DS 223 Marketing Analytics

Bass Model for Innovation Diffusion

Karen Hovhannisyan

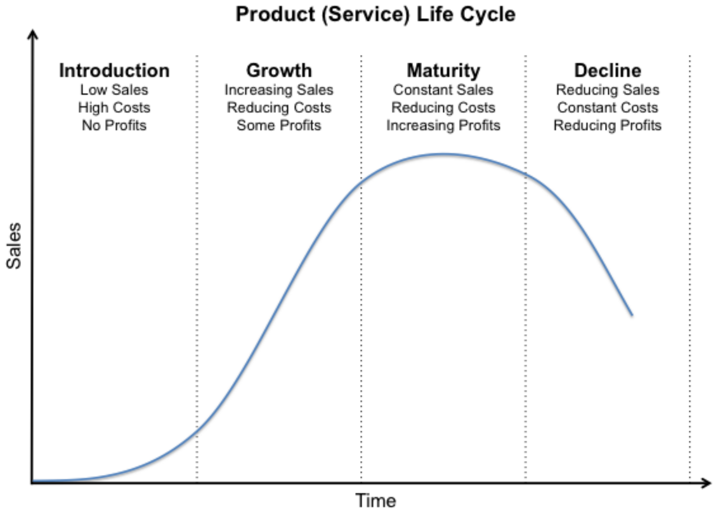
AUA

Libraries

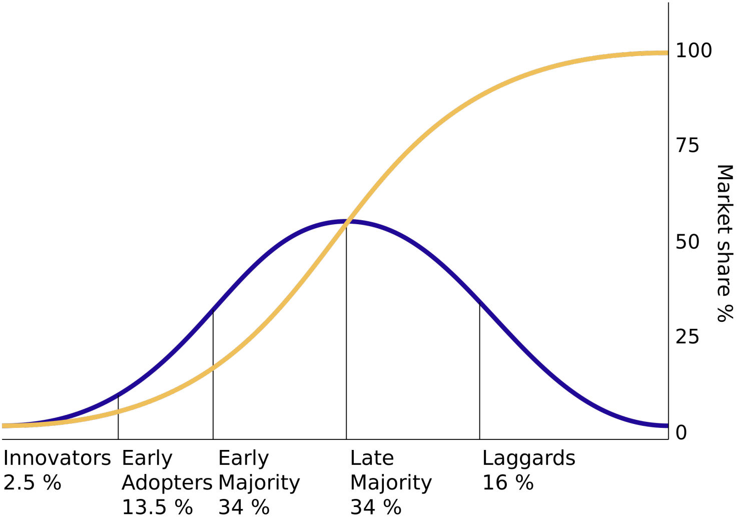
```r
libs<-c('ggplot2','ggpubr','knitr','diffusion')

load_libraries<-function(libs){
  new_libs <- libs[!(libs %in% installed.packages()[,"Package"])]
  if(length(new_libs)>0) {install.packages(new_libs)}
  lapply(libs, library, character.only = TRUE)
}
load_libraries(libs)
```

Diffusion of Innovation Theory

Product Lifecycle



Diffusion of Innovation



Origin

Diffusion of Innovation Theory is developed by E.M. Rogers in 1962. It is one of the oldest social science theories. It originated in communication to explain how, over time, an **idea** or **product** gains momentum and diffuses (or spreads) through a specific population or social system.

The end result of this diffusion is that people, as part of a social system, adopt a new idea, behavior, or product.

Key Elements of Diffusion Research

- Innovators
- Early Adopters
- Early Majority
- Late Majority
- Laggards

Key Elements of Diffusion Research

- Innovators

This group of people who want to be the first to try the innovation. They are willing to take risks, and are often the first to develop new ideas. Very little, if anything, needs to be done to appeal to this population.

