**MET CS544 A1 Foundation of Analytics**
**Homework#2**

Problem#1:
a) From the Bayes' rule example given in Section 3.10, compute the probabilities that a randomly selected non-smoker i) has lung disease and ii) does not have lung disease. Show the calculations without using R. Then, verify with the bayes function provided in the samples.

Solution:
Let L be the person having lung disease and NL be the person who is not having it. The prior is given by

P(L)= 0.07 and P(NL)= 0.93

Let NS be the non smoker and S be the smoker and the likelihood is given by

P(NS | L)= 0.10 and P(NS | NL)=0.75

The probability of selecting a non smoker who has lung disease is given by bayes theorem

P(L | NS)= P(NS | L) *P(L) / ( P(NS | L)*P(L) + P(NS | NL)* P(NL))
=(0.10 * 0.07) / ((0.10 * 0.07) +(0.75*0.93))
= 0.009936

The probability of selecting a non smoker who doesn't have lung disease is

P(NL | NS)= P(NS | NL) *P(NL) / ( P(NS | L)*P(L) + P(NS | NL)* P(NL))
=(0.75 * 0.93) / ((0.10 * 0.07) +(0.75*0.93))
= 0.99

b) Suppose that in a particular state, among the registered voters, 40% are democrats, 50 % are republicans, and the rest are independents. Suppose that a ballot question is whether to impose sales tax on internet purchases or not. Suppose that 70% of democrats, 40% of republicans, and 20% of independents favor the sales tax. If a person is chosen at random that favors the sales tax, what is the probability that the person is i) a democrat? ii) a republican, iii) an independent. Show the solutions with the calculations without using R. Then, verify with the bayes function provided in the samples.

Solution:
Let D be the democrats , R be the republicans and I be the independents.

P(D)=0.40 P(R)=0.50 P(I)=0.10

Let F be the event of favoring the sales tax on internet purchases and A be against it.

P(F | D)=0.70 P(F | R)=0.40 P(F | I)=0.20
P(A | D)=0.30 P(A | R)=0.60 P(A | I)=0.80

The probability of the person selected being a democrat is

P(D | F)= P(F | D)* P(D) / (P(F | D)* P(D) + P(F | R)* P(R)+ P(F | I)* P(I))
= (0.70* 0.40) / ((0.7*0.4)+(0.4*0.5)+(0.2*0.1))
=0.56

The probability of the person selected being a republican is

P(R | F)= P(F | R)* P(R) / (P(F | D)* P(D) + P(F | R)* P(R)+ P(F | I)* P(I))
= (0.40* 0.50) / ((0.7*0.4)+(0.4*0.5)+(0.2*0.1))
=0.4

The probability of the person selected being a independent is

$$P(I \mid F)= P(F \mid I)^* P(I) / (P(F \mid D)^* P(D) + P(F \mid R)^* P(R)+ P(F \mid I)^* P(I))$$

$$= (0.20^* 0.10) / ((0.7^*0.4)+(0.4^*0.5)+(0.2^*0.1))$$

$$=0.04$$

```
> #Bayes rule
> bayes<-function(prior,likelihood){
+   num<-prior*likelihood
+   den<-sum(num)
+   return(num/den)
+ }
> #Problem--1(a)
> prior<-c(0.07,0.93)
> likelihood<-c(0.10,0.75)
> bayes(prior,likelihood)
[1] 0.009936125 0.990063875
> #Problem--1(b)
> prior<-c(0.40,0.50,0.10)
> likelihood<-c(0.70,0.40,0.20)
> bayes(prior, likelihood )
[1] 0.56 0.40 0.04
> 
```

Problem#2:

a) Consider the experiment of rolling a pair of dice. Show how would you define a random variable for the absolute value of the difference of the two rolls. What is the support of this random variable?

Solution:

This experiment takes sample space of size 36. The random variable is defined for the absolute value of difference of the two rolls is defined as

$$X=|X1-X2|$$

where X => random variable for the absolute difference

X1=> events of the first dice

X2=> events of the second dice

which makes the support of the random variable,

$$\text{support}(X)= \{ 0,1,2,3,4,5\}$$

b) Now, use R and set up the probability space and add the random variable. What is the probability that the two rolls differ by 3? What is the probability that the two rolls differ by at most 3? What is the probability that the two rolls differ by at least 4? Use the Prob function.

