Ben Juarez
UID: 2134561

Problem Set 1

1 Problem 1

Taking n random people, we want to determine the probability that at least two of them have a common birthday. Let us first calculate $\mathbb{P}$(no birthday matches between n people) as follows knowing that the probability of any person being born on some day is simply $\frac{1}{365}$:

$$\implies \frac{365}{365} \cdot \frac{364}{365} \cdot ... \cdot \frac{365-n+1}{365} \implies \frac{365 \cdot 364 \cdot ... \cdot (365-n+1)}{365^n}$$

We can see that this probability can also be expressed as $\frac{365!}{365^n \cdot (365-n)!}$. So, since $\mathbb{P}$(at least 2 of n random people share a birthday) = 1 - $\mathbb{P}$(no birthday matches between n people), we arrive are the desired conclusion for this probability as follows:

$$\implies 1 - \frac{365 \cdot 364 \cdot ... \cdot (365-n+1)}{365^n}$$

See MATLAB section for corresponding plot. Furthermore, for $n = 55$, we have $\mathbb{P}(55) = 0.9862$. In other words, for 55 randomly selected people, there is a 98.62% chance that at least two of them share a birthday.

2 Problem 2

We know that this rare disease infects approx. 1 out of 1000 people in a population. We also know that for an infected individual, the test for this disease will come back positive 99% of the time. On the other hand, for an uninfected individual, the test will come back (false) positive 2% of the time. Let us examine the chances of infection for someone who just tested positive. Let us define events in the context of this problem as follows:

- Let I = infected $\implies \mathbb{P}(I) = 0.001$
- Let P = positive test $\implies \mathbb{P}(P) = \mathbb{P}(P \cap I) + \mathbb{P}(P \cap I^c) = \mathbb{P}(P|I)\mathbb{P}(I) + \mathbb{P}(P|I^c)\mathbb{P}(I^c)$
  - This statement holds because the probability of receiving a positive test needs to consider both conditions of being infected or uninfected. So, we take into the account the probability of being infected and testing positive and the probability of being uninfected and testing positive.

Since we want to find the conditional probability of being infected given that a positive test was received, we want to find $\mathbb{P}(I|P)$. Using Baye's Formula, we arrive at the following:

$$\mathbb{P}(I|P) = \frac{\mathbb{P}(P|I) \cdot \mathbb{P}(I)}{\mathbb{P}(P|I)\mathbb{P}(I) + \mathbb{P}(P|I^c)\mathbb{P}(I^c)} = \frac{(0.99)(0.001)}{(0.99)(0.001) + (0.02)(0.999)} = 0.0472$$

Thus, the chances of having the disease after receiving a positive test are 4.72%.

3 Problem 3

Suppose we toss a coin n times and let p denote the probability of heads. Let X be the number of heads.

3.1 Part a

We want to find the expected value of X. Clearly, we have n independent experiments, each of which results in heads (success) with probability p and in tails (failure) with probability $(1-p)$. Thus, we have a binomial distribution. We already have X denoted as the number of successes. Since we have $Bin(n,p)$, we conclude that the expected value of X is np.

$$E[X] = np$$

3.2 Part b

For $p = 0.3$ and $n = 100$, we have $E[X] = 30$. See MATLAB section for corresponding simulation. The computed average was 29.8880 for comparison.

4 Problem 4

We consider tossing a fair die with sample space $S = \{1, ..., 6\}$. Let us simulate $N = 10^4$ draws from the sample space and verify that $\hat{\mathbb{P}}(AB) \approx \hat{\mathbb{P}}(A)\hat{\mathbb{P}}(B)$. See MATLAB section for corresponding script with proper verification.

5 Problem 5

Let $Z \sim U[0,1]$ and let $X = F^{-1}(Z)$ where F^{-1} is the inverse of the CDF F. Let us assume F is strictly increasing.

5.1 Part a

We want to show that $X \sim F$, by showing that $F(x) = \mathbb{P}(X \leq x)$ We also have $X = F^{-1}(Z)$. With these, we can arrive at the following:

$$\mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(Z) \leq x) = \mathbb{P}(Z \leq F(x))$$

With $Z \sim U[0,1]$ we know that $\mathbb{P}(Z \leq F(x))$ must equal $F(x)$ since it is uniform distribution. We can further show this as follows:

$$\mathbb{P}(Z \leq F(x)) \implies \int_0^{F(x)} f(Z)dZ = \int_0^{F(x)} \frac{1}{1-0} dZ = F(x)$$

Therefore, we conclude $X \sim F$ meaning that F is the CDF of X.

