STUDY OF DIFFERENT CONCEPTS OF RENEWAL THEORY, POINT PATTERN ANALYSIS AND COMPARISON OF THE DISTRIBUTIONS OF 2 POINT DATASETS

Submitted by

PRIYABRATA MISHRA

Birla Institute of Technology, Mesra

Guided by

Dr. GAUTAM CHOUDHURY

CCNS, IASST GUWAHATI

Contents

1. Introduction to Renewal Theory

2. Introduction to Point Pattern Analysis

3. Comparison of the Distributions of 2 Point Datasets by using Different Methods

INTRODUCTION TO RENEWAL THEORY

STOCHASTIC PROCESS

- Collection of Random variables indexed by a variable t, which represents time.
- It can also be defined as a 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).
- 4 types of stochastic processes

1. Continuous Time Continuous Space
2. Continuous Time Discrete Space
3. Discrete-Time Continuous Space
4. Discrete Time Discrete Space

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.
- For any i, j $\in \Omega$ and n $\geq$ 0,

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

MARKOV PROCESS

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

RENEWAL PROCESS

A counting process, $\{N(t), t \geq 0\}$, is a renewal process if $\{X1, X2, X3, \ldots\ldots..\}$, the sequence of non-negative random variables of interarrival time are independent and identically distributed.

Ex – Poisson Process

RENEWAL FUNCTION

Renewal Function = M(t) = E(N(t))

$$= \sum_{n=1}^{\infty} P(N(t) \geq n)$$

$$= \sum_{n=1}^{\infty} P(Tn \leq n)$$

$$= \sum_{n=1}^{\infty} F_n(t) \text{ , for } t > 0$$

RENEWAL EQUATION

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

(Volterra Integral Equation of 2nd Kind)

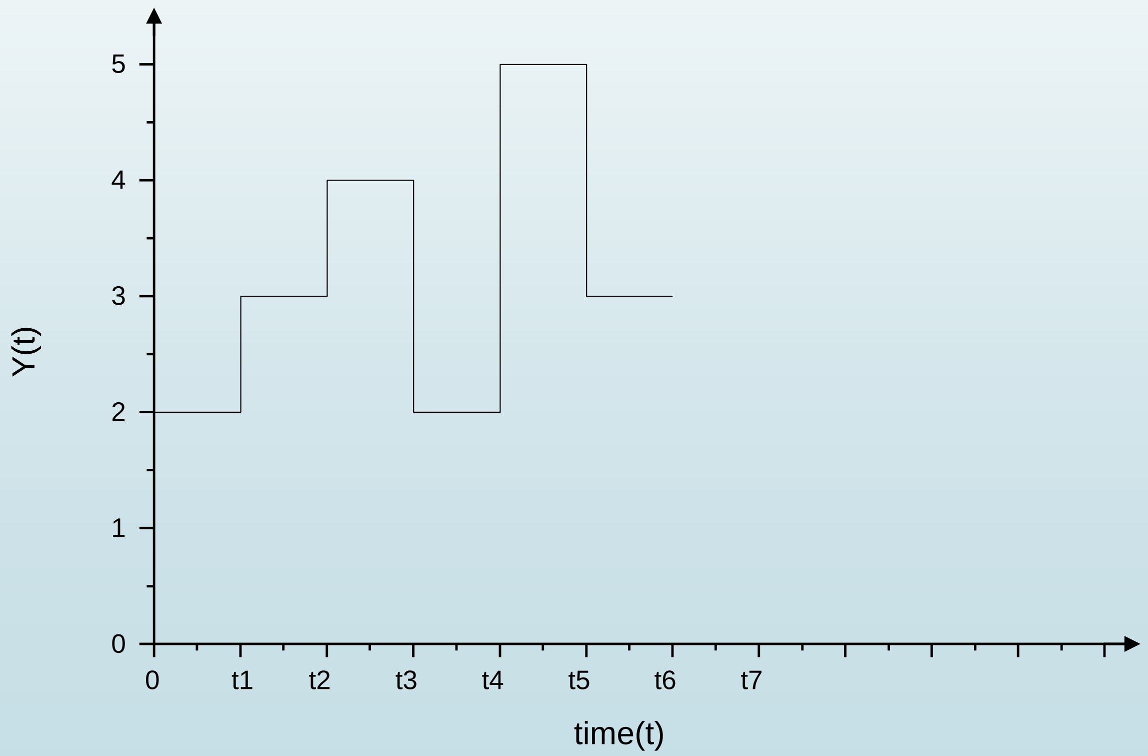
After taking Laplace Transform and solving it, we got

