University of California - Los Angeles

ECE 131A

Probability and Statistics

MATLAB PROJECT

Student: Nha Do

UID: 405314461

Professor: Lara Dolecek

TAs: Lev Tauz & Debarnab Mitra

Winter 2021

1 Tossing a Fair and Unfair Dice

a) The simulation of tossing a 5-sided fair dice for t = 10, 50, 100, 500 and 1000 tosses.

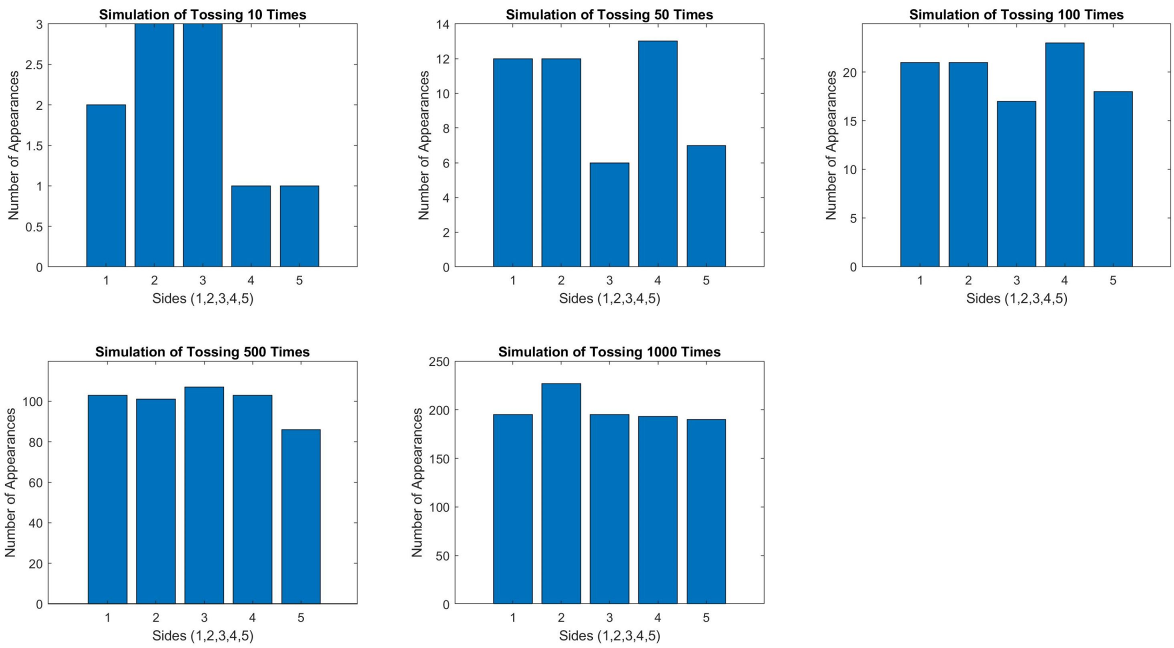

Figure 1: Simulation of Tossing 5-Sided Fair Dice

Based on number of appearances simulated above, we have the estimated probability of odd number below:

t	Odd	Probability(odd)
10	6	0.6
50	25	0.5
100	56	0.56
500	296	0.592
1000	580	0.58

b)

$$\text{P(X has odd value)} = P(X{=}1) + P(X{=}3) + P(X{=}5)$$
$$= \frac{1}{5} + \frac{1}{5} + \frac{1}{5} = \frac{3}{5} = 0.6$$

c)Since they are random and the real probability cannot be exactly the same as mathematical analysis, yet the results in part (a) are very close to part (b). Hence, they agree with the theoretical results and are acceptable.

d) We must have:

$P(X{=}1) + P(X{=}2) + P(X{=}3) + P(X{=}4) + P(X{=}5) = 1$

Based on the problem, we also have:

$$\begin{cases} 2P(X=1)+3P(X=5)=1 \\ P(X=1)=2P(X=5) \end{cases} \Rightarrow \begin{cases} P(X=1)=\frac{2}{7} \\ P(X=5)=\frac{1}{7} \end{cases}$$

Matlab simulation:

