Honor code: I pledge on my honor that: I have completed all steps in the below homework on my own, I have not used any unauthorized materials while completing this homework, and I have not given anyone else access to my homework.

Name and Signature

1. (1 point) Have you read and understood the honor code?

> **Solution:** Yes, I have read and Understood the honor code.

2. You have two identical fair coins. You toss the first coin and if the output is heads then you stay with this coin and toss it again. If the the output is tails then you switch to the other coin and repeat the same process. This process can be summarized as follows:

   Step 1: Select coin 1
   Step 2: Toss the coin
   Step 3: If result = Heads, go to step 2
   Step 4: If result = Tails, switch to the other coin and go to step 2

   (a) (½ point) What is the probability that after 99 tosses you end up with the same coin that you started with?

   > **Solution:** This means that you never switched the coin or switched it multiple times so that you again have coin 1.
   >
   > We will end up with the same coin if we get Tails an even number of times overall i.e. $0, 2, 4, \ldots, 98$.
   >
   > $$P(\text{Same Coin}) = \binom{99}{0}(\tfrac{1}{2})^0(\tfrac{1}{2})^{99} + \binom{99}{2}(\tfrac{1}{2})^2(\tfrac{1}{2})^{97} + \ldots + \binom{99}{98}(\tfrac{1}{2})^{98}(\tfrac{1}{2})^1$$
   >
   > $$= (\tfrac{1}{2})^{99}(\binom{99}{0} + \binom{99}{2} + \ldots + \binom{99}{98})$$
   >
   > $$= (\tfrac{1}{2})^{99}\ 2^{99-1}$$
   >
   > $$= (\tfrac{1}{2})^{99}\ 2^{98}$$
   >
   > $$= \tfrac{1}{2}$$

(b) (½ point) What is the probability that after 100 tosses you end up with the same coin that you started with?

Solution: Again we will end up with the same coin if we get Tails an even number of times overall i.e. $0, 2, 4, \ldots, 98, 100$.

$$P(\text{Same Coin}) = \binom{100}{0}(\tfrac{1}{2})^0(\tfrac{1}{2})^{100}+\binom{100}{2}(\tfrac{1}{2})^2(\tfrac{1}{2})^{98}+\ldots+\binom{100}{98}(\tfrac{1}{2})^{98}(\tfrac{1}{2})^2+\binom{100}{100}(\tfrac{1}{2})^{100}(\tfrac{1}{2})^0$$

$$= (\tfrac{1}{2})^{100}(\binom{100}{0} + \binom{100}{2} + \ldots + \binom{100}{98} + \binom{100}{100})$$

$$= (\tfrac{1}{2})^{100}\ (2^{100-1})$$

$$= (\tfrac{1}{2})^{100}\ (2^{99})$$

$$= \tfrac{1}{2}$$

(c) (1 point) What if instead of fair coins you have identical biased coins with probability of heads $= p$ ($p \neq \frac{1}{2}$)?

Solution: Let us take number of tosses as n in general:

If n is odd:
We will end up with the same coin if we get Tails an even number of times overall i.e. $0, 2, 4, \ldots n-1$.

$$P(\text{Same Coin}) = \binom{n}{0}(1-p)^0(p)^n + \binom{n}{2}(1-p)^2(p)^{n-2} + \ldots + \binom{n}{n-1}(1-p)^{n-1}(p)^1$$

$$= \sum_{i=0}^{\frac{n-1}{2}} \binom{n}{2i}(1-p)^{2i}(p)^{n-2i}$$

If n is even:
We will end up with the same coin if we get Tails an even number of times overall i.e. $0, 2, 4, \ldots n$.

$$P(\text{Same Coin}) = \binom{n}{0}(1-p)^0(p)^n + \binom{n}{2}(1-p)^2(p)^{n-2} + \ldots + \binom{n}{n}(1-p)^n(p)^0$$

$$= \sum_{i=0}^{\frac{n}{2}} \binom{n}{2i}(1-p)^{2i}(p)^{n-2i}$$

Here, $n = 100$ which is even.
So, $P(\text{Same Coin}) = \sum_{i=0}^{\frac{100}{2}} \binom{100}{2i}(1-p)^{2i}(p)^{100-2i}$

3. You are dealt a hand of 5 cards from a standard deck of 52 cards which contains 13 cards of each suite (hearts, diamonds, spades and clubs).

   (a) (½ point) What is the probability that you get an ace, a king, a queen, a jack and a 10 of the same suite? Let us call such a hand as the King's hand.

   > **Solution:**
   > $P(\text{King's hand}) = \frac{\binom{4}{1}}{\binom{52}{5}} = \frac{1}{649740}$

   (b) (½ point) Let n be the number of times you play this game. What is the minimum value of n so that the probability of having no King's hand in these n turns is less than $\frac{1}{e}$?

   > **Solution:**
   > $P(\text{ Not King's hand in one turn}) = 1 - P(\text{King's hand }) = 1 - \frac{1}{649740} = \frac{649739}{649740}$
   >
   > $P(\text{ Not King's hand in } n \text{ turns }) = (\frac{649739}{649740})^n$ (Each Turn is independent).
   >
   > Now for $P(\text{ Not King's hand in } n \text{ turns }) < \frac{1}{e}$
   >
   > $\implies (\frac{649739}{649740})^n < \frac{1}{e}$
   >
   > $\implies \log(\frac{649739}{649740})^n < \log\frac{1}{e}$ Taking log to base e on both sides:
   >
   > $\implies \log(\frac{649739}{649740})^n < -1$
   >
   > $\implies n\log(\frac{649739}{649740}) < -1$
   >
   > $\implies n(-1.539078354 \times 10^{-6}) < -1$
   >
   > $\implies n(1.539078354 \times 10^{-6}) > 1$
   >
   > $\implies n > \frac{10^6}{1.539078354}$
   >
   > $\implies n > 649739.499$
   >
   > Hence smallest value of $n = 649740$

4. (1 point) A spacecraft explodes while entering the earth's atmosphere and disintegrates into 10000 pieces. These pieces then fall on your town which contains 1600 houses. Each

piece is equally likely to fall on every house. What is the probability that no piece falls on your house (assume you have only one house in the town and all houses are of the same size and equally spaced - for example you can assume that the houses are arranged in a 40×40 grid).

Solution: Assuming that the houses are arranged in a 40×40 grid.

A_i: Event that i^{th} piece falls on my house.
$P(A_i) = \frac{1}{1600}$

A_i': Event that i^{th} piece will not fall on my house.
$P(A_i') = 1 - P(A_i) = 1 - \frac{1}{1600} = \frac{1599}{1600}$

B: Event that No piece falls on my house.

Each piece will fall or not fall on my house independent of the others.

Hence, $P(B) = P(\bigcap_{i=1}^{10000} A_i) = \Pi_{i=1}^{10000} P(A_i) = (\frac{1599}{1600})^{10000}$

5. What's in a name?

   (a) (½ point) Why is the hypergeometric distribution called so? (We understand what is geometric but what is "hyper"?)

   Solution: Let us first find the ratio of $p_X(j)$ to $p_X(j-1)$ for a geometric random variable.
   $p_X(j) = (1-p)^{j-1}p$ where p is the probability of success in each trial.
   $p_X(j-1) = (1-p)^{j-2}p$
   $\frac{p_X(j)}{p_X(j-1)} = (1-p)$, which is a constant and hence these forms a geometric progression and hence the distribution is called geometric distribution.

   Let us now find the ratio of $p_X(j)$ to $p_X(j-1)$ for a hypergeometric random variable.
   $p_X(j) = \frac{\binom{J}{j}\binom{N-J}{n-j}}{\binom{N}{n}}$

   N is the population size.
   J is the number of success states in the population.
   n is the number of draws (i.e. quantity drawn in each trial).
   j is the number of observed successes.

$p_X(j-1) = \frac{\binom{J}{j-1}\binom{N-J}{n-j+1}}{\binom{N}{n}}$

$\frac{p_X(j)}{p_X(j-1)} = \frac{(J-j+1)(n-j+1)}{(j)(N-J-n+j)}$

This is a rational function of j which means it is more than constant (as in case of geometric) so in a sense it is over and above geometric distribution and hence called hypergeometric distribution.

