1. **Working with the Law of Large Numbers**

   (a) A fair coin is tossed and you win a prize if there are more than 60% heads. Which is better: 10 tosses or 100 tosses? Explain.

   (b) A fair coin is tossed and you win a prize if there are more than 40% heads. Which is better: 10 tosses or 100 tosses? Explain.

   (c) A coin is tossed and you win a prize if there are between 40% and 60% heads. Which is better: 10 tosses or 100 tosses? Explain.

   (d) A coin is tossed and you win a prize if there are exactly 50% heads. Which is better: 10 tosses or 100 tosses? Explain.

2. **Playing Pollster**

   As an expert in probability, the staff members at the Daily Californian have recruited you to help them conduct a poll to determine the percentage p of Berkeley undergraduates that plan to participate in the student sit-in. They've specified that they want your estimate $\hat{p}$ to have an error of at most ε with confidence $1-\delta$. That is,

   $$P\left(|\hat{p}-p| \leq \varepsilon\right) \geq 1-\delta.$$

   Assume that you've been given the bound

   $$P\left(|\hat{p}-p| \geq \varepsilon\right) \leq \frac{1}{4n\varepsilon^2},$$

   where n is the number of students in your poll.

   (a) Using the formula above, what is the smallest number of students n that you need to poll so that your poll has an error of at most ε with confidence $1-\delta$?

   (b) At Berkeley, there are about 26,000 undergraduates and about $10,000$ graduate students. Suppose you only want to understand the frequency of sitting-in for the undergraduates. If you want to obtain an estimate with error of at most 5% with 98% confidence, how many undergraduate students would you need to poll? Does your answer change if you instead only want to understand the frequency of sitting-in for the graduate students?

   (c) It turns out you just don't have as much time for extracurricular activities as you thought you would this semester. The writers at the Daily Californian insist that your poll results are reported with at least 95% confidence, but you only have enough time to poll 500 students. Based on the bound above, what is the worst-case error with which you can report your results?

3. **Covariance**

We have a bag of 5 red and 5 blue balls. We take two balls from the bag without replacement. Let X_1 and X_2 be indicator random variables for the first and second ball being red. What is $Cov(X_1, X_2)$?

4. **LLSE**

We have two bags of balls. The fractions of red balls and blue balls in bag A are $\frac{2}{3}$ and $\frac{1}{3}$ respectively. The fractions of red balls and blue balls in bag B are $\frac{1}{2}$ and $\frac{1}{2}$ respectively. Someone gives you one of the bags (unmarked) uniformly at random. Then we draw 6 balls from the same bag with replacement. Let X_i be the indicator random variable that ball i is red. Now, let us define $X = \sum_{1 \leq i \leq 3} X_i$ and $Y = \sum_{4 \leq i \leq 6} X_i$. Find $LLSE(Y|X)$. [Hint: recall that $LLSE(Y|X) = E(Y) + \frac{Cov(X,Y)}{Var(X)}(X - E(X))$]