Homework 14

This homework is due December 9, 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. Please 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). Use a screen capture to show us that you finished them.

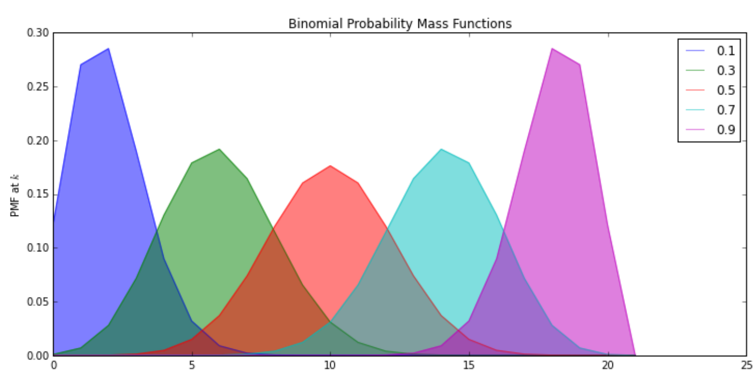
3. **Random Variables and Distributions Lab**

   In this week's lab, we will explore common discrete random variables and their corresponding distributions.

   For each part, students who want to can choose to completely rewrite the question. Basically, you can come up with your own formulation of how to do a series of experiments that result in the same discoveries. Then, write up the results nicely using plots as appropriate to show what you observed. You can also rewrite the entire lab to take a different path through as long as they convey the key insights aimed at in each part.

   Please download the IPython starter code from Piazza or the course webpage, and answer the following questions.

   (a) Plot the PMFs of binomial random variables with $N = 20$, and with success probabilities $p = 0.1, 0.3, 0.5, 0.7$, and 0.9, respectively. You should have 5 different plots in one figure.
   What do you observe as the success probability increases?
   As the success probability increases, the corresponding plot tends more toward the right, and the mean value of the corresponding binomial distribution gets closer to the number of trials.



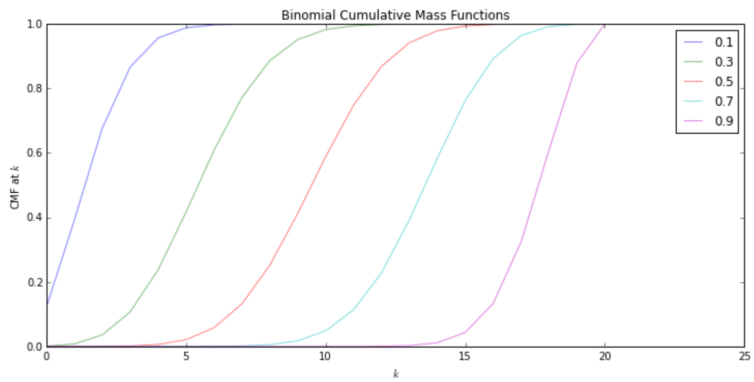
(b) In probability theory and statistics, the cumulative mass function (CMF) describes the probability that a real-valued discrete random variable X with a given probability distribution will be found to have a value less than or equal to x. Mathematically, we define the CMF $F_X(x)$ as

$$F_X(x) = P(X \leq x),$$

where the right-hand side represents the probability that the random variable X takes on a value less than or equal to x.

Plot the CMFs of the five binomial random variables from part (a). These should all be increasing curves. At what value does each CMF plot converge to and stay at 1 (i.e. at what value can you be almost 100% certain that the number of heads (or successful trials) is less than such value)?

Around $6, 12, 15, 19$, and 20, respectively.



(c) Plot the PMF a binomial distribution with parameters $(N = 100, p = 0.2)$ in a bar chart. Then, overlay your plot with the probability density function (PDF) of a normal distribution with parameters $(\mu = 20, \sigma^2 = 16)$. What do you observe?

In a different figure, plot the CMF of the same binomial distribution and overlay it with the cumulative density function (CDF) of the aforementioned normal distribution. Again, what do you observe?

Finally, derive an approximation between the two distributions using a concept you learned in this week's lecture.

The normal curve sits perfectly on top of the binomial distribution in both plots. We have seen this in Homework 11's Virtual Lab, but we now fully understand that this happens because of the Central Limit Theorem (CLT). Let's derive an approximation between the two distributions.

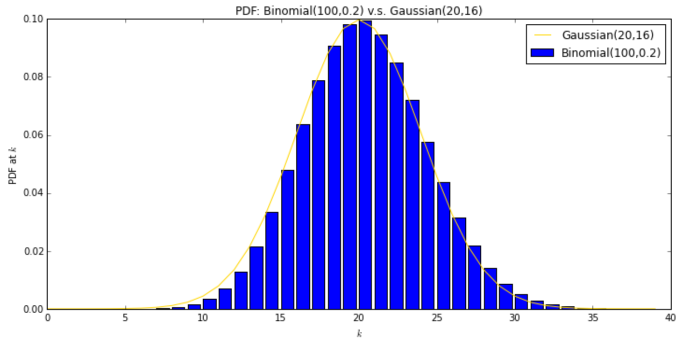
Let X be a binomial random variable with parameters (N, p). We know that X is the sum of i.i.d. Bernoulli random variables, which means $X = X_1 + X_2 + \ldots + X_N$, and each $X_i \sim Bern(p)$. From CLT, we know that:

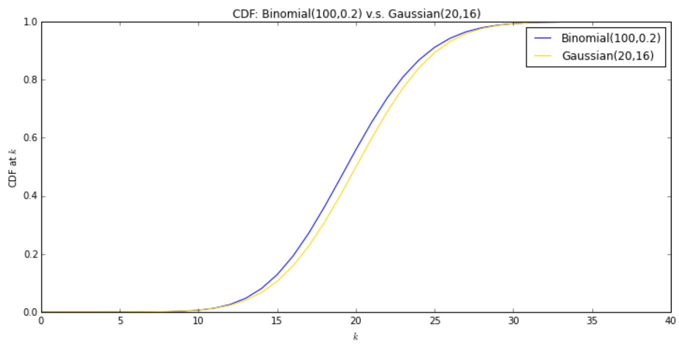
$$\frac{X - Np}{\sqrt{N}} \approx \mathcal{N}(0, \sigma^2),$$

where $\sigma^2 = p(1-p)$, the variance of a Bernoulli random variable. Hence,

$$Bin(N, p) \approx \mathcal{N}(Np, N\sigma^2) = \mathcal{N}(Np, Np(1-p)).$$

With $N = 100$ and $p = 0.2$, we see that $Bin(100, 0.2) \approx \mathcal{N}(20, 16)$, which was confirmed by our plots.





There are other optional parts of this Virtual Lab in the file `vl14.pdf`, which you can find on Piazza or the course website.

Reminder: When you finish, don't forget to convert the notebook to pdf and merge it with your written homework. Please also zip the `ipynb` file and submit it as `hw14.zip`.

4. To Be "Normal"

Suppose a standard 6-sided die (with faces 1 through 6) is rolled n times, and let A be the average of the results.

(a) How does the Central Limit Theorem help us approximate the distribution of A?

Let A_i be the result of the i^{th} toss. Recall that the mean of a single die roll is

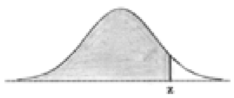
$$\mu = \mathbf{E}[A_i] = \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 2 + \frac{1}{6}\cdot 3 + \frac{1}{6}\cdot 4 + \frac{1}{6}\cdot 5 + \frac{1}{6}\cdot 6 = 3.5,$$

and the standard deviation of a single die roll is

$$\begin{aligned} \sigma &= \sqrt{\text{Var}(A_i)} \\ &= \sqrt{3.5 - \frac{1}{6}(1+4+9+16+25+36)} \\ &= \sqrt{35/12} \approx 1.7078. \end{aligned}$$

Accordingly, the Central Limit Theorem assures us that, for large n, the distribution $(A-3.5)/\sqrt{n}$ approaches a normal distribution with mean zero and standard deviation $\sqrt{35/12} \approx 1.7078$. The distribution of A itself, therefore, is also in some informal sense approximately normal for large n (though note that, as n increases, the standard deviation of A approaches 0. $\square$

(b) Let A' be a random variable drawn from the Gaussian distribution best approximating the distribution of A. If $n = 100$, what are the bounds of an interval $[a,b]$ centered at 3.5 such that $A' \in [a,b]$ with probability exactly 90%?. (See https://statistics.laerd.com/statistical-guides/normal-distribution-calculations.php for normal distribution calculation. You might want to use Table 1.)



