Var}(X_i)$. Your bounds shall not depend on $p_0$ or $p_4$.

**Solutions:**

$$\mathbf{E}[X_i] = -1(\frac{1}{4}(1 - p_4)) + 1(\frac{1}{4}(1 - p_0)) = \frac{1}{4}(p_4 - p_0)$$

The maximum of $\mathbf{E}[X_i]$ is obtained for $p_0 = p_1 = p_2 = p_3 = 0$ and $p_4 = 1$, hence:

$$\mathbf{E}[X_i] \leq \frac{1}{4}$$

$$\begin{aligned}\mathrm{Var}(X_i) &= \mathbf{E}[X_i^2] - \mathbf{E}[X_i]^2 &(9)\\ &= \frac{1}{2} - \frac{1}{4}(p_0+p_4) - \frac{1}{16}(p_0-p_4)^2 &(10)\end{aligned}$$

The maximum of $\mathrm{Var}(X_i)$ happens for any distribution such that $p_0 = p_4 = 0$, so that:

$$\mathrm{Var}(X_i) \leq \frac{1}{2}$$

c) Using Chebyshev's inequality and the above parts, compute the smallest threshold $T$ that we should choose so that $\sum_i X_i \leq T$ for 95% of honest students?

**Solutions:** We have $\mathbf{E}[\sum_{i=1}^{100} X_i] = \sum_{i=1}^{100} \mathbf{E}[X_i] \leq \frac{100}{4} = 25$, and $\mathrm{Var}(\sum_{i=1}^{100} X_i) = 100\mathrm{Var}(X_i) \leq \frac{100}{2} = 50$. Let's transform the probability we are looking for into a Chebyshev type statement.

$$\begin{aligned}\mathbf{P}\left(\sum_i X_i \leq T\right) &= 1 - \mathbf{P}\left(\sum_i X_i > T\right) &(11)\\ &= 1 - \mathbf{P}\left(\sum_i X_i - \mathbf{E}[\sum_i X_i] > T - \mathbf{E}[\sum_i X_i]\right) &(12)\\ &= 1 - \mathbf{P}\left(\sum_i X_i - \mathbf{E}[\sum_i X_i] \geq T - \mathbf{E}[\sum_i X_i] + 1\right) &(13)\\ &\geq 1 - \mathbf{P}\left(|\sum_i X_i - \mathbf{E}[\sum_i X_i]| \geq T - \mathbf{E}[\sum_i X_i] + 1\right) &(14)\\ &\geq 1 - \mathbf{P}\left(|\sum_i X_i - \mathbf{E}[\sum_i X_i]| \geq T - 24\right) &(15)\end{aligned}$$

So by Chebyshev, we have:

$$\mathbf{P}(\sum_i X_i \leq T) \geq 1 - \frac{\mathrm{Var}(\sum_i X_i)}{(T-24)^2} \geq 1 - \frac{50}{(T-24)^2}$$

Setting the right-hand side equal to .95 and solving in $T$, we get:

$$T = 24 + \sqrt{\frac{50}{0.05}} \approx 55.6$$

**6. Those 3407 Votes**

In the aftermath of the 2000 US Presidential Election, many people have claimed that unusually large number of votes cast for Pat Buchanan in Palm Beach County are statistically highly significant, and thus of dubious validity. In this problem, we will examine this claim from a statistical viewpoint.

The total percentage votes cast for each presidential candidate in the entire state of Florida were as follows:

| Gore | Bush | Buchanan | Nader | Browne | Others |
|---|---|---|---|---|---|
| 48.8% | 48.9% | 0.3% | 1.6% | 0.3% | 0.1% |

In Palm Beach County, the actual votes cast (before the recounts began) were as follows:

| Gore | Bush | Buchanan | Nader | Browne | Others | Total |
|---|---|---|---|---|---|---|
| 268945 | 152846 | 3407 | 5564 | 743 | 781 | 432286 |

To model this situation probabilistically, we need to make some assumptions. Let's model the vote cast by each voter in Palm Beach County as a random variable $X_i$, where $X_i$ takes on each of the six possible values (five candidates or "Others") with probabilities corresponding to the Florida percentages. (Thus, e.g., $\Pr[X_i = \text{Gore}] = 0.488$.) There are a total of $n = 432286$ voters, and their votes are assumed to be mutually independent. Let the r.v. $B$ denote the total votes cast for Buchanan in Palm Beach County (i.e., the number of voters $i$ for which $X_i = \text{Buchanan}$).