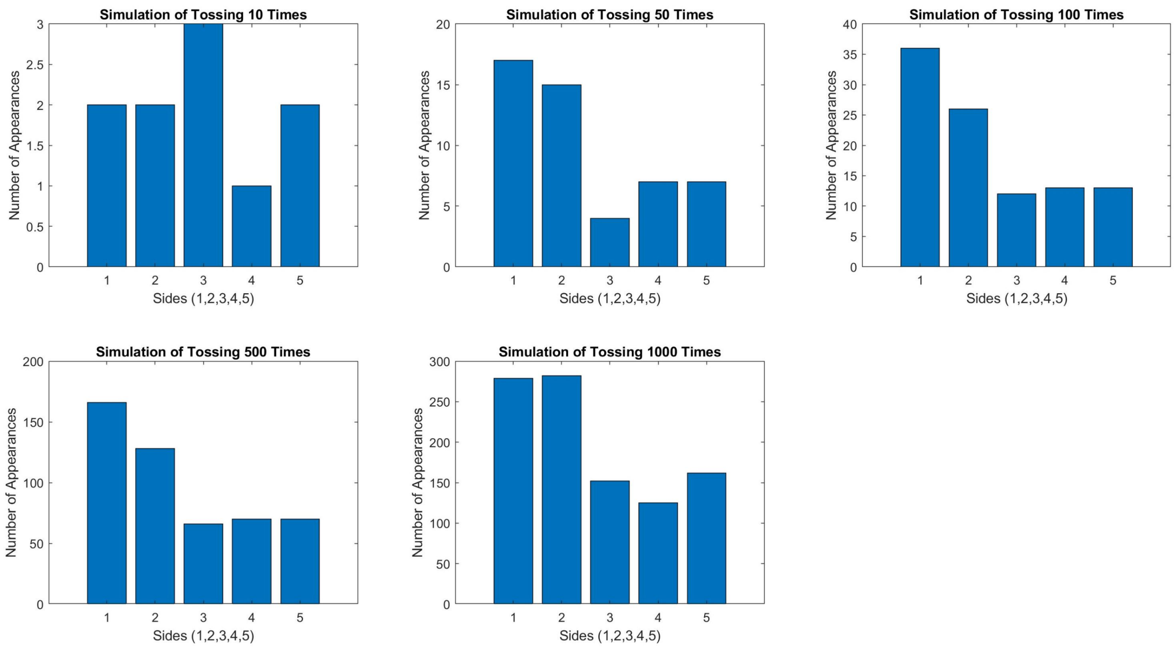

Figure 2: Simulation of Tossing 5-Sided Unfair Dice

Based on number of appearances simulated above, we have the estimated probability of odd number below:

t	Odd	Probability(odd)
10	7	0.7
50	28	0.56
100	61	0.61
500	302	0.604
1000	593	0.593

The exact probability is calculated as:

$$\text{P(X has odd value)} = \text{P(X=1)} + \text{P(X=3)} + \text{P(X=5)}$$

$$= \frac{2}{7} + \frac{1}{7} + \frac{1}{7} = \frac{4}{7} = 0.57$$

Increase the number of tossing makes the probability closer to mathematical analysis, other are also somewhat close. Hence, they agree with the theoretical results and are acceptable.

2 Coding for BEC

a) Matlab simulation for N-repetition code with N = [3 4 5] and p_e = 0.125:0.025:0.40. For a given N, simulate 100000 instances of sending a bit through the BEC.

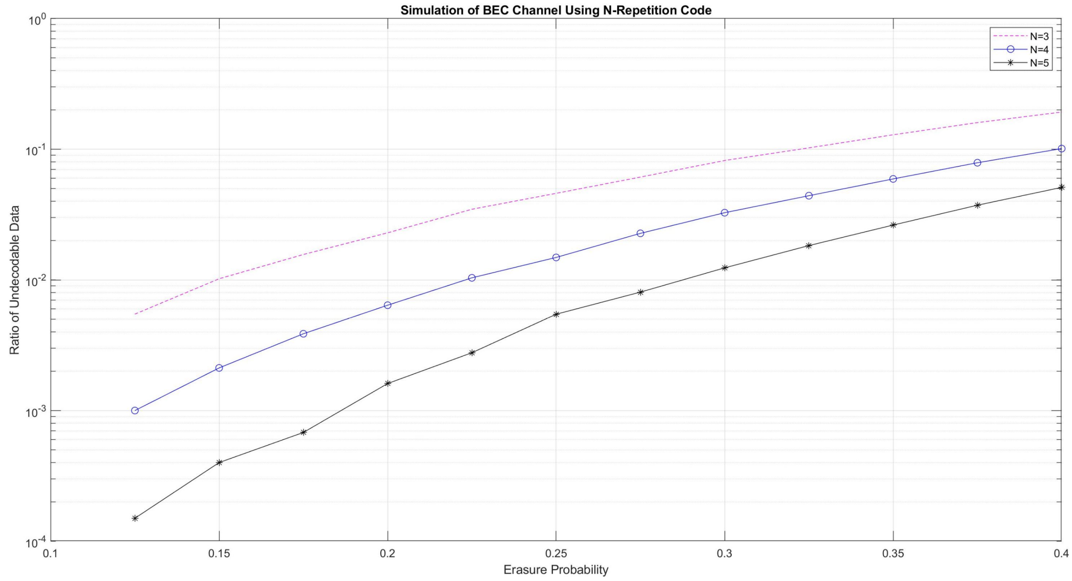

Figure 3: Simulation of BEC Using N-Repetition Code

b) The exact probability in each case is calculated below. It is parameterized by N and p_e:

$$P(undecodable) = Np_e^N$$

Matlab simulation for exact probability:

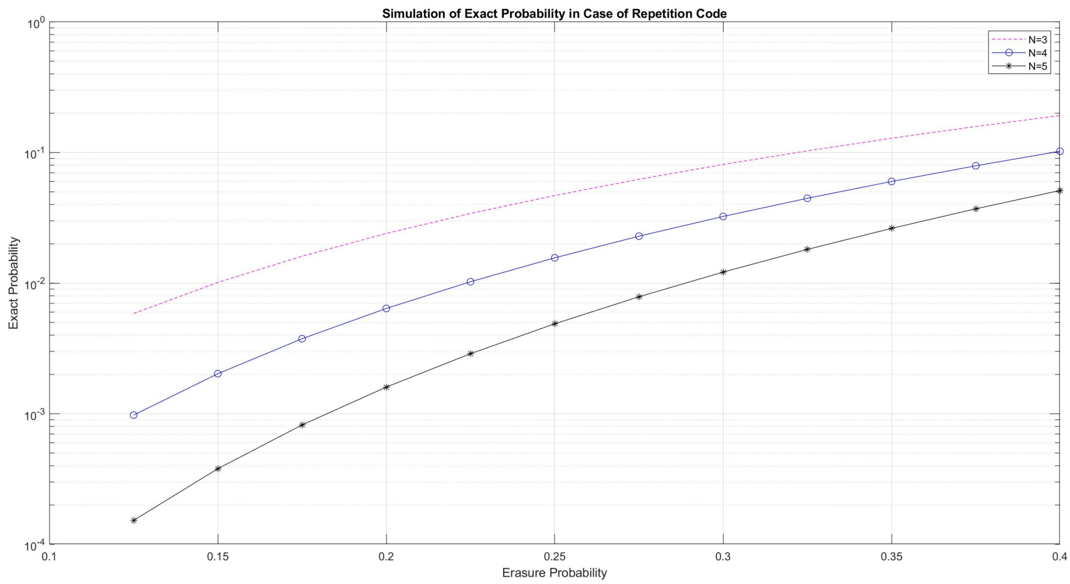

Figure 4: Simulation of Exact Probability for N-Repetition Code

c)Matlab simulation for N-single partiy code with N = [3 4 5] and p_e = 0.125:0.025:0.40. For a given N, simulate 100000 instances of sending a bit through the BEC.