- Early Adopters
- Early Majority
- Late Majority
- Laggards

Key Elements of Diffusion Research

- Innovators
- Early Adopters

This group represents opinion leaders. They enjoy leadership roles, and embrace change opportunities. They are already aware of the need to change and so are very comfortable adopting new ideas. Strategies to appeal to this population include how-to manuals and information sheets on implementation. They do not need information to convince them to change.

- Early Majority
- Late Majority
- Laggards

Key Elements of Diffusion Research

- Innovators
- Early Adopters
- Early Majority

These people are rarely leaders, but they do adopt new ideas before the average person. That said, they typically need to see evidence that the innovation works before they are willing to adopt it. Strategies to appeal to this population include success stories and evidence of the innovation's effectiveness.

- Late Majority
- Laggards

Key Elements of Diffusion Research

- Innovators
- Early Adopters
- Early Majority
- Late Majority

These people are skeptical of change, and will only adopt an innovation after it has been tried by the majority. Strategies to appeal to this population include information on how many other people have tried the innovation and have adopted it successfully.

- Laggards

Key Elements of Diffusion Research

- Innovators
- Early Adopters
- Early Majority
- Late Majority
- Laggards

These people are bound by tradition and very conservative. They are very skeptical of change and are the hardest group to bring on board. Strategies to appeal to this population include statistics, fear appeals, and pressure from people in the other adopter groups.

Bass Model

Idea

The basic idea is that there are two types of customers:

Idea

The basic idea is that there are two types of customers:

- **innovators:** they buy the product using the information about the product, advertisement, etc

Idea

The basic idea is that there are two types of customers:

- **innovators:** they buy the product using the information about the product, advertisement, etc
- **imitators:** their decision is based on experience of other people, their ratings, etc

The Bass Model Equation

$$\frac{f(t)}{1 - F(t)} = p + \frac{q}{M}[A(t)]$$

where:

- $\frac{f(t)}{1-F(t)}$ is called **hazard rate** and shows the probability to purchase the product at time t given that they have not purchased yet
- **p:** innovation rate or coefficient of innovation
- **q:** imitation rate or coefficient of imitation

The parameters of Bass Model

- M: *market Potential*
- t time interval, usually a year
- $f(t)$ the fraction of the total market that adopts at time t
- $F(t)$ the fraction of the total market that has adopted up to and including time t

what will be the relationship between F(t) and f(t)?

The parameters of Bass Model

- M: *market Potential*
- t time interval, usually a year
- $f(t)$ the fraction of the total market that adopts at time t
- $F(t)$ the fraction of the total market that has adopted up to and including time t

The parameters of Bass Model

- *M: market Potential*
- t time interval, usually a year
- $f(t)$ the fraction of the total market that adopts at time t
- $F(t)$ the fraction of the total market that has adopted up to and including time t

- $F(t) = \int_0^T f(t)dt \ \rightarrow \ f(t) = \frac{dF(T)}{dt}$

The parameters of Bass Model

- M: *market Potential*
- t time interval, usually a year
- $f(t)$ the fraction of the total market that adopts at time t
- $F(t)$ the fraction of the total market that has adopted up to and including time t

- $F(t) = \int_0^T f(t)dt \ \rightarrow \ f(t) = \frac{dF(T)}{dt}$
- The number of adopters at time t: $a(t) = M \times f(t)$

The parameters of Bass Model

- M: *market Potential*
- t time interval, usually a year
- $f(t)$ the fraction of the total market that adopts at time t
- $F(t)$ the fraction of the total market that has adopted up to and including time t

- $F(t) = \int_0^T f(t)dt \ \rightarrow \ f(t) = \frac{dF(T)}{dt}$
- The number of adopters at time t: $a(t) = M \times f(t)$
- The cumulative number of previous adopters at the time t: $A(t) = M \times F(t), t > 0$

The parameters of Bass Model

Note, that below two equations are **identical:**

$$\frac{f(t)}{1-F(t)} = p + \frac{q}{M}[A(t)] \iff \frac{f(t)}{1-F(t)} = p + q \times F(t)$$

Formulas of *f(t)* and *F(t)*

Let's do some math:

As $f(t) = a(t)/M$ and $F(t) = A(t)/M$, the equation $\frac{f(t)}{1-F(t)} = p + q \times F(t)$ can be rewritten as:

$$\frac{a(t)/M}{1 - A(t)/M} = p + q \times \frac{A(t)}{M}$$

Solving this for $a(t)$

$$a(t) = pm + (q - p) \times A(t) - \frac{q}{M} \times A^2(t)$$

Formulas of *f(t)* and *F(t)*

$$a(t) = pm + (q - p) \times A(t) - \frac{q}{M} \times A^2(t)$$

By doing the following replacements:

- $pm = \beta_0$,
- $(q - p) = \beta_1$,
- $-\frac{q}{M} = \beta_2$

we get:

$$a(t) = \beta_0 + \beta_1 A(t) + \beta_2 A^2(t)$$

Formulas of *f(t)* and *F(t)*

On the other hand

$\frac{F'(t)}{1-F(t)} = p + q \times F(t)$, thus $\frac{dF}{(1-F)(p+qF)} = dt$, from here

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}}$$

and since $f(t) = F'(t)$,

$$f(t) = \frac{(p+q)^2 e^{-(p+q)t}}{p[1 + \frac{q}{p} e^{-(p+q)t}]^2}$$

Estimates of Number of Products

Estimates of p, q, m by Frank M. Bass

Product	Period covered	a (10^3)	b	c (10^{-7})	R^2	a/σ_a	b/σ_b	c/σ_c	m (10^3)	p	q
Electric refrigerators	1920–1940	104.67	.21305	−.053913	.903	1.164	6.142	−2.548	40,001	.0026167	.21566
Home freezers	1946–1961	308.12	.15298	−.077868	.742	4.195	4.769	−3.619	21,973	.018119	.17110
Black and white television	1946–1961	2,696.2	.22317	−.025957	.576	3.312	3.724	−3.167	96,717	.027877	.25105
Water softeners	1949–1961	.10256	.27925	−512.59	.919	3.593	8.089	−6.451	5,793	.017703	.29695
Room air conditioners	1946–1961	175.69	.40820	−.24777	.911	1.915	8.317	−6.034	16,895	.010399	.41861
Clothes dryers	1948–1961	259.67	.33968	−.23647	.896	2.941	7.427	−5.701	15,092	.017206	.35688
Power lawnmowers	1948–1961	410.98	.32871	−.075506	.932	1.935	7.408	−4.740	44,751	.0091837	.33790
Electric bed coverings	1949–1961	450.04	.23800	−.031842	.976	3.522	6.820	−1.826	76,589	.005876	.24387
Automatic coffee makers	1948–1961	1,008.2	.28435	−.051242	.883	3.109	6.186	−4.353	58,838	.017135	.30145
Steam irons	1949–1960	1,594.7	.29928	−.058875	.828	3.649	5.288	−4.318	55,696	.028632	.32791
Recover players	1952–1961	543.94	.62931	−.29817	.899	1.911	5.194	−3.718	21,937	.024796	.65410

Data Sources: *Economic Almanac, Statistical Abstracts of the U.S., Electrical Merchandising,* and *Electrical Merchandising Week.*

Defining Functions

Define functions for $f(t)$ and $F(t)$:

bass.f() for f(t)

$$f(t) = \frac{(p+q)^2 e^{-(p+q)t}}{p[1 + \frac{q}{p} e^{-(p+q)t}]^2}$$

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

bass.F() for F(t)