Note: $R(z) = \frac{P(z)}{Q(z)}$ is called a rational function of z where $P(z)$ and $Q(z)$ are polynomials in z.

(b) (½ point) Why is the negative binomial distribution called so?

Solution:
The pmf of a negative binomial distribution is given as:

$p_X(k) = \binom{k+r-1}{r-1}(1-p)^k p^r$
$p_X(k)$ is the probability that number of failures is k we observe before we get exactly r successes.
r is the number of predefined successes.
p is the probability of success.
The random variable here is the number of failures we observe before we get exactly r successes.

Let us consider the binomial coefficient in $p_X(x)$

$$\binom{k+r-1}{r-1} = \frac{(k+r-1)!}{(r-1)!k!}$$

$$= \frac{(k+r-1)(k+r-2)...(r)}{k!}$$

$$= (-1)^k \frac{(-r)(-r-1)(-r-2)...(-r-k+1)}{k!}$$

$$= (-1)^k \binom{-r}{k}$$

The word "negative binomial" refers to this $-r$ in the binomial coefficient.

6. Consider a binomial random variable whose distribution $p_X(x)$ is fully specified by the parameters n and p.

(a) (½ point) What is the ratio of $p_X(j)$ to $p_X(j-1)$?

Solution: Let the number of trials be n and probability of success in each trial be p.
$p_X(j) = \binom{n}{j} p^j (1-p)^{n-j}$
$p_X(j-1) = \binom{n}{j-1} p^{j-1} (1-p)^{n-j+1}$

$$\frac{p_X(j)}{p_X(j-1)} = \frac{\binom{n}{j} p^j (1-p)^{n-j}}{\binom{n}{j-1} p^{j-1} (1-p)^{n-j+1}}$$
$$= \frac{(n-j+1)p}{j(1-p)}$$

(b) (½ point) Based on the above ratio can you find the value(s) of j for which $p_X(j)$ will be maximum ?

Solution: From part (a) above:

$$\frac{p_X(j)}{p_X(j-1)} = \frac{(n-j+1)p}{j(1-p)} \qquad (1)$$

$$= \frac{(n-j+1)p}{jq}\text{ , where } q = 1-p$$
$$= \frac{jq+(n-j+1)p-jq}{jq}$$
$$= 1 + \frac{(n+1)p-j}{jq}$$

Let us discuss it in two cases:

Case 1: $(n+1)p$ is not an integer.
Let $(n+1)p = m + f$ where m is an integer and f is a fraction $0 < f < 1$.

Putting this in (1) we have:
$$\frac{p_X(j)}{p_X(j-1)} = 1 + \frac{m+f-j}{qj} \qquad (2)$$

From (2) it is clear that:
$$\frac{p_X(j)}{p_X(j-1)} \begin{cases} > 1 \text{ for } j = 1 \text{ to } m \\ < 1 \text{ for } j = m+1 \text{ onwards} \end{cases}$$

$\therefore p_X(0) < p_X(1) < \ldots < p_X(m-1) < p_X(m) > p_X(m+1) > p_X(m+2) \ldots$

Thus in this case there exists a unique value of j for which $p_X(j)$ is maximum which is $j = m$ i.e. the Integer part of $(n+1)p$

Case 2: $(n+1)p$ is an integer.
Let $(n+1)p = m$ where m is an integer.

Putting this in (1) we have:

$$\frac{p_X(j)}{p_X(j-1)} = 1 + \frac{m-j}{qj} \qquad (3)$$

From (3) it is clear that:

$$\frac{p_X(j)}{p_X(j-1)} \begin{cases} > 1 \text{ for } j = 1 \text{ to } m-1 \\ = 1 \text{ for } j = m \\ < 1 \text{ for } j = m+1 \text{ onwards} \end{cases}$$

$\therefore p_X(0) < p_X(1) < \ldots < p_X(m-1) = p_X(m) > p_X(m+1) > p_X(m+2) \ldots$

Thus in this case there exists a two values of j for which $p_X(j)$ is maximum which are $j = m-1$ and $j = m$ where $m = (n+1)p$

7. Consider a Poisson random variable whose distribution $p_X(x)$ is fully specified by the parameter λ.

   (a) (½ point) What is the ratio of $p_X(j)$ to $p_X(j-1)$?

Solution: The pmf for poisson distribution is given by:

$p_X(x) = \frac{e^{\lambda}\lambda^x}{x!}$

Now, $p_X(j) = \frac{e^{\lambda}\lambda^j}{j!}$

$p_X(j-1) = \frac{e^{\lambda}\lambda^{j-1}}{(j-1)!}$

Dividing these two we get:

$\frac{p_X(j)}{p_X(j-1)} = \frac{\lambda}{j}$

   (b) (½ point) Based on the above ratio can you find the value(s) of j for which $p_X(j)$ will be maximum ?

Solution: From part (a) above:

$\frac{p_X(j)}{p_X(j-1)} = \frac{\lambda}{j}$

Let us discuss it in two cases:

Case 1: λ is not an integer.

Let $\lambda = m + f$ where m is an integer and f is a fraction $0 < f < 1$.

Putting this in (1):

$$\frac{p_X(j)}{p_X(j-1)} = \frac{m}{j} + \frac{f}{j} \qquad (4)$$

From (4) it is clear that:

$$\frac{p_X(j)}{p_X(j-1)} \begin{cases} > 1 \text{ for } j = 1 \text{ to } m \\ < 1 \text{ for } j = m+1 \text{ onwards} \end{cases}$$

$\therefore p_X(0) < p_X(1) < \ldots < p_X(m-1) < p_X(m) > p_X(m+1) > p_X(m+2) \ldots$

Thus in this case there exists a unique value of j for which $p_X(j)$ is maximum which is $j = m$ i.e. the Integer part of λ

Case 2: λ is an integer.
Let $\lambda = m$ where m is an integer.

Putting this in (1) we have:

$$\frac{p_X(j)}{p_X(j-1)} = \frac{m}{j} \qquad (5)$$

From (5) it is clear that:

$$\frac{p_X(j)}{p_X(j-1)} \begin{cases} > 1 \text{ for } j = 1 \text{ to } m-1 \\ = 1 \text{ for } j = m \\ < 1 \text{ for } j = m+1 \text{ onwards} \end{cases}$$

$\therefore p_X(0) < p_X(1) < \ldots < p_X(m-1) = p_X(m) > p_X(m+1) > p_X(m+2) \ldots$

Thus in this case there exists a two values of j for which $p_X(j)$ is maximum which are $j = m-1$ and $j = m$ where $m = \lambda$

8. For each of the following random variables show that the sum of the probabilities of all the values that the random variable can take is 1?

   (a) (½ point) A negative binomial random variable whose distribution is fully specified by p (probability of success) and r (fixed number of desired successes)

Solution: The pmf of a negative binomial distribution is given as:

$p_X(k) = \binom{k+r-1}{r-1}(1-p)^k p^r$
$p_X(k)$ is the probability that number of failures we observe is k before we get exactly r successes.

r is the number of predefined successes.
p is the probability of success.
The random variable here is the number of failures we observe before we get exactly r successes.

Let us consider the binomial coefficient in $p_X(k)$
$\binom{k+r-1}{r-1} = \frac{(k+r-1)!}{(r-1)!k!}$

$$= \frac{(k+r-1)(k+r-2)...(r)}{k!}$$

$$= (-1)^k \frac{(-r)(-r-1)(-r-2)...(-r-k+1)}{k!}$$

$$= (-1)^k \binom{-r}{k}$$

Hence, $p_X(k) = (-1)^k \binom{-r}{k} p^r (1-p)^k$

Now, $\sum_{k=0}^{\infty} (-1)^k \binom{-r}{k} p^r (1-p)^k$

$$= \sum_{k=0}^{\infty} \binom{-r}{k} p^r (p-1)^k$$

$$= p^r \sum_{k=0}^{\infty} \binom{-r}{k} (-q)^k \text{ , where } q = 1-p$$

$$= p^r (1-q)^{-r}$$

$$= p^r (p)^{-r} = 1$$

Hence the sum of probabilities sums up to 1.

