



King Mongkut's University of Technology Thonburi  
 Faculty of Engineering, Department of Computer Engineering  
 CPE 213 Data Model, 2/2020

LAB Lecture 7.5: Monte Carlo Simulation

Assign Date: 2 April 2021 Due Date: 8 April 2021

## Task 1: Pi Estimation

The goal of this problem is to illustrate how  $\pi = 3.1415\dots$  can be computed by random sampling of the unit disk. Starting from the following pseudocode, write R program to calculate Pi.

In your simulation run the code multiple times for  $N=10^i$ ,  $i = 1, 2, 3, \dots$  random numbers. See how the estimate for  $\pi$  improves with increasing  $N$  and compute the deviation from the exact result:  $\text{error} = |\pi - \pi_{\text{estimate}}|$ .

Perform a log-log plot of the error as a function of  $N$  and show that the data can be fit to a straight line of slope  $-1/2$ .

### Code

```
N <- 1000000
x <- runif(n=N, min=0,max=1)
y <- runif(n=N, min=0,max=1)
|
radia <- sqrt(x**2 + y**2)
inbound <- radia < 1

pi_estimate = 4*(1*1)*(sum(inbound)/N)
```

จาก Code ทำการ Random ตัวเลขแบบ Uniform และ ทำการเข้าสมการ  $\sqrt{x^2 + y^2}$  จะได้เป็นรัศมีวงกลม ทำการจุด ที่ค่ารัศมีที่น้อยกว่า 1 จะได้จุดที่ตกอยู่ในพื้นที่วงกลม ต่อจากนั้นทำการเทียบอัตราส่วนจากจุดทั้งหมด ว่าวงกลมมีพื้นที่เป็น อัตราส่วนเท่าใดในสี่เหลี่ยม ทำการคูณกับพื้นที่สี่เหลี่ยมจะได้พื้นที่วงกลมแต่สมการที่เราใช้เป็นแค่  $1/4$  ของ วงกลมดังนั้นต้องคูณ 4 เข้าไปจะได้ พื้นที่วงกลม ซึ่งเราใช้รัศมีเท่ากับ 1 ในการหาค่า Pi ดังนั้นเราสามารถให้พื้นที่ของวงกลมเท่ากับค่า Pi ได้เลยในการประมาณค่า ถ้าต้องการใช้ค่ามีค่าถูกต้องมากขึ้นเราต้องทำการเพิ่ม Trials sample (N) มากขึ้นจะได้ครอบคลุมพื้นที่ของ วงกลมมากขึ้น