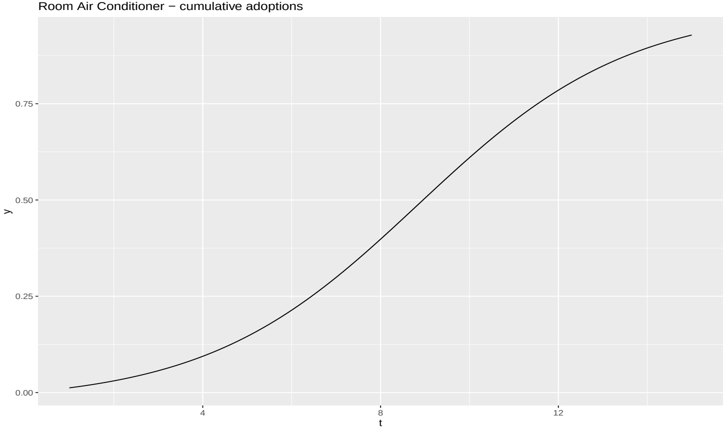
$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}}$$

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

Room Air Conditioner

Cummlative Adoptions

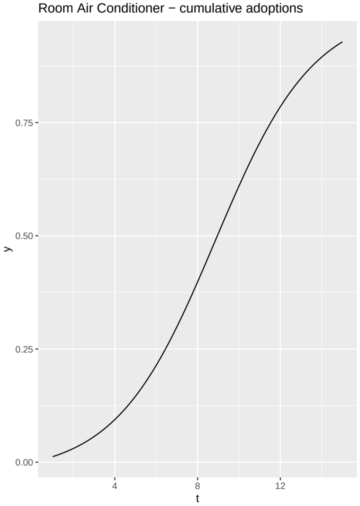
```
ggplot(data.frame(t = c(1, 15)), aes(t)) +
  stat_function(fun = bass.F, args = c(p=0.01, q=0.41)) +
  labs(title = 'Room Air Conditioner - cumulative adoptions')
```

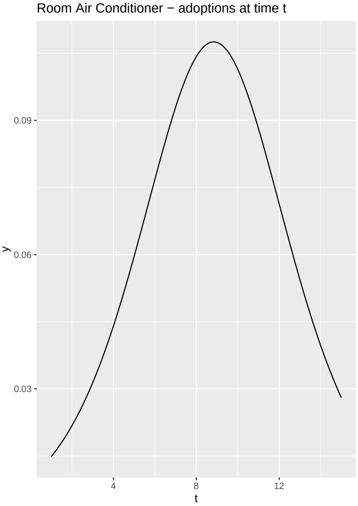


Addoptions and Cumulative Adoptions

```
cum_ad = ggplot(data.frame(t = c(1, 15)), aes(t)) +
  stat_function(fun = bass.F, args = c(p=0.01, q=0.41)) +
  labs(title = 'Room Air Conditioner - cumulative adoptions')

time_ad = ggplot(data.frame(t = c(1, 15)), aes(t)) +
  stat_function(fun = bass.f, args = c(p=0.01, q=0.41)) +
  labs(title = 'Room Air Conditioner - adoptions at time t')

ggarrange(cum_ad, time_ad)
```