Probability Content from -oo to Z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Table 1: Table of the Normal Distribution.
Source: http://cosstatistics.pbworks.com/w/page/27425647/Lesson

The average of n independent die rolls has a mean of

$$\mathbf{E}[A] = \mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n} A_i\right] = \frac{1}{n}\cdot n \cdot \mathbf{E}[A_i] = 3.5,$$

and a standard deviation of

$$\sqrt{\mathrm{Var}(A)} = \sqrt{\mathrm{Var}(\frac{1}{n}\sum_{i=1}^{n} A_i)} = \sqrt{\left(\frac{1}{n}\right)^2 \cdot n\mathrm{Var}(A_i)} = \sqrt{\frac{1}{n}\cdot\frac{35}{12}} \approx 0.17078.$$

So A' should be drawn from the normal distribution with mean 3.5 and standard deviation 0.17078. We want the interval $[a,b]$ centered around $\mathbf{E}[A'] = 3.5$ that

$$\mathbf{P}(a \leq A' \leq b) \approx 0.9.$$

By symmetry, we know 10% of the probability falls outside of $[a,b]$, with 5% below a and 5% above b. Thus, $\mathbf{P}(a \geq A') = \mathbf{P}(A' \geq b) \approx \frac{1-0.9}{2} = 0.05$. We want

$$\mathbf{P}(A' \leq b) \approx 1 - \mathbf{P}(A' \geq b) = 0.95.$$

According to the standard normal distribution table,

$$\mathbf{P}\left(A' \leq b\right) = \mathbf{P}\left(\frac{A' - \mathbf{E}[A']}{\sqrt{\mathrm{Var}(A')}} \leq \frac{b - \mathbf{E}[A']}{\sqrt{\mathrm{Var}(A')}}\right) \approx 0.95$$

when $\dfrac{b - \mathbf{E}[A']}{\sqrt{\mathrm{Var}(A')}} \approx 1.65$. Therefore,

$$\begin{aligned}
\frac{b - \mathbf{E}[A']}{\sqrt{\mathrm{Var}(A')}} &\approx 1.65 \\
\frac{b - 3.5}{0.17078} &= 1.65 \\
b &= 1.65 \cdot 0.17078 + 3.5 \\
b &= 3.781787,
\end{aligned}$$

and

$$a = 3.5 - (b - 3.5) = 3.5 - (3.781787 - 3.5) = 3.218213.$$

Hence, the interval is $[3.22, 3.78]$. □

(c) Approximate the probability that $3 \leq A' \leq 4$, if $n = 30$.
When $n = 30$, $\mathbf{E}[A']$ is still 3.5, whereas the standard deviation is

$$\sqrt{\mathrm{Var}(A')} = \sqrt{\frac{1}{n}\mathrm{Var}(A_i)} = \sqrt{\frac{1}{30} \cdot \frac{35}{12}} \approx 0.3118.$$

We want to know

$$\mathbf{P}\left(A' \leq 4\right) = \mathbf{P}\left(\frac{A' - \mathbf{E}[A']}{\sqrt{\mathrm{Var}(A')}} \leq \frac{4 - \mathbf{E}[A']}{\sqrt{\mathrm{Var}(A')}}\right) = \mathbf{P}\left(\frac{A' - \mathbf{E}[A']}{\sqrt{\mathrm{Var}(A')}} \leq \frac{4 - 3.5}{0.3118}\right) \approx \mathbf{P}\left(\frac{A' - \mathbf{E}[A']}{\sqrt{\mathrm{Var}(A')}} \leq 1.60\right).$$

Looking up $z = 1.60$ from the standard normal distribution table, we find

$$\mathbf{P}(A' \leq 4) \approx 0.9452.$$

By symmetry,

$$\mathbf{P}(A' \leq 3) = \mathbf{P}(A' \geq 4) \approx 1 - 0.9452 = 0.0548.$$

and thus

$$\mathbf{P}(3 \leq A' \leq 4) = \mathbf{P}(A' \leq 4) - \mathbf{P}(A' \leq 3) = 0.9452 - 0.0548 = 0.8904.$$

□

(d) What is the minimum n for which, with probability at least 99%, we have $3 \leq A' \leq 4$?

$\mathbf{E}[A']$ is 3.5 regardless of n. The standard deviation is

$$\sqrt{\operatorname{Var}(A')} = \sqrt{\frac{1}{n}\operatorname{Var}(A_i)} = \sqrt{\frac{35}{12n}}.$$

We want

$$\mathbf{P}(A' \leq 4) = \mathbf{P}\left(\frac{A' - \mathbf{E}[A']}{\sqrt{\operatorname{Var}(A')}} \leq \frac{4 - \mathbf{E}[A']}{\sqrt{\operatorname{Var}(A')}}\right) \approx 0.99.$$

Again, because of symmetry, we look for a z value that gives $99 + \frac{1}{2} = 99.5\%$ probability. $z \approx 2.58$ according to the standard normal distribution table. Now we have,

$$\begin{aligned}
z = \frac{4 - \mathbf{E}[A']}{\sqrt{\operatorname{Var}(A')}} &= 2.58 \\
\frac{4 - 3.5}{\sqrt{\frac{35}{12n}}} &= 2.58 \\
\sqrt{\frac{12n}{35}} &= 5.16 \\
n = \frac{5.16^2 \cdot 35}{12} &= 77.658.
\end{aligned}$$

Therefore, n must be at least 78. $\square$

Using Hoeffding's Inequality: We can also find a proper upper bound with Hoeffding's inequality,

$$\begin{aligned}
\mathbf{P}(A' \geq 4) = \mathbf{P}(A' \geq 3.5 + 0.5) \leq \exp\left(-n \cdot \frac{2(0.5^2)}{(6-1)^2}\right) &\leq 0.005 \\
-0.02n &\leq \ln 0.005 \\
n &\geq \frac{-\ln 0.005}{0.02} = 264.916.
\end{aligned}$$

Hoeffding's inequality says that an n of anything greater than 265 is certainly safe. $\square$

5. Hypothesis testing

We would like to test the hypothesis claiming that a coin is fair, i.e. $P(H) = P(T) = 0.5$. To do this, we flip the coin $n = 100$ times. Let Y be the number of heads in $n = 100$ flips of the coin. We decide to reject the hypothesis if we observe that the number of heads is less than $50 - c$ or larger than $50 + c$. However, we would like to avoid rejecting the hypothesis if it is true; we want to keep the probability of doing so less than 0.05. Please determine c. (*Hints: use the central limit theorem to estimate the probability of rejecting the hypothesis given it is actually true.*)

You might need to use Table 1.

Let X_i be the random variable denoting the result of the i-th flip:

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th flip is head,} \\ 0 & \text{if the } i\text{-th flip is tail.} \end{cases}$$

Then we have $Y = \sum_{i=1}^{n} X_i$. If the hypothesis is true, then $\mu = \mathbf{E}[X_i] = \frac{1}{2}$ and $\sigma^2 = \mathrm{Var}(X_i) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. By central limit theorem, we know that

$$P\left(\frac{Y - n\mu}{\sqrt{n\sigma^2}} \leq z\right) \approx \Phi(z)$$
$$P\left(\frac{Y - 100 \cdot \frac{1}{2}}{\sqrt{100 \cdot \frac{1}{4}}} \leq z\right) \approx \Phi(z)$$
$$P\left(\frac{Y - 50}{5} \leq z\right) \approx \Phi(z)$$

where

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \mathrm{d}x.$$

We will reject the hypothesis when $|Y - 50| > c$. We also want $P(|Y - 50| > c) < 0.05$, or equivalently $P(|Y - 50| \leq c) > 0.95$. We have

$$P(|Y - 50| \leq c) = P\left(\frac{|Y - 50|}{5} \leq \frac{c}{5}\right) \approx 2\Phi(\frac{c}{5}) - 1.$$

Let $2\Phi(\frac{c}{5}) - 1 = 0.95$, so $\Phi(\frac{c}{5}) = 0.975$ or $\frac{c}{5} = 1.96$. That is $c = 9.8$ flips. So we see that if we observe more that $50 + 10 = 60$ or less than $50 - 10 = 40$ heads, we can reject the hypothesis.

This question does not require people using CLT. An alternative solution is by Hoeffding's inequality:
We want $P(|Y - 50| > c) \leq 0.05$. We know that

$$P(|Y - 50| > c) = P(\frac{|Y - 50|}{100} > \frac{c}{100}) \leq 2e^{-100 \cdot \frac{2(\frac{c}{100})^2}{(1-0)^2}}.$$

Let $2e^{-100 \cdot \frac{2(\frac{c}{100})^2}{(1-0)^2}} \leq 0.05$, we get $c \geq 13.58$. So $c = 14$ satisfies our criteria.

6. Simplified Self-Grading (carried over from HW13; only parts (d)-(h) need to be done)

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 2.

R_i \ 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 2: $\mathbf{P}(R_i | S_i)$.

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.

(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.

Refer to HW13's solution

(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.

Refer to HW13's solution

(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?

Refer to HW13's solution

(d) Repeat the above using the Central Limit Theorem to get an approximate answer for T. (You might want to refer to Table 1 for the cumulative normal distribution table.)

$$\mathbf{P}\left(\sum_{i=1}^{100} X_i \leq T\right) = \mathbf{P}\left(\frac{\sum_{i=1}^{100} X_i - 100\mathbf{E}[X_1]}{\sqrt{100\mathrm{Var}(X_1)}} \leq \frac{T - 100\mathbf{E}[X_1]}{\sqrt{100\mathrm{Var}(X_1)}}\right) \tag{1}$$

$$\geq \mathbf{P}\left(\frac{\sum_{i=1}^{100} X_i - 100\mathbf{E}[X_1]}{\sqrt{100\mathrm{Var}(X_1)}} \leq \frac{T-25}{\sqrt{50}}\right) \tag{2}$$

$$\geq \mathbf{P}\left(\frac{\sum_{i=1}^{100} X_i - 100\mathbf{E}[X_1]}{\sqrt{100\mathrm{Var}(X_1)}} \leq \frac{T-25}{7.07}\right) \tag{3}$$

We can apply the CLT approximation as the X_i are i.i.d and of bounded variance. By the CLT, we have that the lower bound is approximately equal to

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{T-25}{7.07}} e^{-\frac{x^2}{2}} dx$$

We can use the cumulative distribution table of the normal distribution to see that $\frac{T-25}{7.07} = 1.65$ will get us 95% probability. Solving back for T, we get

$$T \approx 36.7$$

which is much lower than the conservative Chebyshev bound from question c).

