You must complete 7 of the 9 problems on this exam for full credit. You MUST do the last problem (problem 9). Each problem is of equal weight. Please leave blank, or draw an X through, or write “Do Not Grade,” on the two problems you are eliminating; I will grade the first 7 I get to if I can not figure out your intention. If answers are on the back of the page please tell me so. **Circle final answers.** No calculators allowed, and you may leave *complicated* arithmetic expressions uncomputed, but please do multiply $1/2 \times 1/2$ to get $1/4$ if the occasion presents itself.

In composing your answers, remember that your goal is to show me you understand the techniques presented in the course; if you solve a problem using an inappropriate technique, you may not receive full credit.

**Problem One.** Out of 2000 families with 4 children, how many would you expect to find which have at least one girl and one boy (i.e., not all girls or all boys)? Assume the probability of any particular child being a boy is 0.5.

**Solution:** There are two ways to do this, using the Exclusion Principle, and the other using a binomial distribution.

First way: The probability of all girls is $(1/2)^4 = 1/16$ and for all boys is the same, so $2/16 = 1/8$ of the families would have at least children of only one sex, or $2000/8 = 250$. Thus $2000-250 = 1750$ would have at least one girl and one boy.

Second way: If $X = \text{number of girls in a family with 4 children}$, we are looking for $P(X = 1) + P(X = 2) + P(X = 3)$ for $B(4, 0.5)$, or

$$C(4,1)*(1/2)^4 + C(4,2)*(1/2)^4 + C(4,3)*(1/2)^4$$

$$= 4 \times 1/16 + 6 \times 1/16 + 4 \times 1/16$$

$$= 14/16$$

Then $14/16 \times 2000 = 28000/16 = 1750$. 
**Problem Two.** Suppose that we have the CDF for the normal distribution implemented in Python as shown in lecture:

```python
def Phi(mu, var, x):
    return (1 + math.erf((x-mu)/(var**0.5 * 2.0**0.5))) / 2
```

(a) Complete the following function definition for a similar function that implements the CDF of the standard normal distribution (you may use `Phi(...)` in your answer):

```python
def PhiStandard(mu, var, x):
    return Phi(0, 1, x)
```

(b) Complete the following function definition for a function that approximates the binomial distribution CDF using the normal distribution, using the “continuity correction”:

```python
# for X ~ B(N, p) returns P(X ≤ x)
def binomialCDF(N, p, x):
    return Phi(N*p, N*p*(1-p), x+0.5)
```

(c) Supposing that Math SAT scores are normally distributed with mean 500 and standard deviation 100, give the expression (using `Phi(...)`) which calculates the probability that a random student has a score between 600 and 700.

```
Phi(500, 10000, 700) - Phi(500,10000,600)
```
Problem Three. A spam filter works by looking for common spam phrases in email messages. Suppose that 80% of your email is spam. In 10% of the spam emails, the phrase “bank transfer” is used, whereas it is used in only 1% of non-spam emails. A new email has just arrived, which contains the phrase “bank transfer.” What is the probability that it is spam? [You may give this as a fraction, but have to do the calculation, not just give a formula.]

Solution: There are two ways to do this, using standard conditional probabilities, or using Bayes’s Rule. For the first, the Venn diagram for this would be as follows, where the search space is all email messages, $M$ = spam messages, and $B =$ messages containing the phrase “bank transfer”:

![Venn Diagram](image)

What is given is $P(M) = 0.8$, $P(MB) = P(M) \times 0.1 = 0.08$, and $P(\neg MB) = P(\neg M) \times 0.01 = 0.2 \times 0.01 = 0.002$. We can thus derive that $P(B) = P(\neg MB) + P(MB) = 0.082$.

We want to find out $P( M \mid B ) = P(MB) / P(B) = 0.08 / 0.082 = 80/82 = 40/41$.

To do this using Bayes Rule, we use

$$P( M \mid B ) = P( B \mid M ) \times P( M ) / P( B )$$

$$= 0.1 \times 0.8 / 0.082$$

$$= 40/41$$
**Problem Four.** Suppose a lottery sells 10,000 tickets, labeled with the numbers 0 – 9999. They will select (without replacement) 1 “gold” prize (worth $500), 2 “silver” prizes (worth $100), and 10 “bronze” prizes (worth $10). If they wish to make $20,000 net (after awarding the prizes), assuming all tickets are sold, and assuming there are no costs other than giving out the prize money, what should be the price of an individual ticket?

**Solution:** There are two ways to do this, using simple arithmetic and the other using the expected value. Actually I intended for you to use the second method, but I can’t really say that the first is an “inappropriate technique”!

