A

Report

On

Study of Different Concepts of Renewal Theory, Point Pattern Analysis and Comparison of The Distributions of 2 Point Datasets

Summer Research Internship

at

Institute of Advanced Study in Science and Technology, Guwahati

Under the guidance of

DR. GAUTAM CHOUDHURY

Submitted by

PRIYABRATA MISHRA

IMH/10025/17

Department of Mathematics

BIT MESRA

9th May 2019 – 21st June 2019

Institute of Advanced Study in Science & Technology (IASST)

Guwahati

CERTIFICATE

Certified that the summer internship report "Study of Different Concepts of Renewal Theory, Point Pattern Analysis and Comparison of The Distributions of 2 Point Datasets" is the bonafide work of **Priyabrata Mishra,** Roll number IMH/10025/17, 5th semester Integrated M.Sc. in Mathematics & Computing at Birla Institute of Technology Mesra, Ranchi carried out under my supervision during 9th of May, 2019 to 21st of June, 2019.

Place -

Date -

Signature of the Supervisor

Dr. Gautam Choudhury

Associate Professor

Head of the Department

CCNS

Contents

CHAPTER-1: INTRODUCTION TO RENEWAL THEORY 6
1.1 Stochastic Process 6
1.2 Memoryless Property 6
1.3 Survival Function 7
1.4 Markov Property 8
1.5 Markov Process 8
1.6 Renewal Process 8
1.7 Renewal Function 9
1.8 Renewal equation 10
1.9 Reward Renewal Process 11
1.10 Markov Reward Model 11
1.11 Delayed Renewal Process 12
1.12 Markov Renewal Process 13
1.13 Markov Regenerative Process 15
CHAPTER – 2: INTRODUCTION TO POINT PATTERN ANALYSIS 17
2.1 Pattern Analysis 17
2.2 Point Pattern Analysis 17
2.3 Complete Spatial Randomness (CSR) 17
2.4 Quadrat Count Method 19
2.5 Nearest Neighbour Method 19
2.6 g-function 21
CHAPTER – 3: COMPARISON OF THE DISTRIBUTIONS OF 2 POINT DATASETS BY USING DIFFERENT METHODS 22
RESULTS 33
CONCLUSION 34
REFERENCES 35

ACKNOWLEDGEMENT

I would like to convey my gratitude to my project supervisor Dr. Gautam Choudhury of CCNS, Institute of Advanced Study in Science and Technology, Guwahati for his constant guidance and advice which played a vital role in making the execution of this report. I would also like to thank Miss Priyanka Kalita whose support and mentorship throughout the project made it a success.

Further, I acknowledge the research scholars and other staffs for their valuable suggestions and help.

ABSTRACT

Point Pattern Analysis is important as its applications include the analysis of the distribution of discrete data points. Renewal Theory also deals with discrete data. So in this report numerical data are compared by using the concepts of mean, variance, standard deviation, coefficient of variation and other Pattern Analysis methods.

In the chapter -1 some concepts of Renewal Theory are discussed. Basics of Point Pattern Analysis and different methods to compare the point data are given in chapter – 2. Chapter – 3 is all about taking numerical data such as area of locations, number of hotels in those locations from Google Map and applying the methods discussed in chapter – 1 and chapter – 2 to compare their pattern and decide the better distribution.

CHAPTER-1: INTRODUCTION TO RENEWAL THEORY

1.1 Stochastic Process

A stochastic process is a collection of random variables indexed by a variable t, which represents time.

It can also be defined as the function of two parameters from the sample space and parameter space i.e. $X(w,t)$, where $w \in \Omega$ (sample space) & $t \in T$(parameter time).

If the random variables associated with a stochastic process are discrete, the process is called Discrete Space Stochastic Process and if they are continuous, the process is called Continuous Space Stochastic Process.

Therefore, we have four types of Stochastic Processes:

1. Continuous Time Continuous Space

2. Continuous Time Discrete Space

3.Discrete-Time Continuous Space

4. Discrete Time Discrete Space

1.2 Memoryless Property

The memoryless property states that a probability distribution depends only on the previous state, not the whole history.

Only two types of distributions are memoryless: exponential distribution of non-negative numbers and geometric distribution of non-negative numbers.

1.3 Survival Function

The survival function gives the probability of the survival of an object of interest beyond any given specific time.

Mathematically,

$$G(t) = P(X > t) \tag{1.1}$$

We know that,

$$P(X > t+s \mid X > t) = P(X > s)$$

$$\Rightarrow P(X > s) = \frac{P(X>t+s)}{P(X>t)}$$

(by using conditional probability)

$$\Rightarrow G(s) = \frac{G(t+s)}{G(t)} \qquad \text{(by using (1.1))}$$

$$\Rightarrow G(s).G(t) = G(t+s)$$

From the above relation we can also find that

$$G(a) = G(1)^a = e^{\ln(G(1))\wedge a} = e^{-\lambda a}$$

Where $\lambda = -\ln(G(1))$.

Here G(1) is a probability. So the value of λ cannot be negative. This implies that any memoryless function must be an exponential.

1.4 Markov Property

Markov property refers to the memoryless property of a stochastic process i.e. the conditional probability distribution of the present state depends only on the previous state.