(e) For simplicity, we are going to focus our attention on a hypothetical student who never truly deserves full points and never truly deserves a zero on any question part, i.e., they never give themselves a zero or full points on a question. Recompute better upper bounds on both $\mathbf{E}[X_i]$ and $\mathrm{Var}(X_i)$ that are valid for this student. Recalculate the relevant theshold T using the Central Limit Theorem.

The student's distribution is such that $p_0 = p_4 = 0$. For this student, we have:

$$\mathbf{E}[X_i] = 0$$

and

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

As for the previous question, the relevant threshold is given by T such that:

$$\frac{1}{2\pi} \int_{-\infty}^{\frac{T}{\sqrt{50}}} e^{-\frac{x^2}{2}} dx = 0.95$$

So

$$T \approx 1.65\sqrt{50} \approx 11.7$$

is the new threshold.

(f) Assume this student is inflating their true self-grades S_i by adding 1 point to a question part with probability $1/2$. What is their risk of being caught (i.e., above the threshold T)? (Here, explain how you are modeling things to be true to the spirit of this problem.)

If the student does not cheat, then the X_i are his corresponding score discrepancies with the grader(s). Now, if the student cheats as in the process described above, we can model this by introducing a collection of independent Bernoulli random variables I_i with parameter $1/2$. Now, the scores discrepancies are modeled by $X_i + I_i$. Basically, the S_i retain their “inherent score” (ground truth) signification while we add on top of that a cheating choice to inflate by 1 point the reported score. We have:

$$\mathbf{E}[X_i + I_i] = \mathbf{E}[X_i] + \mathbf{E}[I_i] = 0 + \frac{1}{2} = \frac{1}{2}$$

$$\text{Var}(X_i + I_i) = \text{Var}(X_i) + \text{Var}(I_i) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$$

Let’s approximate the probability of being caught by using the CLT again.

$$\begin{aligned}
\mathbf{P}\left(\sum_{i=1}^{100} X_i + I_i \geq T\right) &= \mathbf{P}\left(\frac{\sum_{i=1}^{100} X_i + I_i - 100\mathbf{E}[X_1 + I_1]}{\sqrt{100\text{Var}(X_1 + I_1)}} \geq \frac{T - 100\mathbf{E}[X_1 + I_1]}{\sqrt{100\text{Var}(X_1 + I_1)}}\right) && (4)\\
&= \mathbf{P}\left(\frac{\sum_{i=1}^{100} X_i + I_i - 100\mathbf{E}[X_1 + I_1]}{\sqrt{100\text{Var}(X_1 + I_1)}} \geq \frac{T - 100\frac{1}{2}}{\sqrt{100\frac{3}{4}}}\right) && (5)\\
&\approx \mathbf{P}\left(\frac{\sum_{i=1}^{100} X_i + I_i - 100\mathbf{E}[X_1 + I_1]}{\sqrt{100\text{Var}(X_1 + I_1)}} \geq -4.43\right) && (6)
\end{aligned}$$

By the CLT, the risk of being caught is about at least

$$\frac{1}{\sqrt{2\pi}} \int_{-4.43}^{\infty} e^{-\frac{x^2}{2}}\, dx \approx 99.997\%$$

(g) If this student is willing to accept a 50% chance of being caught cheating, by how much can they systematically inflate their grade(i.e. inflates his/her grade to every question by some constant number of points)? Assume that they can inflate by no more than 3 points per question part. (Because inflating a 0 to a 4 would get them slammed the first time they did it.) We will assume that $5, 6$ and 7 are allowed as self-reported grades to keep things simple.

We use the same modeling idea as in the previous question. We are hence looking at $X_i + x$ where $x \leq 3$ is some constant number of points the systematic cheater inflates his/her grade by. We have

$$\mathbf{E}[X_i + x] = x$$

$$\text{Var}(X_i + x) = \frac{1}{2}$$

$$\mathbf{P}\left(\sum_{i=1}^{100} X_i + x \geq T\right) = \mathbf{P}\left(\frac{\sum_{i=1}^{100} X_i + x - 100\mathbf{E}[X_1 + x]}{\sqrt{100\text{Var}(X_1 + x)}} \geq \frac{T - 100\mathbf{E}[X_1 + x]}{\sqrt{100\text{Var}(X_1 + x)}}\right) \tag{7}$$

$$= \mathbf{P}\left(\frac{\sum_{i=1}^{100} X_i + x - 100\mathbf{E}[X_1 + x]}{\sqrt{100\text{Var}(X_1 + x)}} \geq \frac{T - 100x}{\sqrt{100\frac{1}{2}}}\right) \tag{8}$$

$$= \mathbf{P}\left(\frac{\sum_{i=1}^{100} X_i + x - 100\mathbf{E}[X_1 + x]}{\sqrt{100\text{Var}(X_1 + x)}} \geq \frac{11.7 - 100x}{\sqrt{50}}\right) \tag{9}$$

By the CLT, the risk of being caught is about at most 50% when x is such that

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{11.7-100x}{8.3}}^{\infty} e^{-\frac{x^2}{2}} dx \leq 50\%$$

which happens when $\frac{11.7-100x}{8.3} \geq 0$, that is $x \leq 0.12$ points.

(h) Is it worth trying to cheat on self-grading, even for a grade-maximizing sociopath[1] student with no internal sense of morality or "decent respect to the opinions of mankind." ?

Systematically cheating at 50% detection rate will give you an increase of at most $500 \times 0.12 = 60$ points on your homework grade. In proportion, this only represents $60/(4*500) = 3\%$ of your total homework grade. Taking into account the fact that your homework grade is only 15% of your total grade for this course, we are talking about $0.15 \times 0.03 = 0.45\%$ of total points you could get by cheating. With a probability one half of getting caught. That doesn't seem reasonably worth it.

7. Tolerating Errors

Assume Alice is trying to send m packets across a noisy channel to her friend Bob. The channel independently has probability p of generating an error on each packet. To account for errors, Alice sends $n > m$ packets. If Alice wants to ensure that Bob can correctly decode her entire message of m packets with probability at least r, how big can m be?

(a) Modeling each error as a coin toss with probability p, what is the probability that Bob cannot correctly decode Alice's message?

Denoting X as the number of errors in the n packets, each with probability p of generating an error. Then $X \sim Binom(n, p)$ with $E[X] = np$ and $Var(X) = np(1-p)$. Since we are dealing with general errors, we want

$$P(\text{no decoding}) = P(X > \frac{n-m}{2}) = \sum_{k=\lfloor \frac{n-m}{2}+1 \rfloor}^{n} \binom{n}{k} p^k (1-p)^{n-k}.$$

[1]This sociopathic model of a selfish maximizer is referred to as a "rational agent" in the formal language of economics. Showing that cheating is not substantially attractive in the context of a mechanism even for a sociopath is one way to show that the mechanism is probably safe against normal humans too — since actual human beings are caring, loving, altruistic, and have senses of integrity and honor.

(b) Assume $n = 100$, $r = 0.9$, and $p = 0.1$. What is the safe bound for m using the Chebyshev bound?
The Chebyshev bound is,

$$\begin{aligned}
P(X > \frac{n-m}{2}) &= P(X - np > \frac{n-m}{2} - np) \\
&\leq P(|X - np| > \frac{n-m}{2} - np) \\
&\leq \frac{np(1-p)}{\left(\frac{n-m}{2} - np\right)^2} < 0.1 \\
&\Rightarrow 10np(1-p) < \left(\frac{n-m}{2} - np\right)^2 \\
&\Rightarrow \sqrt{10np(1-p)} < \frac{n-m}{2} - np \\
&\Rightarrow m < n - 2np - 2\sqrt{10np(1-p)} \approx 61\text{packets.}
\end{aligned}$$

(c) Using the same information as part (b), approximate the safe m using the Central Limit Theorem.
For CLT,

$$P(X > \frac{n-m}{2}) \approx Q\left(\frac{\frac{n-m}{2} - np}{\sqrt{np(1-p)}}\right) \leq 0.1.$$

The corresponding m value can also be found by setting $z^* = 1.28 = \frac{\frac{n-m}{2} - np}{\sqrt{np(1-p)}}$ from a standard Normal distribution table, giving us $m = n - 2np - 2 \cdot 1.28\sqrt{np(1-p)} = 72$ packets.

(d) Using the same information as part (b), what is the safe bound for m using the relevant Chernoff bound for Binomial RV's?
For the Chernoff bound, we know from lecture notes that for unfair coin tosses, $X_1, X_2, \ldots, X_n$,

$$P\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq a\right) \leq e^{-nD(a||p)}$$

where $D(a||p) = aln\frac{a}{p} + (1-a)ln\frac{1-a}{1-p}$. For our case, we need $a = \frac{n-m}{2n}$ and

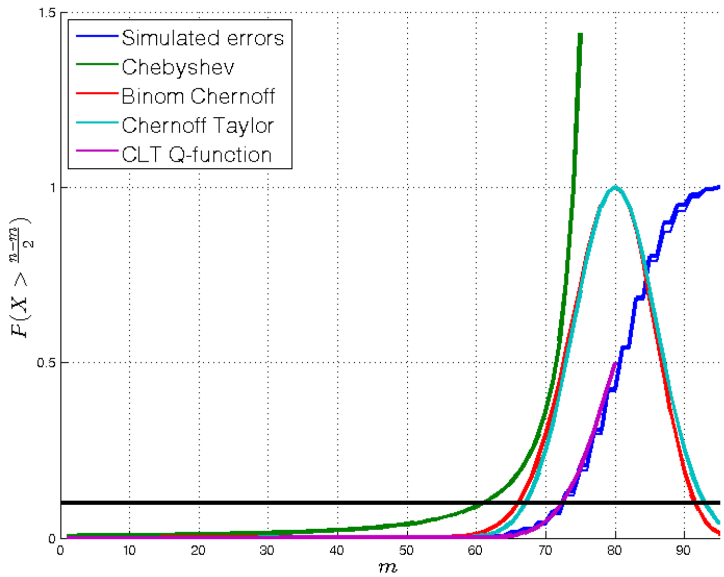
$$\begin{aligned}
P(X > \frac{n-m}{2}) &= P\left(\frac{X}{n} \geq \frac{n-m}{2n}\right) \\
&\leq e^{-nD(\frac{n-m}{2n}||p)} < 0.1.
\end{aligned}$$

Alternatively, assuming the deviation is small, we can use the form in lecture notes derived by the Taylor expansion.

$$P\left(\frac{X}{n} \geq p + \varepsilon\right) \leq exp\left(-n\frac{\varepsilon^2}{2p(1-p)}\right)$$

where $\varepsilon = \frac{n-m}{2n} - p$ for this problem.