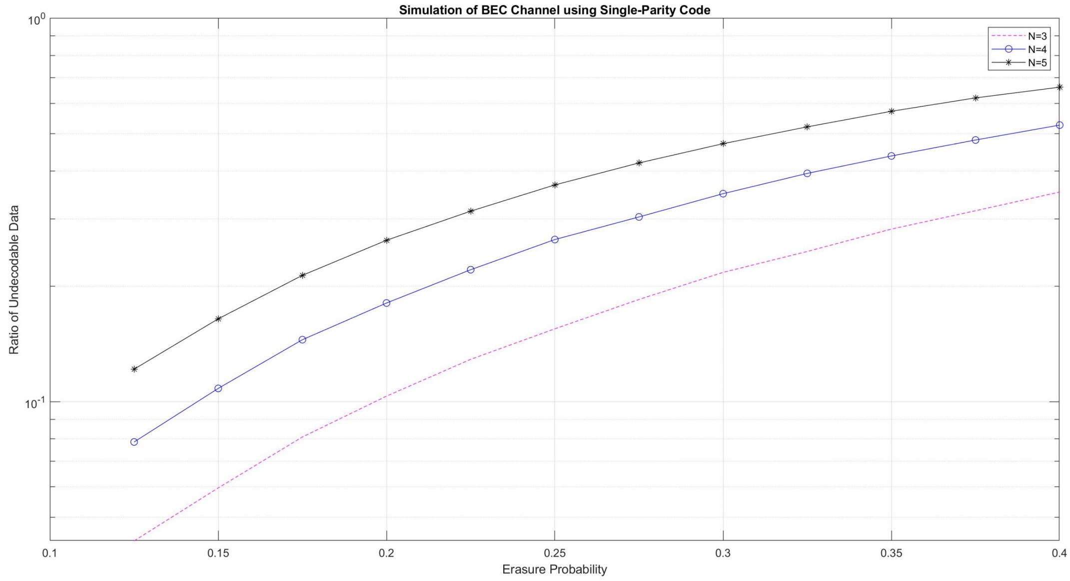

Figure 5: Simulation of BEC Using N-Single-Parity Code

The exact probability in each case is calculated below. It is parameterized by N and p_e:

$$1-(1-p_e)^N - N.p_e.(1-p_e)^{N-1}$$

Matlab simulation for exact probability:

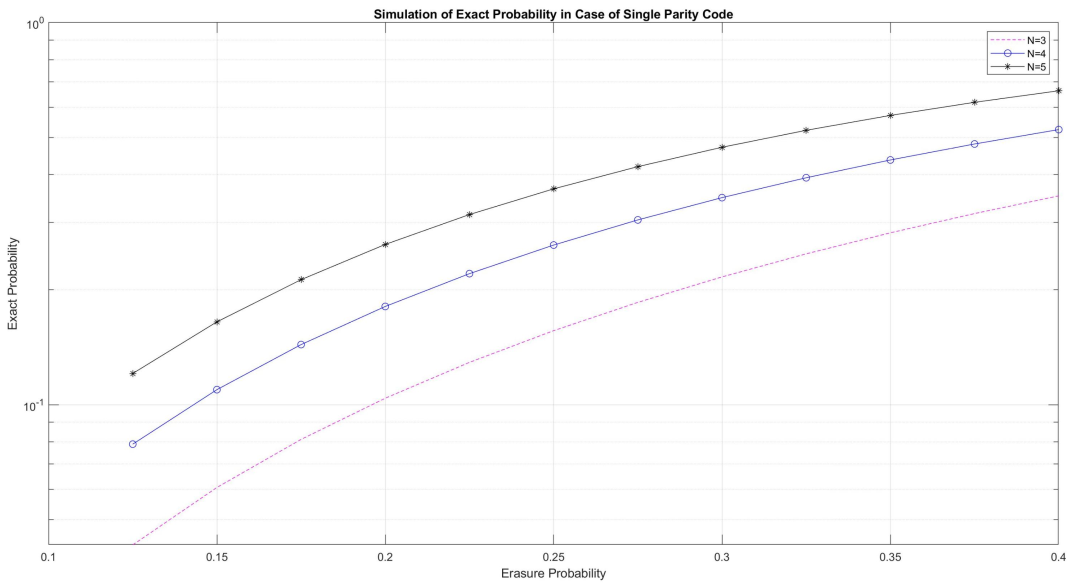

Figure 6: Simulation of Exact Probability for N-Single-Parity Code

d) Compare the plots of N-repetition code and N-single parity code, we can see that the probability of being undecodable in case of N-single parity code is larger. Since, in N-single parity code, a sequence of bit will become undecodable if at least 2 bits are erased. Therefore, when the length of bit sequence becomes larger, the probability of being undecodable also becomes

bigger.
Compare to N-repetition code, we are splitting each bit in the sequence and repeat it N times before transmitting. As long as the received bit sequence has at least 1 bit not erased, we can still receive the correct original bit sequence.

e) When the length of bit sequence is small, such as 1 or 2, we should use N-single parity code. Because the advantage of this kind of encoding is the receiver is capable of reconstructing the signal as long as no more than 1 bit is erased. The speed of this encoding is also faster because we can transmit a whole sequence at one time.
If the length of bit sequence becomes larger, we should use N-repetition code. In this case, the probability of being undecodable is small since the probability of all bits are erased when N large is rare and not likely to happen often. However, for N-repetition code, we have to repeat each bit in a sequence before transmitting, it will slow down the speed. Another disadvantage of N-repetition code compared to N-single parity code is the capable of reconstructing the signal.