Solution:

The probability space is set up and the random variable is added using the R and the following are computed,

```
> #Problem--2(b)
> library(prob)
> S<-rolldie(2,makespace=TRUE)
> s<-addrv(S,X=abs(X1-X2))
> s
   X1 X2 X      probs
1   1  1 0 0.02777778
2   2  1 1 0.02777778
3   3  1 2 0.02777778
4   4  1 3 0.02777778
5   5  1 4 0.02777778
6   6  1 5 0.02777778
7   1  2 1 0.02777778
8   2  2 0 0.02777778
9   3  2 1 0.02777778
10  4  2 2 0.02777778
11  5  2 3 0.02777778
```

```
>
> Prob(s,X== 3)
[1] 0.1666667
> Prob(s,X<= 3)
[1] 0.8333333
> Prob(s,X>= 4)
[1] 0.1666667
>
>
```

c) Show the marginal distribution of the above random variable (using R).

Solution:

```
> #Problem--2(c)
> marginal(s,vars="X")
  X      probs
1 0 0.16666667
2 1 0.27777778
3 2 0.22222222
4 3 0.16666667
5 4 0.11111111
6 5 0.05555556
```

d) Using R, add another random variable to the above probability space using a user defined function. The random variable is TRUE if one roll is twice the other, and FALSE otherwise. What is the probability that one roll is twice the other? Show also the marginal distribution for this random variable.

Solution:

```
> #Problem--2(d)
> rolltwice<-function(x){
+ if(x[[1]]==2*x[[2]] | x[[2]]==2*x[[1]]){
+ return (TRUE)
+ }else{
+ return (FALSE)
+ }
+ }
> s1<-addrv(s,FUN=rolltwice,name="RT")
> s1
   X1 X2 X    RT      probs
1   1  1 0 FALSE 0.02777778
2   2  1 1  TRUE 0.02777778
3   3  1 2 FALSE 0.02777778
4   4  1 3 FALSE 0.02777778
5   5  1 4 FALSE 0.02777778
6   6  1 5 FALSE 0.02777778
7   1  2 1  TRUE 0.02777778
8   2  2 0 FALSE 0.02777778
9   3  2 1 FALSE 0.02777778
10  4  2 2  TRUE 0.02777778
```

```
> Prob(s1,RT==TRUE)
[1] 0.1666667
> marginal(s1,vars="RT")
     RT     probs
1 FALSE 0.8333333
2  TRUE 0.1666667
>
```

Problem#3:
a) Write the function is.prime(x) that returns TRUE/FALSE depending upon if x is a prime number or not. Use the following rules:
i) 2 and 3 are prime numbers
ii) Any even number other than 2 is not a prime number
iii) For any odd number other than 3, take the square root of x (as integer). Check if x is divisible by any odd number from 3 to max(3, square root of x).
Do not use any loops in the above code.

Solution:

The is.prime() function is created and the results are tested,

The loop for the loop is replaced with the vector and the mod operator is applied in it.

```
}else{
  sq.x<-as.integer(sqrt(x))
  p.check<-c(3:max(sq.x,3))
  p.check<-x%%p.check
  if(is.element(0,p.check)){
    return (FALSE)
  } else{
    return (TRUE)
  }
}
```

The result is shown here for some selected values

```
> is.prime(1)
1 is neither prime nor composite!
> is.prime(2)
[1] TRUE
> is.prime(3)
[1] TRUE
> is.prime(4)
[1] FALSE
> is.prime(5)
[1] TRUE
> is.prime(13)
[1] TRUE
> is.prime(15)
[1] FALSE
> is.prime(30)
[1] FALSE
> is.prime(71)
[1] TRUE
> is.prime(75)
[1] FALSE
>
```

b) Using the above function, write a for loop over the sequence 2:100 to print the prime numbers in that range.

Solution:

The values are passed to the function above and the prime values alone are being recorded and stored in a vector which is finally displayed.

```
> #Problem--3(b)
> i=2
> x<-2
> for(i in 2:100){
+   res<-is.prime(i)
+   if(res==TRUE && i!=2){
+     x<-c(x,i)
+   }
+ }
> x
 [1]  2  3  5  7 11 13 17 19 23 29 31 37 41 43 47 53 59 61 67 71 73 79 83 89 97
>
```