5.2 Part b

6 Problem 6

Let us take a stick of unit length and break it at random where X is the length of the longer piece. We want to determine the expected value of X. We can see that this scenario is a case of uniform distribution such that the the breaking point $BP \sim U[0,1]$. Thus, we can further define X as follows:

$$X = \begin{cases} BP & \text{if } BP \geq \frac{1}{2} \\ 1 - BP & \text{if } BP < \frac{1}{2} \end{cases}$$

We define X this way since we know X must be the maximum of BP and $(1-BP)$. Furthermore, for $x \in [0,1]$ we have that the PDF is $f(x) = \frac{1}{1-0} = 1$. At this point, we can now calculate $E[X]$ as follows:

$$E[X] = \int_0^1 x \cdot f(x)\ dx = \int_0^{\frac{1}{2}} (1-x) \cdot 1\ dx + \int_{\frac{1}{2}}^1 x \cdot 1\ dx = \frac{3}{8} + \frac{3}{8} = \frac{6}{8} = \frac{3}{4}$$

7 Problem 7

Let us suppose we want to simulate a fair (unbiased) coin. In other words, we want to generate a random variable X that is equally likely to be 0 or 1. We have a biased coin such that we can generate a random variable Y that equals to 0 with probability p and 1 with probability $(1-p)$. Also, p is unknown. Let us also consider the given algorithms I and II.

Let us first examine algorithm I. Due to its construction, we can imagine four cases. If $Y_1 = Y_2$, we either have $Y_1 = Y_2 = 0$ or $Y_1 = Y_2 = 1$ which corresponds to $\mathbb{P}(Y_1 = Y_2 = 0) = p^2$ and $\mathbb{P}(Y_1 = Y_2 = 1) = (1-p)^2$ using the given information about p. Following algorithm I, we do not consider these first two cases. The cases we consider are $Y_0 = 0, Y_1 = 1$ and $Y_0 = 1, Y_1 = 0$ since these will reach step (4). Examining these cases further gives us the following:

$$\mathbb{P}(X = 0) = \mathbb{P}(Y_1 = 1, Y_2 = 0) = (1-p) \cdot p$$

$$\mathbb{P}(X=1) = \mathbb{P}(Y_1 = 0, Y_2 = 1) = p \cdot (1-p)$$
$$\implies (1-p) \cdot p = p \cdot (1-p) \implies \mathbb{P}(X=0) = \mathbb{P}(X=1)$$

Thus, we can see that with this algorithm, X is equally likely to be either 0 or 1 regardless of what p is. So, algorithm 1 successfully simulates an unbiased coin.

Let us now examine algorithm II. Following its construction, we can see that we essentially have two cases. Say we have n tosses of the biased coin such that the nth toss is different than the first $(n-1)$ tosses which must all be same. We can see that $n \geq 2$. Putting this in terms of $\mathbb{P}(X)$ we arrive at the following:

$$\mathbb{P}(X=1) = \mathbb{P}(Y_1 = 0, ..., Y_{n-1} = 0, Y_n = 1) = p^{n-1} \cdot (1-p)$$
$$\mathbb{P}(X=0) = \mathbb{P}(Y_1 = 1, ..., Y_{n-1} = 1, Y_n = 0) = (1-p)^{n-1} \cdot p$$

Thus, we can see that $\mathbb{P}(X=0) = \mathbb{P}(X=1)$ only if $p = \frac{1}{2}$ or if $n = 2$. We can state that n will not always be 2 because this clearly only happens if the second toss is different than the first. The chances of this happening are $p \cdot (1-p)$, so for any valid value of p, we can that there will always be a chance that $n > 2$.

So, since these conditions ($p = \frac{1}{2}$, $n = 2$) will not always hold, we cannot say that X will always have an equal chance of being 0 or 1. So, we conclude that algorithm II does not simulate a fair (unbiased) coin.

8 Problem 8

Suppose there is an infinite supply of pokemon balls containing pokemons such that each ball contains exactly one pokemon. We have that there are n different types of pokemons and each type is equally likley to appear in any one pokemon ball. We want to find the expected number of pokemon balls needed to open in order to have at least one pokemon of each type.