(a) Compute the expectation $\mathbf{E}[B]$ and the variance $\text{Var}(B)$.

**Solutions:** Let $B_i$ be a random variable representing whether the $i$th person voted for Buchanan. Then $B_i = 1$ if and only if $X_i = \text{Buchanan}$, so $B_i \sim \text{Bernoulli}(0.003)$. Note that the $B_i$'s are independently and identically distributed, with $\mathbf{E}[B_i] = 0.003$ and $\text{Var}(B_i) = 0.003 \times (1 - 0.003) = 0.002991$. Moreover, by linearity of expectation and independence, we find that $\mathbf{E}[B] = \sum_{i=1}^n \mathbf{E}[B_i] = 432286 \times 0.003 \approx 1297$ and $\text{Var}(B) = \sum_{i=1}^n \text{Var}(B_i) = 432286 \times 0.002991 \approx 1293$.

(b) Use Chebyshev's inequality to compute an *upper bound* $b$ on the probability that Buchanan receives at least 3407 votes, i.e., find a number $b$ such that

$$\Pr[B \geq 3407] \leq b.$$

Based on this result, do you think Buchanan's vote is significant?

**Solutions:** Chebyshev's inequality says that

$$\Pr[|B - \mathbf{E}[B]| \geq a] \leq \frac{\text{Var}(B)}{a^2}.$$

In our case $\mathbf{E}[B] = 1297$ and $\text{Var}(B) = 1293$, so if we take $a = 2110$, we find that $\Pr[|B - 1297| \geq 2110] \leq 1293/2110^2 \approx 0.0003$. Now note that the condition $|B - 1297| < 2110$ is equivalent to the condition $-813 < B < 3407$, and since $B$ is non-negative, we find that $\Pr[B > 3407] \leq 0.0003$ (roughly), so we can take $b \approx 0.0003$. In other words, receiving 3407 votes for Buchanan in Palm Beach County seems very unlikely to happen by chance, under this simple model. So yes, this is statistically significant.

(c) Suppose that your bound $b$ in part (b) is exactly accurate, i.e., assume that $\Pr[X \geq 3407]$ is exactly equal to $b$. [*In fact the true value of this probability is much smaller*] Suppose also that all 67 counties in Florida have the same number of voters as Palm Beach County, and that all behave independently according to the same statistical model as Palm Beach County. What is the probability that in *at least one* of the counties, Buchanan receives at least 3407 votes? How would this affect your judgment as to whether the Palm Beach tally is significant?

**Solutions:** Let $p_j$ be the probability that the $j$th county does not receive 3407 votes for Buchanan. We have from part (b) that $p_j = 1 - b \approx 0.9997$. Note that the probability that no county yields at least 3407 votes for Buchanan is $p_1 \times \cdots \times p_{67}$, since the voters in each county behave independently. Thus, the probability that Buchanan does not receive 3407 votes in any county is about $(0.9997)^{67} \approx 0.98$. Consequently, the probability that Buchanan *does* receive at least 3407 votes in some county is about $1 - 0.98 \approx 0.02$. In other words, this seems unlikely to happen by chance.

7. **Median**

Given a list of numbers with an odd length, the median is obtained by sorting the list and seeing which number occupies the middle position. (e.g. the median of $(1,0,1,0,2,1,2)$ is 1 because the list sorts to $(0,0,1,1,1,2,2)$ and there is a 1 in the middle. Meanwhile, the median of $(1,0,2,0,0,1,0)$ is 0 because the list sorts to $(0,0,0,0,1,1,2)$ and there is a 0 in the middle.)

a) Consider an iid sequence of random variables $X_i$ with probability mass function $P_X(0) = \frac{1}{3}$, $P_X(1) = \frac{1}{4}$, $P_X(2) = \frac{5}{12}$. Let $M_i$ be the random variable that is the median of the random list $(X_1, X_2, \ldots, X_{2i+1})$. **Show that $P(M_i \neq 1)$ goes to zero exponentially fast in $i$.**

*(HINT: What has to happen with the $\{X_j\}$ for $M_i = 0$ or $M_i = 2$ ?)*

**Solutions:** The idea for this is as follows. Intuitively, as we sample more and more $X_j$, the fraction of $X_j$ that turn up 0 is approaching $\frac{1}{3}$, and the fraction of $X_j$ that turn up 2 is approaching $\frac{5}{12}$. Since these are both numbers less than 1/2, it is increasingly likely that the list will have the number 1 in the middle when sorted. Note that this is using the fact that the median of the list is not 1 if and only if either at least half the $X_j$ turned up 0 or at least half the $X_j$ turned up 2.