Mathematically,

For any i, j $\in \Omega$ and n ≥ 0,

$$P(X_{n+1} = j \mid X_n = i, \ldots \ldots \ldots, X_0) = P(X_{n+1} = j \mid X_n = i)$$

1.5 Markov Process

A stochastic process satisfying Markov property is called a Markov process.

1.6 Renewal Process

A renewal process is a counting process where the interarrival times, the times between successive events, are independent and identically distributed (iid) random variables with an arbitrary distribution. The Poisson process is a well-known example of this, whose interarrival times are iid exponential random variables.

Definition: A counting process, $\{N(t), t \geq 0\}$, is a renewal process if $\{X_1, X_2, X_3, \ldots\ldots\}$, the sequence of non-negative random variables of interarrival time are independent and identically distributed.

Example: Consider the machine case. When a machine breaks down, we replace it with another machine whose lifetime is independent of the previous one.

Here, renewal is the replacing of a machine i.e. the arrival of an event. X_n is the lifetime of machine n i.e. the n^{th} interarrival time. Then $\{N(t), t \geq 0\}$ is the renewal process and it represents the number of machines that have stopped working by time t i.e. the number of renewals by time t.

1.7 Renewal Function

The renewal function represents the expected value of the distribution N(t).

That is,

$$\text{Renewal Function} = M(t) = E(N(t))$$

The renewal function is unique to its specific renewal process' interarrival distribution.

The renewal function of discrete distribution can be found as below,

$$\begin{aligned} M(t) &= E(N(t)) \\ &= \sum_{n=1}^{\infty} P(N(t) \geq n) \\ &= \sum_{n=1}^{\infty} P(T_n \leq n) \\ &= \sum_{n=1}^{\infty} F_n(t), \text{ for } t > 0 \end{aligned}$$

Renewal density function: The renewal density specifies the mean number of renewals to be expected in a narrow interval near t. The renewal density function is the derivative of renewal function.

Consider m(t) as the renewal density function. Then,

$$m(t) = M'(t)$$

Proof:

$$m(t) = \lim_{h \to 0+} \frac{\text{P(1 or more renewals in (t,t+h))}}{\text{h}}$$

$$= \lim_{h \to 0+} \sum_{n=1}^{\infty} \frac{\text{(n renewals in (t,t+h))}}{\text{h}}$$

$$= \sum_{n=1}^{\infty} \lim_{h \to 0+} \frac{F_n(t+h) - F_n(t)}{h} = M'(t)$$

1.8 Renewal equation

The expression for the renewal equation is given by,

$$M(t) = F(t) + \int_0^t M(t-x)dF(x) \qquad (1.2)$$

with M(t) and F(t) general functions defined for t ≥ 0. Usually, the function F(t) is a known function and the motive is to find a function M(t) that satisfies (1.2) for all t ≥ 0.

We can write the equation (1.2) in another way as

$$M = F + M \,.\, F \qquad (1.3)$$

Taking Laplace on both sides of (1.3)

$$M^* = F^* + M^* \,.\, F^* \qquad (1.4)$$

By the simple Laplace rule, it is known that

$$F^* = \frac{f^*(s)}{s} \qquad (1.5)$$

Using equation (1.5) in equation (1.4), we get the result

$$f^*(s) = \frac{s\,.\,M^*(s)}{1+ s\,.\,M^*(s)}$$

1.9 Reward Renewal Process

Let in an renewal process, X_1 is the time to the first renewal and $\{X_n, n = 2,3,4 \ldots\}$ is the time between (n-1)th and n-th renewal. Assuming $\{X_n, n = 1,2,3,4 \ldots\}$ are independent and identically distributed random variables with distribution function F, expectation of the random variables can be defined as,

$$\mu = E(X_n) = \int_0^\infty x\, dF(x)$$

which is always positive.

Let R_n is the nth reward at the time of the nth renewal and R(t) is the reward earned by the time t.

So, $$R(t) = \int_{n=1}^{N(t)} R_n$$

Note that R_n can negative as well as positive values.

1.10 Markov Reward Model

Markov reward model is a stochastic process which extends a continuous time Markov chain by adding a reward rate to each state.

Here we can define two new terms: State reward structure and Impulse reward structure.

State reward structure: This is a function r that assigns a reward r_s to each state $s \in S$ (state space) if some amount of time t is spent in that particular state. This reward can be a benefit for visiting the state or a loss for spending time in that state. This type of Markov reward model is called rate based markov reward model.

Impulse reward structure: The impulse reward function τ assigns a reward for every transition from s to s' , where s, s' $\in$ S. Unlike the state reward structure this can also be a gain or a loss.

Let $Z(t) = r_{X(t)}$ be the instantaneous reward rate of the Markov reward model at time t. Then the expected instantaneous reward rate at time t is given by,

$$E[Z(t)] = \sum_{i\in S} r_i P_i(t)$$

The expected reward rate in steady-state is given by,

$$E[Z] = \sum_{i\in S} r_i \pi_i$$

where π is the steady-state probability.

1.11 Delayed Renewal Process

There can be a counting process for which the first interarrival time has a different distribution from the remaining ones. This process is called a delayed renewal process or general renewal process.