For each of these bounds and approximations, we plot the corresponding functions vs m and see where the plots cross $(1-r) = 0.1$, assuming $\frac{n-m}{2} > np$. The plot should match the numbers calculated from the bounds and CLT. From the figure below, we see that indeed m can be up to 61 packets using the Chebyshev inequality, 67 packets using the Chernoff bound for binomial RV's, 68 packets using the Taylor expansion form of the Chernoff bound, and 72 packets using CLT.



8. Wrecking Ball

A new startup Milton/Alighieri Games has decided to create a family of games to cater to a previously underserved segment of the video game market. In their quest for a crossover hit, they score a marketing coup by getting the exclusive video game rights to use Miley Cyrus's megahit song, "Lucifer's Lament" which RCA Records had been promoting solely under its alternate title to avoid drawing controversy[2]. This family of game titles are all set in the mythical "War/Rebellion in Heaven" and include real time strategy games (like Starcraft), a Multiplayer Online Battle Arena game (like League of Legends), and an online Collectible Card Game (like Magic: The Gathering or Hearthstone).

You're in charge of one of the main characters, and are proposing a particular attack that you've entitled "Wrecking Ball" to tie in with the song. In this attack, at every turn, one of two things happens: With probability $\frac{2}{3}$, the attack succeeds and the character's health levels double (raising your max if necessary), with that many health points being drained from the opponent. With probability $\frac{1}{3}$, the attack backfires and the character's health gets reduced by a factor of 8, with the lost health being transfered to the opponent. Different turns are independent in whether the attack succeeds or backfires.

This problem is about understanding what happens if this attack is used repeatedly.

Let X_0 be the initial real-valued health of the character. So the health at the end of turn n is $X_n = X_0 \prod_{i=1}^{n} A_i$ where A_i is the random factor that results from the attack in turn i.

(a) Calculate the expected value of A_i.

Note that A_i is $\frac{1}{8}$ with probability 1/3, and 2 with probability 2/3. Hence,

$$\mathbb{E}[A_i] = (1/3)(\frac{1}{8}) + (2/3) * 2 = \frac{1}{24} + \frac{32}{24} = \frac{33}{24}$$

[2] Miley had grown increasingly annoyed since many people weren't appreciating the emotional nuances of her performance without seeing the mythic allusions resonating with it.

(b) Calculate the expected value of X_n. You can assume that $X_0 = x_0$ — a given constant.

We will use the fact that the A_i are independent and identically distributed to compute the expected value of X_n.

$$\begin{aligned}
\mathbb{E}[X_n] &= \mathbb{E}[X_0 \prod_{i=1}^{n} A_i] \\
&= x_0 \mathbb{E}[\prod_{i=1}^{n} A_i] \\
&= x_0 \prod_{i=1}^{n} \mathbb{E}[A_i] \\
&= x_0 \mathbb{E}[A_i]^n \\
&= x_0 \left(\frac{33}{24}\right)^n
\end{aligned}$$

(c) Seeing the previous calculations, the reasonably pious management gets very concerned about your proposed attack and worry that perhaps it is tilting the game in favor of this character.

Explain to them whether they are right or wrong to be worried by explaining to them what the typical range for X_n should be (with probability at least 90%) when n is large.

(Hint: figure out how to invoke the laws of large numbers.)

They are wrong, at least about the proposed attack being overpowered. First, think about the most likely sequences of A_i's. Since A_i is $\frac{1}{8}$ with probability 1/3 and 2 with probability 2/3, if we have a sequence of n A_i's we might expect about 1/3 of them to be $\frac{1}{8}$ and 2/3 of them to be 2. That should be *bad* for the character, since this would cause a net decrease in health. We will use laws of large numbers to formally justify that we are likely to get about 1/3 of the A_i's to be $\frac{1}{8}$ and about 2/3 of them to be 2, and hence with high probability we have an exponential *decrease* in health.

We define the indicator random variables Y_i to indicate the event that $A_i = \frac{1}{8}$.

$$Y_i = \begin{cases} 1 & \text{if } A_i = \frac{1}{8} \\ 0 & \text{if } A_i = 2 \end{cases}$$

Note that since the A_i are independent, the Y_i are iid Ber(1/3). We can use laws of large numbers to bound $\sum Y_i$. We will show how to get bounds with Chebyshevs and CLT; either one of these methods is fine. First, let's calculate the variance of $\sum_{i=1}^{n} Y_i$

$$\mathrm{Var}(\sum_{i=1}^{n} Y_i) = \sum_{i=1}^{n} \mathrm{Var}(Y_i) = \frac{1}{3} \cdot \frac{2}{3} \cdot n = \frac{2n}{9}$$

By Chebyshev's:

$$\Pr(|\sum_{i=1}^{n} Y_i - \frac{1}{3}n| \geq n\varepsilon) \leq \frac{2n}{9\varepsilon^2 n^2} = \frac{2}{9n\varepsilon^2}$$

We want to show that the probability that $\sum_{i=1}^{n} Y_i$ is within $n\varepsilon$ of its mean is at least 0.9. Equivalently, we want to show that the probability that it is farther than $n\varepsilon$ from its mean is at most 0.1. Hence we set the RHS equal to 0.1.

$$\frac{2}{9n\varepsilon^2} = 0.1$$
$$\sqrt{\frac{20}{9n}} = \varepsilon$$

Therefore, with probability at least 0.9, we have between $\frac{1}{3}-\varepsilon$ and $\frac{1}{3}+\varepsilon$ fraction of the A_i's are $\frac{1}{8}$. Note that the more that turn out to be $\frac{1}{8}$, the smaller the product of them all is. Hence, we get that

$$\left(\frac{1}{8}\right)^{(\frac{1}{3}+\varepsilon)n} 2^{(\frac{2}{3}-\varepsilon)n} \leq \prod_{i=1}^{n} A_i \leq \left(\frac{1}{8}\right)^{(\frac{1}{3}-\varepsilon)n} 2^{(\frac{2}{3}+\varepsilon)n}$$

$$\left(\frac{1}{8}\right)^{\frac{1}{3}n+\sqrt{\frac{20n}{9}}} 2^{\frac{2}{3}n-\sqrt{\frac{20n}{9}}} \leq \prod_{i=1}^{n} A_i \leq \left(\frac{1}{8}\right)^{\frac{1}{3}n-\sqrt{\frac{20n}{9}}} 2^{\frac{2}{3}n+\sqrt{\frac{20n}{9}}}$$

$$2^{-3\left(\frac{1}{3}n+\sqrt{\frac{20n}{9}}\right)} 2^{\frac{2}{3}n-\sqrt{\frac{20n}{9}}} \leq \prod_{i=1}^{n} A_i \leq 2^{-3\left(\frac{1}{3}n-\sqrt{\frac{20n}{9}}\right)} 2^{\frac{2}{3}n+\sqrt{\frac{20n}{9}}}$$

$$2^{-\frac{1}{3}n-4\sqrt{\frac{20n}{9}}} \leq \prod_{i=1}^{n} A_i \leq 2^{-\frac{1}{3}n+4\sqrt{\frac{20n}{9}}}$$

$$x_0 2^{-\frac{1}{3}n-4\sqrt{\frac{20n}{9}}} \leq X_n \leq x_0 2^{-\frac{1}{3}n+4\sqrt{\frac{20n}{9}}}$$

So X_n is decaying exponentially.