We can formalize this intuition with a Chernoff bound. Define the random variables $Y_j^0$, $Y_j^1$, and $Y_j^2$ such that

$$Y_j^k = \begin{cases} 1 & \text{if } X_j = k \\ 0 & \text{if } X_j \neq k \end{cases}$$

Then $\mathbb{E}[Y_j^0] = \frac{1}{3}$ for all $j$, and $\mathbb{E}[Y_j^2] = \frac{5}{12}$ for all $j$. From the Chernoff bound for Bernoulli r.v.'s, which was worked out in class, we get that

$$\Pr[\frac{1}{2i+1}\sum_{j=1}^{2i+1} Y_j^0 \geq \frac{1}{2}] \leq e^{-(2i+1)D(\frac{1}{2}||\frac{1}{3})}$$

$$\Pr[\frac{1}{2i+1}\sum_{j=1}^{2i+1} Y_j^2 \geq \frac{1}{2}] \leq e^{-(2i+1)D(\frac{1}{2}||\frac{5}{12})}$$

We note from class that these are both exponentially decaying values, since $D(a||p) = 0$ iff $a = p$. As mentioned above, we know that $M_i \neq 1$ iff $\frac{1}{2i+1}\sum_{j=1}^{2i+1} Y_j^0 \geq \frac{1}{2}$ or $\frac{1}{n}\sum_{j=1}^{2i+1} Y_j^2 \geq \frac{1}{2}$. By the union bound, the probability of either of these events happening is bounded by the sum of the probabilities. Hence we get:

$$\Pr[M_i \neq 1] \leq e^{-(2i+1)D(\frac{1}{2}||\frac{1}{3})} + e^{-(2i+1)D(\frac{1}{2}||\frac{5}{12})}.$$

This is decaying exponentially in $i$.

b) Generalize the argument you have made above to state a law of large numbers for the median of a sequence of iid discrete-valued random variables. Sketch a proof for this law.

*(HINT: Let $Mid(X)$ be that value $t$ for which $P(X < t) < \frac{1}{2}$ and $P(X > t) < \frac{1}{2}$. If such a $t$ doesn't exist, then don't worry about that case at all.)*

**Solutions:** As the number of samples increases, the median of a sequence of iid discrete-valued random variables $\{X_i\}$ converges to the median of the PMF of $X_i$, which is defined in the hint above.

This is basically the same as the last part, except now we do not even have to necessarily worry about exponential convergence. Let

$$Y_i^{X<t} = \begin{cases} 1 & \text{if } X_i < t \\ 0 & \text{if } X_i > t \end{cases}$$

$$Y_i^{X>t} = \begin{cases} 1 & \text{if } X_i > t \\ 0 & \text{if } X_i < t \end{cases}$$

Then by a similar argument to the previous part, the averages of the $Y_i^{X>t}$ and $Y_i^{X<t}$ are going to converge to their means, $\Pr[X_i < t]$ and $\Pr[X_i > t]$ respectively, as the number of samples approaches infinity. Since both these values are less than 1/2, this means that as the number of samples approaches infinity, the probability of the frequency exceeding $\frac{1}{2}$ is going to go to zero exponentially.

Consequently, with high probability, the middle of the sequence will be $t$. This is because $P(X_i > t) + P(X_i = t) + P(X_i < t) = 1$ and so the probability that $X_i$ takes value $t$ is nonzero. Hence, if we let $M_n$ be the median of $n$ trials, $\Pr[M_n \neq Mid(X)] \to 0$, so $\Pr[M_n = Mid(X)] \to 1$.

**8. (Optional) Best Question NA**

In the game League of Legends, two champions fight each other and the one with greater prowess wins. Each champion can carry items to enhance their fighting ability. In this problem, we will do a (slightly simplified) analysis of three items in the game.