Definition: Let $\{X_n, n = 1,2,\ldots.\}$ be a sequence of independent non-negative random variables with X_1 having distribution G, and X_n having distribution F, n > 1.

Let $S_0 = 0$, $S_n = \sum_1^n X_i$, $n \geq 1$, and define a function

$$N_D(t) = \sup\{n: S_n \leq t\}$$

Then the stochastic process $\{N_D(t), t \geq 0\}$ is called a delayed renewal process.

Note that when the distribution G = F, the process is an ordinary renewal process.

1.12 Markov Renewal Process

Sojourn time: Amount of time spent by an object in a system before leaving it is called sojourn time.

Holding time: This is the time between 2 jumps.

There are 2 properties:

1. Sojourn times are exponential distributions.

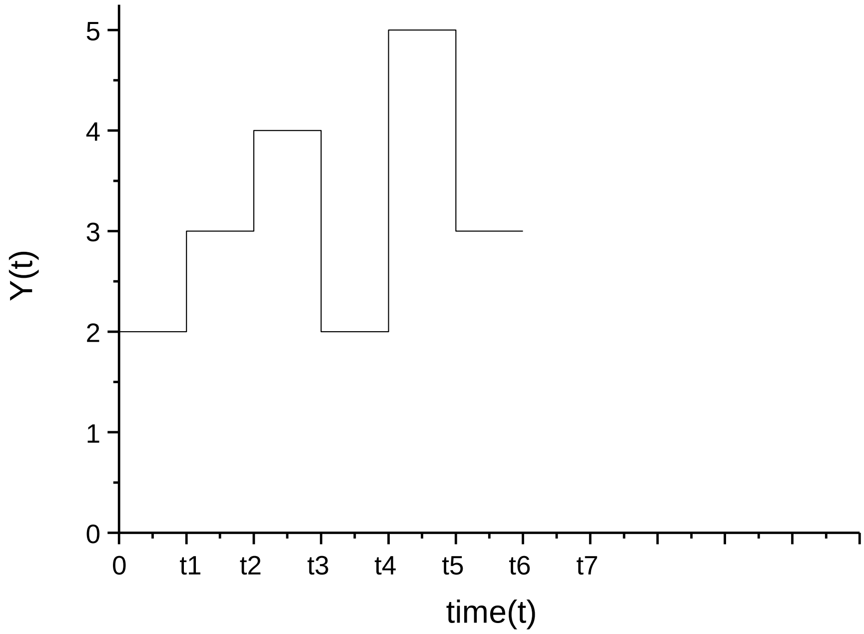
2. Every state change represents a markovian regeneration instant.

If any stochastic process satisfies the second property it is called Markov Renewal Process or Semi Markov Process (SMP).

Any stochastic process satisfying the first property is called Markov regenerative process (MRGP). We will discuss it later.

Definition: Let Y(t) denotes the state of the system at a time t, then $Y(t_n) = X_n$ for n = {0,1,2,….}. The process $\{Y(t), t \geq 0\}$ is called a semi markov process when all the time instants t_n, n = {0,1,2,…..} satisfy the markov property.

That is, the state at the time instant $t = t_n$ depends only on the past history of the system till $t = t_n$.



(Figure - 1)

The sample path of a semi Markov process is given in Figure - 1. The time points (t1,t2,…) shown in the figure are the renewal points and they follow the Markov property i.e., for example, the state at time t5 is directly dependent only on the state at time t4.

Note that, in case of semi Markov process the Markov property is satisfied only at each of the transition epochs $\{t_n\}$, not at all points. If it is satisfied at all points, the process is called Markov process.

1.13 Markov Regenerative Process

Consider a stochastic process $\{Z(t), t \geq 0\}$ with state space $S = \{0,1,2,3,\ldots.\}$. Suppose in the process $Z(t)$ some time points exist at which the process restarts itself i.e. the future of the process Z after every time becomes a probabilistic replica of the future after time zero. Such time points are called regeneration points and the process is called regenerative process.

Note that in an MRGP the stochastic evolution between two successive regeneration points depends only on the state at regeneration, not before regeneration.

Markov renewal sequence: A sequence of bivariate random variables $\{(Y_n,S_n), n \geq 0\}$ is called a Markov renewal sequence if

(i) $S_0 = 0$, $S_{n+1} = S_n$, $Y_n \in S'$
(where S' is a subset of the state space S)

(ii) for all $n \geq 0$,

$$P\{Y_{n+1} = j, S_{n+1} - S_n \leq t \mid Y_n = i, S_n, Y_{n-1}, S_{n-1}, .. Y_0, S_0\}$$