(b) (½ point) A hypergeometric random variable whose distribution is fully specified by N (number of objects in the given source), a (number of favorable objects in the source) and n (size of the sample that you want to select)

Solution: The pmf of a hypergeometric distribution is given by $p_X(x) = \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$, where x is the number of observed favorable objects.

Now, $\sum_{x=0}^{n} p_X(x) = \sum_{x=0}^{n} \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$

$$= \frac{1}{\binom{N}{n}} \sum_{x=0}^{n} \binom{a}{x}\binom{N-a}{n-x}$$

Consider the quantity $\sum_{x=0}^{n} \binom{a}{x}\binom{N-a}{n-x}$.
This quantity can be interpreted as :
We have a collection of N objects out of which a are of one kind and $N-a$ of second kind and we need to selct n objects.

For $x=0$ it means we are taking 0 objects of first kind and n objects of second kind.
For $x=1$ it means we are taking 0 objects of first kind and $n-1$ objects of second kind.
$\vdots$
For $x=n$ it means we are taking n objects of first kind and 0 objects of second kind.

Thus it enumerates all possible ways of chosing n objects from N objects and hence it is also equal to $\binom{N}{n}$

Hence , $\sum_{x=0}^{n} p_X(x) = \frac{\binom{N}{n}}{\binom{N}{n}} = 1$

(c) (½ point) A Poisson random variable whose distribution is fully specified by λ (i.e., arrival rate in unit time)

Solution: The pmf of a poisson random variable is given by $p_X(k) = \frac{\lambda^k}{k!}e^{-\lambda}$

Now, $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda}$

$= e^{-\lambda}\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$

$= e^{-\lambda}\sum_{k=0}^{\infty}(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \dots)$

$= e^{-\lambda}\sum_{k=0}^{\infty}(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \dots)$

$= e^{-\lambda}e^{\lambda}$ Using the expansion of e^{λ}

$= 1$

9. There are 100 seats in a movie theatre. Customers can buy tickets online. Based on past data, the theatre owner knows that 5% of the people that book tickets do not show up (of course, he gets to keep the money they paid for the ticket). To make more money

he decides to sell more tickets than the number of seats. For example, if he sells 102 tickets, then as long as at least 2 customers don't show up, he will be able to make extra money while not dissatisfying any customers.

(a) (½ point) If he sells 105 tickets what is the probability that no customer would be denied a seat on arrival.

Solution:

A: Event that a customer does not show up.
$P(A) = p = \frac{1}{20}$ 5% of customers do not show up.

Number of tickets sold = 105

Let X be the random variable: No of customers that do not show up out of these 105 customers. It follows a binomial distribution with $p = \frac{1}{20} = 0.05$

B: Event that no customer is denied a seat i.e. atleast 5 customers do not out of these 105 do not show show up.
C: Event that less than 5 customers do not show up out of these 105 machines.

$$P(C) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$$

$$= \binom{105}{0}(0.05)^0(0.95)^{105} + \binom{105}{1}(0.05)^1(0.95)^{104} + \binom{105}{2}(0.05)^0(0.95)^{103} + \binom{105}{3}(0.05)^3(0.95)^{102} + \binom{105}{4}(0.05)^4(0.95)^{101}$$

$$= (0.95)^{101}((0.95)^4 + 105(0.05)^1(0.95)^3 + 5460(0.05)^2(0.95)^2 + 187460(0.05)^3(0.95)^{102} + 4780320(0.05)^4)$$

$$= (0.95)^{101}(69.7721625) = 0.392433$$

Now, $P(B) = 1 - P(C) = 1 - 0.39777 = 0.607566$.

(b) (½ point) What is the maximum number of seats that he can sell so that there is at least a 90% chance that every customer will get a seat on arrival?

Solution:

Let the maximum number of customers be n.
A; Event that a customer does not show up.
$P(A) = p = \frac{1}{20} = 0.05$ 5% of customers do not show up.

Let X be the random variable: No of customers that do not show up out of

these n customers.
It follows a binomial distribution with $p = \frac{1}{20} = 0.05$

The question can be alternatively stated that what is the the largest value of n such that the probability that every customer gets a seat is atleast 0.9

This can be further restated as what is the the largest value of n such that the probability that less than or equal to 100 customers show up is atleast 0.9, where $n \geq 100$

This can be further stated that what is the the largest value of n such that the probability that at least $n - 100$ customers do not show up is atleast 0.9, where $n \geq 100$
So, the condition is $1 - [\sum_{k=0}^{n-100} \binom{n}{k} (0.05)^k (0.95)^{n-k}] \geq 0.9$

Using the Code below the largest value of n that satisfies this is $n = 102$.

Python Code

```
from scipy.stats import binom
n = 100
while True:
    summ = 0
    for i in range(0, n-100):
        summ = summ + binom.pmf(i, n , 0.05)

    if(1 - summ >= 0.9):
        n = n+1
    else:
        break
print(n-1)
```

(c) (1 point) Suppose he makes a profit of 5 INR for every satisfied customer and a loss of 50 INR (as penalty) for every dissatisfied customer (i.e., a customer who does not get a seat). What is his expected gain/loss if he sells 105 tickets?

Solution: Let X be the random variable that represents the Number of customers that do not show up.
It follows a binomial distribution with $p = \frac{1}{20} = 0.05$
Let Y be the random variable representing the gain of the theatre owner.

Some portion of the probability distributions of X and Y are given below:
$P(X=0)=\binom{105}{0}(0.05)^0(0.95)^{105}=0.00458$
$P(X=1)=\binom{105}{1}(0.05)^1(0.95)^{104}=0.025317$
$P(X=2)=\binom{105}{2}(0.05)^2(0.95)^{103}=0.069288$
$P(X=3)=\binom{105}{3}(0.05)^3(0.95)^{102}=0.125206$
$P(X=4)=\binom{105}{4}(0.05)^4(0.95)^{101}=0.168040$
$\vdots$
$\vdots$
$P(X=105)=\binom{105}{105}(0.05)^{105}(0.95)^0=2.465190328815676e-137$

Probability distribution of Y is shown below:

x	y	$P(Y=y)$
0	$5(100)-50(5)=250$	0.00458
1	$5(100)-50(4)=300$	0.025317
2	$5(100)-50(3)=350$	0.069288
3	$5(100)-50(2)=400$	0.125206
4	$5(100)-50(1)=450$	0.168040
$\vdots$		
$\vdots$		
105	$5(0)=0$	$2.465190328815676e-137$

Hence, Expected gain $=E(Y)$
$=250(0.00458)+300(0.025317)+350(0.069288)+400(0.125206)+450(0.168040)+\ldots+0(2.465190328815676e-137)$

$=457.47282444948553$ Using below Code

Code

```
import numpy as np
from scipy.stats import binom
```

```
s = 105
totalgain = 0
for i in range(0,106):
    if (i >= 5):
        gain = (s-i)*5
    elif i < 5:
        gain = 100*5 - (s-i-100)*50
    totalgain += gain* binom.pmf(i , s , 0.05)
print(totalgain)
```

P.S.: This is what many international airlines do. They often sell more tickets than the number of available seats thereby profiting twice from the same seat!

10. In recently conducted elections, there were a total of 100 counting centres. The losing party claims that some of the counting machines were rigged by a hacker. To verify these allegations, the Election Commission decides to manually recount the votes in some centres (obviously, manual recounting in all centres would be prohibitively expensive so it can only do so in some centres).

    (a) (½ point) If 5% of the machines were rigged then in how many centres should recounting be ordered so that there is a 50% chance that rigging would be detected (i.e., in at least one of the selected centres the number of votes counted manually will not match the number of votes counted by the machine)

    Solution:

    Assuming one center has one machine only.

    A: Event that a particular center is rigged.

    $P(A) = \frac{1}{20}$ 5% of machines are rigged.
    $P(A') = \frac{19}{20}$
    Let the minimum number of centers required be n.
    B: Event that atleast one centre out of these n is rigged.
    C: Event that No centre is rigged out of these n .