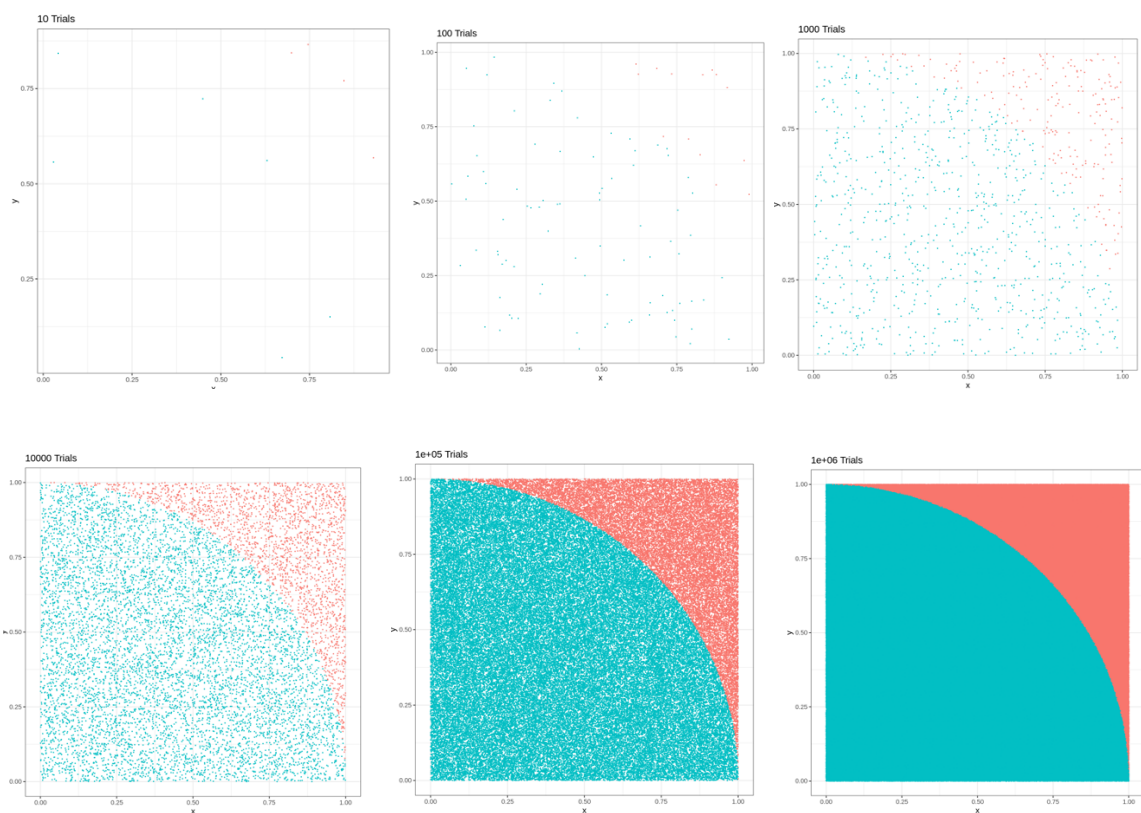


King Mongkut's University of Technology Thonburi  
Faculty of Engineering, Department of Computer Engineering  
CPE 213 Data Model, 2/2020

LAB Lecture 7.5: Monte Carlo Simulation

Assign Date: 2 April 2021 Due Date: 8 April 2021

## Result



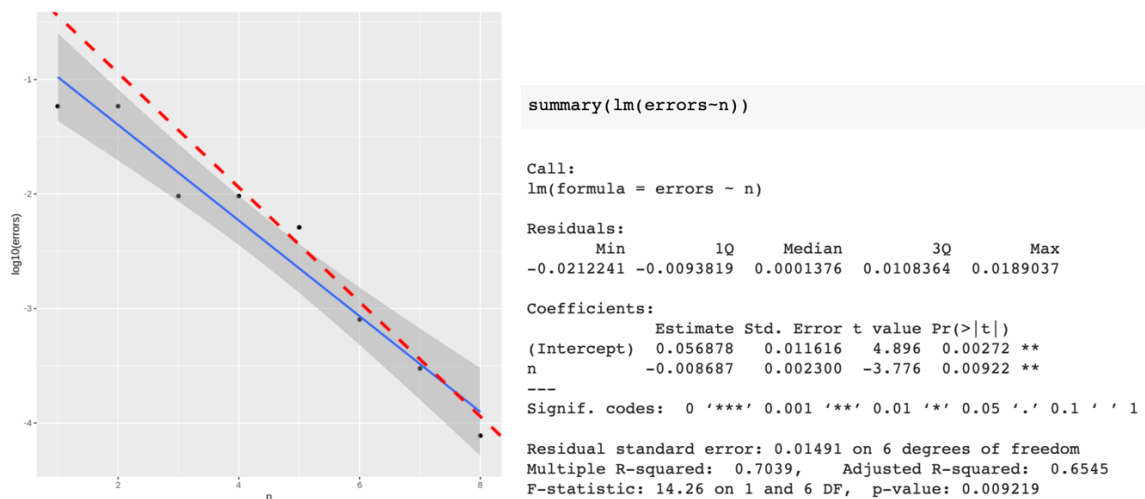
จากการทดลองเราจะเห็นได้ว่า ยิ่งเราทำการเพิ่มค่า Sampling point (N) เราจะสามารถครอบคลุมพื้นที่ของวงกลมได้มากขึ้นทำให้เราสามารถ ประเมินค่า Pi ได้ดียิ่งขึ้น จะเห็นได้จากที่รูปที่ทำการให้ค่า Trials sample เท่ากับ  $10^6$  จะได้รูปวงกลมที่ชัดเจนมาก



