Introduction to Simulations using R

Karthik Sriram
Indian Institute of Management Ahmedabad
karthiks@iima.ac.in

Outline

A Simple Example

Suppose I toss a coin (with two faces), what is the probability of Heads?

A Simple Example

Suppose I toss a coin (with two faces), what is the probability of Heads?

The outcome of coin toss depends on

"Likelihood of getting a Heads"

or

"Probability of Heads"

A Simple Example ctd

How do we find the likelihood of getting heads for a given coin?

A Simple Example ctd

How do we find the likelihood of getting heads for a given coin?

One approach

Toss the coin a large number of times

Compute the average number of times one gets heads.

A Simple Example - Formally

1. Simulate the "Coin Toss"
2. Record the required outcome, viz. Heads or Tails
3. Repeat steps 1 and 2 a large number of times
4. Obtain the distribution of outcomes, i.e. % of outcomes that are Heads, % of outcomes that are Tails

Idea of Simulation

1. Simulate an experiment that involves random outcomes
2. Record the outcome each time you do the experiment
3. Repeat steps 1 and 2 a large number of times
4. Obtain the probability distribution of outcomes

Simulation Example: Dice

If I want to find out the probabilities of different faces of dice

1. Throw the dice
2. Record the outcome i.e. 1,2 ...6
3. Repeat steps 1 and 2 a large number of times
4. Obtain the probability distribution of outcomes,
   % of times 1 occurs, % of times 2 occurs ... % times 6 occurs

What is a Probability Distribution?

A Probability Distribution consists of two parts

1. Set of possible outcomes (e.g. coin toss Heads and Tails)
2. Likelihoods of outcomes (P(Heads) = 0.5, P(Tails)=0.5)

Probability Distribution Example: Fair Dice

A convenient way to write the probability distribution of outcomes from a coin toss

Outcomes	0	1
Probability	$\frac{1}{2}$	$\frac{1}{2}$

SImilarly, Throwing a Dice of 6 faces

Outcomes	1	2	3	4	5	6
Probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Probability Distribution Example: Biased Dice

Note that the coin or dice need not be fair!,

e.g.

Outcomes	0	1
Probability	$\frac{1}{3}$	$\frac{2}{3}$

Outcomes	1	2	3	4	5	6
Probability	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{8}$	0	0

Some examples

The coin or dice examples are simple but extremely useful models!

e.g. Whether an insurance policy brings a claim or not.

e.g. Loan defaults or not

e.g. Number of people out of 6 applicants who get selected for a job

Some examples

The coin or dice examples are simple but extremely useful models!

e.g. Whether an insurance policy brings a claim or not.

e.g. Loan defaults or not

e.g. Number of people out of 6 applicants who get selected for a job

Dice can be *k*-faced in general !

e.g. Number of people out of 100 bookings who show up to board a flight.
100 faced dice!

Types of Distributions

1. Discrete Distributions :

   e.g. Possible outcomes are discrete: Number of customers arriving at a bank counter in a day

2. Continuous Distributions :

   e.g. Possible outcomes are continuous–distribution of heights, salaries, wind speed

Store Example

On any given day, a store is able to sell either 0, 1, 2, or 3 items of a product. An analysis of historical data reveals that the distribution of number of items sold is as follows

Outcomes	0	1	2	3
Probability	0.6	0.2	0.15	0.05

What is the distribution of total number of items sold in 100 days?

Store Example- one simulation

```
# Generating outcome for 1 day
X<-sample(c(0,1,2,3), size=1, replace=TRUE, prob=c(.6,.2,.15,.05))
X

# Generate outcomes for 100 days
X<-sample(c(0,1,2,3), size=100, replace=TRUE, prob=c(.6,.2,.15,.05))
X

# 1 simulation of Total items sold in 100 days
Y<-sum(X)
```

To know the distribution of Y, i.e. possible ooutcomes for Y and % frequency

We need to repeat the above a large number of times

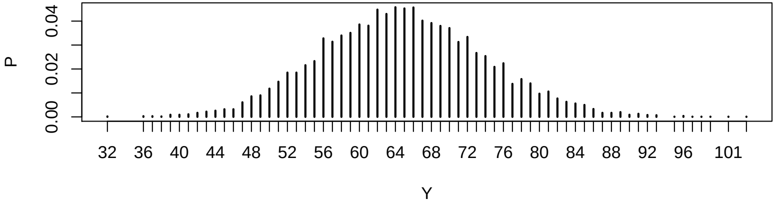
Store Example- Large number if simulations

```
# Function for computing total sales in m days

sales<-function(m){
  X<-sample(c(0,1,2,3), size=m, replace=TRUE, prob=c(.6,.2,.15,.05))
  sum(X)
}
# number of simulations
N<-10000
Y<-replicate(N,sales(100))
```

The vector Y contains a simulation of possiblities for total sales in 100 days

Store Example- Distribution of total sales in 100 days

```
# Vector containing large number (N) of simulations of possiblities
Y<-replicate(N,sales(100))
# Probability distribution
P<-table(Y)/N
plot(P)
```



The vector Y contains a simulation of possiblities for total sales in 100 days

Store Example- How to use the distribution?

