Due on Friday, April 23 at 1:00PM (283 Soda)

1. **(30 pts.) More on the Birthday Paradox**
   Suppose we invite n people to a party, on a planet with k days in a year. We assume that each person's birthday is uniformly and independently at random among each of the k days of the year. The birthday paradox problem asks for the probability that some pair of people have the same birthday.

   Let the random variable X denote the number of pairs of people at the party who have the same birthday. We showed in lecture that if $n \approx \sqrt{2k}$, then $\mathbf{E}[X] \approx 1$. I then argued (waving my hands rapidly) that this suggests that we might expect to find the first repeated birthday once about $\sqrt{2k}$ people have arrived at the party. Of course, this was an informal heuristic that came without proof. In this problem, you will show how to justify this estimate without using any hand-waving.

   (a) Define the random variable $X_{i,j}$ as follows: it is 1 if person i has the same birthday as person j and 0 otherwise. As a warmup, show that $\mathbf{E}[X] = n(n-1)/2k$. Hence conclude that if $n \approx \sqrt{2k}$, then $\mathbf{E}[X] \approx 1$.

   (b) A set of random variables $Y_1, \ldots, Y_m$ are called *pairwise independent* if every pair of random variables are independent, i.e., $Y_\ell, Y_{\ell'}$ are independent for each pair of distinct indices $\ell < \ell'$. Show that the $X_{i,j}$'s defined above give a set of random variables that are pairwise independent but not mutually independent.

   (c) Show that, for a set of pairwise independent random variables, the variance of the sum is the same as the sum of the variances. In other words, show that $\mathrm{Var}[\sum_{1 \le i < j \le n} X_{i,j}] = \sum_{1 \le i < j \le n} \mathrm{Var}[X_{i,j}]$. Use this to calculate $\mathrm{Var}[X]$ in terms of k and n.

   (d) Show how to calculate an upper bound b on the probability that the partiers' birthdays are all different, i.e., find a value b so that $\Pr[X = 0] \le b$. Your answer for b will be an expression in terms of k and n.

   (e) If we fix k, how large does n have to be (in terms of k) to ensure that some pair of party-goers will have the same birthday with probability at least $1/2$?

2. **(20 pts.) Another Card Game** This problem concerns the following game. A deck of six cards is randomly shuffled. Your task is to guess the order of the cards. Once you have written down your guess, the cards are turned up. You receive \$1 for each card whose position you have guessed correctly. You pay \$1 for the privilege of playing the game. Here are three possible guessing strategies you might use:

   (i) Write down card 1 six times.

   (ii) Write down a random permutation of cards 1–6.

   (iii) Roll a fair die six times, and write down the sequence of six scores.

   Let's analyze these strategies.

(a) For each strategy, determine the expected number of dollars you win.

(b) For each strategy, determine the variance of your winnings.

(c) Which strategy would you use if you wanted to minimize the chance of losing any money? Which strategy would you use if you wanted the most “exciting” game (i.e., with the largest wins and losses)? Justify your answers with reference to part (b).

3. (15 pts.) Confidence Interval

You own a telephone company that services two cities A and B, each having 5000 customers. You would like to link your exchange with the more distant city C. You estimate that during the busiest time each customer will require a line to C with probability .01. You want to be sure that there are enough lines to C so that there is only a 1% chance that at the busiest time some customer will be unable to get a line to C. Each trunkline to C will cost $10,000 and can carry one call at a time. You have two options. Either link A and B (to C) as if they were separate exchanges or link the entire exchange to C. In the second option, additional equipment costing $50,000 would be needed. Which option is cheaper?

4. (15 pts.) Hypothesis Testing

Suppose that it is hypothesized that twice as many automobile accidents resulting in deaths occur on Saturday and Sunday as on other days of the week. That is, the probability that such accidents occur on Saturday is 2/9, on Sunday is 2/9, and on each other day of the week is 1/9. From the national record file, cards for 90 accidents are selected at random. These yield the following distribution of accidents according to the days of the week:

- Sunday, 30
- Monday, 6
- Tuesday, 8
- Wednesday, 11
- Thursday, 7
- Friday, 10
- Saturday, 18

Do these data tend to support or contradict the hypothesis? Use a 5% significance level.

5. (15 pts.) Poisson Distribution

A small insurance company sells just one kind of insurance. Every year they charge each client $100, and each client receives $1,000,000 with probability 5×10^{-5}. The company has 20,000 clients and assets of only $1,000,000. What is the probability that the company will go bankrupt in a given year?