Look Alike Analysis

Nane Mambreyan

Future Innovation

GPS is great, unless you're in a blind spot, inside a building, or under water. Professor Hiroyuki Tanaka at the University of Tokyo has developed a system that's more reliable and precise. It uses muons, particles that are a natural form of radiation, which constantly bombard the Earth's surface and are absorbed by the land or water they fall on. Tanaka's muon positioning system tracks the level of radiation that reaches a receiver and computes how much has been absorbed along the way. This allows systems to map walls and floors or detect the presence of people. "The current positioning accuracy is 3.5 centimeters indoors," says Tanaka. In March, the tech was used to uncover a hidden room within a 4,500-year-old Egyptian pyramid.

Existing Innovation

Existing innovations from the past that have parallels with muons positioning system are positioning technologies like GNSS(Global navigation satellite system), Indoor Positioning and Navigation systems and surveying technologies.

While the specific methods and technologies may differ, the mentioned existing innovations share common goals and functionalities with Professor Tanaka's muon positioning system, particularly in providing accurate and reliable location information across various environments.

Data description and manipulation

This means that to conduct look alike analysis for the new positioning system with Bass model, we can use the data of existing ones and make some predictions. For that, I am using 'GNSS & positioning market revenue by technology 2013-2020' dataset from Statista. This dataset gives information on the revenue generated from different positioning systems from 2013 to 2020. Since for Bass model, we need units of adopted product rather than the generated revenue, I did some research to find the unit price for adoption of each technology(demonstrated below). Dividing the overall generated revenue by the unit price, we get the volume of the adopted product/technology.

- GNSS receivers can range anywhere from $500 to $20,000 or more.
- An indoor navigation app could cost between $8000 and $30000.
- The price of survey instruments can vary widely — from as little as $200 to as much as $1,400.