A Notable Alternative Solution: This method focuses on the exponents of the multipliers. Note that we can say $A_i = 2^{B_i}$ where

$$B_i = \begin{cases} -3 & \text{if } A_i = \frac{1}{8} \\ 1 & \text{if } A_i = 2 \end{cases}$$

Then we can use Chebyshev's to bound the probability that $\sum_{i=1}^{n} B_i$ deviates from $\frac{-1}{3}n$, which is its mean. It turns out that this method actually gives the same answer for both the Chebyshev and CLT methods. We will briefly walk through this method with the B_i's for Chebyshev's. Note that it is also valid with CLT, and can be easily checked against the CLT solution we have written here, since they use the same method, just a different random variable, and a fixed value of n will result in the same bound.

$$\mathbb{E}[\sum_{i=1}^{n} B_i] = n * (-3(\frac{1}{3}) + 1(\frac{2}{3})) = -\frac{1}{3}n$$

$$\text{Var}(\sum_{i=1}^{n} B_i) = n * \text{Var}(B_i) = n(9(\frac{1}{3}) + 1(\frac{2}{3}) - \left(-\frac{1}{3}\right)^2) = n(3 + \frac{2}{3} - \frac{1}{9}) = \frac{32}{9}n$$

Then Chebyshev's gives us that

$$\Pr(|\sum_{i=1}^{n} B_i - (-\frac{1}{3}n)| \geq n\varepsilon) \leq \frac{32}{9n\varepsilon^2}$$

Hence, we get

$$0.1 = \frac{32}{9n\varepsilon^2}$$
$$\varepsilon^2 = \frac{320}{9n}$$
$$\varepsilon = \sqrt{\frac{320}{9n}}$$

Exponentiating all the terms preserves the inequality, since 2^x is a monotonic function.

$$2^{-\frac{1}{3}n-n\varepsilon} \le \prod_{i=1}^{n} A_i \le 2^{-\frac{1}{3}n+n\varepsilon}$$

$$2^{-\frac{1}{3}n-\sqrt{\frac{320n}{9}}} \le \prod_{i=1}^{n} A_i \le 2^{-\frac{1}{3}n+\sqrt{\frac{320n}{9}}}$$

Multiplying by x_0:

$$x_0 2^{-\frac{1}{3}n-\sqrt{\frac{320n}{9}}} \le X_n \le x_0 2^{-\frac{1}{3}n+\sqrt{\frac{320n}{9}}}$$

Just to verify that we have the same bound as with the other method, substitute $\sqrt{\frac{320n}{9}} = 4\sqrt{\frac{20n}{9}}$:

$$x_0 2^{-\frac{1}{3}n-4\sqrt{\frac{20n}{9}}} \le X_n \le x_0 2^{-\frac{1}{3}n+4\sqrt{\frac{20n}{9}}}$$

By CLT:

We want to estimate $\Pr(|\sum_{i=1}^{n} Y_i - \frac{1}{3}n| \le n\varepsilon) = \Pr((\frac{1}{3}-\varepsilon)n \le \sum_{i=1}^{n} Y_i \le (\frac{1}{3}+\varepsilon)n)$, in order to get a bound on the number of attacks we need for $\sum_{i=1}^{n} Y_i$ to be close to the mean with high probability. Remember from your virtual labs that the sum of the independent Bernoulli random variables Y_i is approaching a normal (Gaussian) distribution, which is why we can apply CLT.

We will approach this problem in the standard way that we approach CLT problems: manipulate the random variable in question to have expectation 0 and variance 1, then find the values of n for which the area under the normal curve that you are interested in attains the probability that you want. This will be clearer when we start the manipulations:

$$\begin{aligned}
&\Pr((\frac{1}{3}-\varepsilon)n \le \sum_{i=1}^{n} Y_i \le (\frac{1}{3}+\varepsilon)n) \\
=&\Pr(-n\varepsilon \le \sum_{i=1}^{n} Y_i - \frac{1}{3}n \le n\varepsilon) \\
=&\Pr\left(\frac{-n\varepsilon}{\sqrt{\frac{2}{9}n}} \le \frac{\sum_{i=1}^{n} Y_i - \frac{1}{3}n}{\sqrt{\frac{2}{9}n}} \le \frac{n\varepsilon}{\sqrt{\frac{2}{9}n}}\right)
\end{aligned}$$

By the CLT, $\frac{\sum_{i=1}^{n} Y_i - \frac{1}{3}n}{\sqrt{\frac{2}{9}n}} \to N(0,1)$. Hence, we get that

$$\Pr\left(\frac{-n\varepsilon}{\sqrt{\frac{2}{9}n}} \le \frac{\sum_{i=1}^{n} Y_i - \frac{1}{3}n}{\sqrt{\frac{2}{9}n}} \le \frac{n\varepsilon}{\sqrt{\frac{2}{9}n}}\right) \approx \int_{-\frac{3}{\sqrt{2}}\sqrt{n}\varepsilon}^{\frac{3}{\sqrt{2}}\sqrt{n}\varepsilon} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

We want $\Pr(|\sum_{i=1}^{n} Y_i - \frac{1}{3}n| \le n\varepsilon) \ge 0.9$, so we set:

$$\int_{-\frac{3}{\sqrt{2}}\sqrt{n}\varepsilon}^{\frac{3}{\sqrt{2}}\sqrt{n}\varepsilon} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0.9$$

Since the probability that we are 2 standard deviations away from the mean of a normal distribution is roughly 95%, and we would expect that we want $\frac{3}{\sqrt{2}}\sqrt{n}\varepsilon$ to be a little under 2. Using a look-up table, we realize that we want to set $\frac{3}{\sqrt{2}}\sqrt{n}\varepsilon \approx 1.645$ to get the probability of being in this region to be 90%.

$$\frac{3}{\sqrt{2}}\sqrt{n}\varepsilon \approx 1.645$$
$$\varepsilon \approx \frac{1.645\sqrt{2}}{3\sqrt{n}}$$

Therefore, we get that:

$$\left(\frac{1}{8}\right)^{(\frac{1}{3}+\varepsilon)n} 2^{(\frac{2}{3}-\varepsilon)n} \le \prod_{i=1}^{n} A_i \le \left(\frac{1}{8}\right)^{(\frac{1}{3}-\varepsilon)n} 2^{(\frac{2}{3}+\varepsilon)n}$$

$$2^{-3n(\frac{1}{3}+\varepsilon)} 2^{n(\frac{2}{3}-\varepsilon)} \le \prod_{i=1}^{n} A_i \le 2^{-3n(\frac{1}{3}-\varepsilon)} 2^{n(\frac{2}{3}+\varepsilon)}$$

$$2^{-3n(\frac{1}{3}+\frac{1.645\sqrt{2}}{3\sqrt{n}})} 2^{n(\frac{2}{3}-\frac{1.645\sqrt{2}}{3\sqrt{n}})} \le \prod_{i=1}^{n} A_i \le 2^{-3n(\frac{1}{3}-\frac{1.645\sqrt{2}}{3\sqrt{n}})} 2^{n(\frac{2}{3}+\frac{1.645\sqrt{2}}{3\sqrt{n}})}$$

$$2^{-\frac{1}{3}n-\frac{4(1.645)\sqrt{2n}}{3}} \le \prod_{i=1}^{n} A_i \le 2^{-\frac{1}{3}n+\frac{4(1.645)\sqrt{2n}}{3}}$$

$$x_0 2^{-\frac{1}{3}n-\frac{4(1.645)\sqrt{2n}}{3}} \le X_n \le x_0 2^{-\frac{1}{3}n+\frac{4(1.645)\sqrt{2n}}{3}}$$

So as n grows large, X_n is decaying exponentially.

A notable *incorrect* solution arises if you try to use Chebyshev's directly on X_n. Why might this be? Well, one of the problems is that the variance of the random variable X_n is growing exponentially with n. In particular, it is a faster growing exponential than the expected value of X_n, or even the square of the expected value of X_n. Hence, the only way to get a useful bound on the probability (i.e. not upper bounding with 1) from Chebyshev's is to choose ε to be larger than the expected value of X_n. Then the region whose probability you are bounding must include 0, as well as extremely large values that grow exponentially with n. This makes the bound effectively useless. It should intuitively make sense why this method is doomed to fail: in this case, the expected value of X_n is *not a value of X_n that we get with high probability*. Hence, trying to justify how X_n is very likely close to its mean is futile, because it is simply *not*. This type of solution should get at most half credit.

9. A Chernoff Bound

In this problem, you will show that the probability that the average of specific 3-valued independent random variables is "far away" from its expectation decays exponentially in the number of random variables in the sum. You have already seen how to do this for independent Bernoulli random variables via the Chernoff bound in the notes; now, we will explore how to do this for the following independent 3-valued random variables. Let X_i be a sequence of independent identically distributed random variables such that:

$$X_1 = \begin{cases} 2 & \text{with prob. } 1/3 \\ 1 & \text{with prob. } 1/3 \\ 0 & \text{with prob. } 1/3 \end{cases}$$

Let $X = \sum_{i=1}^{n} X_i$.

a) Argue why $\mathbf{P}(X \geq na) = \mathbf{P}(e^{sX} \geq e^{nsa})$ for all values $s \geq 0$. Our goal is to come up with an upper bound for this quantity that decays exponentially with n. Argue why we should only concern ourselves with $1 < a < 2$ (Is $a \leq 1$ interesting for this bound?).

   We realize that $X \geq na \iff sX \geq nsa$, for $s \geq 0$. Hence $\mathbf{P}(X \geq na) = \mathbf{P}(sX \geq nsa)$. Furthermore, since e^x is a monotonic increasing function, we have that $sX \geq nsa \iff e^{sX} \geq e^{nsa}$. Hence $\mathbf{P}(X \geq na) = \mathbf{P}(sX \geq nsa) = \mathbf{P}(e^{sX} \geq e^{nsa})$.

   We only care about $a > 1$ because if $a \leq 1$, since 1 is the mean X_i, then $\mathbf{P}(X \geq a)$ will get fairly large. This is because we know from our virtual labs and the WLLN that $\frac{1}{n}X$ will tend to cluster about the mean of X_i, which is 1, as n increases. So if $a \leq 1$, the probability in question is not actually small. Furthermore, if $a > 2$, since X_i can take value at most 2 we get $\mathbf{P}(X \geq na) = 0$. If $a = 2$, then we get $\mathbf{P}(X \geq 2n) = \mathbf{P}(X = 2n) = \prod_{i=1}^{n} \mathbf{P}(X_i = 2) = (\frac{1}{3})^n$, so we already know that the probability is either decaying exponentially or 0 for $a \geq 2$.

b) Argue why $\mathbf{P}(X \leq na) = \mathbf{P}(e^{sX} \geq e^{nsa})$ for all values $s \leq 0$. Argue why we should only concern ourselves with $0 \leq a < 1$.