**First way:** The cost of the prizes is $500 + $200 + $100 = $800. Therefore they need to make $20,800 in order to net $20,000. Therefore each ticket should cost $20,800/10,000 = $2.08. (Duh)

**Second way:** If you pay nothing for a ticket then the expected value is 0.0001*$500 + 0.0002*$100 + 0.001*$10 = 0.05+0.02+0.01 = 0.08. Therefore if you have to pay $20,000/10,000 = $2 for the ticket and $0.08 for the expected value, giving $2.08.
**Problem Five.** From an ordinary deck of 52 cards we draw cards at random, with replacement, and successively until a face card (Jack, Queen, or King) is drawn. What is the probability that at least 5 draws are needed?

**Solution:** This is the Geometric Distribution $G(3/13)$ and we seek the tail probability $P(X > 4)$. The simplest calculation is to find the probability of 4 non-face cards or $(10/13)^4 = 0.3501$. Or, you could use the Exclusion Principle and count the probabilities of the cases to exclude:

\[
1.0 - ( P(X=1) + P(X=2) + P(X=3) + P(X=4))
\]

\[
= 1.0 - ( \frac{3}{13} + \left( \frac{10}{13} \right) \left( \frac{3}{13} \right) + \left( \frac{10}{13} \right)^2 \left( \frac{3}{13} \right) + \left( \frac{10}{13} \right)^3 \left( \frac{3}{13} \right) )
\]

\[
= 1.0 - (0.2308 + 0.1775 + 0.1365 + 0.1050)
\]

\[
= 0.3501
\]
**Problem Six.** A fair die is rolled twice. Let A denote the event that the sum of the outcomes is odd, and B denote the event that on the first toss it shows a 2. Are A and B independent? Why or why not? (You must answer this mathematically.)

**Solution:** The two events are independent. For A, there are 2 ways to get a 3, 2 ways to get 11, 4 ways to get 5, 4 ways to get 9, and 6 ways to get a 7, so \( P(A) = \frac{18}{36} = \frac{1}{2} \).

\( P(B) = \frac{1}{6} \). We can establish independence in several (equivalent) ways:

1. There are 3 outcomes where the first roll is a 2, and the total is odd, so
   \[ P(AB) = \frac{3}{36} = \frac{1}{12} = \frac{1}{2} \times \frac{1}{6} = P(A) \times P(B) \]

2. The conditional probability \( P(A|B) = \frac{1}{2} \) (since half the second roll’s outcomes are odd), and thus
   \[ P(A|B) = P(A) \]

3. Finally, \( P(B|A) = \frac{3}{18} = \frac{1}{6} \) (since, of the 18 odd outcomes, 3 are due to a 2 on the first roll) and so
   \[ P(B|A) = P(B) \]

This can be illustrated in the following diagram, where A has red numbers and B has grey backgrounds; the probability of an odd roll, or 1/2, is the same overall and inside the grey boxes:

![Outcomes from Two Dice](attachment:image.png)
Problem Seven. This problem is about the “memory-less property.” (i) State the property in mathematical terms, (ii) Which distributions have this property? (iii) Explain the application of this property in practical terms.

Solution:

(i) A distribution has the memory-less property if

\[ P(X > s+t | X > s) = P(X > t) \]

(ii) Only the Geometric and Exponential Distributions have this property.

(iii) Suppose the inter-arrival times of the T are distributed according to the Exponential Distribution; then the probability that a T train will arrive in the next five minutes is the same whether you start to time the five minutes as soon as you arrive at the T stop, if you wait 15 minutes with no train arriving, and then start to time the five minutes; any time you arrive, you will have the same probabilities and the same expected wait time.
**Problem Eight.** In a certain town, crimes occur at a Poisson rate of five per month. What is the probability of having exactly two months (not necessarily consecutive) with no crimes during the next year?

**Solution:** The key point here is that you have to calculate the probability of no crimes in a random month using the Poisson, and then use the Binomial to calculate the probability of 2 months out of 12 having no crimes:

Choose one month as the unit of time. Then $\lambda = 5$ and the probability of no crimes during any given month of a year is $P(N(1) = 0) = e^{-5} = 0.0067$. Hence the desired probability is

$$\binom{12}{2} (0.0067)^2 (1 - 0.0067)^{10} = 0.0028.$$
Problem Nine (Required). Write a Python function `nextBinomial(N,p)` to generate random variates for the binomial distribution B(N,p) using the “inverse method” (by inverting the CDF).

Solution: This is just recasting the method we used for generating Poisson-distributed and Exponentially-distributed variates in Lab 07 to use the Binomial distribution, which just means replacing the call to `poisson(lamb, x)` or `exponential(lambda, x)` with `binomial(N,p,x)`: 

```python
def nextBinomial(N,p):
    d = random.random()
    sum = 0.0
    for x in range(N+1):
        sum = sum + binomial(N,p,x)  # P(X=k) for B(N,p)
        if sum > d:
            return x
    return x
```

The syntax in red is the only thing that needs to change when you use the inverse method for generating random variates for different distributions.