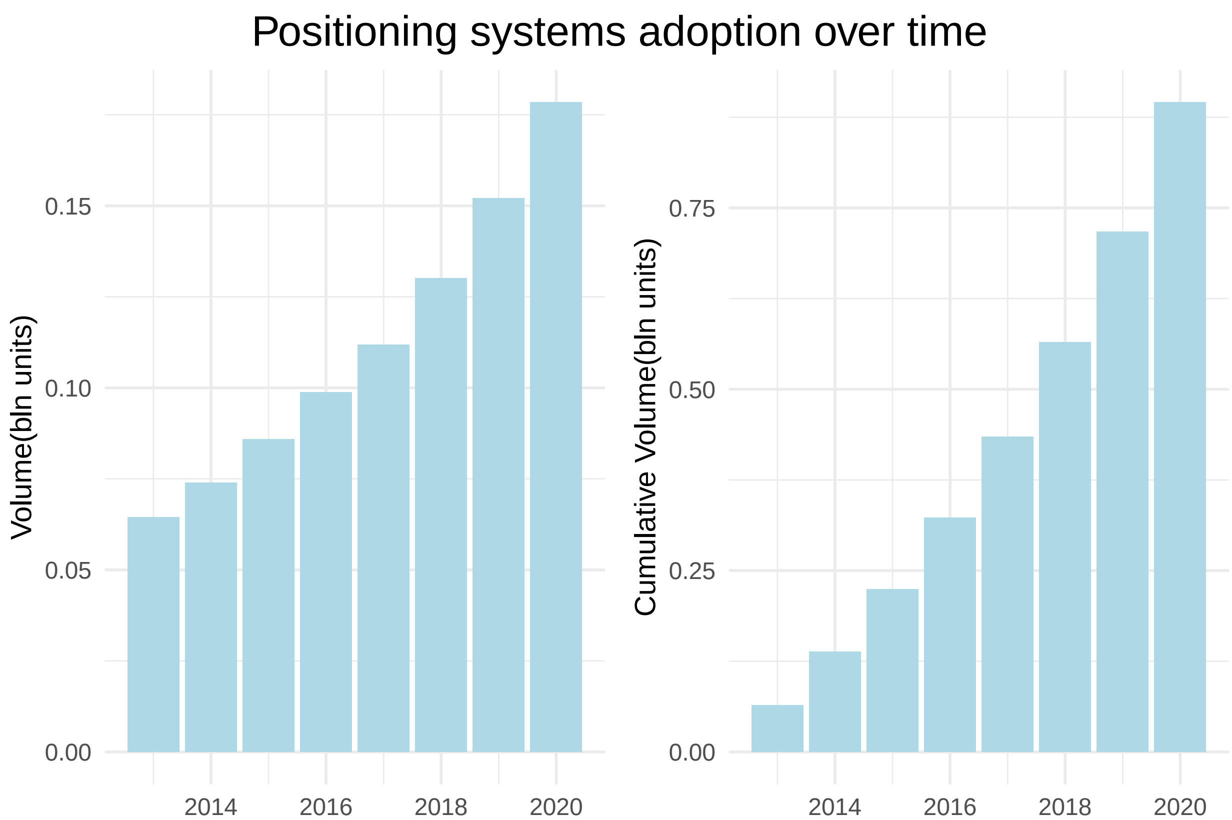
```
gnss_unit_pirce <- mean(c(500, 2000)) # taking the mean unit price for each
indoor_mapping_tech_unit_price <- mean(c(8000, 30000))
surveying_equipment_unit_price <- mean(c(200, 1400))


gnss <- read_excel('gnss.xlsx', sheet='Data') %>%
  slice(-c(1:2)) %>% # remove unnecessary rows
  rename("Year" =`GNSS & positioning market revenue by technology 2013-2020`,
```

```r
        "GNSS" = `...2`, "Indoor" = `...3` , "Surveying" = `...4`) %>%
  mutate(across(everything(), ~gsub("\\*$", "", .))) %>% # remove * to convert into int
  mutate(Year = as.integer(Year), # type conversions
         GNSS = as.double(GNSS) ,
         Indoor = as.double(Indoor) ,
         Surveying = as.double(Surveying)) %>%
  mutate(GNSS = GNSS / gnss_unit_pirce, # getting quantity for each
         Indoor = Indoor / indoor_mapping_tech_unit_price,
         Surveying = Surveying / surveying_equipment_unit_price) %>%
  group_by(Year) %>%
  mutate(Volume = sum(GNSS, Indoor, Surveying))%>%
  ungroup() %>%
  mutate(Cumulative_Volume = cumsum(Volume))
  # grouping by the year and calculating the total quantity over all segments
  # then grouping to calculate the cumulative quantity

head(gnss, 3)
```

```
## # A tibble: 3 x 6
##    Year   GNSS  Indoor Surveying Volume Cumulative_Volume
##   <int>  <dbl>   <dbl>     <dbl>  <dbl>             <dbl>
## 1  2013 0.0618 0.0019    0.00075 0.0644            0.0644
## 2  2014 0.0707 0.00199   0.00125 0.0740            0.138
## 3  2015 0.0818 0.00209   0.002   0.0858            0.224
```

I begin with plotting the histogram of the products' volume over time.

```r
g0 <- ggplot(data = gnss, aes(x = Year, y = Volume)) +
  geom_bar(stat = 'identity', fill = 'lightblue') +
  labs(y = 'Volume(bln units)', x = '' ) +
  theme(panel.background = element_blank()) + theme_minimal()



g1 <- ggplot(data = gnss, aes(x = Year, y = Cumulative_Volume)) +
  geom_bar(stat = 'identity', fill = 'lightblue') +
  labs(y = 'Cumulative Volume(bln units)' , x = '') +
  theme(panel.background = element_blank()) + theme_minimal()



grid.arrange(g0, g1, ncol = 2,
             top = textGrob('Positioning systems adoption over time',
                            gp = gpar(fontsize = 16)))
```



The main objective of look alike analysis with Bass model will be to find such innovation and imitation rates (p, q) using which we can get a similar trend of the new product diffusion, i.e. the plot of a(t) = m * f(t) will be a good fit for the Volume histogram and A(t) = m * F(t) will be a good fit for the Cumulative Volume histogram where

- A(t) - number of people who have adopted the technology up to time t(inclusive).
- a(t) - number of people who adopt the technology at time t
- F(t) - fraction of people who have adopted the technology up to time t(inclusive).
- f(t) - fraction of people who adopt the technology at time t
- m - potential market size

This means that we deal with probability density function(f(t)) and probability distribution function(F(t)).

Satisfying the above mentioned criteria, we can make predictions on the future similar product diffusion.