King Mongkut's University of Technology Thonburi  
Faculty of Engineering, Department of Computer Engineering  
CPE 213 Data Model, 2/2020

LAB Lecture 7.5: Monte Carlo Simulation

Assign Date: 2 April 2021 Due Date: 8 April 2021



(สีน้ำเงินคือการทำการ Regression fit, สีแดงคือ ทำการวาดเส้นที่จุดตัดเดียวกัน กับ Regression fit แต่ Slope คือ  $-1/2$ )

จากกราฟเราจะสามารถเห็นได้ว่าถ้าเราทำการเพิ่ม Trails sample ไปมากขึ้น เราจะได้ค่าที่ใกล้เคียงกับค่า  $\pi$  จริง ๆ มากขึ้น จะเห็นได้จากค่า Error ที่ลดลง เมื่อทำการ Regression fit เราจะได้ว่าตัวข้อมูล มีแนวโน้ม Slope อยู่ที่  $-1/2$  โดยประมาณ



King Mongkut's University of Technology Thonburi  
 Faculty of Engineering, Department of Computer Engineering  
**CPE 213 Data Model, 2/2020**

LAB Lecture 7.5: Monte Carlo Simulation

Assign Date: 2 April 2021 Due Date: 8 April 2021

## Task 2: Sales Simulation

Jimmy prints a neighborhood newspaper with 10 subscribers. He also sells it to whoever comes by from his front lawn on Friday afternoons. His mother has kept track of his demand (including requests made after he had sold out) for the past 100 weeks, and observed the pattern shown below.

Paper Demanded	Number of weeks	Probability	Cumulative Probability
13	1	0.01	0.01
14	2	0.02	0.03
15	4	0.04	0.07
16	9	0.09	0.16
17	10	0.1	0.26
18	15	0.15	0.41
19	16	0.16	0.57
20	15	0.15	0.72
21	12	0.12	0.84
22	9	0.09	0.93
23	4	0.04	0.97
24	2	0.02	0.99
25	1	0.01	1
26	0	0	1
<b>Total</b>	<b>100</b>	<b>1</b>	<b>1</b>



King Mongkut's University of Technology Thonburi  
 Faculty of Engineering, Department of Computer Engineering  
**CPE 213 Data Model, 2/2020**

LAB Lecture 7.5: Monte Carlo Simulation

Assign Date: 2 April 2021 Due Date: 8 April 2021

The papers cost 30 cents to print and Jimmy sells them for 50 cents. Assume that he prints 20 copies a week. Mom makes him throw away unsold copies. Simulate his sales for the next 12 weeks and determine his earnings.

Week	Random Number	Number Demanded	Number Sold	Number Thrown	Revenue	Cost	Profit
1	0.4175	19	19	1	950	600	350
2	0.8434	22	20	0	1000	600	400
3	0.5227	19	19	1	950	600	350
4	0.1624	17	17	3	850	600	250
5	0.0149	14	14	6	700	600	100
6	0.0067	14	14	6	700	600	100
7	0.0957	16	16	4	800	600	200
8	0.6233	20	20	0	1000	600	400
9	0.9990	25	20	0	1000	600	400
10	0.0391	15	15	5	750	600	150
11	0.2901	18	18	2	900	600	300
12	0.0779	16	16	4	800	600	200
Total							3200

จากการ Simulation เราจะได้กำไร 3200 cent หรือ 32 USD



Should Jimmy increase his paper copies to 30? Would it increase the earning? Why?