Let X_i represent the random variable associated with the number of pokemon balls opened until the ith distinct pokemon is found (after the $(i-1)$th distinct pokemon was found). We can also see that we have geometric distribution in this case because we are perform independent experiments until a success occurs such that a success is the discovery of the next distinct pokemon. Let us examine the probability of this ith distinct pokemon being found after $(i-1)$ pokemon have already been found:

$$\mathbb{P}(i\text{th distinct pokemon found in next box}) = \frac{n-(i-1)}{n} = \frac{n-i+1}{n} \approx p$$

This probability (denoted as p) is justified because with n total distinct pokemon, there are $n-(i-1)$ distinct pokemon remaining after the first $(i-1)$ are found.

Furthermore, let us now examine the expected value of this geometric distribution using p:

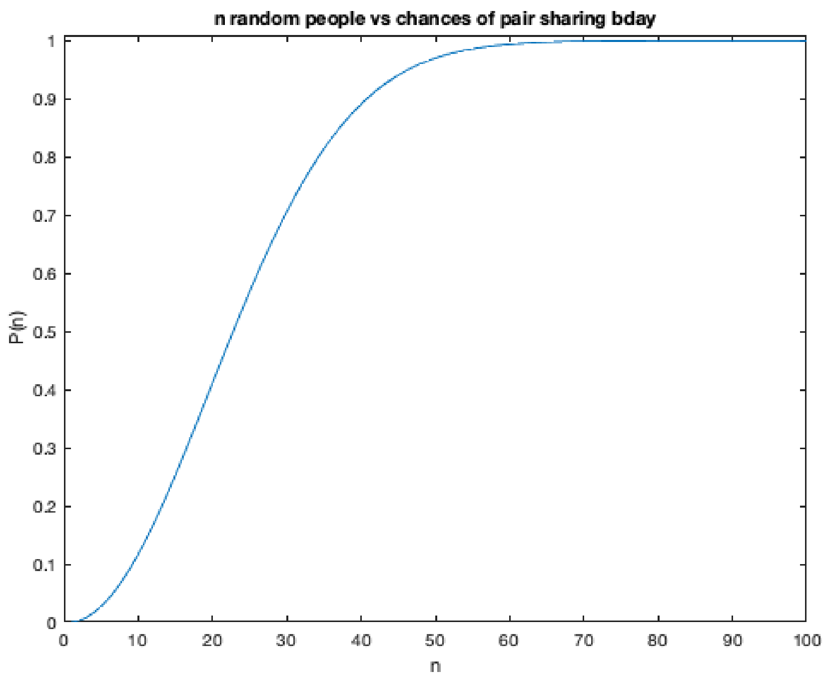
$$E[X_i] = \frac{1}{p} = \frac{n}{n-m+1} \implies E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{n}{n-i+1}$$

Thus, we have found the expected number of pokemon balls needed to open in order to have at least one pokemon of each type.

9 MATLAB

```
% Ben Juarez   ACM 116    PS1

bday = ones(100, 1);
b = 1;
for n = 1:100
    b = b .* ((365-n+1)/365);
    bday(n) = 1 - b;
end

bday(55)
plot(1:1:100, bday)
xlabel("n");
ylabel("P(n)");
ylim([0 1.01]);
title("n random people vs chances of pair sharing bday");
snapnow
```