Defining Bass diffusion model and getting parameters

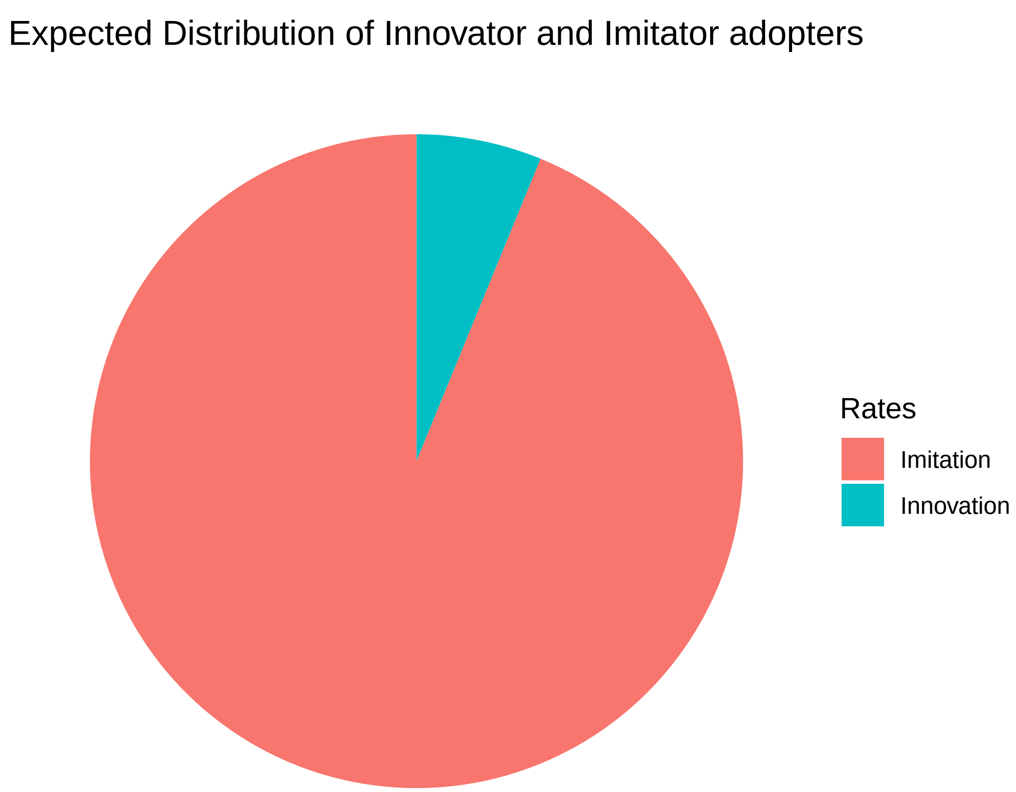
```
diff_m = diffusion(gnss$Volume) # fit the model
p=round(diff_m$w,4)[1] # get each parameter from w={p,q,m} parameter set
q=round(diff_m$w,4)[2]
m=round(diff_m$w,4)[3]
diff_m
```

```
## bass model
```

```
##
## Parameters:
##                                Estimate p-value
## p - Coefficient of innovation   0.0122      NA
## q - Coefficient of imitation    0.1844      NA
## m - Market potential            4.6531      NA
##
## sigma: 0.0038
```

We might define some KPI related to the rate of innovation and imitation in the future, hence we can start our predictions by visualizing those.

```r
pie_data <- data.frame(Rates = c("Innovation", "Imitation"),
                       Value = c(p, q))

ggplot(pie_data, aes(x = "", y = Value, fill = Rates)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +
  labs(title = "Expected Distribution of Innovator and Imitator adopters") +
  theme_void()
```