จากการทดลองถ้าเราทำการ เพิ่มจำนวน Copies ให้เป็น 30 ฉบับเราจะ ขาดทุน 50 cent สามารถสังเกตได้จาก ตารางข้างล่างนี้ เพราะ การเพิ่มเป็น 30 ฉบับจะต้องใช้เงินลงทุนที่มากขึ้นและเราต้องทิ้งมากขึ้นด้วยถ้าขายไม่ออก จากการทำ Simulation เราจะได้ว่า วิธีการขายนี้จะทำให้ Jimmy ขาดทุน 50 cent ใน ระยะเวลาการขาย 12 สัปดาห์

[illegible]



King Mongkut's University of Technology Thonburi  
 Faculty of Engineering, Department of Computer Engineering  
 CPE 213 Data Model, 2/2020

LAB Lecture 7.5: Monte Carlo Simulation

Assign Date: 2 April 2021 Due Date: 8 April 2021

## Code

```
copies = 20
cost_price = 30
sold_price = 50
# Observed Data
weeks <- c(1, 2, 4, 9, 10, 15, 16, 15, 12, 9, 4, 2, 1, 0)
demand <- c(13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26)
# Create Prob and Cum table
data <- data.frame(
  demand,
  weeks,
  prob = weeks/sum(weeks),
  cum = cumsum(weeks/sum(weeks))
)
rawdata <- rep(demand, weeks)

# Simulate next 12 weeks
sim_table <- data.frame(
  rad = c(0.4175, 0.8434, 0.5227, 0.1624, 0.0149, 0.0067, 0.0957, 0.6233, 0.9990, 0.0391, 0.2901, 0.0779)
)
sim_table$demand = ceiling(quantile(rawdata, sim_table$rad))
sim_table$sold = ifelse(sim_table$demand <= copies, sim_table$demand, copies)
sim_table$thrown = ifelse(sim_table$demand > copies, 0, copies - sim_table$demand)
sim_table$revenue = sim_table$sold * sold_price
sim_table$cost = copies * cost_price
sim_table$profit = sim_table$revenue - sim_table$cost
```

A data.frame: 14 x 4

demand	weeks	prob	cum
<dbl>	<dbl>	<dbl>	<dbl>
13	1	0.01	0.01
14	2	0.02	0.03
15	4	0.04	0.07
16	9	0.09	0.16
17	10	0.10	0.26
18	15	0.15	0.41
19	16	0.16	0.57
20	15	0.15	0.72
21	12	0.12	0.84
22	9	0.09	0.93
23	4	0.04	0.97
24	2	0.02	0.99
25	1	0.01	1.00
26	0	0.00	1.00

A data.frame: 12 x 7

rad	demand	sold	thrown	revenue	cost	profit
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
0.4175	19	19	1	950	600	350
0.8434	22	20	0	1000	600	400
0.5227	19	19	1	950	600	350
0.1624	17	17	3	850	600	250
0.0149	14	14	6	700	600	100
0.0067	14	14	6	700	600	100
0.0957	16	16	4	800	600	200
0.6233	20	20	0	1000	600	400
0.9990	25	20	0	1000	600	400
0.0391	15	15	5	750	600	150
0.2901	18	18	2	900	600	300
0.0779	16	16	4	800	600	200



King Mongkut's University of Technology Thonburi  
 Faculty of Engineering, Department of Computer Engineering  
 CPE 213 Data Model, 2/2020

LAB Lecture 7.5: Monte Carlo Simulation  
 Assign Date: 2 April 2021 Due Date: 8 April 2021

### Task 3: Simulation from Data

Using the Superstore data,

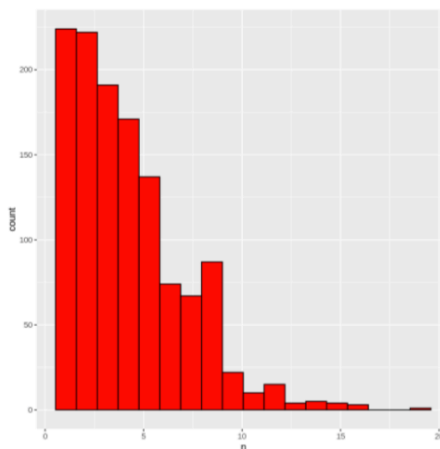
1. Analyze the distribution of daily demand of each order.