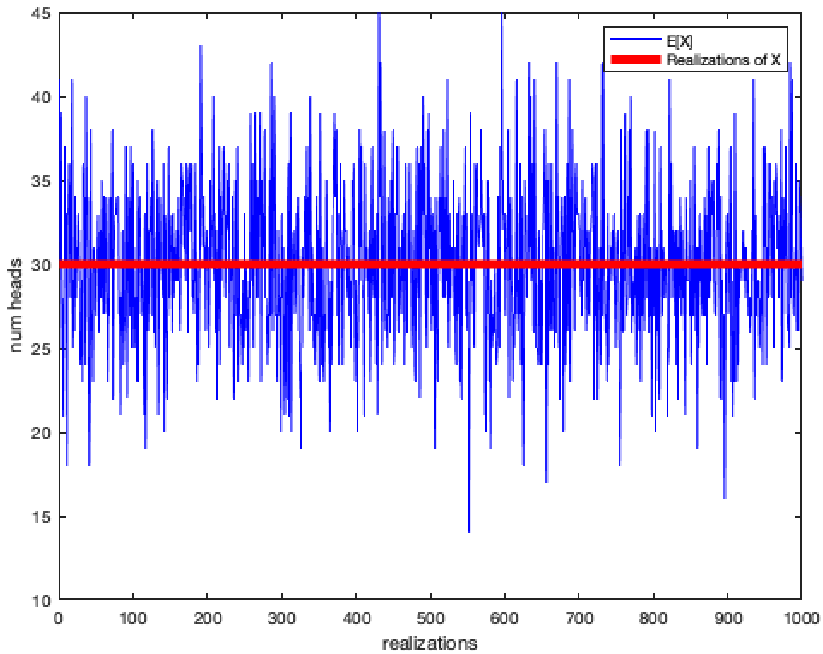
```
ans =

    0.9863
```



Published with MATLAB® R2021b

```
% Ben Juarez   ACM 116    PS1Q3b

line = zeros(1000,1);
X = zeros(1000,1);
sum = 0;
for i = 1:1000
    line(i) = 30;
    X(i) = binornd(100,0.3);
    sum = sum + X(i);
end
average = sum/1000
plot(X, 'blue', 'LineWidth', .5)
hold on
plot(line, 'red', 'LineWidth', 5)
legend("E[X]", "Realizations of X");
ylabel("num heads");
xlabel("realizations");
```

```
average =

   29.9950
```



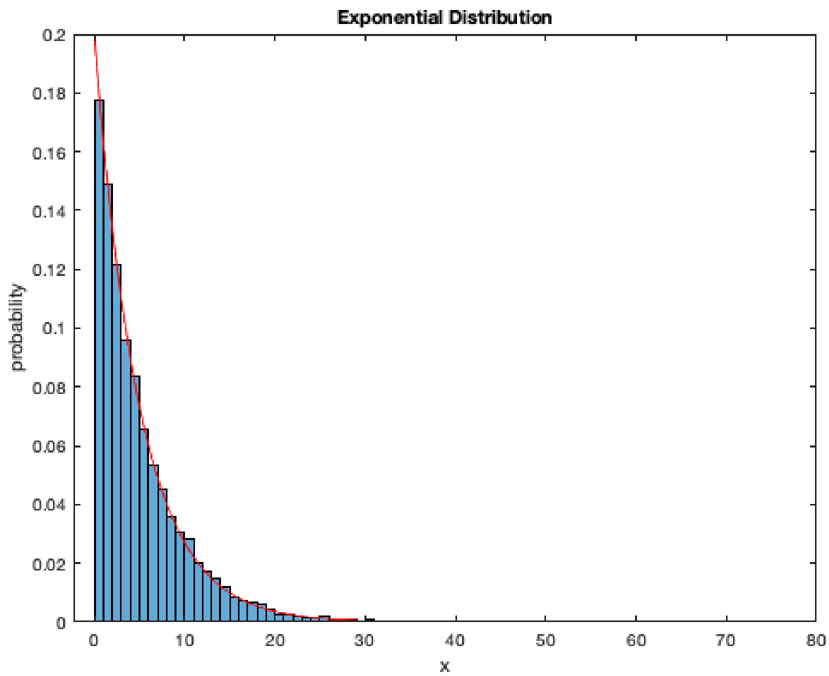
Published with MATLAB® R2021b

```
% Ben Juarez   ACM 116    PS1Q4
clear;
A = [2,4,6];
B = [1,2,3,4];
a = zeros(10^4,1);
b = zeros(10^4,1);
ab = zeros(10^4,1);
N = 10^4;
for i = 1:N
    roll_dice = unidrnd(6);
    if ismember(roll_dice, A) == 1 && ismember(roll_dice, B) == 1
        ab(i) = 1;
    end
    if ismember(roll_dice, A) == 1
        a(i) = 1;
    end
    if ismember(roll_dice, B) == 1
        b(i) = 1;
    end
end
P_A = sum(a)/10^4
P_B = sum(b)/10^4
P_AB = sum(ab)/10^4
P_A_P_B = P_A * P_B % since P(A)P(B) is approx. P(AB), we arrive at desired
 verification.
```

```
P_A =

    0.5056


P_B =

    0.6685


P_AB =

    0.3353


P_A_P_B =

    0.3380
```

Published with MATLAB® R2021b

```
% Ben Juarez   ACM 116    PS1Q5b
clear;
N = 10^4;
X = zeros(N, 1);
for i = 1:N
    X(i) = expinv(rand, 5);
end
histogram(X, "Normalization", "pdf")
hold on
x = 0:0.1:75;
plot(x, exppdf(x, 5), 'red')
xlabel("x");
ylabel("probability");
title("Exponential Distribution");
```



Published with MATLAB® R2021b