**STM4PSD ASSIGNMENT 1**

In submitting my work, you are consenting that it may be copied and transmitted by the University for the detection of plagiarism. Submission is your guarantee that the below statement of originality is correct.

*“This is my own work. I have not copied any of it from anyone else.”*

NAME: **Maninderpreet Singh Puri**

STUDENT NUMBER: **20494381**

* 1. According to the given information officer will correctly detain someone driving under the influence 75% of the time, and correctly release a driver who is not intoxicated 75%. So probability that the officer will be making a right decision is P (A) = 0.75.

So say probability that a person is not drunk and the officer makes a wrong decision and detains the driver is P ()

P (A) +P () = 1

P () = 1-0.75= 0.25

* 1. For finding the probability of any given driver will be detained we can use the law of total probability that is,

P (A) = P (A|B) . P (B) + P (A|) . P ()

According to the given information we can assume P (A|B) = 0.75.

P (A|) = 1- P (A|B) = 1- 0.75= 0.25

Probability of drivers intoxicated in all the population is P (B) = 0.14

P () = 1- P (B) = 1 – 0.14 =0.86

So,

= 0.14 \* 0.75 + 0.86 \* 0.25=

0.32

c. Using Bayes Theorem:

P (A|B) =

P (B|A) = 0.14

P (A) = 0.75

P (B) = 0.32

= = 0.328

1. P () = 0.25

P (B|A) = 0.14

=0.14 \* 0.25=0.035

2.

a.

As we know that, Support

Where M is the total number of transactions, and) is A and B occurring together in a sample space. So in the question we have to find the answer using the above formula.

Support

That is = as F and I occur 4 times together out of 6 total transactions. So Support =0.66

Now for the calculating the Confidence we use the formula,

Confidence

So using the above formula we can say that,

Confidence

Support (F) = = 0.83

Confidence = = 0.79

By using the above formula for Support we can say that,

Support

That is = as T and I occur two times together out of six total transactions. So Support =0.33

Now for the calculating the Confidence we can say that,

Confidence

Support (T) = = 0.50

Confidence = = 0.66



So calculating the lift for and

Lift =

= = 0.97

Lift () =

= = 0.8

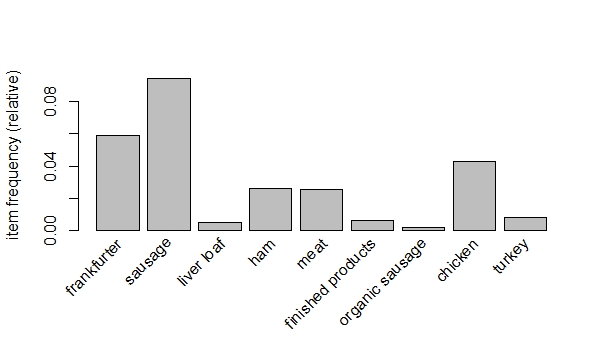
So here rule has a higher lift value and is close to one so it is better association rule for cross-sales.

a.

Input in R:

itemFrequencyPlot(Groceries[topN=9])

Output:

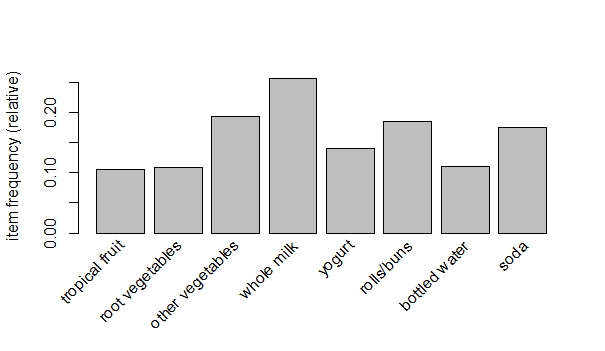


b.

Input in R:

itemFrequencyPlot(Groceries, support= 0.1)

Output:



c.

Input in R:

data(Groceries)

second.rules <- apriori(Groceries,

parameter = list(support = 0.025, confidence = 0.025))