#### Code

```
orderByDate <- df %>%
  group_by(Order.Date, Order.ID ) %>%
  summarise(n = 1) %>%
  summarise(n = sum(n))

ggplot(orderByDate) +
  geom_histogram(aes(x = n), fill='red',color='black',bins=18)
```

#### Result



จาก Histogram เราจะเห็นได้ว่าจำนวน Order ในแต่ละวันจะมีการกระจายใกล้เคียงกับ Poisson ซึ่งเราจะเห็นการกระจุกตัวของข้อมูลทางด้านซ้าย แปลว่าในหนึ่งวัน ไม่ได้มีจำนวน Order ที่มากกว่า 10 Order มากนัก





King Mongkut's University of Technology Thonburi  
Faculty of Engineering, Department of Computer Engineering  
CPE 213 Data Model, 2/2020

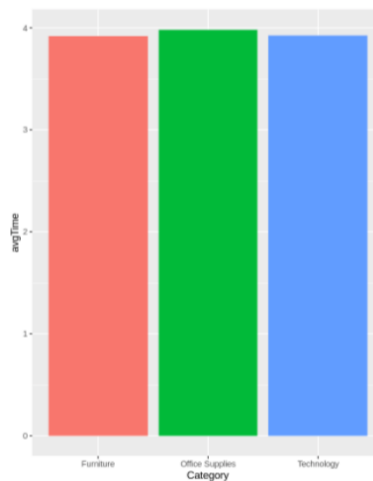
LAB Lecture 7.5: Monte Carlo Simulation  
Assign Date: 2 April 2021 Due Date: 8 April 2021

2. Analyze the time between order of different product category.

### Code

```
processingTimeEachCat <- df %>%  
  mutate(processingTime = as.Date(Ship.Date) - as.Date(Order.Date)) %>%  
  group_by(Category) %>%  
  summarise(avgTime = mean(processingTime))  
  
ggplot(processingTimeEachCat) +  
  geom_col(aes(x = Category, y = avgTime, fill=Category))
```

### Result



จากกราฟเราจะสามารถเห็นได้ว่า เวลาเฉลี่ยของ Order ในแต่ละ Category นั้นไม่ได้แตกต่างกันมาก ซึ่งอาจจะหมายความว่า Category ของ Order ไม่ได้ส่งผลต่อเวลาในการดำเนินการ



King Mongkut's University of Technology Thonburi  
Faculty of Engineering, Department of Computer Engineering  
CPE 213 Data Model, 2/2020

LAB Lecture 7.5: Monte Carlo Simulation

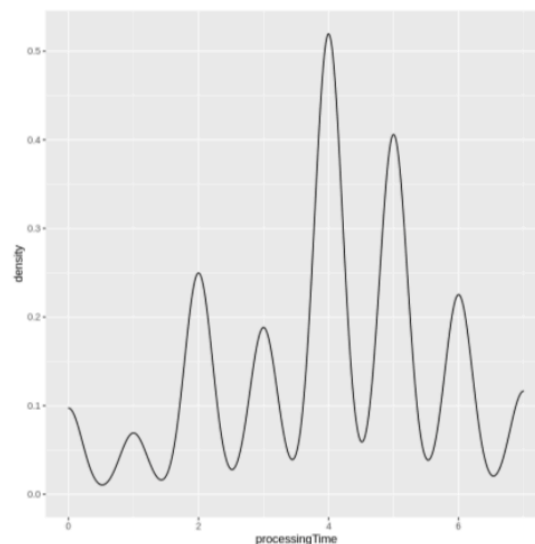
Assign Date: 2 April 2021 Due Date: 8 April 2021

3. Analyze the distribution of order processing time of each order.

### Code

```
processingTimeDist <- df %>%  
  mutate(processingTime = as.Date(Ship.Date) - as.Date(Order.Date))  
  
ggplot(processingTimeDist) +  
  geom_density(aes(x = processingTime))
```