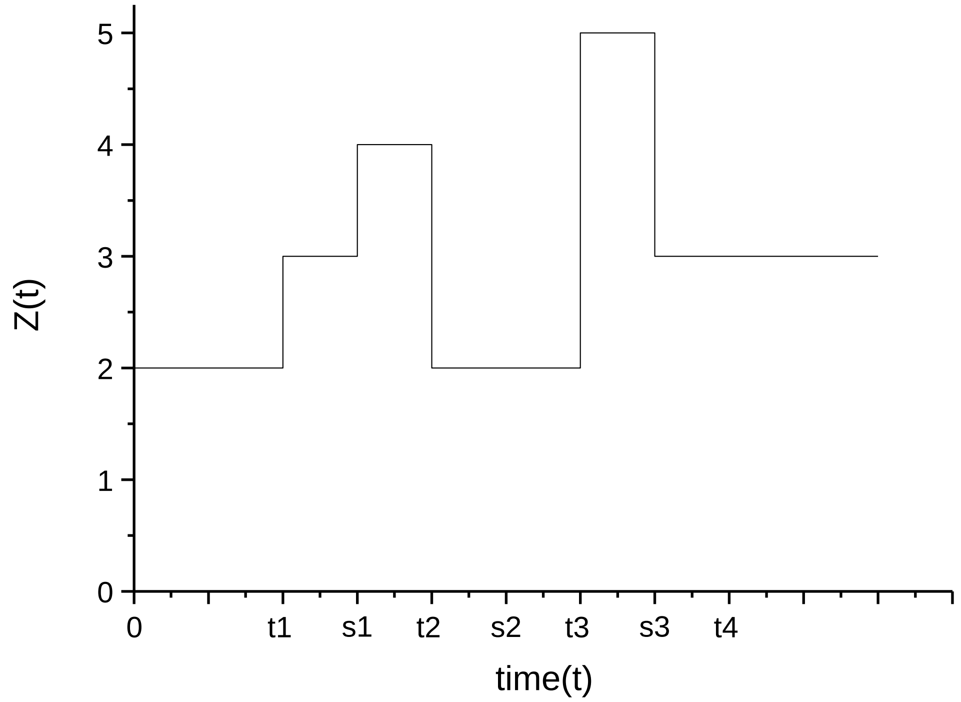
$$= P\{Y_{n+1} = j, S_{n+1} - S_n \leq t \mid Y_n = i, S_n\} \text{ (Markov property)}$$

$$= P\{Y_1 = j, S_1 - S_0 \leq t \mid Y_0 = I, S_0\} \text{ (Time Homogeneity)}$$

$$= P\{Y_1 = j, S_1 \leq t \mid Y_0 = i\} \quad \{\text{as } S_0 = 0\}$$

Definition of MRGP: A stochastic process $\{Z(t), t \geq 0\}$ on S is called Markov regenerative process if there exists a Markov renewal sequence $\{(Y_n,S_n), n \geq 0\}$ of random variables such that all conditional distributions of $\{Z(S_n + t), t \geq 0\}$ are the same as those of $\{Z(t), t \geq 0\}$.

In the Figure -2 the sample path of a Markov regenerative process is shown where {t1,t2,t3,t4} are the regeneration points. We can observe that in this case, the graph goes up or down in between the regeneration points also. The points {s1,s2,s3} where this incidence occurs are called service completion times.



(Figure - 2)

CHAPTER – 2: INTRODUCTION TO POINT PATTERN ANALYSIS

2.1 Pattern Analysis

Pattern analysis is finding a particular pattern from a given random data.

2.2 Point Pattern Analysis

Point pattern analysis is the study of spatial arrangements in geographical space.

Point pattern process is of 3 types:

(i) Deterministic Process – If the spatial process gives the same outcome every time it is executed, it is called deterministic process. All mathematical equations are an example of this process.

(ii) Stochastic Process – Spatial process giving different outcomes every time is called stochastic process. These outcomes are unpredictable and this process has no particular pattern. Ex- disease spread in a location.

(iii) Independent Random Process – This process is independent of location.

2.3 Complete Spatial Randomness (CSR)

This is a point process where point events occur within a given study area in a completely random fashion.

2 conditions are satisfied in CSR:

(i) Independence of event positions.

(ii) The probability of events or points being in any position is equal.

The frequency distribution of complete spatial randomness satisfies the Poisson distribution.

The expected proportion of quadrats with k events is approximated through the Poisson distribution. (quadrats are the equally divided areas)

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Where λ is the average rate of events per quadrats in the study area. i.e. $\lambda = \frac{total\ number\ of\ events}{total\ number\ of\ quadrats}$

Point process under CSR is often referred to as Spatial Poisson Process.

Note that CSR is the standard case i.e. we compare other patterns to this to find they are clustered or evenly distributed. Basically, we consider CSR as the null hypothesis.

But the real world processes are not always CSR. They differ from it due to 2 effects.

1st order effect – It produces a variation in point density in response to some casual variables. Ex- Soil type affects the presence of plant species.

2nd order effect – Results due to the interaction between events. Ex- the spread of contagious diseases.

2.4 Quadrat Count Method

This method detects if a point pattern deviated from randomness.

Step-1: Divide the whole area into a number of quadrats.

Step-2: Count the number of events in each quadrat.

Step-3: Analyse the expected and observed number of events.

There are 2 types of quadrat count methods.

(i) Exhaustive Census: No overlapping of patterns is there in this case. This method uses measured event data.

(ii) random Sampling: Quadrats are overlapped in this case. This method is mainly applied in field works.

2.5 Nearest Neighbour Method

Mean nearest neighbour distance: It finds the distance between each point or event and its closest neighbour, then averages these distance.