   Very similar to the last part. $X \leq na \iff sX \geq nsa$ for $s \leq 0$, and $sX \geq nsa \iff e^{sX} \geq nsa$ since e^x is monotonic increasing. Hence, we get $\mathbf{P}(X \leq na) = \mathbf{P}(e^{sX} \geq e^{nsa})$. Similar to the last part, since the mean of X_i is 1, if we had $a \geq 1$ then $\mathbf{P}(X \leq na)$ would be large as we know from the WLLN that $\frac{1}{n}X$ would tend to cluster around 1. Hence we should concern ourselves with $a < 1$. Furthermore, since the X_i are nonnegative, if $a < 0$ we get $\mathbf{P}(X \leq na) = 0$. All the X_i have to be 0 in order to have $X = na$ if $a = 0$, so if $a = 0$ we get $\mathbf{P}(X \leq na) = \left(\frac{1}{3}\right)^n$. Therefore, we know the probability is either 0 or exponentially decreasing if $a \leq 0$.

c) Since it was the right hand side of both the above equalities, let's focus on bounding $\mathbf{P}(e^{sX} \geq e^{nsa})$. Show that $\mathbf{P}(e^{sX} \geq e^{nsa}) \leq e^{-n(sa - \ln M(s))}$, where $M(s) = \mathbf{E}[e^{sX_i}]$.

   Apply Markov's inequality. This directly gives $\mathbf{P}(e^{sX} \geq e^{nsa}) \leq \frac{\mathbf{E}[e^{sX}]}{e^{nsa}}$. We have from the previous part that $\mathbf{P}(e^{sX} \geq e^{nsa}) \leq \frac{\mathbf{E}[e^{sX}]}{e^{nsa}}$. It seems we already have part of the expression – the main concern is to turn $\mathbf{E}[e^{sX}]$ into something in terms of $\mathbf{E}[e^{sX}i]$. We can write:

$$\begin{aligned} \mathbf{E}[e^{sX}] &= \mathbf{E}[e^{s\sum_{i=1}^{n} X_i}] \\ &= \mathbf{E}[\prod_{i=1}^{n} e^{sXi}] \end{aligned}$$

By noting that the X_i are independent, we know that the e^{sXi} are independent. Hence:

$$\begin{aligned}
&= \prod_{i=1}^{n} \mathbf{E}[e^{sXi}] \\
&= \mathbf{E}[e^{sXi}]^n \\
&= M(s)^n
\end{aligned}$$

Now we substitute this into our bound: $\mathbf{P}(e^{sX} \geq e^{nsa}) \leq \frac{\mathbf{E}[e^{sX}]}{e^{nsa}} = \frac{M(s)^n}{e^{nsa}} = e^{-nsa} e^{n \ln M(s)} = e^{-n(sa - \ln M(s))}$.

d) Compute $M(s) = \mathbf{E}[e^{sX_i}]$.

Using the definition of expectation, we get that

$$\begin{aligned}
M(s) = \mathbf{E}[e^{sX_i}] &= (\frac{1}{3})e^{s*0} + (\frac{1}{3}e^{s*1}) + (\frac{1}{3}e^{s*2}) \\
&= \frac{1}{3} + \frac{1}{3}e^{s} + \frac{1}{3}e^{2s}
\end{aligned}$$

.

e) Now we have the tools to continue part a) and start to bound $\mathbf{P}(X \geq na)$. Use the parts above to conclude that $\mathbf{P}(X \geq na) = \mathbf{P}(e^{sX} \geq e^{nsa}) \leq e^{-n \max\limits_{s \geq 0}(sa - \ln M(s))}$.

We know that $\mathbf{P}(X \geq na) = \mathbf{P}(e^{sX} \geq e^{nsa}) \leq e^{-n(sa - \ln M(s))}$ for all $s \geq 0$. Since this holds for $s \geq 0$, it also holds for the $s \geq 0$ that maximizes the function $sa - \ln M(s)$. Hence, we get $\mathbf{P}(X \geq na) \leq e^{-n \max\limits_{s \geq 0}(sa - \ln M(s))}$

f) Plug $s = 0$ into $sa - \ln M(s)$. Given this value, what can you conclude about the *maximum* value of $sa - \ln M(s)$ for $s \geq 0$?

When we plug $s = 0$ into $sa - \ln M(s)$, we get

$$0(a) - \ln M(0) = 0 - \ln(1/3 + 1/3 + 1/3) = 0.$$

Hence, the maximum value of $sa - \ln M(s)$ over $s \geq 0$ must be nonnegative.

g) Give the value of s that maximizes $sa - \ln M(s)$ for $s \geq 0$. Show that this is a positive value for s given that $1 < a < 2$. What does the fact that this is a *positive* value for s tell you about the value of $sa - \ln M(s)$ when maximized over $s \geq 0$? Potentially Useful Hint: If you want to show that $(x+y)/z > 1$, you can start by determining whether x, y and z are nonnegative or negative. Once you know this, you can manipulate the inequality $(x+y)/z > 1$ to get an equivalent inequality that you can verify more easily. Also, if α and β are positive, then $\alpha > \beta \iff \alpha^2 > \beta^2$.

We will take the derivative of $sa - \ln M(s)$ with respect to s, set that to 0 and solve for s. This will give the value of s that maximizes this expression. Then we use the fact that $1 < a < 2$ to demonstrate that this maximum value of s is greater than 0, which means that $sa - \ln M(s) > 0$ for this value of s, by the previous part.

$$\begin{aligned}
\frac{d}{ds}(sa - \ln M(s)) &= 0 \\
a - \frac{1}{(1/3)(1 + e^s + e^{2s})}(1/3)(e^s + 2e^{2s}) &= 0 \\
a(1 + e^s + e^{2s}) &= e^s + 2e^{2s} \\
(2-a)e^{2s} + (1-a)e^s - a &= 0
\end{aligned}$$

We apply the quadratic equation to solve for e^s:

$$e^s = \frac{-(1-a) \pm \sqrt{(1-a)^2 + 4(2-a)(a)}}{2(2-a)}$$

Applying $1 < a < 2$, we see that $-(1-a) > 0$, $\sqrt{(1-a)^2 + 4(2-a)(a)}$ is real and greater than 0, and $2(2-a) > 0$. Since $\sqrt{(1-a)^2 + 4(2-a)(a)} > 0$, choosing $e^s = \frac{-(1-a)+\sqrt{(1-a)^2+4(2-a)(a)}}{2(2-a)}$ is the larger solution, and hence is our best candidate for a solution having $s > 0$. So we will focus on this choice of e^s. Our goal is now to show that $e^s > 1$, since $e^s > 1 \iff s > 0$, due to the fact that e^x is a monotonic increasing function.

In accordance with the hint, we want to show that $e^s = \frac{x+y}{z} > 1$, where $x = -(1-a) > 0$, $y = \sqrt{(1-a)^2 + 4(2-a)(a)} > 0$, and $z = 2(2-a) > 0$. Now that we have this information, we can manipulate this inequality to come up with an equivalent inequality that is easier to prove:

$$\begin{aligned} \frac{x+y}{z} &> 1 \\ x+y &> z \\ y &> z-x \end{aligned}$$

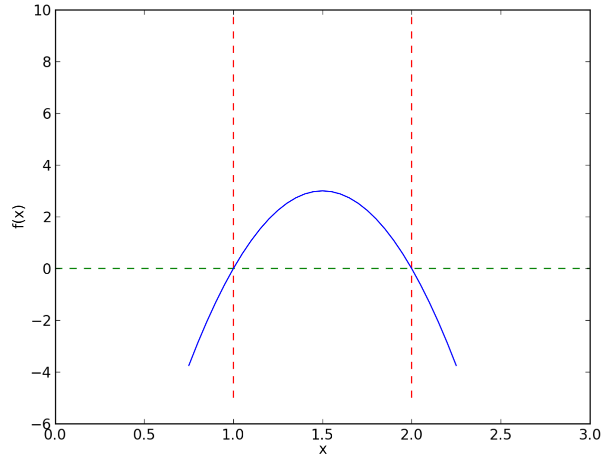
Note here that if $x > z$ then this inequality is trivially true, hence we are done. Otherwise, $z - x \geq 0$ and we continue:

$$y^2 > (z-x)^2$$

Now we substitute back in for x, y and z:

$$\begin{aligned} (1-a)^2 + 4(2-a)a &> (4-2a+1-a)^2 \\ 1-2a+a^2+8a-4a^2 &> (5-3a)^2 \\ 1+6a-3a^2 &> 25-30a+9a^2 \\ -12a^2+36a-24 &> 0 \end{aligned} \tag{10}$$

We notice that at $a = 1$ and $a = 2$ the LHS is 0. Since the LHS is concave down, we know it has a single local maximum and no local minimums, so its value on $a \in [1,2]$ will be minimized at either $a = 1$ or $a = 2$. In order to verify that its value on on $(1,2)$ satisfies this inequality, we take derivatives with respect to a at $a = 1$ and $a = 2$. We just need to make sure that the LHS is non-zero when $1 < a < 2$. This is apparent by the shape of the parabola, and also follows from the definition of being a concave function, namely that any line segment between two points on the graph lies below the graph, as illustrated below.



Hence we know that $-12a^2+36a-24>0$ for $1<a<2$, and since the manipulations to the inequality were reversible we conclude that the maximizing value of $e^s>1$ for $1<a<2$. Therefore $s=\ln\frac{-(1-a)+\sqrt{(1-a)^2+4(2-a)(a)}}{2(2-a)}>0$ for $1<a<2$.