3 Naiive Bayes Classifer

a) Bought: 0 for did not buy and 1 for did buy:

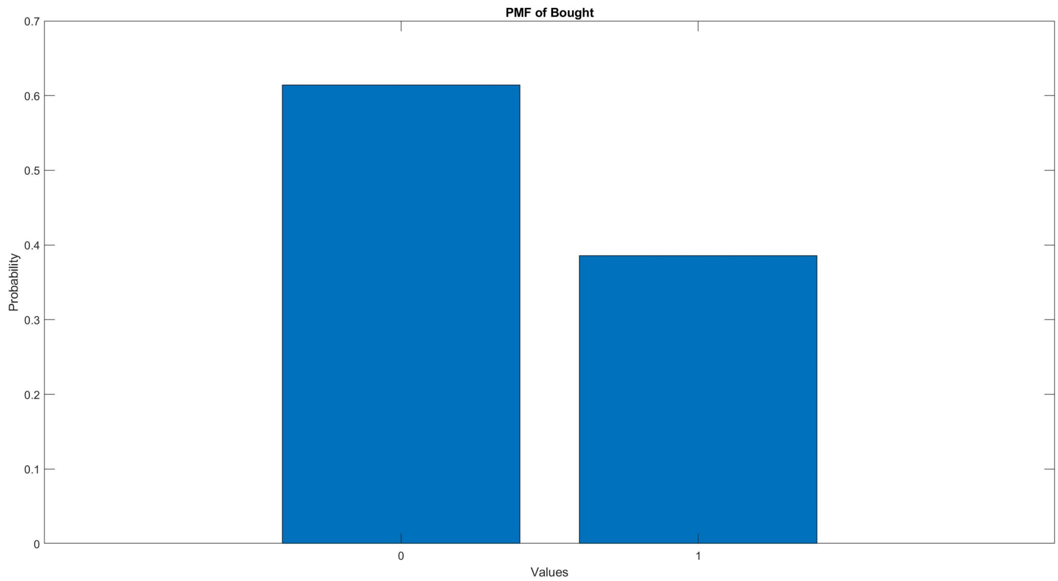

Figure 7: PMF of Bought

Type of spender: 1 for larger spender, 2 for medium spender, 3 for small spender

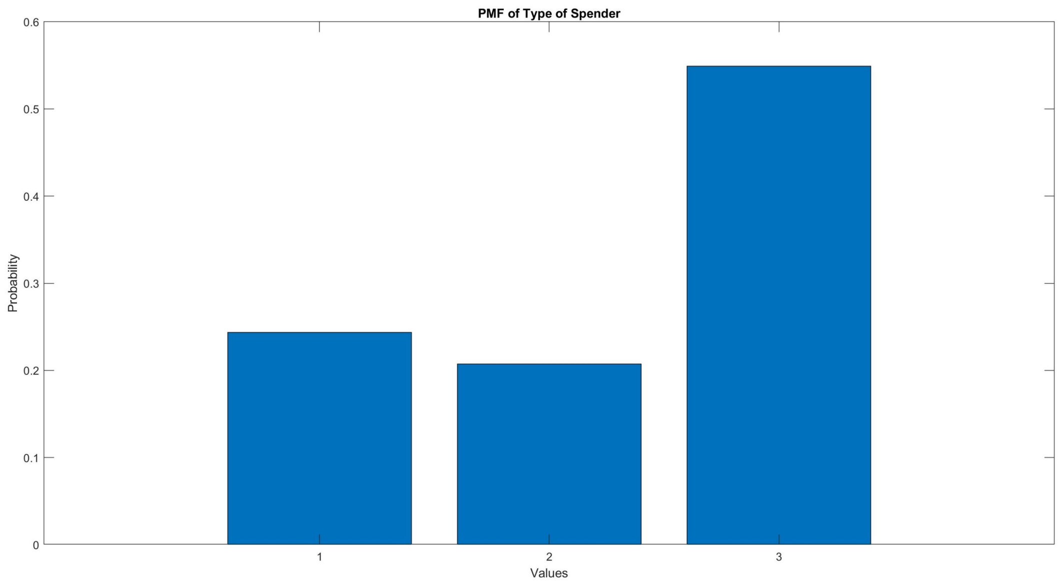

Figure 8: PMF of Type of Spender

Sex of user: 1 for Male and 0 for Female

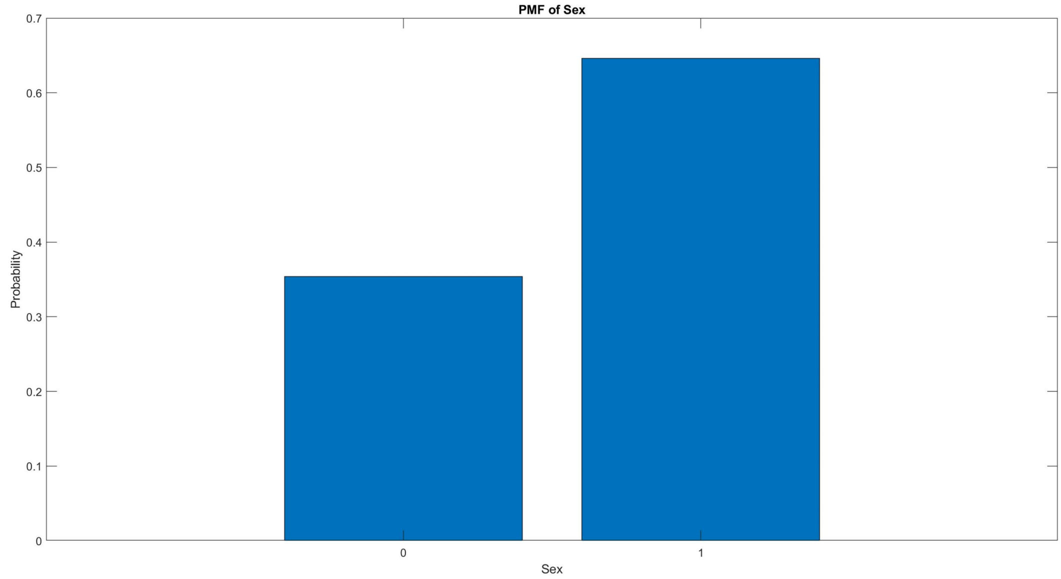

Figure 9: PMF of Sex of User

Age of user:

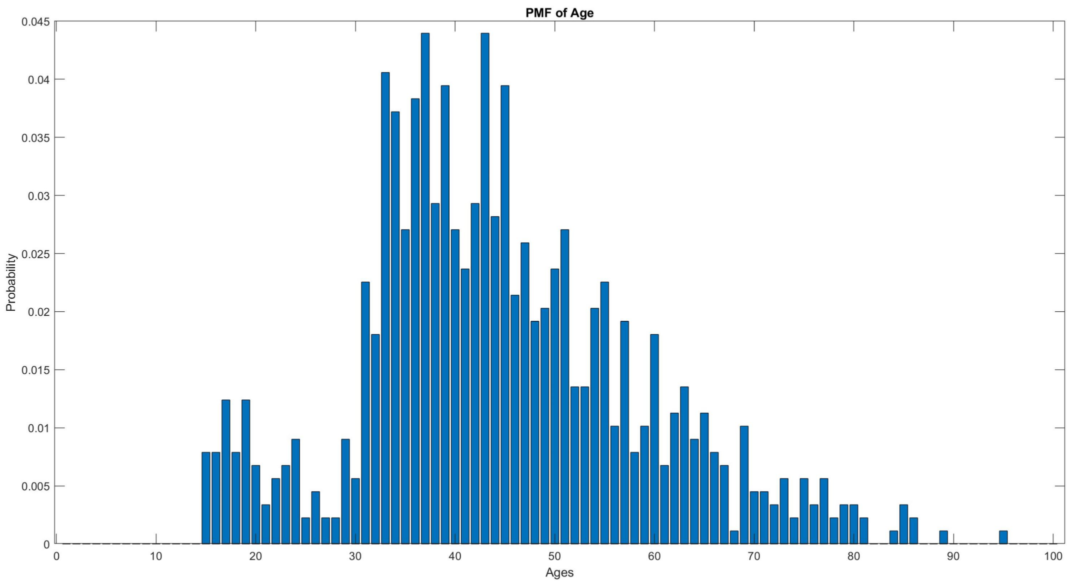

Figure 10: PMF of Age of User

b) PMF of T given B:

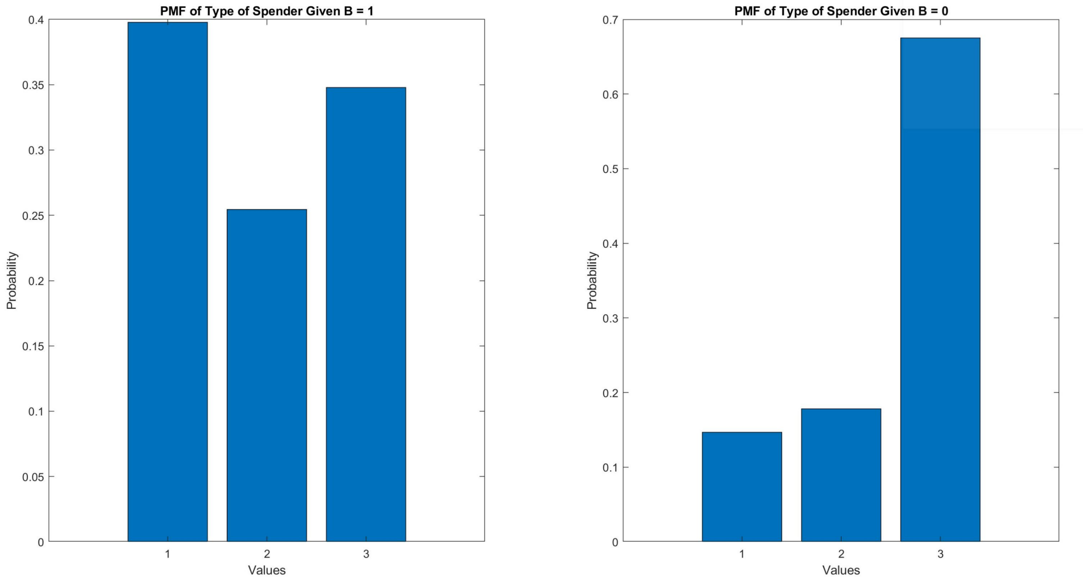

Figure 11: PMF of T given B = 1(left) and B = 0(right)

PMF of S given B:

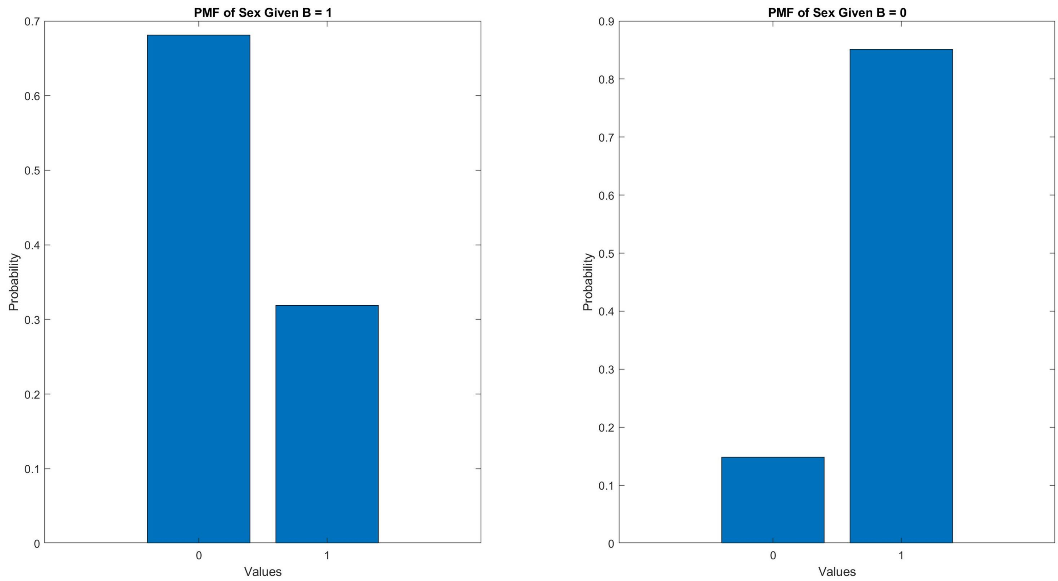

Figure 12: PMF of S given B = 1(left) and B = 0 (right)

PMF of A given B:

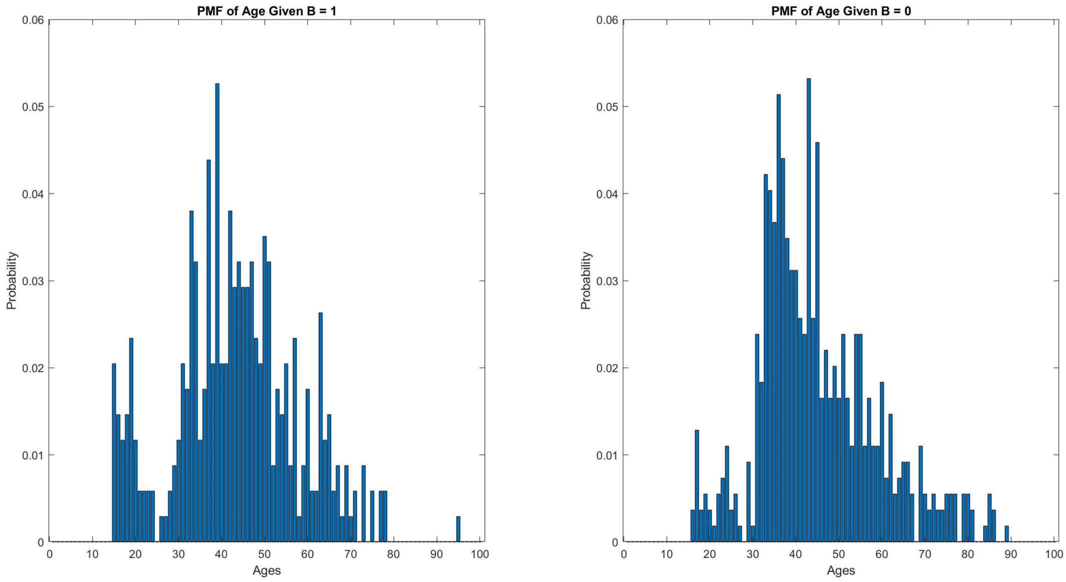

Figure 13: PMF of A given B = 1(left) and B = 0(right)

c)Based on results in question (a) and (b). I wrote the code to generate the values of some conditional probability. The code is attached in Appendix.

$P(T = 1 \mid B = 0)$	0.146789
$P(S = 0 \mid B = 0)$	0.148624
$P(A \leq 55 \mid B = 0)$	0.7963303
$P(T = 1 \mid B = 1)$	0.397661
$P(S = 0 \mid B = 1)$	0.681287
$P(A \leq 55 \mid B = 1)$	0.8099415

Hence:

$$P(T = 1, S = 0, A \leq 55 \mid B = 0) = P(T = 1 \mid B = 0).P(S = 0 \mid B = 0).P(A \leq 55 \mid B = 0)$$

$$= (0.146789).(0.148624).(0.7963303) = 0.01737$$

$$P(T = 1, S = 0, A \leq 55 \mid B = 1) = P(T = 1 \mid B = 1).P(S = 0 \mid B = 1).P(A \leq 55 \mid B = 1)$$

$$= (0.397661).(0.681287).(0.8099415) = 0.21943$$

$$P(B = 0, T = 1, S = 0, A \leq 55) = P(T = 1, S = 0, A \leq 55 \mid B = 0).P(B = 0)$$

$$= (0.01737).(0.6144307) = 0.01067$$

$$P(B = 1, T = 1, S = 0, A \leq 55) = P(T = 1, S = 0, A \leq 55 \mid B = 1).P(B = 1)$$

$$= (0.21943).(0.3855693) = 0.08461$$

d)

$$\begin{aligned}P(T=1, S=0, A \leq 55) &= P(T=1, S=0, A \leq 55 \mid B=0).P(B=0) \\ &+P(T=1, S=0, A \leq 55 \mid B=1).P(B=1) \\ &= (0.01737).(0.6144307) + (0.21943).(0.3855693) = 0.095\end{aligned}$$

Hence, using Bayes rule to make prediction:

$P(B=0 \mid T=1, S=0, A \leq 55) = \frac{P(T=1,S=0,A\leq 55|B=0).P(B=0)}{P(T=1,S=0,A\leq 55)} = \frac{(0.01737).(0.6144307)}{0.095} = 0.11$

$P(B=1 \mid T=1, S=0, A \leq 55) = \frac{P(T=1,S=0,A\leq 55|B=1).P(B=1)}{P(T=1,S=0,A\leq 55)} = \frac{(0.21943).(0.3855693)}{0.095} = 0.89$

Therefore, a female whose age is below 55 and who is a large spender will likely buy this product with probability of 0.89.