Observed Mean NN distance = $d_{obs}(\text{mean}) = \frac{\sum_{i=1}^{n} d_{min}(s_i)}{n}$

where n = Number of events.

This value can be compared to the expected mean nearest neighbour (NN) distance under CSR.

Expected Mean NN distance = $d_{exp}(\text{mean}) = \frac{1}{2}\sqrt{\frac{\text{Area}}{n}}$

Area = Total area of the considered location.

Nearest neighbour index: This measures the similarity between the observed mean nearest neighbour distance and the expected mean nearest neighbour distance of a random point pattern.

The nearest neighbour index can be defined as a difference or as a ratio. The expressions for both are given below,

Nearest neighbour index 'd' as a difference = d_{obs}(mean) - d_{exp}(mean)

Nearest neighbour index 'r' as a ratio = $\frac{d_{obs}(\text{mean})}{d_{exp}(\text{mean})}$

We can find the pattern by observing these indices. The results are given in the table below (Table – 2.1).

d	r	Pattern
$d < 0$	$r < 1$	Clustered
$d > 0$	$r > 1$	Dispersed
$d = 0$	$r = 1$	Random

(Table – 2.1)

2.6 g-function

The nearest neighbour method hides many of the nearest neighbour information and presents only a single mean value. So g-function is considered for the similar test. It is basically an extension of the nearest neighbour method.

g-function finds the cumulative frequency distribution of nearest neighbour distances.

The formula for the g-function is given by,

$$g(d) = \frac{\text{Number of events having minimum distance} \leq d}{\text{Total number of events}}$$

We then plot this value on a graph and observe the behaviour. For the clustered distribution g-function increases rapidly at short distances and for evenly spaced distribution it slowly increases up to event spacing.

CHAPTER – 3: COMPARISON OF THE DISTRIBUTIONS OF 2 POINT DATASETS BY USING DIFFERENT METHODS

Two locations of Guwahati city were considered and the number of hotels was counted in these two regions. Hence hotels are events or points in our case. Then each location is divided into 5 quadrats. All the data regarding the number of hotels and areas are shown in the tables below. Google Map was used as the source of all data.

Location – 1 : (Total Area = 0.54 km^2)

Sr. No.	Area (km^2)	No of Hotels	Cumulative Area (km^2)	Cumulative No of Hotels	Cumulative Probability
1	0.108	5	0.108	5	0.454
2	0.108	5	0.216	10	0.909
3	0.108	0	0.324	10	0.909
4	0.108	1	0.432	11	1.000
5	0.108	0	0.540	11	1.000

(Table – 3.1: Data for location - 1)

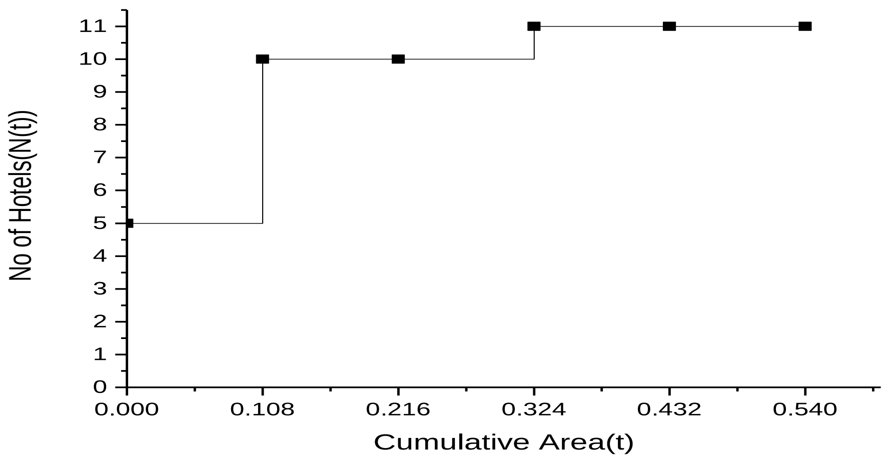
Location – 2: (Total Area = 0.29 km^2)

Sr. No.	Area (km^2)	No of Hotels	Cumulative Area (km^2)	Cumulative No of Hotels	Cumulative probability
1	0.058	2	0.058	2	0.111
2	0.058	6	0.116	8	0.444
3	0.058	1	0.174	9	0.500
4	0.058	3	0.232	12	0.667
5	0.058	6	0.290	18	1.000

(Table – 3.2: Data for location - 2)

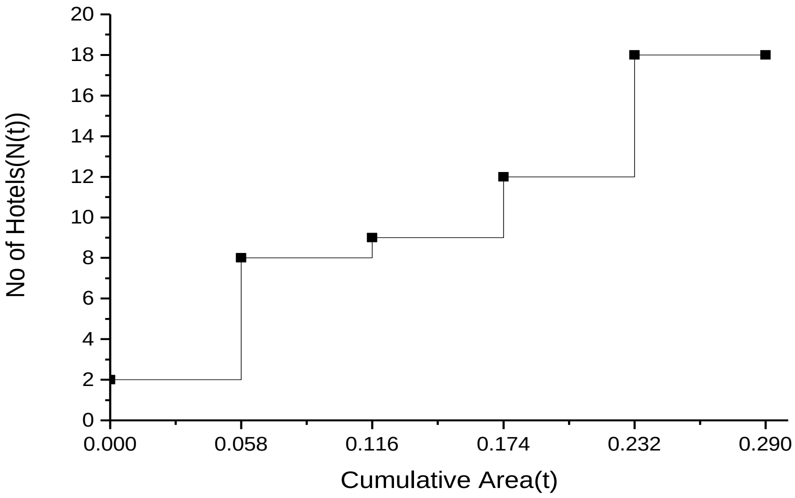
The renewal sample paths of the distributions are given below. Here cumulative areas are taken as the regeneration points.