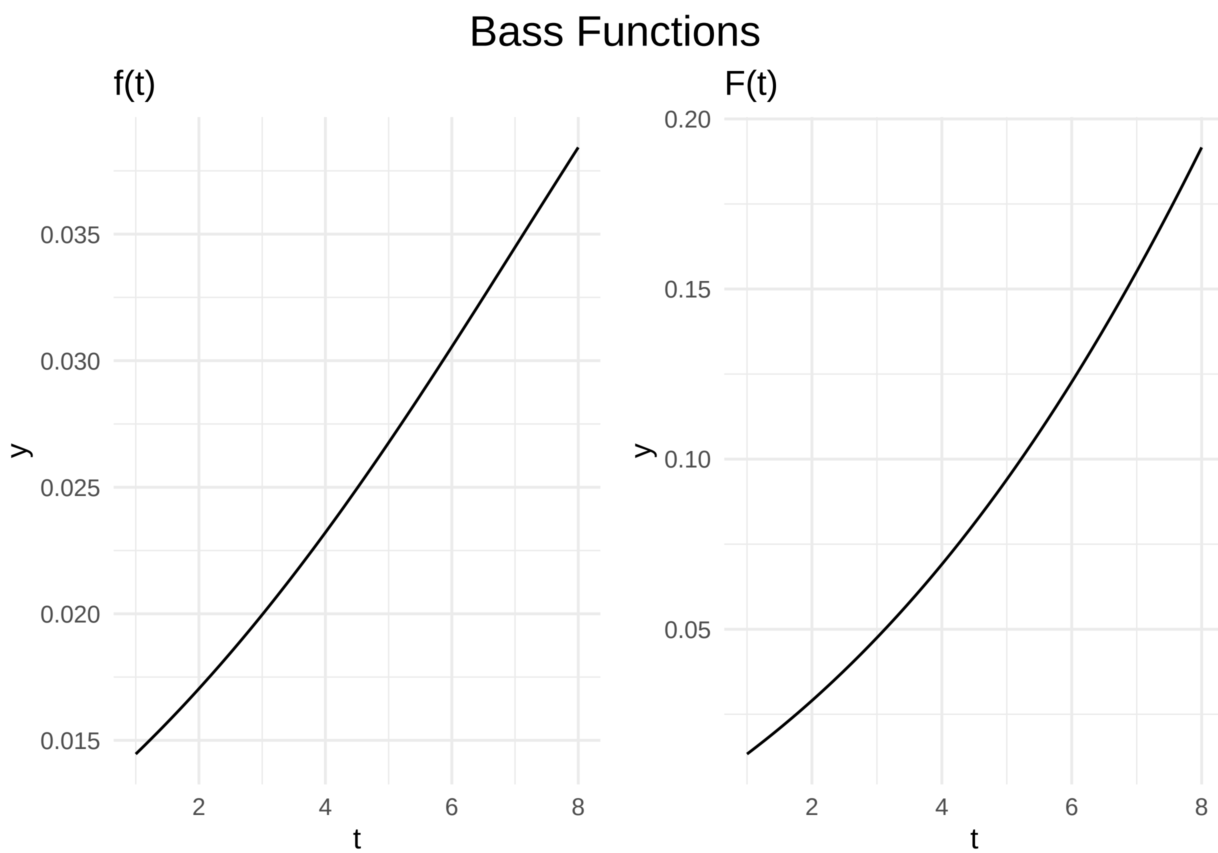


We see that the vast majority of adopters will be imitators.