### Result



จากการกระจายของข้อมูล เราจะเห็นได้ว่าข้อมูลการกระจายที่สม่ำเสมอ ใกล้เคียงกับ Normal Distribution ซึ่งจะทำให้เราเห็นว่า Processing time ใช้ของแต่ละ Order ไม่ได้แตกต่างกันมาก



King Mongkut's University of Technology Thonburi  
 Faculty of Engineering, Department of Computer Engineering  
 CPE 213 Data Model, 2/2020

LAB Lecture 7.5: Monte Carlo Simulation

Assign Date: 2 April 2021 Due Date: 8 April 2021

#### 4. Simulate the superstore order and processing time.

##### Code

```
orderProcessingTime <- df %>%
  mutate(processingTime = as.Date(Ship.Date) - as.Date(Order.Date)) %>%
  select(processingTime)

orderByDate <- df %>%
  group_by(Order.Date) %>%
  summarise(n = n())

x <- quantile(orderByDate$n, runif(10))
y <- quantile(orderProcessingTime$processingTime, runif(10))

df_sim <- data.frame(x, y)
names(df_sim) <- c('numberOfOrder', 'processingTime')
df_sim
print(paste0("Average number of order: ", mean(x)))
print(paste0("Average Processing time: ", mean(y)))
```

##### Result

	numberOfOrder <dbl>	processingTime <drtm>
75.2358%	12	3 days
59.78235%	8	6 days
28.26439%	4	2 days
16.04654%	2	4 days
52.8008%	7	4 days
65.87386%	9	4 days
44.75241%	6	4 days
78.71847%	13	6 days
81.48462%	13	4 days
71.80698%	11	5 days

```
[1] "Average number of order: 8.5"
[1] "Average Processing time: 4.2"
```

จากการทำ Simulation ถ้าเราทำการ Simulate อีก 10 วันข้างหน้าเราจะได้ ผลลัพธ์ตามดังนี้



King Mongkut's University of Technology Thonburi  
 Faculty of Engineering, Department of Computer Engineering  
 CPE 213 Data Model, 2/2020

LAB Lecture 7.5: Monte Carlo Simulation

Assign Date: 2 April 2021 Due Date: 8 April 2021

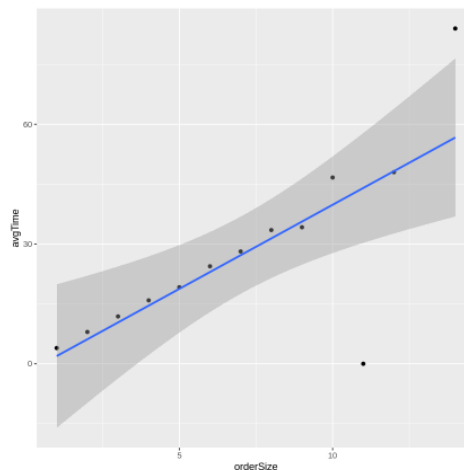
5. Is there relationship between order size and average processing time?

### Code

```
orderSizeATime <- df %>%
  group_by(Order.ID) %>%
  mutate(processingTime = as.Date(Ship.Date) - as.Date(Order.Date)) %>%
  summarise(orderSize = n(), processingTime = sum(processingTime)) %>%
  group_by(orderSize) %>%
  summarise(avgTime = mean(processingTime))

ggplot(orderSizeATime) +
  geom_point(aes(x = orderSize, y = avgTime)) +
  geom_smooth(aes(x = orderSize, y = avgTime), method='lm')
```

### Result



เมื่อลอง Plot ความสัมพันธ์ระหว่าง Order size กับ Processing time เราจะเห็นได้ว่า 2 ตัวแปรนี้ มีความสัมพันธ์กันเชิงผกผันกัน ถ้า Order size ยิ่งมาก จะทำให้ Processing time มากตามขึ้นด้วย



King Mongkut's University of Technology Thonburi  
 Faculty of Engineering, Department of Computer Engineering  
 CPE 213 Data Model, 2/2020