    Let X be the random variable: No of centers that are rigged out of these n centers. It follows a binomial distribution with $p = \frac{1}{20}$.

    So, $P(C) = P(X = 0) = \binom{n}{0}(\frac{1}{20})^0(\frac{19}{20})^n = (\frac{19}{20})^n$

    $P(B) = P(X >= 1) = 1 - P(X = 0) = 1 - P(C) = 1 - (\frac{19}{20})^n$

For atleast 50% chance that rigging would be detected $P(B) \geq 0.5$

$\implies 1 - (\frac{19}{20})^n \geq 0.5$
$\implies (\frac{19}{20})^n \leq 0.5$

Using the code below $n = 14$.

Python Code:

```
n = 1
while True:
    if (0.95 ** n <= 0.5):
        break
    else:
        n = n+1
print(n)
```

Note: Here I am solving for atleast 50% chance , there is no solution for exact 50%.

(b) (½ point) If the hacker knows that the Election commission can only afford to do a recounting in 10 randomly sampled centres then what is the maximum number of machines he/she can rig so that there is less than 50% chance that the rigging will get detected.

Solution:
Let the maximum number of machines he can rig be n
A: Event that a particular center is rigged.

$P(A) = \frac{n}{100}$.
$P(A') = 1 - \frac{n}{100}$
B: Event that atleast one center out of the 10 randomly chosen centers is rigged.
C: Event that No machine is rigged out of these 10 randomly chosen centers .

Let X be the random variable: No of machines that are rigged out of the 10 randomly chosen centers. It follows a binomial distribution with $p = \frac{n}{100}$.

So, $P(C) = P(X = 0) = \binom{10}{0}(\frac{n}{100})^0(1 - \frac{n}{100})^{10} = (1 - \frac{n}{100})^{10}$

$P(B) = P(X >= 1) = 1 - P(X = 0) = 1 - P(C) = 1 - (1 - \frac{n}{100})^{10}$

For less than 50% chance that rigging would be detected $P(B) < 0.5$

$1 - (1 - \frac{n}{100})^{10} < 0.5$
$\implies (1 - \frac{n}{100})^{10} > 0.5$
Using the code below $n = 6$.

Python Code:

```
n = 1
while True:
    if ((1-(n/100)) ** 10 > 0.5):
        n = n+1
    else:
        break
print(n-1)
```

11. (1 point) Amar and Bala are two insurance agents. Their manager has given them a list of 40 potential customers and a target of selling a total of 5 policies by the end of the day. They decide to split the list in half and each one of them talks to 20 people on the list. Amar is a better salesman and has a probability p_1 of selling a policy when he talks to customer. On the other hand, Bala has a probability p_2 ($< p_1$) of selling a policy when he talks to a customer. The customers do not know each other and hence one customer does not influence another. What is the probability that they will be able to meet their target by the end of the day? (it doesn't matter if Amar sells more policies than Bala or the other way round - the only thing that matters is that the total should be **exactly** 5).

Solution:

Let X_1 be the random variable denoting the number of policies sold by Amar and X_2 be the random variable denoting the number of policies sold by Bala.

Clearly X_1 follows a binomial distribution with $n = 20$ and probability of success as p_1 and X_2 also follows a binomial distribution with $n = 20$ and probability of success as p_2.

The required probability is $P(X_1 + X_2 = 5)$

Now $P(X_1 + X_2 = 5) = \sum_{i=0}^{5} P(X_1 = i)P(X_2 = 5 - i)$

$$= \sum_{i=0}^{5} \binom{20}{i} p_1^i (1-p_1)^{20-i} \binom{20}{5-i} p_2^{5-i} (1-p_2)^{20-5+i}$$

The above question was solved by assuming that the Insurance agents should just complete the target.

Now If we assume that at the end of the day means that the the target should be completed at the end of the day only and not in between the the questions becomes a question of negative binomial distribution:

Now using this assumption both X_1 and X_2 follow negative binomial distribution.

Now we have 6 cases:
1. Amar sells 0 policies, Bala sells 5.
2. Amar sells 1 policies, Bala sells 4.
3. Amar sells 2 policies, Bala sells 3.
4. Amar sells 3 policies, Bala sells 2.
5. Amar sells 4 policies, Bala sells 1.
6. Amar sells 5 policies, Bala sells 0.

All these possibilities are included in the following summation:

So, $P(\text{Meeting the target}) = (1-p_1)^{20} \times \binom{19}{4} p_2^5 (1-p_2)^{15} + [\sum_{i=1}^{4} \{ \binom{19}{i-1} p_1^i (1-p_1)^{20-i} \times \binom{19}{4-i} p_2^{5-i} (1-p_2)^{15+i} \}] + \binom{19}{4} p_1^5 (1-p_1)^{15} \times (1-p_2)^{20}$

12. Find the expectation and variance of the following discrete random variables

(a) (½ point) A binomial random variable whose distribution is fully specified by p (probability of success) and n (number of trials)

Solution:
The pmf of a binomial random variable is : $p_X(x) = \binom{n}{x} p^x q^{n-x}$, where $q = 1-p$

For Expectation: $E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x}$

$= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} q^{n-x}$

$= np(q+p)^{n-1} = np \qquad\qquad q+p=1$

For variance:
$E(X^2) = \sum_{x=0}^{n} x^2 \binom{n}{x} p^x q^{n-x}$

$= \sum_{x=0}^{n}(x(x-1)+x)\frac{n(n-1)}{x(x-1)}\binom{n-2}{x-2}p^x q^{n-x}$

$= n(n-1)p^2[\sum_{x=2}^{n}\binom{n-2}{x-2}p^{x-2}q^{n-x}] + np\sum_{x=1}^{n}\binom{n-1}{x-1}p^{x-1}q^{n-x}$

$= n(n-1)p^2(q+p)^{n-2} + np$

$= n(n-1)p^2 + np$

Now, variance $= V(X) = E(X^2) - [E(X)]^2$
$= n(n-1)p^2 + np - n^2p^2 = np(1-p)$

(b) (½ point) A negative binomial random variable whose distribution is fully specified by p (probability of success) and r (fixed number of desired successes)

Solution: The pmf of a negative binomial random variable is given as:

$p_X(k) = \binom{k+r-1}{r-1}(1-p)^k p^r$
$p_X(k)$ is the probability that number of failures we observe is k before we get exactly r successes.
r is the number of predefined successes.
p is the probability of success.
The random variable here is the number of failures we observe before we get exactly r successes.

As derived in Q5 part b this pmf can also be written as:
$p_X(k) = (-1)^k\binom{-r}{k}(1-p)^k p^r$

$= \binom{-r}{k}(p-1)^k p^r$

$= \binom{-r}{k}(-q)^k p^r$, where $q = 1-p$

For Expectation:
$E(X) = \sum_{k=0}^{\infty} k\binom{-r}{k}(-q)^k p^r$

$= p^r(-r)(-q)\sum_{k=1}^{\infty}\binom{-r-1}{k-1}(-q)^{k-1}$

$= p^r(-r)(-q)(1-q)^{-r-1}$

$= p^r(-r)(-q)(1-q)^{-r}(1-q)^{-1}$

$= p^r(-r)(-q)(p)^{-r}(p)^{-1}$

$= \frac{rq}{p}$

$= \frac{r(1-p)}{p}$

For variance:
$E(X^2) = \sum_{k=0}^{\infty} k^2 \binom{-r}{k}(-q)^k p^r$

$= \sum_{k=0}^{\infty}[k(k-1)+k]\binom{-r}{k}(-q)^k p^r$

$= p^r(-q)^2(-r)(-r-1)[\sum_{k=2}^{\infty}\binom{-r-2}{k-2}(-q)^{k-2}] + p^r(-q)(-r)[\sum_{k=1}^{\infty}\binom{-r-1}{k-1}(-q)^{k-1}]$

$= p^r(-q)^2(-r)(-r-1)(1-q)^{-r-2} + p^r(-q)(-r)(1-q)^{-r-1}$

$= p^r q^2 r(r+1)(1-q)^{-r}(1-q)^{-2} + p^r(-r)(-q)(1-q)^{-r}(1-q)^{-1}$

$= p^r q^2 r(r+1)(p)^{-r}(1-q)^{-2} + p^r(-r)(-q)(p)^{-r}(1-q)^{-1}$

$= q^2 r(r+1)(1-q)^{-2} + (-r)(-q)(1-q)^{-1}$

$= \frac{r^2q^2+rq^2}{p^2} + \frac{rq}{p}$

Now, variance $= V(X) = E(X^2) - [E(X)]^2$
$= \frac{r^2q^2+rq^2}{p^2} + \frac{rq}{p} - (\frac{rq}{p})^2$

$= \frac{rq^2}{p^2} + \frac{rq}{p}$

$= \frac{rq(p+q)}{p^2} = \frac{rq}{p^2}$

(c) (½ point) A hypergeometric random variable whose distribution is fully specified by N (number of objects in the given source), a (number of favorable objects in the source) and n (size of the sample that you want to select)

Solution:
The pmf of a hypergeometric distribution is given by $p_X(x) = \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$, where x is the number of observed successes.