What is the probability that more than 50 items will be sold in 100 days?

```
mean(Y>50)

## [1] 0.9477
```

One interpretation of such probabilities. Approximately 9 out of 10 such stores would be able to sell more than 50 items in 100 days.

Store Example-modified

On any given day, a store is able to sell either 0, 1, 2, or 3 items of a product. An analysis of historical data reveals that the distribution of number of items sold is as follows

Outcomes	0	1	2	3
Probability	0.6	0.2	0.15	0.05

Also, the price of the item on any day is subject to random fluctuation and has the following distribution:

Price(Rs.)	10	15	20
Probability	.15	0.8	0.05

Assuming that Price and Demand are independent for the product, What is the distribution of total sales (Rs.) in 100 days?

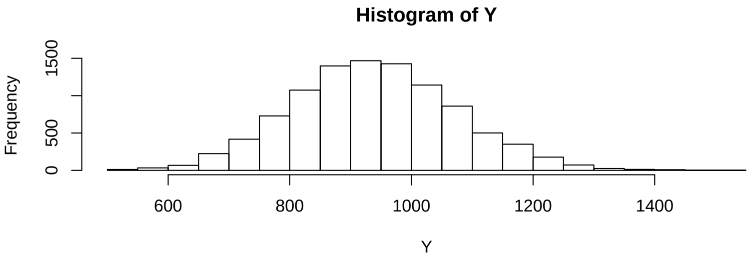
Store Example- Sales in Rupees

```
# Function for computing total sales in m days
sales<-function(m){
  X<-sample(c(0,1,2,3), size=m, replace=TRUE, prob=c(.6,.2,.15,.05))
  Price<- sample(c(10,15,20), size=m, replace=TRUE, prob=c(.15,.8, .05))
  sum(Price*X)
}
# number of simulations
N<-10000
Y<-replicate(N,sales(100))
```

The vector Y contains a simulation of possiblities for total sales (Rupees) in 100 days

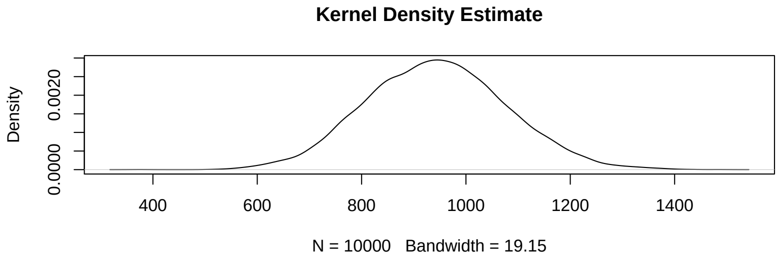
Store Example- Plotting Distribution for Continuous outcomes

```
# Vector containing large number (N) of simulations of possiblities
Y<-replicate(N,sales(100))

# Histogram: Estimate of Probability Distribution
hist(Y)
```



Store Example- Plotting Distribution for Continuous outcomes

```
# Vector containing large number (N) of simulations of possiblities
Y<-replicate(N,sales(100))

#Kernel Density :(Continuous) Estimate of Probability Distribution
plot(density(Y), main="Kernel Density Estimate")
```



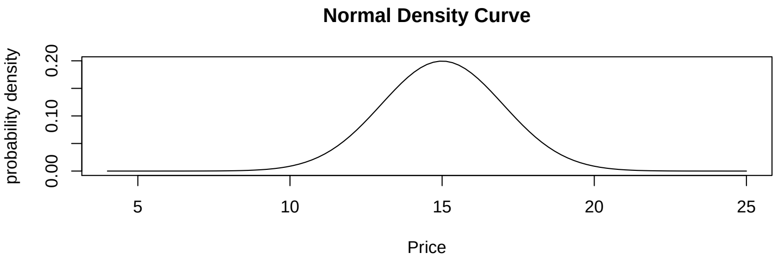
Simulation from known standard distributions