Alternative way to verify that $-12a^2+36a-24>0$:
We can simplify Inequality (10) a little further.

$$\begin{aligned}
-12a^2+36a-24 &> 0\\
a^2-3a+2 &< 0\\
(a-2)(a-1) &< 0
\end{aligned}$$

Since $1<a<2$, $(a-2)$ must be negative and $(a-1)$ must be positive, hence their product is less than zero as we wanted.

What does this mean? It means that the unique maximum value of $sa-\ln M(s)$ for $s\geq 0$ occurs at a nonzero value of s. Since $sa-\ln M(s)=0$ when $s=0$, and the maximum value occurs elsewhere, the maximum value must be strictly greater than 0. Therefore, we can conclude that the maximum value of $sa-\ln M(s)$ for $s\geq 0$ is positive.

h) Give an upper bound that decays exponentially with increasing n for $\mathbf{P}(X\geq na)$, using your previous parts to justify it.
Plugging in the answer from above, we get

$$\begin{aligned}
\mathbf{P}(X\geq na) &\leq e^{-n\max\limits_{s\geq 0}(sa-\ln M(s))}\\
&= e^{-n\left(a\ln\frac{-(1-a)+\sqrt{(1-a)^2+4(2-a)(a)}}{2(2-a)}-\ln\left(\left(\frac{1}{3}\right)\left(1+\frac{-(1-a)+\sqrt{(1-a)^2+4(2-a)(a)}}{2(2-a)}+\left(\frac{-(1-a)+\sqrt{(1-a)^2+4(2-a)(a)}}{2(2-a)}\right)^2\right)\right)\right)}
\end{aligned}$$

We know that this is an exponentially decaying bound, since we used a value of s that makes $sa-\ln M(s)>0$, which we justified in the previous parts.

i) **(Optional)** Now complete the exponential upper bound for b). In part b), $s\leq 0$ and we want to bound $\mathbf{P}(X\leq na)$. With this in mind, repeat similar arguments to those in the previous parts to come up with a bound that decays exponentially with n for $\mathbf{P}(X\leq na)$, where $a<1$.
This is very similar to the previous parts. The important part for people who did this optional part is to note the differences from the previous parts. We still end up with the same Chernoff bound, but

the maximization is now over $s \leq 0$ rather than $s \geq 0$. Furthermore, we have $0 < a < 1$ rather than $1 < a < 2$. These are both restrictions that you justified in part b). Note that the RHS of the inequality in b) is the same as the RHS in a), so we can reuse a lot of the work. Everything is the same until we get to the step where we have to justify that our choice of s that maximizes $sa - \ln M(s)$.

After taking the derivative to maximize $sa - \ln M(s)$, we get the same expression, namely $e^s = \frac{-(1-a) \pm \sqrt{(1-a)^2 + 4(2-a}}{2(2-a)}$

We choose the '+' solution since e^s should be a positive number. So we start with $e^s = \frac{-(1-a)+\sqrt{(1-a)^2+4(2-a)a}}{2(2-a)}$, and want to apply $0 < a < 1$ to show that $e^s < 1$, and therefore that $s < 0$. We employ the same method as before, manipulating the inequality we want to prove until it becomes something we can prove more easily.

$$\frac{-1(1-a) + \sqrt{(1-a)^2 + 4(2-a)a}}{2(2-a)} < 1$$

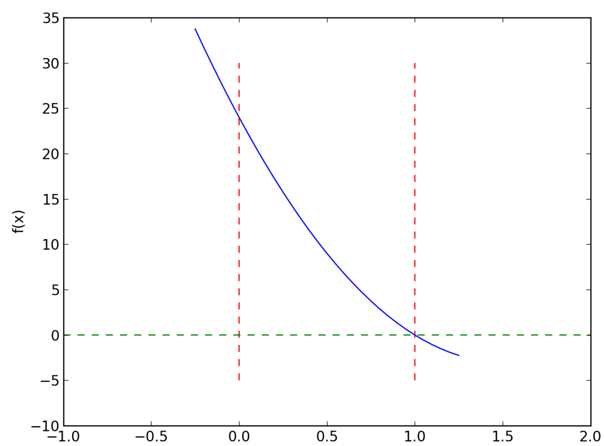
Since $a < 1$, we have that $2(2-a) > 0$. Multiply both sides by it.

$$-(1-a) + \sqrt{(1-a)^2 + 4(2-a)a} < 2(2-a)$$
$$\sqrt{(1-a)^2 + 4(2-a)a} < 5 - 3a$$

Since $a < 1$, we know that both sides are positive. Square both sides.

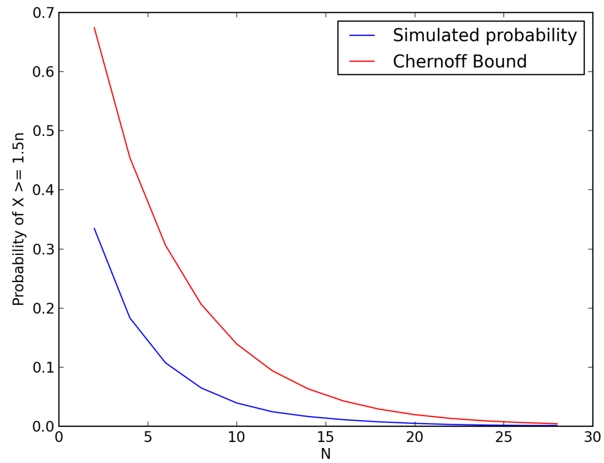
$$(1-a)^2 + 4(2-a)a < (5-3a)^2$$
$$1 - 2a + a^2 + 8a - 4a^2 < 9a^2 - 30a + 25$$
$$1 + 6a - 3a^2 < 9a^2 - 30a + 25$$
$$0 < 12a^2 - 36a + 24$$

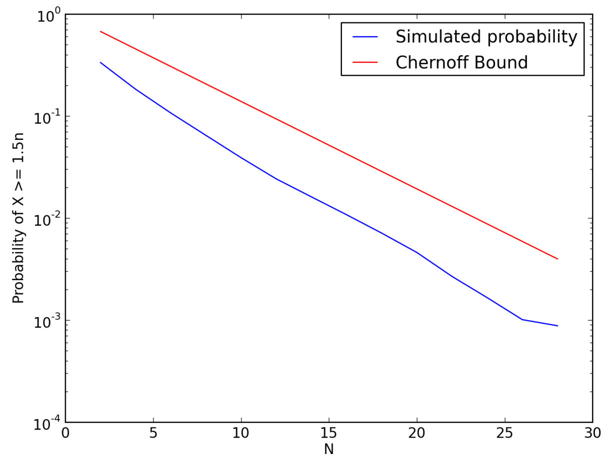
Now we have a much easier expression to analyze. Note that the RHS is convex (concave up), and hence any line segment between two points lies above the graph. We note that the graph has its local minimum at $a = 1.5$, and since it is concave up this means that it is strictly decreasing on the interval $(0,1)$. Therefore, since its value at 1 is 0, its value on $(0,1)$ must be strictly positive. Plot:



j) **(Virtual Lab, Optional)** For $a = 1.5$, do an appropriate simulation using a computer and plot the actual probability of having this sort of large deviation happen as compared to what your bounds above say. Use the appropriate kind of axes to judge the quality of the bound.

Once you fill in the appropriate function for $M(s)$ as well as the value for s that maximizes $sa - \ln M(s)$ that you have computed for previous parts, you should get plots similar to the following:





The plot with linear axes is heartening, as it shows that Chernoff is indeed an exponential upper bound for the simulated probability. The quality of the bound, however, is best gauged by the plot with the log-scaled y-axis. Note that the slopes of the Chernoff bound and the simulated probability are quite similar – the Chernoff bound is decreasing *slightly* less than the simulated probability. However, the similarity of these slopes indicates that the exponential decay in the Chernoff bound is quite close to the exponential decay in the real probability. Hence, we can conclude that the Chernoff bound, while not decaying exactly like the real probability, is fairly good.

10. (Optional) Binomial CLT

In this question we will explicitly see why the central limit theorem holds for the binomial distribution as the number of coin tosses grows.

Let X be the random variable showing the total number of heads in n independent coin tosses.

(a) Compute the mean and variance of X. Show that $\mu = E[X] = n/2$ and $\sigma^2 = \text{Var}[X] = n/4$.

We can write X as a sum: $X = Y_1 + \cdots + Y_n$ where each Y_i is a Bernoulli random variable; i.e. $Y_i = 1$ when the i-th coin toss is heads and 0 if it is tails. Then from linearity of expectation we have

$$E[X] = E[Y_1] + \cdots + E[Y_n] = nE[Y_1] = n/2$$

where we used the fact that all Y_i have the same expectation which is $1/2$.

To compute the variance, note that because $Y_1, \dots, Y_n$ are independent, we can decompose the variance into a sum of variances. Therefore we have the following:

$$\mathrm{Var}[X] = \mathrm{Var}[Y_1] + \cdots + \mathrm{Var}[Y_n]$$

Now in order to compute $\mathrm{Var}[Y_i]$, note that by definition $\mathrm{Var}[Y_i] = E[(Y_i - E[Y_i])^2]$. We know that $E[Y_i] = 1/2$, so $\mathrm{Var}[Y_i] = E[(Y_i - 1/2)^2]$. But note that Y_i takes the values 0 and 1, therefore $(Y_i - 1/2)^2$ always takes the value $1/4$. So its expectation is also $1/4$. This means that

$$\mathrm{Var}[X] = 1/4 + \cdots + 1/4 = n/4$$

(b) Prove that $\Pr[X = k] = \binom{n}{k}/2^n$.