For Expectation:

$E(X) = \sum_{x=0}^{n} x \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$

$= \frac{a}{\binom{N}{n}} \sum_{x=1}^{n} \binom{a-1}{x-1}\binom{N-a}{n-x}$

$= \frac{a}{\binom{N}{n}} \sum_{x=0}^{m} \binom{A}{k}\binom{N-A-1}{m-k}$, where $k = x-1, m = n-1, a-1 = A$,

$= \frac{a}{\binom{N}{n}} \binom{N-1}{m}$

Because $\sum_{x=0}^{m} \binom{A}{k}\binom{N-A-1}{m-k}$ is the sum of numerators all values of a geometric random variable with paramters $N-1, A, m$ and should be equal to denominator because total sum is 1.

$= \frac{a}{\binom{N}{n}} \binom{N-1}{n-1}$

$= \frac{na}{N}$

For variance:

$E(X^2) = \sum_{x=0}^{n} x^2 \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$

$= \sum_{x=0}^{n} [x(x-1)+x] \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$

$= \frac{a(a-1)}{\binom{N}{n}} \sum_{x=2}^{n} \binom{a-2}{x-2}\binom{N-a}{n-x} + \sum_{x=0}^{n} x \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$

$= \frac{a(a-1)}{\binom{N}{n}} \binom{N-2}{n-2} + E(X)$

$= \frac{a(a-1)n(n-1)}{N(N-1)} \binom{N-2}{n-2} + \frac{na}{N}$

$= \frac{a(a-1)n(n-1)}{N(N-1)} + \frac{na}{N}$

Now, variance $= V(X) = E(X^2) - [E(X)]^2$
$= \frac{a(a-1)n(n-1)}{N(N-1)} + \frac{na}{N} + (\frac{na}{N})^2$

On further simplification:

$V(X) = \frac{na(N-a)(N-n)}{N^2(N-1)}$

(d) (½ point) A Poisson random variable whose distribution is fully specified by λ (i.e., arrival rate in unit time)

Solution:
The pmf of a poisson random variable is: $p_X(x) = \frac{e^{\lambda}\lambda^x}{x!}$

For Expectation:
$E(X) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!}$

$= \lambda e^{-\lambda} \sum_{x=1}^{\infty} x \frac{\lambda^{x-1}}{(x-1)!}$

$= \lambda e^{-\lambda}[1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots]$

$= \lambda e^{-\lambda} e^{\lambda}$ Using Expansion of e^{λ}
$= \lambda$

For Variance:
$E(X^2) = \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!}$

$= [x(x-1) + x]\frac{e^{-\lambda}\lambda^x}{x!}$

$= e^{-\lambda} \sum_{x=0}^{\infty} x(x-1)\frac{\lambda^x}{x!} + \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!}$

$= e^{-\lambda}\lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda$

$= e^{-\lambda}\lambda^2[1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots] + \lambda$

$= e^{-\lambda}\lambda^2 e^{\lambda} + \lambda$

$= \lambda^2 + \lambda$

Now, variance $= V(X) = E(X^2) - [E(X)]^2$
$= \lambda^2 + \lambda - \lambda^2 = \lambda$

13. (1 point) Consider a language which has only 5 words w_1, w_2, w_3, w_4, w_5. The way you construct a sentence in this language is by selecting one of the 5 words with probabilities p_1, p_2, p_3, p_4, p_5 respectively $(\sum_{i=1}^{5} p_i = 1)$. This word will be the first word in the sentence. You will then repeat the same process for the second word and continue to form a sentence of arbitrary length. As should be obvious, the i-th word in the sentence is independent of all words which appear before it (and after it). What is the expected position at which the word w_2 will appear for the first time?

Solution:

Let X be the random variable that denotes the position word w_2 appears for the first time.

Clearly X can take values in $\{1, 2, 3, \ldots\}$.

Clearly X is a geometric random variable that takes value 1 with probability p_2, 2 with probability $(1-p_2)p_2$, 3 with probability $(1-p_2)^2p_2$ and so on.
So, $p_X(x) = (1-p_2)^{x-1}p_2$

The expected value of this random variable is:

$E(X) = \sum_{x=1}^{\infty} x(1-p_2)^{x-1}p_2$

$= p_2 \sum_{x=1}^{\infty} x(1-p_2)^{x-1}$

$= p_2 \frac{d}{dp_2}[-\sum_{x=1}^{\infty}(1-p_2)^x]$

$= p_2 \frac{d}{dp_2}[-\frac{1}{p_2}]$

$= p_2(\frac{1}{p_2})^2 = \frac{1}{p_2}$

14. Two fair dice are rolled. Let X be the sum of the two numbers that show up and let Y be the difference between the two numbers that show up (number on first dice minus number on second dice).

(a) (½ point) Show that $E[XY] = E[X]E[Y]$

Solution:

Let A and B be independently distributed random variables denoting the outcome of these two fair dice rolled.

Both A and B can take values in $\{1, \ldots, 6\}$. The random variables take these values with the same probability of $\frac{1}{6}$.
Clearly $E(A) = E(B)$.
Also, $E(A^2) = E(B^2)$.

Now, $X = A + B$ and $Y = A - B$.

Now $E(X)E(Y) = E(X)(E(A) - E(B)) = 0$

Also, $E(XY) = E((A - B)(A + B)) = E(A^2 - B^2) = E(A^2) - E(B^2) = 0 = E(X)E(Y)$

Hence shown.

(b) (½ point) Are X and Y independent? Explain your answer.

Solution: No, X and Y are not independent.
Consider $X = 2, Y = 0$
Now, $P_{XY}(xy) = P(X = 2, Y = 0) = \frac{1}{36}$ Only Favourable event being $(1, 1)$
But $P(X)P(Y) = \frac{1}{36}\frac{6}{36} = \frac{6}{36} \neq P(XY)$

Hence X and Y are not independent.

15. The *martingale doubling system* is a betting strategy in which a player doubles his bet each time he loses. Suppose that you are playing roulette in a fair casino where the roulette contains only 36 numbers (no 0 or 00). You bet on red each time and hence your probability of winning each time is 1/2. Assume that you enter the casino with 100 rupees, start with a 1-rupee bet and employ the martingale system. Your strategy is to stop as soon as you have won one bet or you do not have enough money to double the previous bet.

(a) (½ point) Under what condition will you not have enough money to double your previous bet?

Solution:
We will run out of money if we lose bets continuously and money left will be less than the double of previous betted money.
Let k be the maximum number of straight losses we can afford.
Then , $100 - (1 + 2 + 2^2 + \ldots + 2^{k-1}) < 2^k$

$\implies 100 - (2^k - 1) < 2^k$

$\implies 101 < 2.2^k$

$\implies 2^{k+1} > 101$

$\implies k+1 > \log_2 101$

$\implies k+1 > 6.65$
$\implies k > 5.65$
Smallest k that satisfies this is 6.

Hence the condition is we start playing the game and lose 6 games continuously at the begining.

(b) (½ point) What would your expected winnings be under this system? (for every 1 INR you bet you get 2 INR if you win)

Solution: $p = \frac{1}{2}$, the probability of winning a game.