Sample path for Location – 1:



(Figure – 3.1)

Sample path for location – 2:



(Figure – 3.2)

From the above two sample paths it is clear that the distribution of hotels over areas satisfies the renewal process. We will compare both the distributions by measuring their mean, variance, standard deviation and coefficient of variation. The expressions are given by,

$$\text{Mean} = \mu = x.P(X = x)$$

$$\text{Variance} = \text{Var}(X) = (x - \mu)^2 .P(X = x)$$

$$\text{Standard Deviation} = \text{SD}(X) = \sqrt{\text{Var}(X)}$$

$$\text{Coefficient of Variation} = \frac{\text{Standard Deviation}}{\text{Mean}}$$

$$= \frac{\text{SD}(X)}{\mu}$$

Calculating Mean, Variance, Standard Deviation and Coefficient of Variance for both the locations:

Location – 1:

Mean $= 0.108 \times 0.454 + 0.216 \times 0.909 + 0.324 \times 0.909 + 0.432 \times 1.000 + 0.540 \times 1.000 = 1.512$

Variance $= \text{Var}(X) = (0.108 - 1.512)^2 \times 0.454 + (0.216 - 1.512)^2 \times 0.909 + (0.324 - 1.512)^2 \times 0.909 + (0.432 - 1.512)^2 \times 1.000 + (0.540 - 1.512)^2 \times 1.000 = 5.816$

Standard Deviation $= \text{SD}(X) = \sqrt{5.816} = 2.412$

Coefficient of Variation $= \frac{2.412}{1.512} = 1.595$

Location – 2:

Mean $= 0.058 \times 0.111 + 0.116 \times 0.444 + 0.174 \times 0.5 + 0.232 \times 0.667 + 0.290 \times 1.000 = 0.467$

Variance $= \text{Var}(X) = (0.058 - 0.467)^2 \times 0.111 + (0.116 - 0.467)^2 \times 0.444 + (0.174 - 0.467)^2 \times 0.5 + (0.232 - 0.467)^2 \times 0.667 + (0.290 - 0.467)^2 \times 1.000 = 0.184$

Standard Deviation $= \text{SD}(X) = \sqrt{0.184} = 0.429$

Coefficient of Variation $= \frac{0.429}{0.467} = 0.919$

Comparing the distributions by Nearest Neighbour method:

Tables below show each hotel in both locations and the distance to their nearest hotel. All distances are measured from Google Map.

Location -1:

Sr. No.	Hotel	Nearest Hotel	Distance (km)
1	H1	H2	0.09
2	H2	H4	0.05
3	H3	H2	0.1
4	H4	H2	0.05
5	H5	H4	0.03
6	H6	H5	0.07
7	H7	H8	0.06
8	H8	H7	0.06
9	H9	H10	0.1
10	H10	H9	0.1
11	H11	H7	0.34

(Table – 3.3)

In both the tables H_i represents hotel number where i belongs to the set of positive integers.

Location – 2:

Sr. No.	Hotel	Nearest Hotel	Distance (km)
1	H1	H2	0.06
2	H2	H1	0.06
3	H3	H4	0.05
4	H4	H6	0.05
5	H5	H6	0.05
6	H6	H4	0.05
7	H7	H6	0.12
8	H8	H9	0.09
9	H9	H10	0.07
10	H10	H11	0.06
11	H11	H12	0.05
12	H12	H11	0.05
13	H13	H14	0.03
14	H14	H13	0.03
15	H15	H17	0.05
16	H16	H17	0.03
17	H17	H16	0.03
18	H18	H16	0.12

(Table – 3.4)

By using the data in the table -3.3, table - 3.4 and the formula discussed in section 2.5 we can find the nearest neighbour index for both the locations and see the behaviour.

Location – 1:

$d_{exp}(\text{mean}) = \frac{1}{2}\sqrt{\frac{\text{Area}}{n}} = \frac{1}{2}\sqrt{\frac{0.540}{11}} = 0.1108$

$d_{obs}(\text{mean}) = \frac{\sum_{i=1}^{n} d_{min}(s_i)}{n}$

$= \frac{0.09+0.05+0.1+0.05+0.03+0.07+0.06+0.06+0.1+0.1+0.34}{11}$

$= \frac{1.05}{11} = 0.0954$

Nearest Neighbour Index as a difference = d_1 = d_{obs}(mean) - d_{exp}(mean) = 0.0954 – 0.1108 = -0.0226

As the index is negative, that is d < 0, we found that the distribution of hotels in this location is clustered.

