1. (30 points) **(Designing a Tournament)** This question will help you understand the applications of Chernoff bound and Union bound more, as you'll deal with tuning parameters to optimize the efficiency of the system. You'll also get an understanding of why NBA tournaments are designed the way they are, with more repeated matches being played as we get closer to the finals. For example, in the NBA, the early rounds are *best-of-five*, and the later rounds become *best-of-seven*. You will understand why, and also learn how to design tournaments with n participants, for large n.

   Suppose there are n teams, and they are totally ranked. That is, there is a well-defined best team, second ranked team and so on. It's just that we (the tournament designer) don't know the ranking. Moreover, assume that for any given match between two players, the better ranked team will win the match with probability $p = \frac{1}{2} + \delta$, *independent of all other matches between these players and all other players also*. Here δ is a small positive constant. All your answers below will depend on the value of δ.

   (a) (5 points) Let n be a power of two, and fix an *arbitrary tournament tree* starting with $n/2$ matches, then $n/4$ matches and so on. That is, initially, team 1 plays team 2, team 3 plays team 4, and so on (each series is only 1 game). The winners advance, and pair up and play each other once, and so on until one team remains. What is the probability that the best team wins the tournament?

   > **Solution:**
   >
   > *Proof.* $1 + 1 = 2$. □

   (b) (10 points) Now change the tournament to make each series as a *best-of-k* series. How large should k be, and so how many games do you end up conducting in total to get a $1 - \epsilon$ probability of the best team winning the overall series?

   > **Solution:**
   >
   > *Proof.* $1 + 1 = 2$. □

   (c) (15 points) Can you get better dependence on n if you allow different number of games in each round: example, try having k_1 games in the first series, k_2 games in the next series, etc. and optimize for these values to get a total of $O_\epsilon(n)$ games which still retains $1 - \epsilon$ probability of the best team winning eventually. Here $O_\epsilon(n)$ means $O(n)$ for all constant $\epsilon > 0$.

   > **Solution:**
   >
   > *Proof.* $1 + 1 = 2$. □

2. (15 points) **(Who wins the election? Exit Poll Design)** Imagine there are only two parties standing in the national election, and you have access to sampling and calling up uniformly random people from the electorate to find out who they're going to vote for.

If the total population size of India is N and an unknown $\frac{1}{4} \leq p \leq \frac{3}{4}$ fraction prefer BJP and $(1-p)$ prefer Congress, approximately how many people do you need to sample in order to estimate p upto an additive error of ϵ? Don't try too hard to optimize the constants, so feel free to use whichever concentration bound you see fit.