Addoptions and Cumulative Adoptions





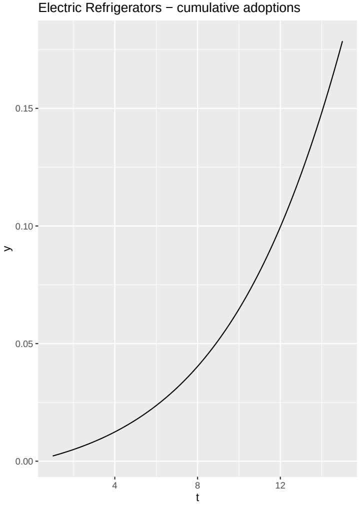
Electric Refrigerators

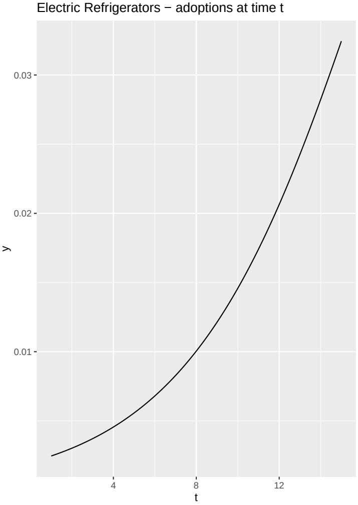
Addoptions and Cumulative Adoptions

```
cum_ad = ggplot(data.frame(t = c(1, 15)), aes(t)) +
  stat_function(fun = bass.F, args = c(p=0.002, q=0.21)) +
  labs(title = 'Electric Refrigerators  - cumulative adoptions')

time_ad =  ggplot(data.frame(t = c(1, 15)), aes(t)) +
  stat_function(fun = bass.f, args = c(p=0.002, q=0.21)) +
  labs(title = 'Electric Refrigerators  - adoptions at time t')

ggarrange(cum_ad, time_ad)
```

Addoptions and Cumulative Adoptions

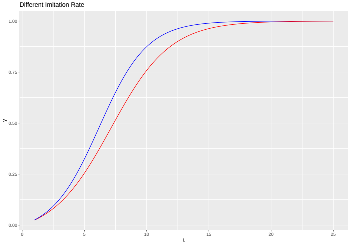
Try to change the horizon of t



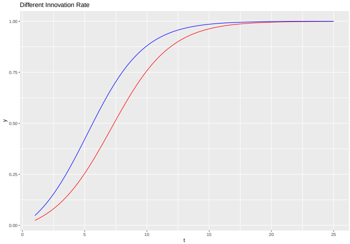


Different Innovation/Imitation Rates

```
ggplot(data.frame(t = c(1, 25)), aes(t)) +
  stat_function(fun=bass.F,args=c(p=0.02,q=0.4),color='red')+
  stat_function(fun=bass.F,args=c(p=0.02,q=0.5),color='blue')+
  labs(title = "Different Imitation Rate")
```




```
ggplot(data.frame(t = c(1, 25)), aes(t)) +
  stat_function(fun=bass.F,args=c(p=0.02,q=0.4),color='red')+
  stat_function(fun=bass.F,args=c(p=0.04,q=0.4),color='blue')+
  labs(title = "Different Innovation Rate")
```



Smartphone sales

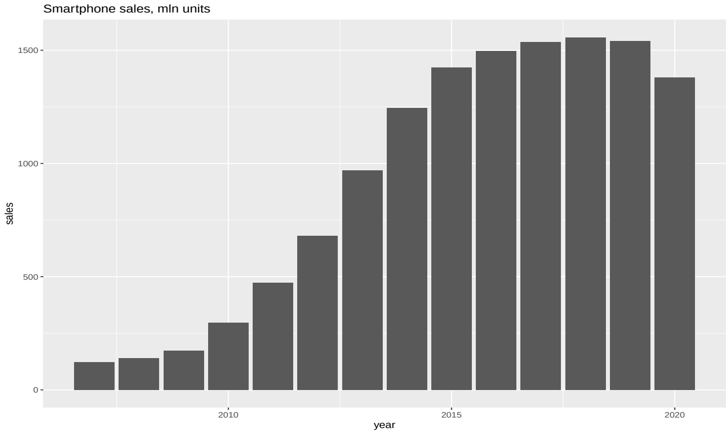
Smartphone sales data

```
smart = read.csv('Data/smartphones.csv', fileEncoding="UTF-8-BOM")
head(smart))
```

year	sales
2007	122.32
2008	139.29
2009	172.38
2010	296.65
2011	472.00
2012	680.11

Historical Sales

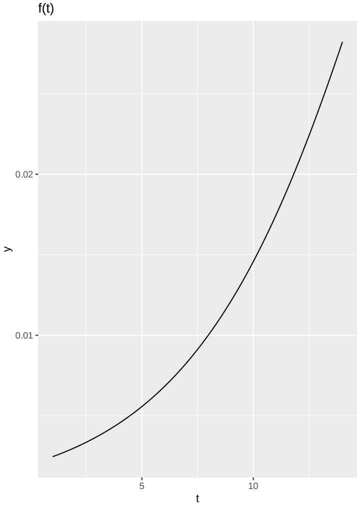
```
ggplot(data = smart, aes(x = year, y = sales)) +
  geom_bar(stat = 'identity') +
  ggtitle('Smartphone sales, mln units')
```