Each champion has

- $H_0$: Health (HP) - how much damage they can take before dying. e.g. 3000 HP
- $A_0$: Base attack damage (AD) - how much damage they do without any items every time they attack. e.g. 50 AD
- $C_0$: Base critical strike percentage (crit chance) - Every attack has a certain chance of dealing a critical strike, which doubles the amount of damage dealt. $C_0$ is the probability of this happening without items. e.g. 0.05 crit chance
- $S_0$: Base attack speed (AS) - how many times the champion attacks per second. e.g. 1.1 AS

We will analyze three common items: Infinity Edge, Bloodthirster, and Blade of the Ruined King. Their abilities are outlined below:

**Infinity Edge**

- +70 attack damage
- +25% critical strike percentage
- Critical strikes will deal 250% damage, instead of 200%

**Bloodthirster**

- +100 attack damage

**Blade of the Ruined King**

- +25 attack damage
- Will grant additional attack damage equal to 5% of the opponents current health.
- +0.4 attack speed

Here is an example if you are still confused as to how this all works: Suppose I am fighting against an enemy champion with 1000 HP. I have 100 base AD, 0 base crit chance, and 1 base attack speed. Without any items, it would take 10 seconds for me to defeat him. With a Bloodthirster, I would have 200 AD and would defeat him 5 seconds. With an Infinity Edge, I would have 170 AD and a critical strike (which would happen 25% of the time) would deal $170 \times 2.5 = 425$ damage. With Blade of the Ruined King, my first attack would deal $125 + 0.05(1000) = 175$ damage, leaving him with 825 HP; my second attack would deal 166.25 damage.

In order to compare the items, we will estimate our damage per second (DPS) with each of the three items. Let $H_0 = \infty, A_0, C_0, S_0$ denote our champion's statistics and $H_0', A_0', C_0', S_0'$ denote our opponent's statistics.

(a) In terms of the above variables, what is our expected damage per second with an Infinity Edge?

**Solutions:** My damage for a non-critical strike is $A_0 + 70$. For a critical strike, which occurs with probability $C_0 + 0.25$, I deal $2.5(A_0 + 70)$ damage. Thus my expected damage per strike with IE is

$$(1 - C_0 - 0.25)(A_0 + 70) + (C_0 + 0.25)(2.5(A_0 + 70)) = 1.5A_0C_0 + 1.375A_0 + 105C_0 + 96.25$$

The DPS is the above quantity times our attack speed $S_0$

$$S_0(1.5A_0C_0 + 1.375A_0 + 105C_0 + 96.25)$$

(b) What is our expected damage per second with a Bloodthirster?

**Solutions:** My damage for a non-critical strike is $A_0 + 100$. For a critical strike, which occurs with probability $C_0$, I deal $2(A_0 + 100)$ damage. Thus my expected damage per strike with BT is

$$(1 - C_0)(A_0 + 100) + C_0(2(A_0 + 100)) = A_0C_0 + A_0 + 100C_0 + 100$$

The DPS is the above quantity times our attack speed $S_0$

$$S_0(A_0C_0 + A_0 + 100C_0 + 100)$$

(c) What is our average damage per second with a Blade of the Ruined King? (Your damage will be lower after every hit against the enemy since his HP will go down. Compute the average amount of damage you deal until the enemy is dead.)

(Hint(s): First, take the average damage per second to mean the average damage dealt per second till the enemy is defeated. So, it would be $H_0'$ divided by number of seconds to defeat the enemy.

Secondly, in order to actually calculate this, set up a recurrence relation of the enemy's health after each attack. This would be of the form $h_n = A \cdot C^n + B$ where $h_n$ is the health after the $n^{th}$ attack. To solve this, write the enemy's health after the $n + 1^{th}$ attack as a function of their health after the $n^{th}$ attack. Given that you know the form, plug in a couple of numbers for $n$ and solve for $A, B$ and $C$.

Finally, incorporate your attack speed back into this as you calculated it per attack. Now you have your DPS!)

**Solutions:** Since the order between the critical strike and the 5% additional attack damage was unclear in the question and has caused some confusion, we will accept answers from either of the interpretation.