4 Central Limit Theorem

a) X_i be a uniform continuous random variable in the interval (3,7). The result below is generated by Matlab with n = 1,2,3,10,30,100 and number of samples is 10000.

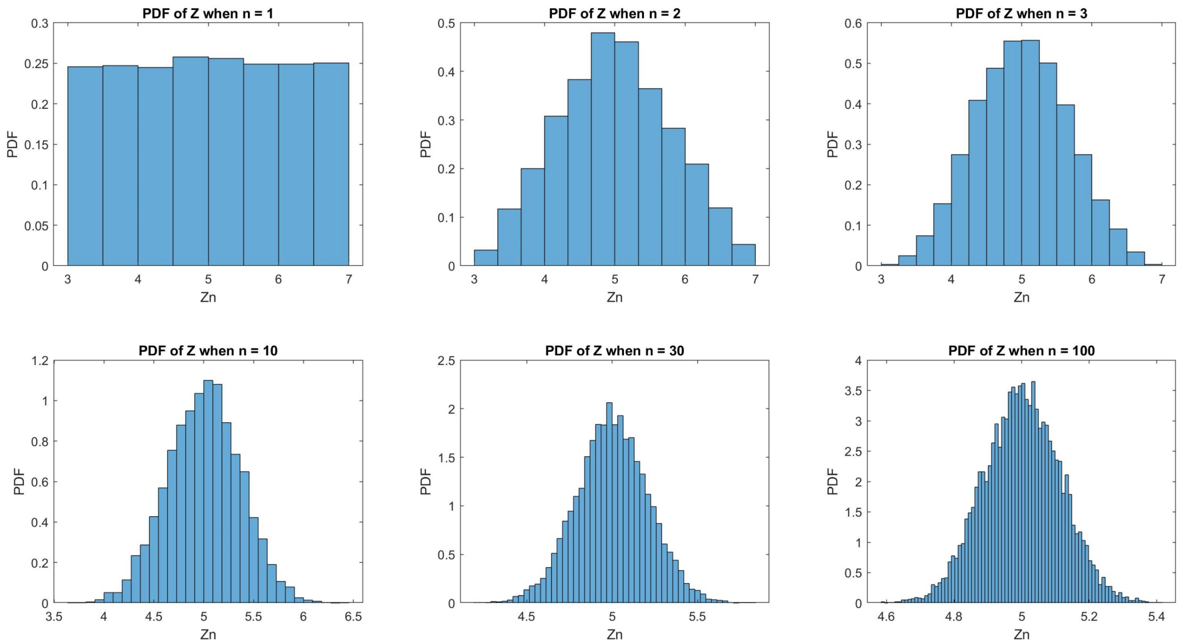

Figure 14: PDF of Z in different n

Observation: The larger n, the closer shape of PDF to the normal distribution (the bell curve).

b) Since Xi is a uniform continuous random variable in the interval (3,7), so:

$$E[X_i] = \frac{3+7}{2} = 5$$

$$VAR(X_i) = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{9+21+49}{3} - \frac{10^2}{4} = \frac{4}{3}$$

$$\Rightarrow E[Z_n] = \frac{E[X_1]+E[X_2]+...+E[X_n]}{n} = \frac{nE[X_1]}{n} = E[X_1] = 5$$

$$\Rightarrow VAR(Z_n) = \frac{\sigma^2}{n} = \frac{4}{3n}$$

c) Based on mean and variance of Z in question (b), I generated corresponding Gaussian distribution on each plots in (a) using normpdf. When n is high enough, we see that it comes closer to the Gaussian distribution.

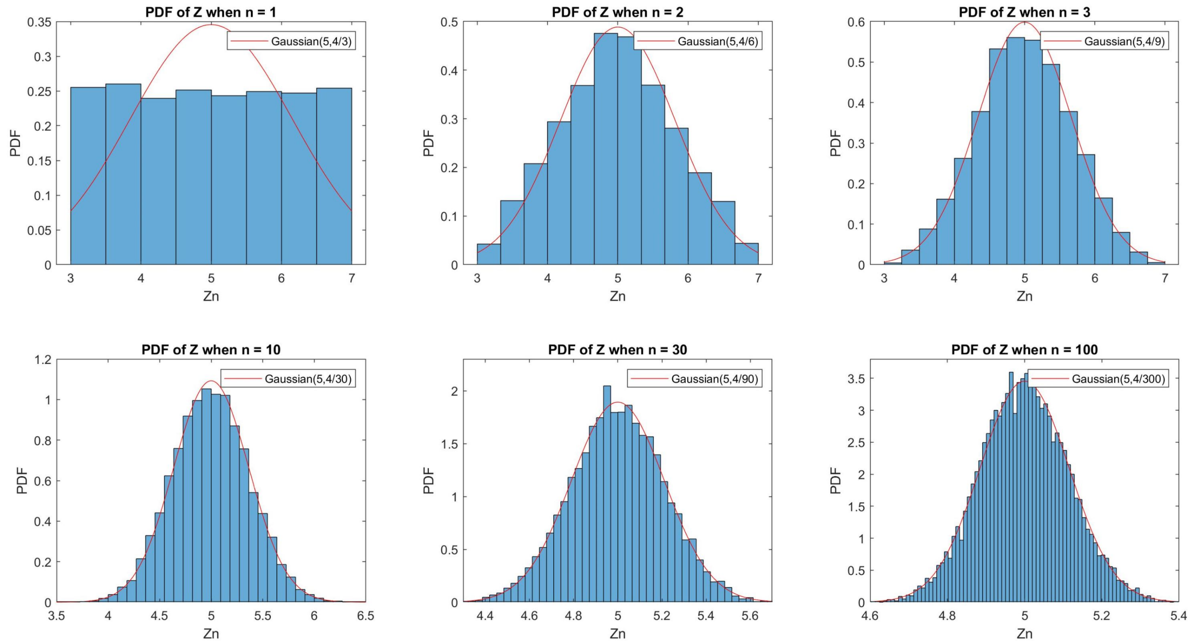

Figure 15: Superimpose Gaussian distribution on the previous plots

d) Now, Xi representing a toss of 5-sided unfair dice in question 1, number of samples = 10000. Again, we have:

$$P(X=1)=\tfrac{2}{7}, P(X=5)=\tfrac{1}{7}$$

X	1	2	3	4	5
P(X)	2/7	2/7	1/7	1/7	1/7

$$E[X]=\frac{2}{7}+\frac{4}{7}+\frac{3}{7}+\frac{4}{7}+\frac{5}{7}=\frac{18}{7}$$

$$E[X^2]=\frac{2}{7}+\frac{8}{7}+\frac{9}{7}+\frac{16}{7}+\frac{25}{7}=\frac{60}{7}$$

$$VAR(X)=E[X^2]-(E[X])^2=\frac{60}{7}-(\frac{18}{7})^2=\frac{96}{49}$$

$$\Rightarrow E[Z_n]=E[X]=\frac{18}{7}$$

$$\Rightarrow VAR(Z_n)=\frac{VAR(X)}{n}=\frac{96}{49n}$$

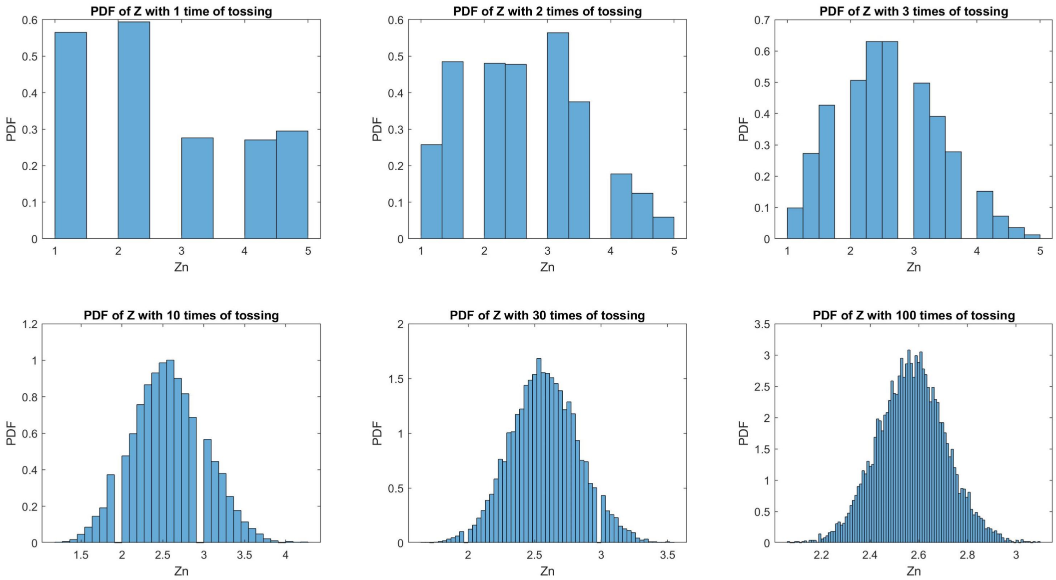

Figure 16: Histogram of Z with different n

Observation: Again, although the dice is unfair, when the number of tossing increases, the distribution comes closer to the normal distribution.

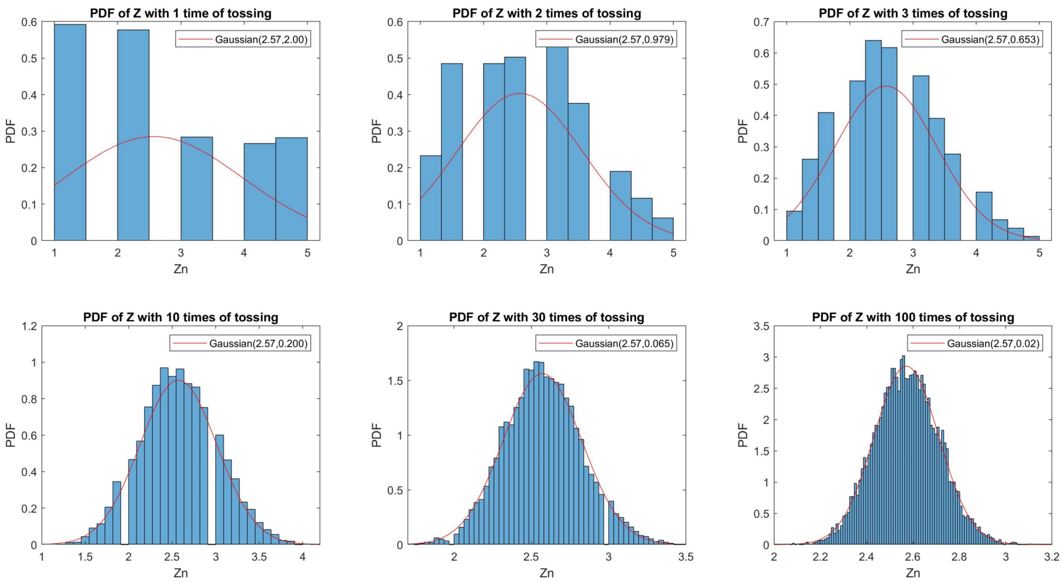

Figure 17: Superimpose Gaussian distribution on the previous plots