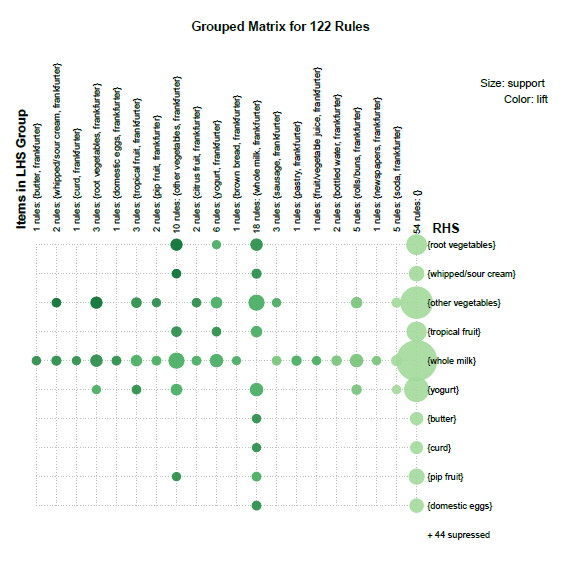
pdf(file="fig\_market\_basket\_rules\_matrix.pdf", width = 8.5, height = 8.5)

plot(second.rules, method="grouped",

control=list(col = rev(brewer.pal(9, "Greens")[4:9])))

dev.off()

Output:



The antecedent {tropical fruit, frankfurter} represents three association rules with {other vegetables}, {Whole milk}, {yogurt}.

The color of the bubble represents the lift of the association rule and size of the bubble represents the support of the association rule.

The lift of the association rule {tropical fruit, frankfurter} with {other vegetables}, {yogurt} is almost same and with {Whole milk} it is less.

The support of the association rule {tropical fruit, frankfurter} with {other vegetables} and {Whole milk} is almost same and with {yogurt} it is less.

2. Given F(T) =

Now derivative of the F (T) can be written as F (T)

So F (T) =

Or

=

= 0- ()

=

So turns out to be F (T) =

1. As the messages sent are measured in bytes and mean should also be described as Bytes.
2. E(T)=

As given in the question

So, E (T) = =

So according to the Question, a message will have size less than or equal to E (T), when α = 1.5.

So we can say that Probability of message having size less than or equal to E (T) which is lies in the range,

P (0t)

Or

P (0t)

As (1 < α < 2), so we can integrate f (t)

=

Using the R code calculating the value of f (t):

fn <- function(t) {

1.5 \* (1+t) ^ (-2.5)

}

c <- integrate (fn, lower=0, upper=2)$value

Output we get for ‘c’ is



2. R code:

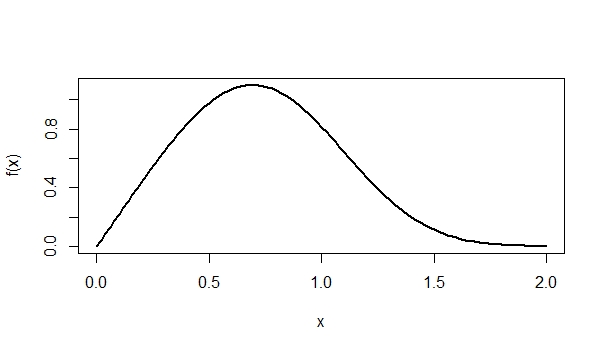
f <-function (x){

(3/gamma(2/3)\*x\*exp(-x^3))

}

curve(f, from = 0, to = 2, xlab = "x", ylab = "f(x)", lwd = 2)

Output plot:

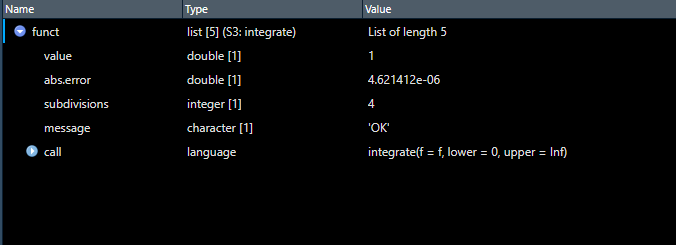


1. R code input:

funct <- integrate(f, lower=0, upper= Inf)

funct$value

Output:



1. R code input:

functm <- function (x) {

x\* (3/gamma(2/3)\*x\*exp(-x^3))

}

mean <- integrate(functm, lower=0, upper= Inf)

mean $ value

We get probability= 0.7384881

1. E(x)=

=

=

=

Let t = , dt= 3 dx

= = [-]

= [-]= []

= = 0.7384

1. Using R code:

demandexceed <- integrate(fm, lower=1.2, upper= Inf)

demandexceed $ value

We get probability= 0.09605142

1. Using R code:

demandmeet <- integrate(fm, lower=0.064, upper= 1.2)

demandmeet $ value

We get probability= 0.8994118

2. So from the given table we can say that P(X=1000 and Y=500) = 0.075
3. P(X=750|Y=500)=

So = = 0.475

1. (750, 250)= F(X)

=P (500,200) +P (750,250)

=0.25 + 0.175=0.425

1. So the marginal probability of Y is the sum of the values when y=250 and y=500

(x)=

1. T=X+Y

So ( t)=

1. E(T)= 750 (0.25) +1000 (0.375) + 1250 (0.3) +1500 (0.075)= 1050