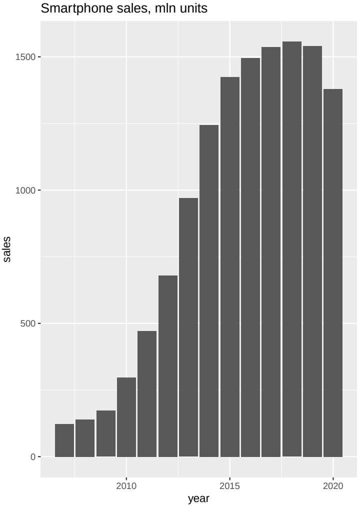


Modeling *f(t)*

```
time_ad =  ggplot(data.frame(t = c(1:14)), aes(t)) +
  stat_function(fun = bass.f, args = c(p=0.002, q=0.21)) +
  labs(title = 'f(t)')

sm_sales = ggplot(data = smart, aes(x = year, y = sales)) +
  geom_bar(stat = 'identity') +
  ggtitle('Smartphone sales, mln units')

ggarrange(time_ad, sm_sales)
```

Visualizing *f(t)*





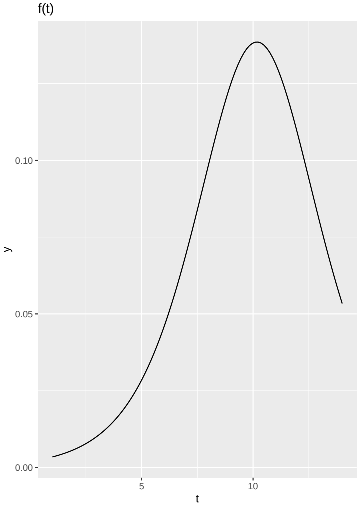
Changing Imitation Rate

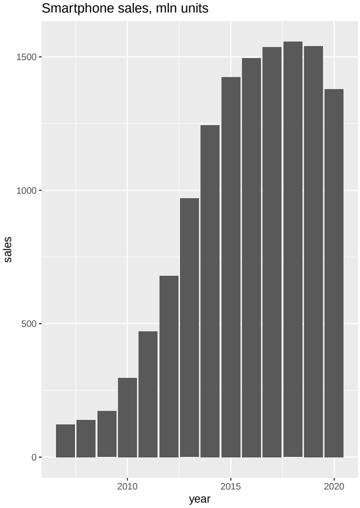
Let's play with q by **increasing** it, so the peak:

- $p = 0.002$
- $q = 0.55$

```
time_ad =  ggplot(data.frame(t = c(1:14)), aes(t)) +
  stat_function(fun = bass.f, args = c(p=0.002, q=0.55)) +
  labs(title = 'f(t)')

ggarrange(time_ad, sm_sales)
```

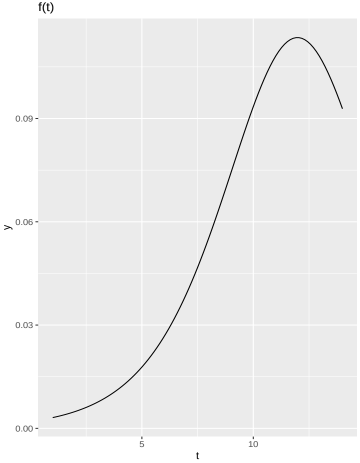
Changing Imitation Rate

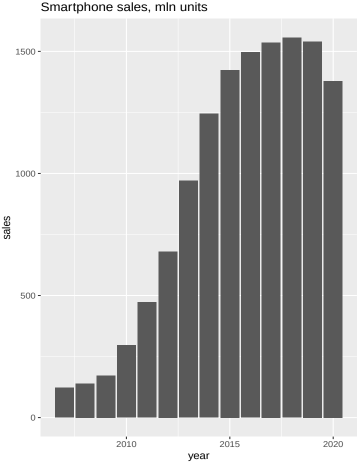




Changing Imitation Rate

Decrease q to 0.45





Parameter Estimation | Method 1

We can estimate m, p and q with different approaches.