$q = \frac{1}{2}$, the probability of losing a game.

Assuming that the 2 INR I get on winning include the 1 INR i used for betting such that i get 1 INR as the profit i.e. on every win i get 1 INR as profit for every 1 INR i bet.

Now if I win at the k^{th} bet then my net gain is:

$-(1+2+2^2+\ldots+2^{k-1})+2^k = -(2^k-1)+2^k = 1$
This means that whenever I win I will end up with 1 INR as the profit.

Now I can afford only 6 straight losses at the starting of game.
So, probability that I cannot continue betting is $(\frac{1}{2})^6$

Let X be the the random variable representing my total gain.

This takes value $-(1+2+\ldots+2^5) = -63$ with probability $(\frac{1}{2})^6$ and the value 1 with probability $1-(\frac{1}{2})^6$.

Hence my total expected gain $= E(X) = 1(1-(\frac{1}{2})^6) - 63(\frac{1}{2})^6$
$= 0$

In this derivation I have considered winnings as my gain which is more appropriate.(Because when we consider winnings it is basically gain e.g. If I won 10

INR and lost 5 INR it means I won only 5 INR).

Now, If we consider winning as only the money I win without considering my loss (as said in discussion forum) then expected winnings becomes:
$= 2 \times \frac{1}{2} + 4 \times \frac{1}{4} + 8 \times \frac{1}{8} + \ldots + 0 \times (1 - (\frac{1}{2} + \frac{1}{4} + \ldots + \frac{1}{32})) = 6$

16. (1 point) You have 800 rupees and you play the following game. A box contains two green hats and two red hats. You pull out the hats out one at a time without replacement until all the hats are removed. Each time you pull out a hat you bet half of your present fortune that the pulled hat will be a green hat. What is your expected final fortune?

Solution: The only 4 possible sequences of draws are as below:
White - White - Black - Black
White - Black - White - Black
Black - White - Black - White
Black - Black - White - White

In each sequence our fortune gets divided by half twice as we lose twice and multiplied by $\frac{3}{2}$ twice. (Current fortune + $\frac{1}{2}$ of current fortune for winning).

Let X be the random variable that denotes the final fortune.
It can only one value as all the 4 sequences have two losses and two wins.
The value it takes is $800 \times \frac{1}{2} \times \times \frac{1}{2} \times \frac{3}{2} \times \frac{3}{2} = 450$ and it takes this value with probability 1.

Hence Expected final fortune $= E(X) = 1 \times 450 = 450$

17. There are 6 dice. Each dice has 0 on five sides and on the 6th side it has a number between 1 and 6 such that no two dice have the same number (i.e, if dice 2 has the number 3 on the 6th side then no other dice can have the number 3 on the sixth side). All the 6 dice are rolled and let X be the sum of the numbers on the faces which show up.

(a) Find $E[X]$ and $Var(X)$

Solution:
Without loss of generality let us number the dice.
i^{th} dice has the number i on one of the faces and 0 on the rest of the faces.

$i \in \{1,2,3,4,5,6\}$

Let X_i be the random variable denoting the number appearing on the i^{th} dice.

Now, $E(X_i) = 0 \times \frac{5}{6} + i \times \frac{1}{6} = \frac{i}{6}$

Now, $X = \sum_{i=1}^{6} X_i$

$E(X) = \sum_{i=1}^{6} E(X_i)$ Linearity of Expectation

$= \frac{1}{6}\sum_{i=1}^{6} i$

$= \frac{1}{6}(\frac{6(6+1)}{2}) = \frac{7}{2}$

For variance:

Let us first calculate $E(X_i^2)$
$E(X_i^2) = 0^2 \times \frac{5}{6} + i^2 \times \frac{1}{6} = \frac{i^2}{6}$

Now let us consider X_iX_j.
It takes only two values 0 and ij.
To calculate expectation we need only the probability of ij which will be $(\frac{1}{6})^2$

So, $E(X_iX_j) = 0 \times P(X_iX_j) + ij \times (\frac{1}{6})^2 = \frac{ij}{36}$

Now, $X = \sum_{i=1}^{6} X_i$

So, $E(X^2) = E(\sum_{i=1}^{6} X_i)^2$

$= \sum_{i=1}^{6} E(X_i^2) + \sum_{i=1}^{6}\sum_{j=1(j\neq i)}^{6} E(X_iX_j)$ Linearity of Expectation

$= \sum_{i=1}^{6} \frac{i^2}{6} + \sum_{i=1}^{6}\sum_{j\neq i} \frac{ij}{36}$

$= \frac{1}{6}\sum_{i=1}^{6} i^2 + \frac{1}{36}\sum_{i=1}^{6}\sum_{j\neq i} ij$

$= \frac{1}{6}\frac{6(6+1)(2\times 6+1)}{6} + \frac{1}{36}\sum_{i=1}^{6}\sum_{j\neq i} ij$

$= \frac{91}{6} + \frac{1}{36}\sum_{i=1}^{6}\sum_{j\neq i} ij$

$= \frac{91}{6} + \frac{1}{36}(350) = \frac{896}{36} = \frac{224}{9}$

Now $Var(X) = E(X^2) - [E(X)]^2$
$= \frac{224}{9} - (\frac{7}{2})^2 = \frac{455}{36}$

(b) Suppose you are the owner of a casino which has a game involving these 6 dice. The players can bet on the sum of the numbers that will show up. If I bet on the number 21 what should the payoff be so that the bet looks as attractive as possible to me but in the long run the casino will not lose (e.g., in the game of roulette a payoff of 1:35 looks attractive while still protecting the interests of the casino).

Solution: Probability of getting a sum of 21 = $(\frac{1}{6})^6$

So, probability of not getting 21 as sum = $1 - (\frac{1}{6})^6$

Let the most attractive payoff for player be $1 : x$. (1 is for casino and x is for player)

Let X be the random variable denoting the gain of the player.

Then Expected Gain = $E(X) = x(\frac{1}{6})^6 - 1(1 - (\frac{1}{6})^6)$

To make the bet as attractive as possible and still protect the interests of the casino, we have to find the largest value of x such that $E(X) \leq 0$.

This condition ensures that the casino never loses in long run although the gain may be zero.

So, $x(\frac{1}{6})^6 - 1(1 - (\frac{1}{6})^6) \leq 0$

$\implies x(\frac{1}{6})^6 \leq (1 - (\frac{1}{6})^6)$

$\implies x \leq 6^6 - 1$

$\implies x \leq 60466175.$

So, largest value is $x = 46655$.

So the most attractive bet is $1 : 46655$.

18. (1 point) A friend invites you to play the following game. He will toss a fair coin till the first heads appears. If the first head appears on the k-th toss then he will give you 2^k rupees. His condition is that you should first pay him a 100 million rupees to get a chance to play this game. Would you be willing to pay this amount to get a chance to play this game? Explain your answer.

Solution:

Let X be the random variable that denotes my net gain.

Now the probability of getting a head at the k^{th} toss is $(1 - \frac{2}{2})^{k-1}\frac{1}{2} = (\frac{1}{2})^k$

X takes the values of the form $2^k - 100M$ where $k \in \{1, 2, 3, \ldots\}$

Now X takes the value $2^k - 100M$ with probability $(\frac{1}{2})^k$

$E(X)$ represents my expected gain.
Now, $E(X) = \sum_{k=1}^{\infty}(2^k - 100M)(\frac{1}{2})^k$

$= \sum_{k=1}^{\infty} 2^k(\frac{1}{2})^k - \sum_{k=1}^{\infty} 100M(\frac{1}{2})^k$

$= \sum_{k=1}^{\infty} 1 - 100M \sum_{k=1}^{\infty}(\frac{1}{2})^k$

$= \sum_{k=1}^{\infty} 1 - 100M \times 1$
$= \infty$

This is a divergent sum and hence Expectation does not exist.
The value infinity means that the game is favorable for me now matter how much I pay.

Let us now consider a particular run of the game. Now my gain will be positive if $2^k > 100,000,000$ which gives $k = 27$.
This means that tail should come 26 times first and then head and the probability of this happening is $(\frac{1}{2})^{27}$ which is extremely less.
Hence my chances of gain in a particular game is extremely less.