Location – 2:

$d_{exp}(\text{mean}) = \frac{1}{2}\sqrt{\frac{\text{Area}}{n}} = \frac{1}{2}\sqrt{\frac{0.290}{18}} = 0.0635$

$d_{obs}(\text{mean}) = \frac{\sum_{i=1}^{n} d_{min}(s_i)}{n} = \frac{1.05}{18} = 0.0583$

Nearest Neighbour Index as a difference = d_2 = d_{obs}(mean) - d_{exp}(mean) = 0.0583 – 0.0635 = -0.0052

We found the index in case of location – 2 is also negative. So the distribution of hotels in this location is also clustered.

But we observed that the absolute value of NN index for location – 1 is greater than the absolute value of NN index for location – 2. That means the distribution in location – 2 is closer to the CSR case.

g-function comparison:

The g-function values for both the locations are given in the tables below. These values can be calculated by using the formula given in section 2.6.

Location – 1:

Sr. No.	No of Hotels	g-value
1	2	0.182
2	3	0.273
3	5	0.455
4	6	0.545
5	7	0.636
6	10	0.909
7	11	1.000

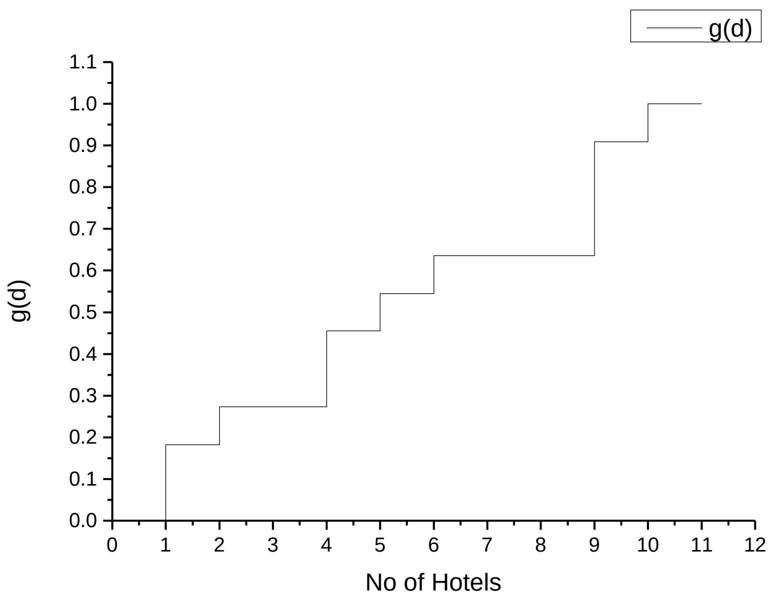
(Table – 3.5)

Location – 2:

Sr. No.	No of Hotels	g-value
1	4	0.222
2	11	0.611
3	14	0.778
4	15	0.833
5	16	0.889
6	18	1.000

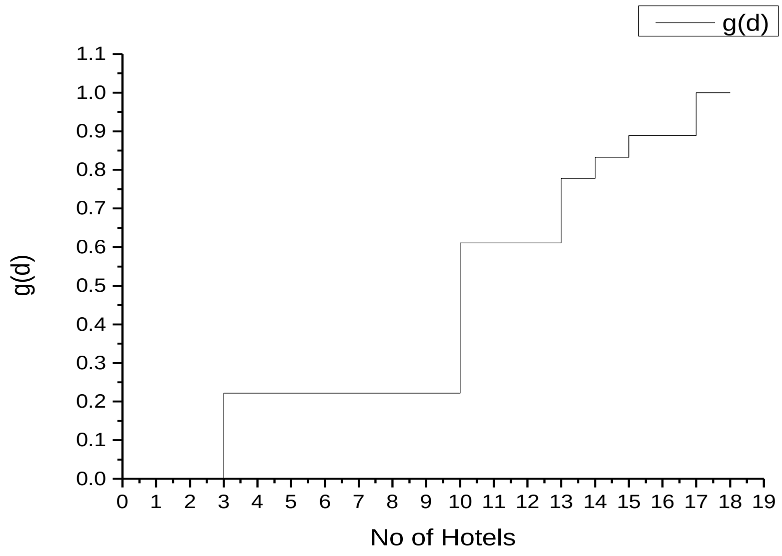
(Table – 3.6)

When the data given in Table – 3.5 and Table – 3.6 were plotted we found the following graphs.