Use **nls()** - Non-linear Least Squares

$$a(t) = m\frac{(p+q)^2 e^{-(p+q)t}}{p[1+\frac{q}{p}e^{-(p+q)t}]^2}$$

Now this is a number of adopters rather than the probability of adoptions (multiplied by m)

Define starting parameters:

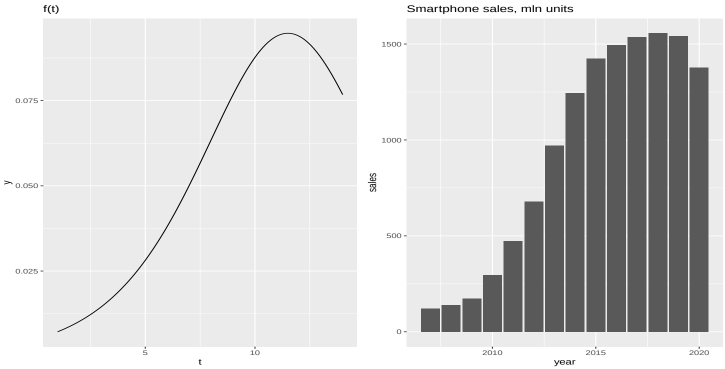
```
sales = smart$sales
t = 1:length(sales)
bass_m = nls(sales ~ m*(((p+q)^2/p)*exp(-(p+q)*t))/
               (1+(q/p)*exp(-(p+q)*t))^2,
              start=c(list(m=sum(sales),p=0.02,q=0.4)))
```

Model Summary | Method 1

```
summary(bass_m)
##
## Formula: sales ~ m * (((p + q)^2/p) * exp(-(p + q) * t))/(1 + (q/
##     exp(-(p + q) * t))^2
##
## Parameters:
##          Estimate    Std. Error t value        Pr(>|t|)
## m 17449.7605423   660.3010917  26.427 0.0000000000264 ***
## p     0.0053274     0.0006362   8.374 0.0000042206387 ***
## q     0.3659114     0.0179672  20.366 0.0000000004400 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 75.55 on 11 degrees of freedom
##
## Number of iterations to convergence: 11
## Achieved convergence tolerance: 0.000007999
```

With Estimated Parameters | Method 1

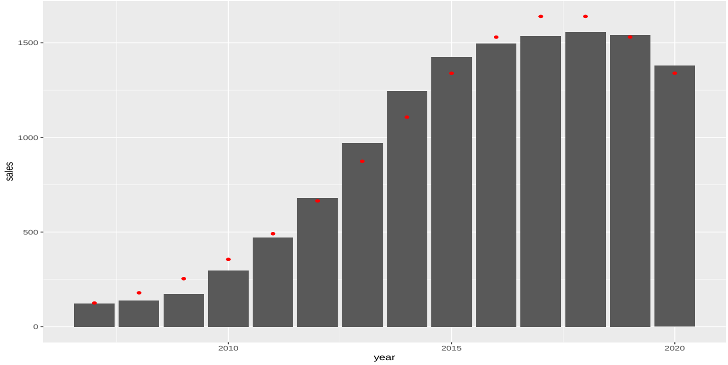
```
time_ad =  ggplot(data.frame(t = c(1:14)), aes(t)) +
  stat_function(fun = bass.f, args = c(p=0.005, q=0.369)) +
  labs(title = 'f(t)')

ggarrange(time_ad, sm_sales)
```