These two things, one that expectation is ∞(basically doesn't exist) and the other that having the chances of gaining on a particular run of the game is extremely very less lead me to the conclusion that I will play the game if I have arbitrary large money, otherwise not.

19. Suppose you are playing with a deck of 20 cards which contain 10 red cards and 10 black cards. The dealer opens the cards one by one but you cannot see a card before he opens it. Before he opens a card you are supposed to guess the color of the card.

(a) (½ point) If you are guessing randomly then what is the expected number of correct guesses that you will make?

> **Solution:**
> Let X be the random variable denoting the number of correct guesses.
>
> Then the probability of getting the guess right is $\frac{1}{2}$ throughout.
>
> Let us show it for a guess where number of black cards remaining is b and number of red cards remaining is r.
> On this guess we are equally likely to chose any color.
>
> So, the probability of getting the guess correct $= (\frac{1}{2}\frac{b}{b+r} + \frac{1}{2}\frac{r}{b+r}) = \frac{1}{2}$.
>
> Hence X follows a binomial distribution with $n = 20$ and $p = \frac{1}{2}$.
>
> so, $E(X) = np = 20 \times \frac{1}{2} = 10$

(b) (½ point) Can you think of a better strategy than random guessing?

> **Solution:**
> A better strategy would be:
>
> 1. Count the cards dealt . then at the i^{th} deal:
>
> If the number of black cards remaining is more than red cards then guess black.
>
> If the number of black cards remaining is less than red cards then guess red.
>
> If the number of black cards remaining is equal to the number of red cards then guess randomly (equivalent to tossing a fair coin).

(c) (1 point) What is the expected number of correct guesses under this intelligent strategy? It is hard (but possible) to come up with an analytical solution. However, it is easy to do a simulation. Write a program to play this game a 1000 times and note down the number of correct guesses each time. Based on this simulation calculate the estimate number of correct guesses.

Solution:

Code

```
import numpy as np
from random import randint

# 0 represents black and 1 represents red
deck_array = np.array([ i % 2 for i in range(20) ])
total_correct = 0
for simulations in range(1000):
    current_deck =  np.random.permutation(deck_array)
    count_black = 0
    count_red = 0
    for j in range(20):
        if( count_black == count_red):
            guess = randint(0, 1)

        elif(count_black > count_red):
            guess = 1

        else:
            guess = 0

        if(current_deck[j] == 0):
            count_black += 1
        else:
            count_red += 1

        if (guess == current_deck[j]):
            total_correct = total_correct + 1

expectation = total_correct/1000
print(expectation)
```

Results

Estimate number of correct guesses one particular run of 1000 simulations = 12.342

20. Suppose every morning the front page of Chennai Times contains a photo of exactly one of the n celebrities of Kollywood. You are a movie buff and collect these photos. Of course, on some days the paper may publish the photograph of a celebrity which is already in your collection. Suppose you have already obtained photos of $k-1$ celebrities. Let X_k be the random variable indicating that number of days you have to wait before you obtain the next new picture (after obtaining the first $k-1$ pictures).

(a) (½ point) Show that X_k has a geometric distribution with $p = (n-k+1)/n$

Solution:

Clearly, $P(X_k = x) = P($ I have old pictures on $(x-1)$ days and on day x I have a new picture$)$.

$\implies P(X_k = x) = (\frac{k-1}{n})^{x-1}(\frac{n-k+1}{n})$

$$= (1 - \tfrac{n-k+1}{n})^{x-1}(\tfrac{n-k+1}{n})$$
$$= (1-p)^{x-1}(p), \text{ where } p = \tfrac{n-k+1}{n}$$

Hence, X_k has a geometric distribution with $p = \frac{n-k+1}{n}$.

(b) (½ point) Simulate this experiment with 50 celebrities. Carry out a large number of simulations and estimate the expected number of days required to get the photos of the first 25 celebrities and the next 25 celebrities. Paste your code and estimates of the two expected values below.

Solution:

Code

```
import numpy as np
from random import randint

n_celebrities = 50
n_simulations = 1000
count_first25 = 0
count_next25 = 0
count_check = 0

for i in range(n_simulations):
    current_celebrity_array = []
    count_check = 0
    while(len(current_celebrity_array) != 50):
```

```
        todays_photo = randint(1,n_celebrities)
        if todays_photo not in current_celebrity_array:
            current_celebrity_array.append(todays_photo)
            count_check  += 1
        if (count_check < 25):
            count_first25 += 1

        else:
            count_next25 += 1

print(count_first25/n_simulations )
print(count_next25/n_simulations)
```

Results

Expected number of days required to get the photos of the first 25 celebrities on one particular run of 1000 simulations = 33.321

Expected number of days required to get the photos of the next 25 celebrities on one particular run of 1000 simulations = 190.462

21. You want to test a large population of N people for COVID19. The probability that a person may be infected is p and it is the same for every person in the population. Instead of independently testing each person you decide to do pool testing wherein you collect the blood samples of k people and test them together (N is divisible by k). If the test is negative then you conclude that all are negative and no further tests are required for these k people. However, if the test is positive then you do k more tests (one for each person).

    (a) ($1/2$ point) What is the probability that the test for a given pool of k people will be positive?

    Solution: The test for the selected k people will be positive if at least one person among these k people is positive.
    Let X be the random variable denoting the number of persons positive in this group of k people.
    Then X follows a binomial distribution with probability of success as p and number of trials as k. (Assuming a person may be positive or negative independent of the others.)

    Now probability of test being positive $= P(X \geq 1) = 1 - P(X = 0)$

$$= 1 - \binom{k}{0} p^0 (1-p)^k = 1 - (1-p)^k$$

(b) (1 point) What is the expected number of tests required under this strategy to conclusively test the entire population?

Solution: No of groups than can be made= $\frac{N}{k}$
Let X_1 be a random variable denoting number of positive groups.
Then the random variable X_1 will be a binomial random variable with $n = \frac{N}{k}$ and probability of success $= 1 - (1-p)^k$

$E(X_1) = \frac{N}{k}(1 - (1-p)^k)$

Let Y be a random variable denoting the number of tests.

Then Y can be modelled as:
$Y = \frac{N}{k} + kX_1$

Now, $E(Y) = \frac{N}{k} + kE(X_1)$

$$= \frac{N}{k} + k(\frac{N}{k}(1 - (1-p)^k))$$

$$= \frac{N}{k}(1 + k - k(1-p)^k))$$

(c) (½ point) When would such a pooling strategy be beneficial?

Solution: Such a pooling strategy would be beneficial if the expected number of tests under this strategy is less than the number of individual tests required. The condition is:

$$\frac{N}{k}(1 + k - k(1-p)^k)) < N$$

22. A family decides to have children until they have a girl or until there are 3 children, whichever happens first. Let X be the random variable indicating the number of girls in the family and let Y be the random variable indicating the number of boys in the family. Assume that the probability of having a girl child is the same as that of having a boy child.

(a) (½ point) Find $E[X]$ and $Var[X]$.

Solution:

X: Random variable indicating the number of girls in the family.
Let the probability of having a girl child = Probability of having a boy child be p.
Since both are equal $p = \frac{1}{2}$(Assuming they will definitely have 3 children).

Now, X can take values 0 or 1.
It takes the value 0 when all the first 3 child are boys. This happens with probability p^3.
It takes value 1 when either first child is a girl or second child is a girl or third child is a girl.
This happens with probability $p + p(p) + p^2(p) = p + p^2 + p^3$

$$\text{Hence, } E(X) = 0(p^3) + 1(p + p^2 + p^3) = p + p^2 + p^3$$
$$= \tfrac{1}{2} + (\tfrac{1}{2})^2 + (\tfrac{1}{2})^3 = \tfrac{7}{8}$$

$$\text{Now, } E(X^2) = 0^2(p^3) + 1^2(p + p^2 + p^3) = p + p^2 + p^3$$

$$\text{Hence, } Var(X) = E(X^2) - E(X)^2 = (p + p^2 + p^3) - (p + p^2 + p^3)^2$$
$$= \tfrac{7}{8} - (\tfrac{7}{8})^2 = \tfrac{7}{64}$$

(b) (½ point) Find $E[Y]$ and $Var[Y]$.