- Sometimes we may want to simulate from standard distributions.
- e.g. Price may have Normal Distribution mean= 15 and standard deviation = 2
- To work with Normal distribution, we need to know mean and sd.
- Notation: Price N(15, sd= 2)

Probability Density of Normal

Price N(15, 2)
The probability distribution is continuous and would look like this:

```
curve(dnorm(x,15,2),xlim=c(4,25),main='Normal Density Curve',
      xlab="Price", ylab="probability density")
```

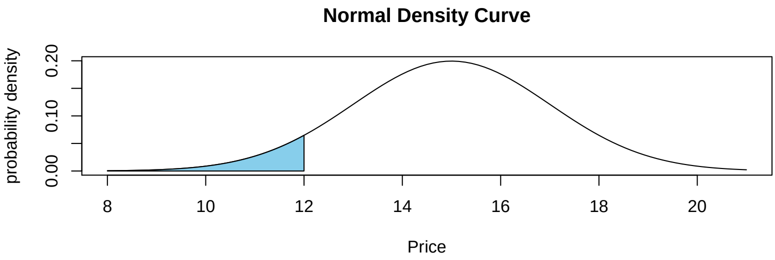



```
polygon(cord.x,cord.y,col='skyblue')

## Error in xy.coords(x, y):  object 'cord.y' not found

# For help on normal distribution in R
```

Probability Density of Normal

Notation: Price∼ $N(mean = 15, sd = 2)$

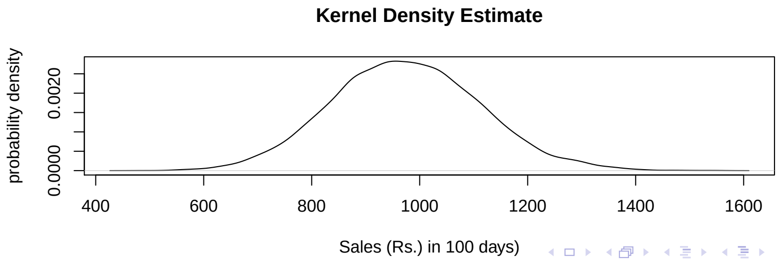
```
cord.x <- c(8,seq(8,12,0.01), 12)
 cord.y <- c(0,dnorm(seq(8,12,0.01),15,2),0)
 curve(dnorm(x,15,2),xlim=c(8,21),main='Normal Density Curve', xlab="Price", yl
 polygon(cord.x,cord.y,col='skyblue')
```




```
 #Probability of the Shaded Region
 pnorm(12, mean=15, sd=2)
```

Store Example- Sales in Rupees, Normal dist

```
# Function for computing total sales in m days
sales<-function(m){
  X<-sample(c(0,1,2,3), size=m, replace=TRUE, prob=c(.6,.2,.15,.05))
  Price<- rnorm(m, 15, 2)    # Random Numbers from Normal Distribution
  sum(Price*X)
}
N<-10000  # number of simulations
Y<-replicate(N,sales(100))  ## Contains simulation of possiblities for sales in

plot(density(Y), main="Kernel Density Estimate", xlab="Sales (Rs.) in 100 days)
```



Insurance Example- Exercise

A insurance company has 10000 auto insurance policies. It has determined that the probability of any policy bringing a claim in a year is 5%. Any such claim can be in various amounts and has been modeled to follow a Normal distribution with mean = 25000 (Rs.) and sd=2000($Rs.$).

What is the 95^{th} percentile of claim payments for the portfolio?
(Assume single claim per policy and independence between policies)

Insurance Example- Exercise 2

A insurance company has 10000 auto insurance policies. It has determined that the probability of any policy bringing a claim in a year is 5%. Any such claim can be in various amounts and has been modeled to follow a Normal distribution with mean $= 25000$ (Rs.) and sd=2000($Rs.$). Suppose that a limit of 30000 applies to each claim.

What is the 95^{th} percentile of claim payments for the portfolio?
(Assume single claim per policy and independence between policies)

Electricity Consumption- Exercise 3

Suppose the electricity consumption (in rupees) per year as a function of square footage of certain commercial buildings can be modeled as follows:

$$\text{Electricity Consumption} \sim N(100 \times \text{Sq Footage}, sd = 25 \times \text{Sq Footage})$$

What is the probability that the total electricity consumption by 5 different commercial buildings, with square footage 1200, 800, 900, 1000 and 500 square foot repectively exceeds Rs. 5,40,000?

(Assume that consumption by buildings are independent of each other)