Predicting sales | Method 1

Multiplying $f(t)$ by m

```
smart$pred_sales = bass.f(1:14, p = 0.005, q = 0.369)*17450
ggplot(data = smart, aes(x = year, y = sales)) +
  geom_bar(stat = 'identity') +
  geom_point(mapping = aes(x=year, y=pred_sales), color = 'red')
```



Parameter Estimation | Method 2

Using *library diffusion*, where the p, q, m parameters are estimated as suggested by **Bass**

```
diff_m = diffusion(sales)
p=round(diff_m$w,4)[1]
q=round(diff_m$w,4)[2]
m=round(diff_m$w,4)[3]
diff_m
## bass model
##
## Parameters:
##                                    Estimate p-value
## p - Coefficient of innovation        0.0054      NA
## q - Coefficient of imitation         0.3990      NA
## m - Market potential             16317.8950      NA
##
## sigma: 77.594
```

Reaching to the peak

We can also find the period when the sales will reach to the peak.

Recall:

$$a(t) = m\frac{(p+q)^2 e^{-(p+q)t}}{p[1+\frac{q}{p}e^{-(p+q)t}]^2} \rightarrow t^* = \frac{ln(\frac{q}{p})}{p+q}$$

For the smartphone sales:

```
data.frame(Predicted=log(q/p)/(p+q),
                 Actual=which.max(smart$sales))
```

Predicted	Actual
11.50096	12

New Product Estimation

If you have a new product to market and want to forecast sales/adoptions:

- Find similar products (looks-like analysis)
- Estimate p and q
- Get your own estimate of m

Digital Cameras and Lenses

Digital Cameras and Lenses Data

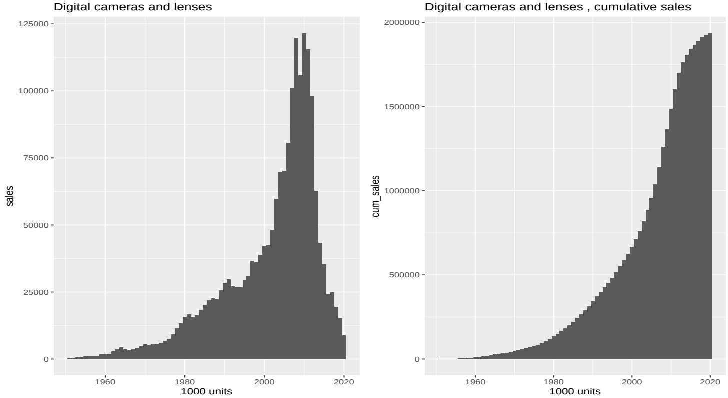
Sales of digital cameras and lenses, 1000 units
source

```
cameras = read.csv('Data/cameras and lenses.csv',
                   fileEncoding="UTF-8-BOM")
cameras$cum_sales = cumsum(cameras$sales)
kable(head(cameras))
```

year	sales	cum_sales
1951	258	258
1952	376	634
1953	586	1220
1954	787	2007
1955	949	2956
1956	1178	4134

Historical Sales

```
sales =  ggplot(cameras, aes(x = year, y = sales)) +
  geom_bar(stat='identity') +
  labs(title = 'Digital cameras and lenses', x = '1000 units')
cum_sales =   ggplot(cameras, aes(x = year, y = cum_sales)) +
  geom_bar(stat='identity') +
  labs(title = 'Digital cameras and lenses , cumulative sales', x = '1000 units')
ggarrange(sales, cum_sales)
```



Bass Model

```
diffusion(cameras$sales)
## bass model
##
## Parameters:
##                                    Estimate p-value
## p - Coefficient of innovation        0.0000      NA
## q - Coefficient of imitation         0.1332      NA
## m - Market potential           4596167.0408      NA
##
## sigma: 39001.1817
```

Bass Model

```
sales = cameras$sales[10:60]
diffusion(sales, verbose = T)$parameters
## Estimation iteration: 1
##   Estimate p-value
## p  20.9775      NA
## q   1.2405      NA
## m  63.0909      NA
## Estimation completed
## NULL
```