Defining Bass functions

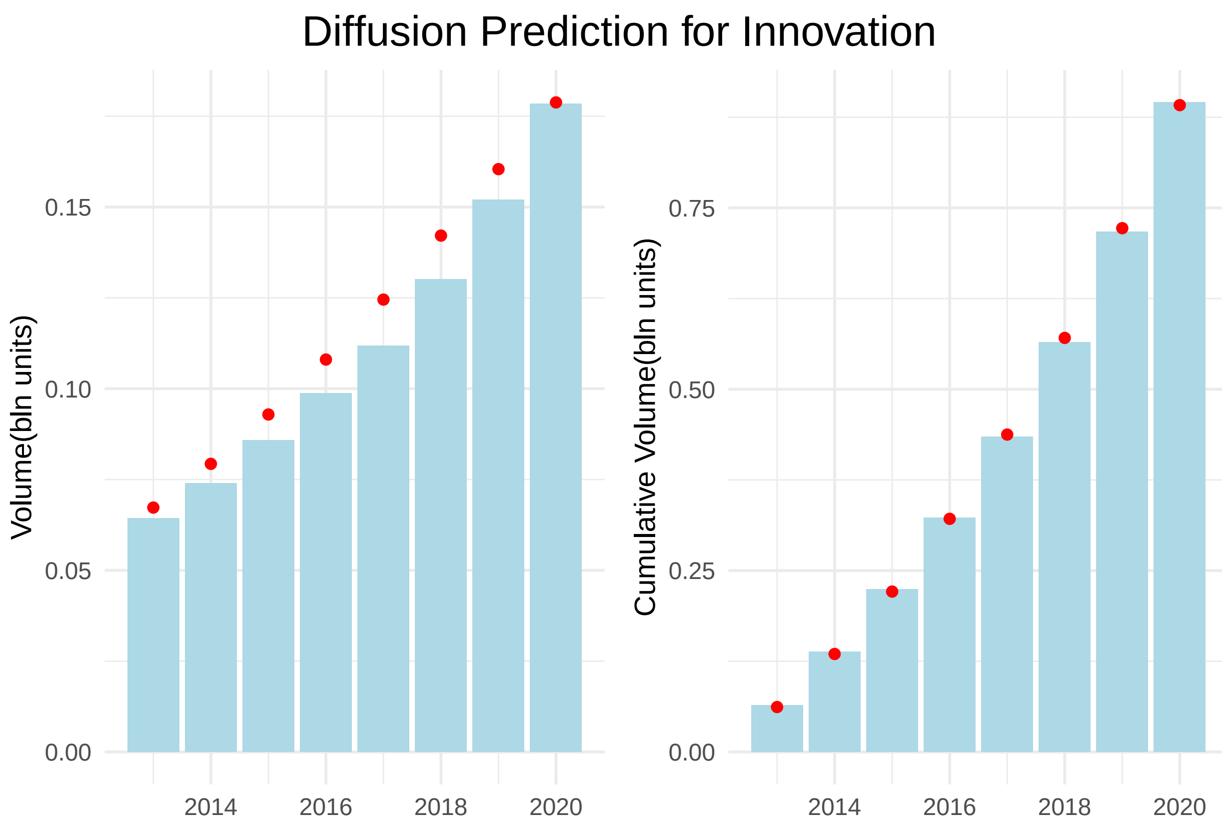
```
bass.f <- function(t,p,q){
((p+q)^2/p)*exp(-(p+q)*t)/
(1+(q/p)*exp(-(p+q)*t))^2
}


bass.F <- function(t,p,q){
(1-exp(-(p+q)*t))/
(1+(q/p)*exp(-(p+q)*t))
}
```

```
g2 <- ggplot(data.frame(t = c(1:length(gnss$Year))), aes(t)) +
  stat_function(fun = bass.f, args = c(p, q)) +
  labs(title = 'f(t)') +
  theme(panel.background = element_blank()) + theme_minimal()


g3 <- ggplot(data.frame(t = c(1:length(gnss$Year))), aes(t)) +
  stat_function(fun = bass.F, args = c(p, q)) +
  labs(title = 'F(t)') +
  theme(panel.background = element_blank()) + theme_minimal()



grid.arrange(g2,g3, ncol = 2,
             top = textGrob('Bass Functions',
                            gp = gpar(fontsize = 16)))
```