The number of configurations of heads/tails for the coins that result in k coins being heads is $\binom{n}{k}$, since there are this many ways to pick the positions of the heads. Each configuration of heads/tails is equally likely and they each have probability $1/2^n$, because the coins are independent and the probability of each coin being in a specific state is $1/2$. So the total probability for the event $X = k$ is $\binom{n}{k}/2^n$.

(c) Show by using Stirling's formula that $\Pr[X = k] \simeq \frac{1}{\sqrt{2\pi}}(\frac{n}{2k})^k(\frac{n}{2(n-k)})^{n-k}\sqrt{\frac{n}{k(n-k)}}$.

In general we expect $2k$ and $2(n-k)$ to be close to n for the probability to be non-negligible. When this happens we expect $\sqrt{\frac{n}{k(n-k)}}$ to be close to $\sqrt{\frac{n}{(n/2)\times(n/2)}} = 2/\sqrt{n}$. So replace that part of the formula by $2/\sqrt{n}$.

We need Stirling's formula to approximate the $\binom{n}{k}$ part. Remember that $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, and Stirling's approximation says that $m! \simeq \sqrt{2\pi m}(m/e)^m$. Therefore we have

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \simeq \frac{\sqrt{2\pi n}(n/e)^n}{\sqrt{2\pi k}(k/e)^k\sqrt{2\pi(n-k)}((n-k)/e)^{n-k}}$$

We can break the $(n/e)^n$ part into $(n/e)^k(n/e)^{n-k}$ and then combine these with the denominator. By doing this the part $\frac{(n/e)^k}{(k/e)^k}$ becomes $(n/k)^k$ and the part $\frac{(n/e)^{n-k}}{((n-k)/e)^{n-k}}$ becomes $(n/(n-k))^{n-k}$.

As for the parts under the square root, one of the $\sqrt{2\pi}$'s in the denominator cancels the one in the numerator and therefore only one remains in the denominator. We also get $\frac{\sqrt{n}}{\sqrt{k}\sqrt{n-k}} = \sqrt{\frac{n}{k(n-k)}}$.

Therefore we have

$$\binom{n}{k} \simeq \frac{1}{\sqrt{2\pi}}(n/k)^k(n/(n-k))^{n-k}\sqrt{\frac{n}{k(n-k)}}$$

Now we need to divide both sides by 2^n to get to $\Pr[X = k]$. We can write $2^n = 2^k 2^{n-k}$ and merge each term into the corresponding power. We get

$$\binom{n}{k}/2^n \simeq \frac{1}{\sqrt{2\pi}}(n/2k)^k(n/2(n-k))^{n-k}\sqrt{\frac{n}{k(n-k)}}$$

which is what we wanted to prove.

If we replace the $\sqrt{\frac{n}{k(n-k)}}$ part with $2/\sqrt{n}$ we get

$$\binom{n}{k}/2^n \simeq \frac{1}{\sqrt{2\pi}}(n/2k)^k(n/2(n-k))^{n-k}\frac{2}{\sqrt{n}}$$

(d) In order to normalize X, we need to subtract the mean, and divide by the standard deviation. Let $Y = (X-\mu)/\sigma$ be the normalized version of X. Note that Y is a discrete random variable. Determine the set of values that Y can take. What is the distance d between two consecutive values?

The set of values that X can take is $\{0,\dots,n\}$. Therefore the set of values that Y can take is $(i-n/2)/(\sqrt{n}/2) = (2i-n)/\sqrt{n}$ for $i = 0,\dots,n$. Originally (for X) the distance between consecutive values is 1, but since we are dividing by $\sigma = \sqrt{n}/2$, this distance becomes $1/(\sigma) = 2/\sqrt{n}$. Note that subtracting the mean has no effect on the distance between consecutive points.

(e) Let $X = k$ correspond to the event $Y = t$. Then $X \in [k-0.5, k+0.5]$ corresponds to $Y \in [t-d/2, t+d/2]$. For conceptual simplicity, it is reasonable to assume that the mass at point t is distributed uniformly on the interval $[t-d/2, t+d/2]$. We can capture this with the idea of a "probability density" and say that the probability density on this interval is just $\Pr[Y=t]/d = \Pr[X=k]/d$.

Compute k as a function of t. Then substitute that for k in the approximation you have from part (c) to find an approximation for $\Pr[Y=t]/d$. Show that the end result is equivalent to

$$\frac{1}{\sqrt{2\pi}}\left((1+\frac{t}{\sqrt{n}})^{1+\frac{t}{\sqrt{n}}}(1-\frac{t}{\sqrt{n}})^{1-\frac{t}{\sqrt{n}}}\right)^{-n/2}$$

We know how to compute t as a function of k. We simply do what we do to X to get to Y, i.e. subtract the mean of X and divide by its standard deviation. Therefore $t = (k-n/2)/(2/\sqrt{n}) = \frac{2k-n}{\sqrt{n}}$. Now to reverse this process and go from t to k we need to do the reverse, i.e. first multiply by σ and then add the mean of X. Therefore $k = \sqrt{n}t/2 + n/2 = \frac{\sqrt{n}t+n}{2}$.

Now note that $n/(2k) = n/(\sqrt{n}t+n) = ((\sqrt{n}t+n)/n)^{-1} = (1+\frac{t}{\sqrt{n}})^{-1}$. Similarly we have $n/2(n-k) = n/(2n-n-\sqrt{n}t) = ((n-\sqrt{n}t)/n)^{-1} = (1-\frac{t}{\sqrt{n}})^{-1}$.

Now we can write $(n/2k)^k$ as $(1+\frac{t}{\sqrt{n}})^{-k}$ and $(n/2(n-k))^{n-k}$ as $(1-\frac{t}{\sqrt{n}})^{-(n-k)}$. To get rid of k even in the exponent we need to write it in terms of t. We have $-k = -(\sqrt{n}t+n)/2 = -(n/2)(1+\frac{t}{\sqrt{n}})$. Similarly we have $-(n-k) = -(n-n/2-\sqrt{n}t/2) = -(n/2)(1-\frac{t}{\sqrt{n}})$.

Now it's time to assemble the pieces. Remember that we had

$$\Pr[X=k] = \Pr[Y=t] \simeq \frac{1}{\sqrt{2\pi}}(n/2k)^k(n/2(n-k))^{n-k}\frac{2}{\sqrt{n}}$$

Replacing the parts $(n/2k)^k$ and $(n/2(n-k))^{n-k}$ the way we described gives us

$$\Pr[Y=t] \simeq \frac{1}{\sqrt{2\pi}}(1+\frac{t}{\sqrt{n}})^{-(n/2)(1+\frac{t}{\sqrt{n}})}(1-\frac{t}{\sqrt{n}})^{-(n/2)(1-\frac{t}{\sqrt{n}})}\frac{2}{\sqrt{n}}$$

We need to approximate $\Pr[Y=t]/d$, and note that $d = 2/\sqrt{n}$ which is exactly the last term appearing in our approximation of $\Pr[Y=t]$. So by dividing by d, that term simply cancels out and we get

$$\Pr[Y=t]/d \simeq \frac{1}{\sqrt{2\pi}}\left((1+\frac{t}{\sqrt{n}})^{1+\frac{t}{\sqrt{n}}}(1-\frac{t}{\sqrt{n}})^{1-\frac{t}{\sqrt{n}}}\right)^{-n/2}$$

(f) As you can see, we have expressions of the form $(1+x)^{1+x}$ in our approximation. To simplify them, write $(1+x)^{1+x}$ as $\exp(\ln(1+x)(1+x))$ and then replace $\ln(1+x)(1+x)$ by its Taylor series.
The Taylor series up to the x^2 term is $\ln(1+x)(1+x) \simeq x + x^2/2 + \dots$ (feel free to verify this by hand). Use this to simplify the approximation from the last part. In the end you should get the familiar formula that appears inside the CLT:

$$\frac{1}{\sqrt{2\pi}}e^{-t^2/2}$$

(The CLT is essentially taking a sum with lots of tiny slices and approximating it by an integral of this function. Because the slices are tiny, dropping all the higher-order terms in the Taylor expansion is justified.)

The term $(1+x)^{1+x}$ as suggested can be written as $\exp(\ln(1+x)(1+x))$ and then $(1+x)\ln(1+x)$ can be replaced by its Taylor series up to the first few terms, i.e. by $x+x^2/2$. Now if we also do this for $-x$, we get $(1-x)^{1-x} = \exp(\ln(1-x)(1-x)) \simeq \exp(-x+x^2/2)$. By multiplying our approximation for x and $-x$ we get

$$(1+x)^{1+x}(1-x)^{1-x} \simeq \exp(x+x^2/2)\exp(-x+x^2/2) = \exp(x^2)$$

Now if we let $x = \frac{t}{\sqrt{n}}$ we get an approximation for the term inside parenthesis from last part. We get

$$(1+\frac{t}{\sqrt{n}})^{1+\frac{t}{\sqrt{n}}}(1-\frac{t}{\sqrt{n}})^{1-\frac{t}{\sqrt{n}}} \simeq \exp((t/\sqrt{n})^2) = e^{t^2/n}$$

Therefore we have

$$\Pr[Y=t]/d \simeq \frac{1}{\sqrt{2\pi}}(e^{t^2/n})^{-n/2} = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$$

which is the formula for the probability density function of the standard normal random variable.

11. 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?