**Interpretation 1: Critical strike *before* 5% additional attack damage**

The average damage per strike is going to be given by the total HP of our opponent divided by the number of hits needed to defeat him. Let $h_n$ denote the HP after attack $n$. The expected damage is $A_0 + 25 + 0.05h_n$ with probability $1 - C_0$ (critical strike doesn't happen) and $2(A_0 + 25) + 0.05h_n$ with probability $C_0$ (critical strike happens before the 5% additional damage). Therefore, the opponent's HP after attack $n+1$ is,

$$\begin{aligned} h_{n+1} &= h_n - (1-C_0)(A_0+25+0.05h_n) - C_0(2(A_0+25)+0.05h_n) \\ &= 0.95h_n - (1+C_0)(A_0+25). \end{aligned}$$

Then we have the following recurrence relation:

$$\begin{aligned} h_0 &= H_0' \\ h_{n+1} &= 0.95h_n - (1+C_0)(A_0+25) \end{aligned} \tag{16}$$

For brevity, let $x = 0.95$ and $y = -(1+C_0)(A_0+25)$ so we can rewrite Equation (16) to

$$h_{n+1} = xh_n + y \tag{17}$$

We take $h_n$ to be of the following form, from the hint:

$$h_n = A \cdot C^n + B$$

Plugging in $n = 0$, $n = 1$ and $n = 2$, we get the system of equations

$$\begin{aligned} h_0 = H_0' &= A + B \\ h_1 = xH_0' + y &= A \cdot C + B \\ h_2 = xh_1 + y = x(xH_0' + y) + y &= A \cdot C^2 + B \end{aligned}$$

Solving this system gives us

$$A = H_0' + \frac{y}{x-1},\ B = -\frac{y}{x-1},\ C = x \tag{18}$$

**Proof**: We can prove that this is the correct formula by using induction on $n$.
*Base Case:* $h_0 = H_0' = H_0' + \frac{y}{x-1} - \frac{y}{x-1} = AC^0 + B$; The formula works for $n = 0$.
*Inductive Hypothesis:* Assume $h_k = AC^k + B$ for $0 \leq k$.
*Inductive Step:* We will show that the formula works for $n = k+1$.

$$\begin{aligned} h_{k+1} &= xh_k + y \\ &= x(AC^k + B) + y && \text{(Induction Hypothesis)} \\ &= x\left(\left(H_0' + \frac{y}{x-1}\right)x^k - \frac{y}{x-1}\right) + y \\ &= \left(H_0' + \frac{y}{x-1}\right)x^{k+1} + \frac{(-xy) + y(x-1)}{x-1} \\ &= \left(H_0' + \frac{y}{x-1}\right)x^{k+1} + \frac{y}{x-1} \\ &= AC^{k+1} + B \end{aligned}$$

$\square$

Now that we have an explicit expression for the health of our opponent over time, we want to find the minimum $n$ such that $h_n \leq 0$.

$$\begin{aligned}
h_n &\leq 0 \\
AC^n + B &\leq 0 \\
C^n &\leq \frac{-B}{A} \\
n \log C &\leq \log\left(\frac{-B}{A}\right) \\
n \log x &\leq \log\left(\frac{\frac{y}{x-1}}{H_0' + \frac{y}{x-1}}\right) \qquad \text{(From Equation (18))}
\end{aligned} \tag{19}$$

Plugging in values of $x$ and $y$, we get,

$$\begin{aligned}
n \log 0.95 &\leq \log\left(\frac{\frac{-(1+C_0)(A_0+25)}{0.95-1}}{H_0' + \frac{-(1+C_0)(A_0+25)}{0.95-1}}\right) \\
n \log 0.95 &\leq \log\left(\frac{20(1+C_0)(A_0+25)}{H_0' + 20(1+C_0)(A_0+25)}\right) \\
n &\geq \frac{1}{\log 0.95} \log\left(\frac{20(1+C_0)(A_0+25)}{H_0' + 20(1+C_0)(A_0+25)}\right)
\end{aligned}$$

Note that the inequality flips because $\log(0.95)$ is negative. Since $n$ must be integer, we take the ceiling of the above expression

$$n = \left\lceil \frac{1}{\log(0.95)} \log \frac{20(1+C_0)(A_0+25)}{H_0' + 20(1+C_0)(A_0+25)} \right\rceil$$

The average damage per strike is then

$$\frac{H_0'}{n}$$

And since our attack speed is $S_0 + 0.4$, the average DPS is

$$(S_0 + 0.4) \frac{H_0'}{\left\lceil \frac{1}{\log(0.95)} \log \frac{20(1+C_0)(A_0+25)}{H_0' + 20(1+C_0)(A_0+25)} \right\rceil}$$

