1. Dice Try! (15 points)
You are playing a game that uses a fair 12-sided die whose faces are numbered 1, 2, \ldots, 12. The value of a roll is the number showing on the top of the die when it comes to rest. Give all answers as simplified fractions.
(a) [5 Points] Let $X$ be the value of one roll of the die. Compute $E[X]$ and $\text{Var}(X)$.
(b) [5 Points] Let $Y$ be the sum of the values of 4 independent rolls of the die. Compute $E[Y]$ and $\text{Var}(Y)$. Use independence, and state precisely where in your computation you are using it.
(c) [5 Points] Let $Z$ be the average of the values of 4 independent rolls of the die. Compute $E[Z]$ and $\text{Var}(Z)$. Use independence, and state precisely where in your computation you are using it.

2. It's The Standard (20 points)
Let $X$ be a random variable with expected value $\mu$ and variance $\sigma^2$. Find the expected value and variance of $Y = (X - \mu)/\sigma$. (Your answer will help explain why $Y$ is called the “standardized” version of $X$: standardizing two random variables puts them on more of an equal footing for comparing, as your answer will show.)

3. Binomial From Nowhere (20 points)
Consider repeatedly rolling a fair 6-sided die, each roll being independent of the others. Define the random variable $Y$ to be the number of rolls until (and including) the first roll of a 6, and define the random variable $X$ to be the number of 1's rolled before the first 6 is rolled. Show that $\Pr(X = j \mid Y = i)$, as $j$ ranges over its possible values, is the probability mass function of a binomially distributed random variable and determine its parameters $n$ and $p$.

4. Sample Sampling Algorithm (20 points)
Consider the following algorithm for generating a random sample of size $n$ from the set of integers $\{1, 2, \ldots, N\}$, where $0 < n < N$. 
Sample($N$, $n$):

1. $I = 0$
2. chosen = {} // chosen is a set of distinct integers, initially an empty set
3. while $|\text{chosen}| < n$:
4.   $I += 1$ // $I$ is counting the total number of rolls of the die
5.   chosen.add($\text{RollDie}(N)$) // if the roll of the die (which is random in 1 ... $N$)
6.     // is not in chosen, then add it to chosen.
7. return chosen

(a) [10 Points] Calculate $E[I]$ in terms of the harmonic numbers $H_m$. Recall that $H_m = \sum_{i=1}^{m} \frac{1}{i}$.

(b) [10 Points] Calculate $\text{Var}(I)$ as a summation.

5. Binomial Conditioned on Binomial (10 points)
Suppose that $X$ is a binomial random variable with parameters $n$ and $p$, and $Y$ is another independent binomial random variable with parameters $m$ and $p$. What kind of random variable is $X + Y$? Compute the probability that $X = j$ conditioned on $X + Y = r$.

6. Random Grades (15 points)
Every week, 20,000 students flip a 10,000-sided fair dice, numbered 1 to 10,000, to see if they can get their GPA changed to a 4.0. If they roll a 1, they win (they get their GPA changed). You may assume each student’s roll is independent. Let $X$ be the number of students who win.

(a) [5 Points] For any given week, give the appropriate probability distribution (including parameter(s)), and find the expected number of students who win.

(b) [5 Points] For any given week, find the exact probability that at least 2 students win. Give your answer to 5 decimal places.

(c) [5 Points] For any given week, estimate the probability that at least 2 students win, using the Poisson approximation. Give your answer to 5 decimal places.

7. Instagram (20 points)
A photo-sharing startup offers the following service. A client may upload any number $N$ of photos and the server will compare each of the $\binom{N}{2}$ pairs of photos with their proprietary image matching algorithms to see if there is any person that is in both pictures. Testing shows that the matching algorithm is the slowest part of the service, taking about 100 milliseconds of CPU time per photo pair. Hence, estimating the number of photos uploaded by each client is a key part of sizing their data center. The people in charge say that their gut feeling is that $N = 10$. You (the chief technical officer) say, “but $N$ is a random variable”. What will the average CPU demand per client (as a function of $N$, $p$ or $\lambda$) if $N$ follows

- the “distribution” where $N$ is the same fixed number with probability 1?
- the Poisson distribution with parameter $\lambda$?
- the geometric distribution with parameter $p$?
- $N = 80X + 5$, where $X$ is a Bernoulli random variable with parameter $p$?

In each case, include as part of your answer the expected value of $N$ and the variance of $N$. 