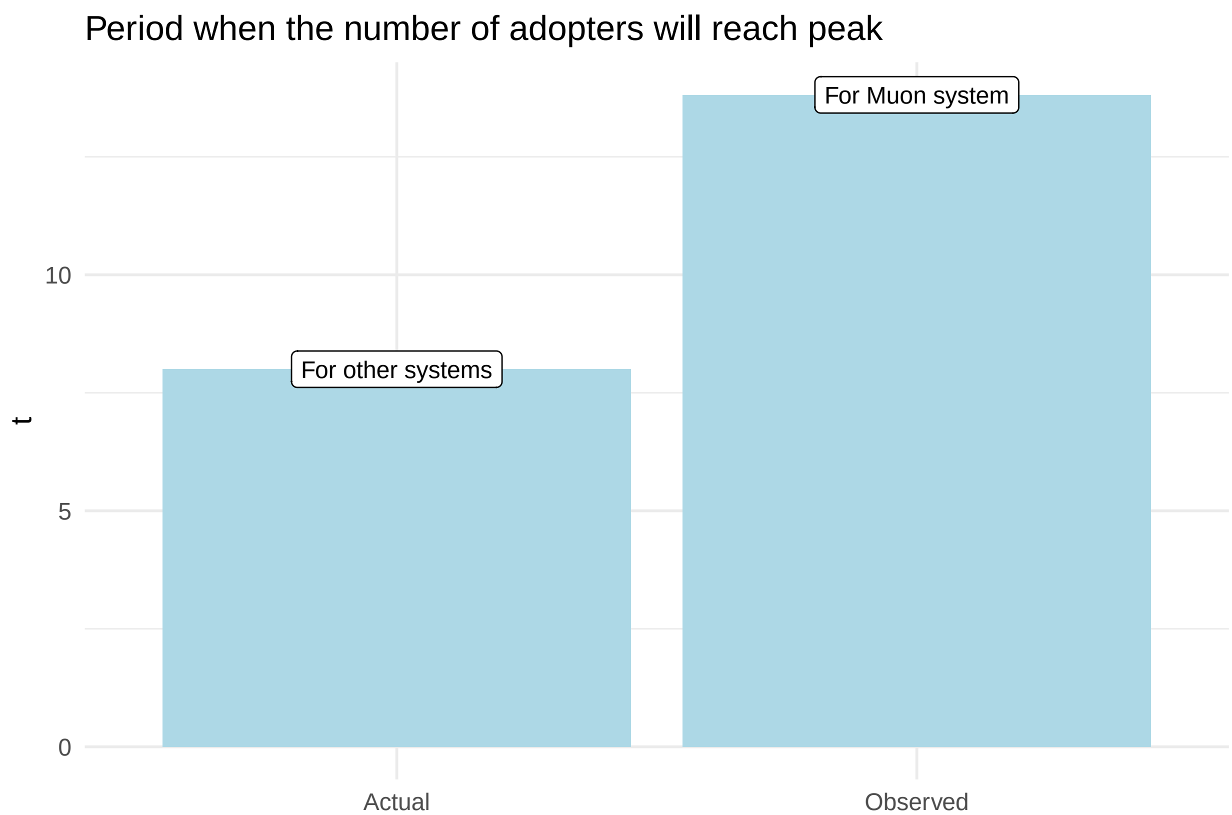
Since a(t) is f(t) but scaled with some constant m, it(f(t)) should demonstrate similar trend as the Volume histogram. Same holds for A(t). So to compare the histograms and probability functions, we need to multiple probability functions with m to get the same scale.

```
gnss$a_t = bass.f(1:length(gnss$Year), p , q )*m
gnss$A_t = bass.F(1:length(gnss$Year), p , q )*m

g4 <- ggplot(data = gnss, aes(x = Year, y = Volume)) +
geom_bar(stat = 'identity', fill = 'lightblue') +
geom_point(mapping = aes(x=Year, y=a_t), color = 'red') +
labs(y = 'Volume(bln units)', x = '') +
theme(panel.background = element_blank(),
      axis.text.y = element_text(size = 8)) + theme_minimal()

g5 <- ggplot(data = gnss, aes(x = Year, y = Cumulative_Volume)) +
geom_bar(stat = 'identity', fill = 'lightblue') +
geom_point(mapping = aes(x=Year, y=A_t), color = 'red') +
labs(y='Cumulative Volume(bln units)', x='') +
theme(panel.background = element_blank(),
      axis.text.y = element_text(size = 8)) + theme_minimal()

grid.arrange(g4, g5, ncol = 2,
             top = textGrob('Diffusion Prediction for Innovation',
                            gp = gpar(fontsize = 16)))
```



The red dots represent our anticipation of how the diffusion of Muon positioning system would have been within given 8 years (t = 8 time units) if it had been created in 2013, so we can assume similar trend for the future. We can further extend our predictions by finding the period when (in which unit time t) the number of adopters will reach the peak, based on the other available positioning systems.

```
peak_df <- data.frame(Type = c("Observed", "Actual"),
                      t = c(log(q/p)/(p+q), which.max(gnss$Volume)),
                      Label = c("For Muon system", "For other systems"))

ggplot(data = peak_df, aes(x = Type, y = t)) +
  geom_bar(stat = "identity",  fill = 'lightblue') +
  geom_label(aes(label = Label), size = 3, fill = "white") +
  labs(title = "Period when the number of adopters will reach peak",
       x = '') +
  theme_minimal()
```



We see that in case of Muon we will observe the peak later over time compared to other systems.