LAB Lecture 7.5: Monte Carlo Simulation

Assign Date: 2 April 2021 Due Date: 8 April 2021

6. What would happen, if every Friday, the number of orders is boosted by 50%?

### Code

วิเคราะห์ Profit ที่จะได้มากขึ้น

```
originFriday <- df %>%
  mutate(day = format(as.Date(Order.Date), format="%a")) %>%
  group_by(day) %>%
  summarise(n = n(), profit = sum(Profit)) %>%
  filter(day == 'Fri')

boostedFriday <- df %>%
  mutate(day = format(as.Date(Order.Date), format="%a")) %>%
  group_by(day) %>%
  summarise(n = n(), profit = sum(Profit)) %>%
  mutate(n = ifelse(day == 'Fri', n * 1.5, n)) %>%
  mutate(profit = ifelse(day == 'Fri', profit * 1.5, profit)) %>%
  filter(day == 'Fri') %>%
  mutate(day = 'BoostedFri')

rbind(originFriday, boostedFriday) %>%
  ggplot() +
  geom_col(aes(x = day, y = profit, fill = profit))
```

วิเคราะห์การกระจายตัวของจำนวน Order ในแต่ละวัน

```
bootsFri <- df %>%
  group_by(Order.Date, Order.ID) %>%
  summarise(n = 1) %>%
  summarise(n = sum(n)) %>%
  mutate(day = format(as.Date(Order.Date), format="%a")) %>%
  mutate(n = ifelse(day == 'Fri', n * 1.5, n))

`summarise()` has grouped output by 'Order.Date'. You can override using the `.groups` argument.

originFri <- df %>%
  group_by(Order.Date, Order.ID) %>%
  summarise(n = 1) %>%
  summarise(n = sum(n))

`summarise()` has grouped output by 'Order.Date'. You can override using the `.groups` argument.

ggplot(bootsFri) +
  geom_histogram(aes(x = n), fill='red',color='black',bins=18)
```

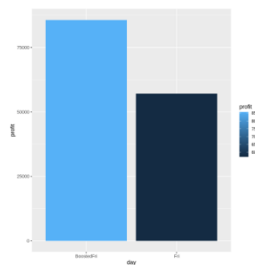


King Mongkut's University of Technology Thonburi  
Faculty of Engineering, Department of Computer Engineering  
CPE 213 Data Model, 2/2020

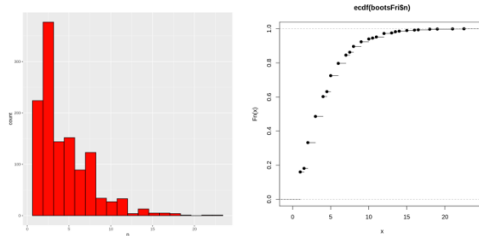
LAB Lecture 7.5: Monte Carlo Simulation

Assign Date: 2 April 2021 Due Date: 8 April 2021

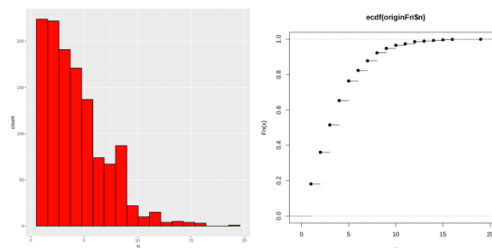
## Result



การที่เราเพิ่มจำนวน Order ให้มากขึ้น 50% ทุกวันศุกร์ เราจะได้ Profit มากขึ้นด้วยประมาณ 50% จากจำนวน Profit เดิมที่มี



เพิ่มจำนวนของ Order 50 %



จำนวน Order ปกติ

ถ้าเราดูที่การกระจายตัวของ Order (PDF) กับ CDF จะเห็นได้ว่าจำนวนของ Order ในแต่ละวัน จะไปจุกตัวเข้าใกล้กับตรงกลางมากขึ้น และ จะเห็นได้ว่าค่า CDF จะถูกบีบให้แคบลงด้วย แต่อย่างไรก็ตาม ไม่ได้ ทำให้การกระจายตัวแตกต่างไปจากเดิมมาก ยังคงเป็นการกระจายตัวแบบ Poisson อยู่เหมือนเดิม