Solution:
Y: Random variable indicating the number of boys in the family.
Let the probability of having a girl child = Probability of having a boy child be p.
Since both are equal $p = \frac{1}{2}$(Assuming they will definitely have 3 children).

Now, Y can take values 0 or 1 or 2 or 3.
It takes the value 0 when the first child is a girl. This happens with probability p.
It takes value 1 when first child is a boy and second child is a girl.
This happens with probability $p(p) = p^2$

It takes value 2 when first two child are boys and third is a girl.
This happens with probability $p^2(p) = p^3$

It takes value 3 when all the 3 child are boys.

This happens with probability p^3

Hence, $E(X) = 0(p) + 1(p^2) + 2(p^3) + 3(p^3) = p^2 + 2p^3 + 3p^3$
$= (\frac{1}{2})^2 + 2(\frac{1}{2})^3 + 3(\frac{1}{2})^3 = \frac{7}{8}$

Now, $E(X^2) = 0^2(p) + 1^2(p^2) + 2^2(p^3) + 3^2(p^3) = p^2 + 4p^3 + 9p^3$
$= (\frac{1}{2})^2 + 4(\frac{1}{2})^3 + 9(\frac{1}{2})^3 = \frac{15}{8}$

Hence, $Var(X) = E(X^2) - E(X)^2$
$= \frac{15}{8} - (\frac{7}{8})^2 = \frac{71}{64}$

23. Suppose n people bring their umbrellas to a meeting. While returning back each one randomly picks up an umbrella and walks out.

Let X_i be the random variable indicating whether the i-th person walked out with his own umbrella (i.e., the umbrella that he walks out with is the same as the umbrella that he walked in with).

(a) (½ point) Find $E[X_i^2]$

Solution: X_i be the random variable indicating whether the i-th person walked out with his own umbrella.

The probability that a particular person gets his own umbrella is $\frac{1}{n}$

Now, X_i can take only 2 values 0 and 1.
It takes value 0 when i^{th} person does not get his own umbrella. This happens with probability $1 - \frac{1}{n}$
It takes value 1 when i^{th} person does gets his umbrella. This happens with probability $\frac{1}{n}$

Hence, $E(X_i^2) = 0^2(1 - \frac{1}{n}) + 1^2(\frac{1}{n}) = \frac{1}{n}$

(b) (½ point) Find $E[X_iX_j]$ (for $i \neq j$)

Solution:
The random variable X_iX_j also can take only values 0 or 1.

It takes the value 0 when either the i^{th} person or the j^{th} or both do not get their own umbrella.

For calculating the Expectation we need only the probability that $X_iX_j = 1$

It takes the value 1 when both get their own umbrella.
This happens with probability $\frac{1}{n}\frac{1}{n-1} = \frac{}{n(n-1)}$

Hence, $E(X_iX_j) = 0 \times P(X_iX_j = 0) + 1 \times \frac{1}{n(n-1)} = \frac{1}{n(n-1)}$ for $i \neq j$

Let S be the random variable indicating the number of people who walk out with their own umbrella.

(c) (½ point) Find E[S]

Solution:
$E(S) = \sum_{k=1}^{n} kP(S = k)$

Let let S_i be an indicator for the event that the i^{th} man gets his own umbrella.

That is, $S_i = 1$ is the event that he gets his own umbrella, and $S_i = 0$ is the event that he gets the wrong umbrella.

$$E(S_i) = 1.P(S_i = 1) + 0.P(S_i = 0) \\ = \frac{1}{n}$$

Now S is the sum of these Indicator variables.
So, $S = S_1 + S_2 + \ldots + S_n$

Now, $E(S) = E(S_1 + S_2 + \ldots + S_n)$
$= E(S_1) + E(S_2) + \ldots + E(S_n)$ Linearity of Expectation
$= n.\frac{1}{n} = 1$

(d) (½ point) Var[S]

Solution:
$E(S^2) = E[(S_1 + S_2 + \ldots + S_n)^2]$

$= \sum_{i=1}^{n} E(S_i^2) + \sum_{i=1}^{n} \sum_{j=1(j\neq i)}^{n} E(S_iS_j)$ Linearity of Expectation

$= n.\frac{1}{n} + n(n-1)\frac{1}{n(n-1)} = 2$

Now, $Var(S) = E(S^2) - [E(S)]^2$
$= 2 - 1^2 = 1$

24. (½ point) The covariance of two random variables X and Y is defined as $Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$. Show that if X and Y are independent then $Cov(X, Y) = 0$.

Solution:

First we will prove if X and Y are independent, then $E(XY) = E(X)E(Y)$
We will do it for discrete random variables. Similar procedure applies to continuous random variables.

$E(XY) = \sum_i \sum_j xy f_{XY}(x_i, y_i)$

Since X and Y are independent the joint distribution factorizes.

So, $E(XY) = (\sum_i x_i f_x(x_i))(\sum_j y_j f_y(y_j) = E(X)E(Y)$

Now, $Cov(X, Y) = E[(X - E(X))(Y - E(Y])]$

$= E[XY - XE(Y) - YE(X) + E(X)E(Y)]$

$= E(XY) - E[XE(Y)] - E[YE(X)] + E[E(X)E(Y)]$

$= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y)$

$= E(X)E(Y) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y)$

$= 0$ Since X and Y are Independent

25. Consider a die which is loaded such that the probability of a face is proportional to the number on the face. Let X be the random variable indicating the outcome of the die.
    (a) (½ point) Find $E[X]$ and $Var(X)$

Solution:
X: Random variable indicating the outcome of the die.

Now X can take value i with probability ki where $i \in \{1, 2, 3, 4, 5, 6\}$ and k is a constant.

Now the sum of probabilities of a Random Variable is 1.

$$\sum_{i=1}^{6} ki = 1$$

$$\implies k \sum_{i=1}^{6} i = 1$$

$$\implies k \frac{6.7}{2} = 1$$

$$\implies k = \frac{2}{42} = \frac{1}{21}$$

Now $E(X) = \sum_{i=1}^{6} i(\frac{1}{21} i)$

$$= \frac{1}{21} \sum_{i=1}^{6} i^2$$

$$= \frac{1}{21} \frac{6(6+1)(2\times 6+1)}{6}$$

$$= \frac{91}{21} = \frac{13}{3}$$

Now $E(X^2) = \sum_{i=1}^{6} i^2(\frac{1}{21} i)$

$$= \frac{1}{21} \sum_{i=1}^{6} i^3$$
$$= \frac{1}{21} (\frac{6(6+1)}{2})^2$$

$$= 21$$

$Var(X) = E(X^2) - [E(X)]^2$

$$= 21 - (\frac{91}{21})^2$$

$$= \frac{980}{441} = \frac{20}{9}$$

(b) (½ point) Write code to simulate such a die. Run 1000 simulations and note down the number of times each face shows. Calculate the empirical expectation and see if it matches the theoretical expectation you have computed above.

Solution:

Code

```
from random import randint

count_one = 0
count_two = 0
count_three = 0
count_four = 0
count_five = 0
count_six = 0
for simulation in range(1000):
    number = randint(1,21)
    if(number == 1):
        count_one += 1
    elif(number >= 2 and number <=3):
        count_two += 1
    elif(number >= 4 and number <=6):
        count_three += 1
    elif(number >= 7 and number <=10):
        count_four += 1
    elif(number >= 11 and number <=15):
        count_five += 1
    else:
        count_six += 1


empirical_expectation = (1)*(count_one/1000) + (2)*(count_two/1000)\
+ (3)*(count_three/1000)+(4)*(count_four/1000)\
+(5)*(count_five/1000)+ (6)*(count_six/1000)

print(empirical_expectation)
```

Results

Empirical Expectation on one particular run of 1000 simulations = 4.22
Theoretical Expectation = $\frac{13}{3} = 4.33$

The two values are quite close to each other and as we increase the number of simulations the empirical expectation will get closer to theoretical expectation.