g-function for the data of location-1

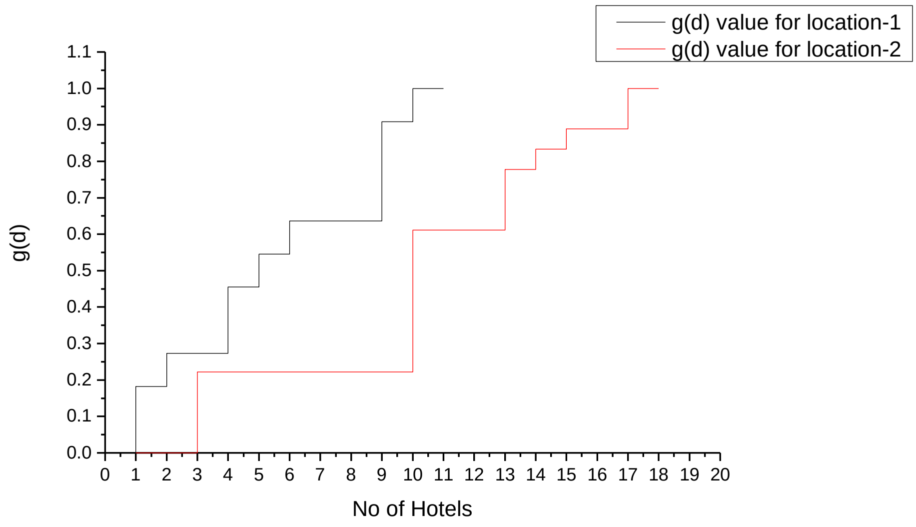
(Figure – 3.3)



g-function for the data of location-2

(Figure – 3.4)

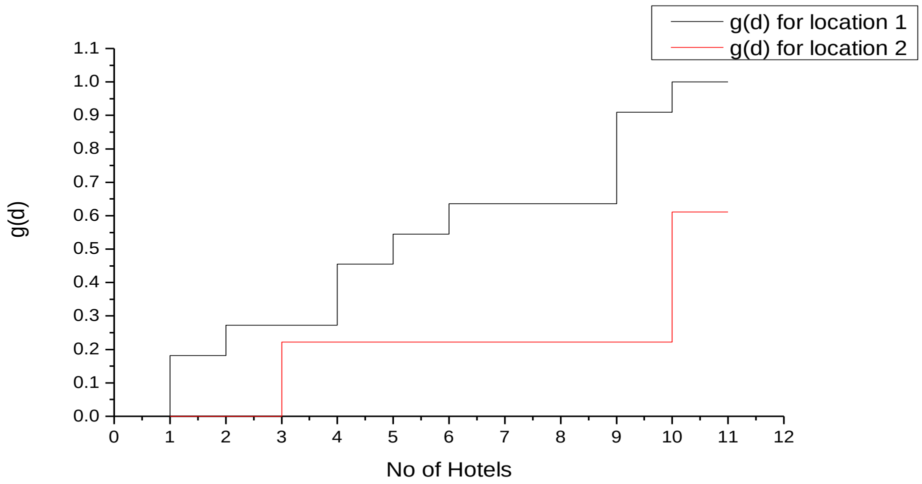
Plotting the g values for both the locations we got,



Comparison of g(d) values

(Figure – 3.5)

But the issue is that the number of hotels in both locations is not the same. So we will consider only the first 11 hotels of location – 2 for the g value comparison.



Comparison of g(d) values of first 11 points

(Figure – 3.6)

In figure – 3.6 the distribution of first 11 hotels of location – 2 is compared to the distribution of the hotels of location – 1. We know if the points are closely located the graph rises up rapidly. Here the graph for location – 1 goes up faster than the graph of location – 2. This means the hotels in location – 2 are more evenly distributed than the hotels in location – 1.

RESULTS

1. We found for location – 1 standard deviation (2.412) is greater than the mean (1.512) and for location – 2 standard deviation (0.429) is less than mean (0.467). It says that the distribution of hotels is normal in case of location – 2 as for a good distribution of data standard deviation should be less than the mean value.
2. Coefficient of variation gives the degree of variation and the less numerical value of this gives a better result. In our case coefficient of variation of location – 1 is 1.595 and that for location – 2 is 0.919.
3. From the Nearest Neighbour Method, we got the NN indices are -0.0226 and -0.0052 respectively. The absolute value of the NN index for location – 2 is less, so the distribution is dispersed as compared to the location – 1.
4. The graphs of g-functions show that the hotels in location – 1 are closely located as compared to the hotels in location – 2.

CONCLUSION

In the studies mentioned in this report, we have started with describing the stochastic process and ended with different methods to compare the distributions of point data. When we compared the data which are given in this report, we found the location – 2 has a better distribution. Hotels are not concentrated in only one place here, which helps people to find them more easily. It also helps the owners to analyse and decide where to start a hotel business.

In real life, these methods are applicable for not only hotels but also for any kind of shops, shopping malls etc. These methods are also applicable to analyse the distribution of natural resources in a particular location, to identify individual stands of specific types of trees, the location of diseased trees in forests.

We are currently working on a paper showing the application of Renewal Theory on the Point Pattern Analysis. This is to be done by analysing the pattern of the hotels in Guwahati City. We are just mentioning it as we have not got all the results.

REFERENCES

[1] Medhi, J. *Stochastic Processes*.

[2] Ross, M., Sheldon. *Stochastic Processes*.

[3] Cox, D.R., Miller, H.D., *The Theory of Stochastic Processes.*

[4] Hochmair, H., Hartwig, Benjamin, R.,Adam, *Point Pattern Analysis: Spatial Process.*

[5] Dharmaraj, S., Dr., *Stochastic Processes (NPTEL)*.

[6] Kleiner, L., Hallie, *Application of Renewal Theory to Pattern Analysis.*