Midterm 1

No collaboration permitted.
One sheet of notes, both sides, permitted. Turn it in with your exam.
Be clear and precise in your answers. Illegible exams cannot be graded.
Write your name and student ID on EVERY sheet.
Come to the front if you have a question.
There are a total of 120 points on this exam. Anything higher than an 70 is a great score.
Goodluck!

Figure 1: Gangnam Style

1. (20 points) True or False? Prove or give a counterexample.

   (a) (10 points) If X and Y are two random variables, then $\text{var}(X) + \text{var}(Y) = \text{var}(X + Y)$.

   (b) (10 points) If X is a general real-valued random variable, then $P(|X| \geq a) \leq \frac{E[|X|]}{a}$.

2. (50 points) An online advertiser would like to estimate the effectiveness of an ad for hair products, i.e. the probability p that a customer will buy the product after watching the ad.

   (a) (10 points) Ad effectiveness is estimated by showing the ad to many people and then observing how many of them buy the product. **If you would like to estimate p with 99% confidence to within an error of 0.01 how many people would you choose to show the ad to as a test?**

   (b) (10 points) We complete the estimation experiment above and find that $p = 0.6$. Now suppose customers can also independently (without being shown the ad) discover and buy the product with probability $q = 0.3$, and the ad is shown to 50% of people. **What is the probability that someone who bought the product had seen the ad?**

(c) (30 points) Advertisers would like to market the product to both people with long hair and people with short hair. However, they are unsure about the fraction of long-haired people, h, in the population, but know that $0.4 \le h \le 0.6$. (Assume long and short haired form a partition of the space.). Both sets respond to the ad differently, and the advertisers would like to estimate the effectiveness of the ad for long-haired people, l. **How many people should we sample to estimate l to within 0.01 with 98% confidence?**

HINT: The 98% confidence tells us that we are allowed to be outside our error margin 2% of the time. Use this error margin in two parts.

First, think of a set of just long-haired people. How large would you choose that set so that we have a small estimation error? Use part of the 2% error to bound this. Feel free to reuse work you have already done. Then, think about how you could guarantee enough long-haired people in the sample (this is the second part of your error margin).

Blank page for work

3. (50 points) Psy's Gangnam style is already one of the most viral videos on YouTube today. As of this weekend it had 305,016,460 likes on YouTube, and is now apparently in the Guinness Book of World Records for the most likes on YouTube. As an engineer at YouTube, you'd need to have some means of estimating how viral videos are distributed, so that you can plan to devote adequate resources. This question plays with some toy models to estimate this growth.

   Let us consider a social network like Facebook and see how this video might spread. To start, we will have a simple model, with Facebook friends A_i, $0 \leq i$ arranged on a linear graph as below.

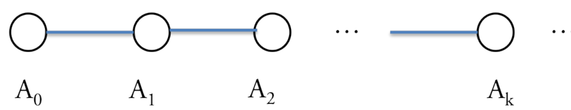

Figure 2: Graph

   The root-node A_0 is friends with A_1. A_1 is friends with A_0 and A_2 and so on. **The root-node A_0 discovers the Gangnam style video and posts it to Facebook with probability p. Assume, the remaining nodes have probability 0 of discovering the video independently. Given that A_0 posted the video, A_1 posts it to her wall with probability q. Similarly, this continues down the chain: given that A_{i-1} posted the video, A_i posts it with probability q.**

   Let X_i be an indicator random variable such that $X_i = 1$ if A_i posted the video.

   (a) (5 points) Find the probability that A_3 posts the video given that A_0 posted it.

   (b) (5 points) Find the probability that A_k posts the video given that A_0 posted it.

(c) (10 points) Find the joint PMF $P_{X_0,X_1}(x_0, x_1)$.

(d) (10 points) Consider an infinite number of people in the chain, i.e. $k \to \infty$. What is the expected number of people who post the video?

Blank page for work.

(e) (20 points) Now consider a tree-like graph instead of a linear graph. Each person in this graph has two new friends. A_0 is the root at level-0, and has two friends at level-1, A_{11} and A_{12} and so on. The video posting rule remains the same A_{ij} posts with probability q given that his friend on level $i-1$ posted the video, and none of the nodes other than A_0 independently discovers the video. The root node, A_0 posts the video with probability p. **Let the number of levels in the graph go to infinity. What is the expected number of people who will post the video? How does the value of q affect this expectation?**

HINT: What is the probability that A_{31} posts the video? How about A_{k1}?

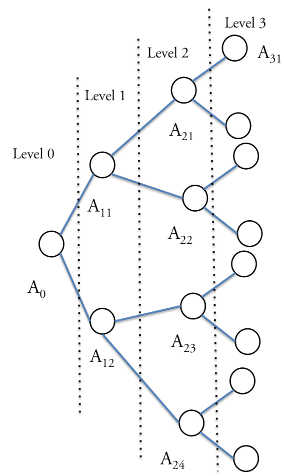

Figure 3: Graph

Blank page for work.