$\square$

**Interpretation 2: 5% additional attack damage *before* critical strike**

The expected damage is $A_0 + 25 + 0.05h_n$ with probability $1 - C_0$ (critical strike doesn't happen) and $2(A_0 + 25 + 0.05h_n)$ with probability $C_0$ (5% additional damage happens before the critical strike). Therefore, the opponent's HP after attack $n+1$ is,

$$\begin{aligned}
h_{n+1} &= h_n - (1 - C_0)(A_0 + 25 + 0.05h_n) - C_0 \cdot 2(A_0 + 25 + 0.05h_n) \\
&= (0.95 - C_0)h_n - (1 + C_0)(A_0 + 25).
\end{aligned}$$

The recurrence relation for this case would be

$$\begin{aligned}
h_0 &= H_0' \\
h_{n+1} &= (0.95 - C_0)h_n - (1 + C_0)(A_0 + 25)
\end{aligned} \tag{20}$$

Let $x = 0.95 - C_0$ and $y = -(1+C_0)(A_0+25)$. Again, we have Equation (20) in the form,

$$h_{n+1} = xh_n + y,$$

which we already know the solution. Plugging $x$ and $y$ into Inequality (19), we get,

$$\begin{aligned}
n\log x &\le \log\left(\frac{\frac{y}{x-1}}{H_0' + \frac{y}{x-1}}\right) \\
n\log(0.95 - C_0) &\le \log\left(\frac{\frac{-(1+C_0)(A_0+25)}{0.95-C_0-1}}{H_0' + \frac{-(1+C_0)(A_0+25)}{0.95-C_0-1}}\right) \\
n\log(0.95 - C_0) &\le \log\left(\frac{\frac{(1+C_0)(A_0+25)}{0.05+C_0}}{H_0' + \frac{(1+C_0)(A_0+25)}{0.05+C_0}}\right) \\
n\log(0.95 - C_0) &\le \log\left(\frac{(1+C_0)(A_0+25)}{(0.05+C_0)H_0' + (1+C_0)(A_0+25)}\right) \\
n &\ge \frac{1}{\log(0.95-C_0)}\log\left(\frac{(1+C_0)(A_0+25)}{(0.05+C_0)H_0' + (1+C_0)(A_0+25)}\right).
\end{aligned}$$

Note that, again, the inequality flips because $\log(0.95 - C_0)$ is negative. Since $n$ must be integer, we take the ceiling of the above expression

$$n = \left\lceil \frac{1}{\log(0.95-C_0)}\log\frac{(1+C_0)(A_0+25)}{(0.05+C_0)H_0' + (1+C_0)(A_0+25)} \right\rceil$$

The average damage per strike is then

$$\frac{H_0'}{n}$$

And since our attack speed is $S_0 + 0.4$, the average DPS is

$$(S_0+0.4)\frac{H_0'}{\left\lceil \frac{1}{\log(0.95-C_0)}\log\frac{(1+C_0)(A_0+25)}{(0.05+C_0)H_0' + (1+C_0)(A_0+25)} \right\rceil}$$

□

(d) An item is better if it has higher expected DPS. Come up with 3 different scenarios (values for $H_0, A_0$, etc.) where each of the three items is the optimal choice.

**Solutions:** Comparing the expressions for IE and BT, it is clear that IE is almost always better. However, there is a tiny window where BT is better if $C_0 \approx A_0 \approx 0$. The expression for BotRK suggests that the DPS increases as the health of our opponent increases. Thus against opponents with large amounts of health, BotRK is optimal. There are many possible answers for this question. Here is one possible answer:

$$C_0 = 0.75, A_0 = 100, S_0 = 1, H_0' = 100$$

The DPS are 425, 350, 140 for IE,BT,BotRK, respectively (IE is best).

$$C_0 = 0, A_0 = 0, S_0 = 1, H_0' = 100$$

The DPS are 96.25, 100, 35 for IE,BT,BotRK, respectively (BT is best).

$$C_0 = 0, A_0 = 0, S_0 = 0.1, H_0' = 100$$

The DPS are 9.625, 10, 12.5 for IE,BT,BotRK, respectively (BotRK is best).

**For LoL players:** Now you know which item is best for your AD carry in different situations! Granted, the above analysis leaves out some of the finer details (e.g. life steal, builds, actives, etc.), but as far as raw damage output goes, this is fairly accurate.

**9. Write your own problem**

Write your own problem related to this week's material and solve it. You may still work in groups to brainstorm problems, but each student should submit a unique problem. What is the problem? How to formulate it? How to solve it? What is the solution?

*You might have noticed that this homework is shorter than usual. If you want more practice questions, please come talk to us. Enjoy your Thanksgiving!*