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

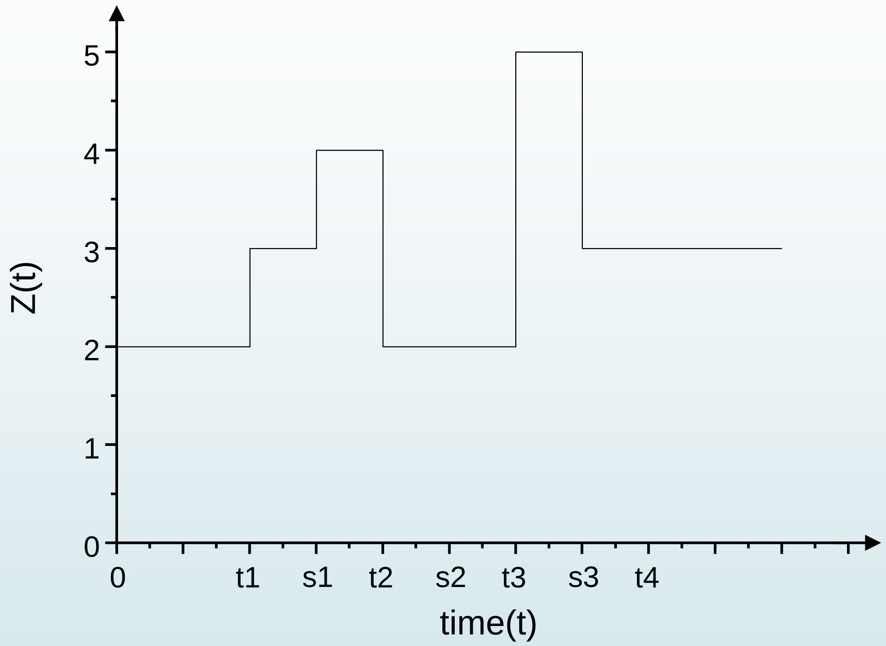
{By applying $F^*(s) = \frac{f*(s)}{s}$}

MARKOV RENEWAL PROCESS

- **Also known as Semi Markov Process.**
- **Time instants t_{n}, {n = 0,1,2,…} satisfy the Markov Property.**
- **If the Markov Property is satisfied everywhere, it is called Markov Process.**



Markov Regenerative Process(MRGP)



- t1,t2,t3,….. are the regeneration points.
- s1,s2,s3,… are the service completion times.

INTRODUCTION TO POINT PATTERN ANALYSIS

- Study of spatial arrangements in geographical space.
- 3 types: Deterministic – Outcomes are same every time.

  Stochastic – Outcomes are different.

  Independent Random Process – Independent of location.

COMPLETE SPATIAL RANDOMNESS (CSR)

- Point events occur within a given study area in a completely random fashion.
- Frequency distribution of CSR satisfies the Poisson distribution.
- Standard case or Null Hypothesis.

NEAREST NEIGHBOUR METHOD

- Finds the distance between each point and its closest neighbour, then averages these distance.
- Observed Mean NN distance = $d_{obs}(mean) = \frac{\sum_{i=1}^{n} d_{min}(si)}{n}$
- Expected Mean NN distance = $d_{exp}(mean) = \frac{1}{2}\sqrt{\frac{Area}{n}}$
- Nearest neighbour index 'd' as a difference = $d_{obs}(mean) - d_{exp}(mean)$
- Nearest neighbour index 'r' as a ratio = $\frac{d_{obs}(mean)}{d_{exp}(mean)}$

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

G-FUNCTION

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

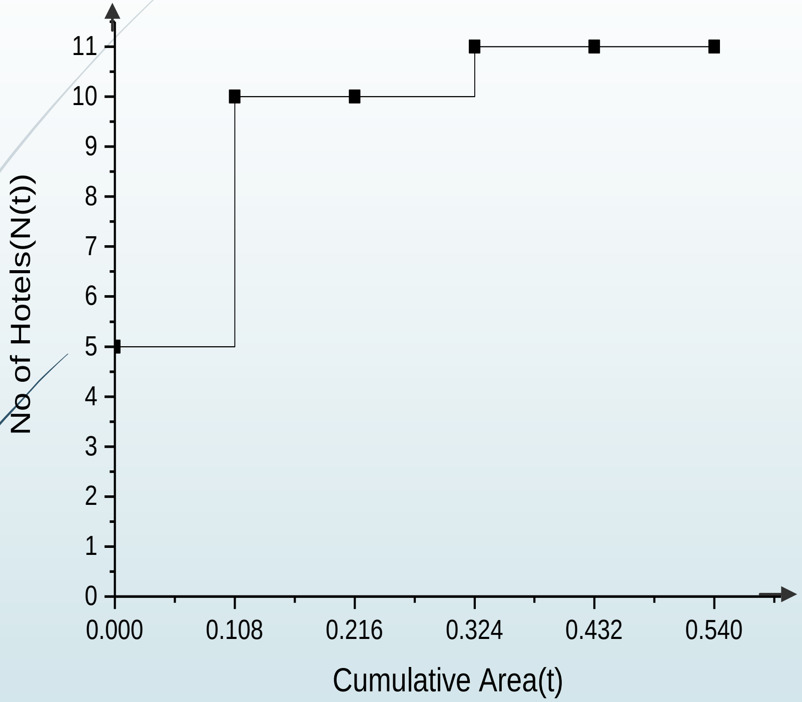
- For the clustered distribution g-function increases rapidly at short distances.
- For evenly spaced distribution it slowly increases up to event spacing.

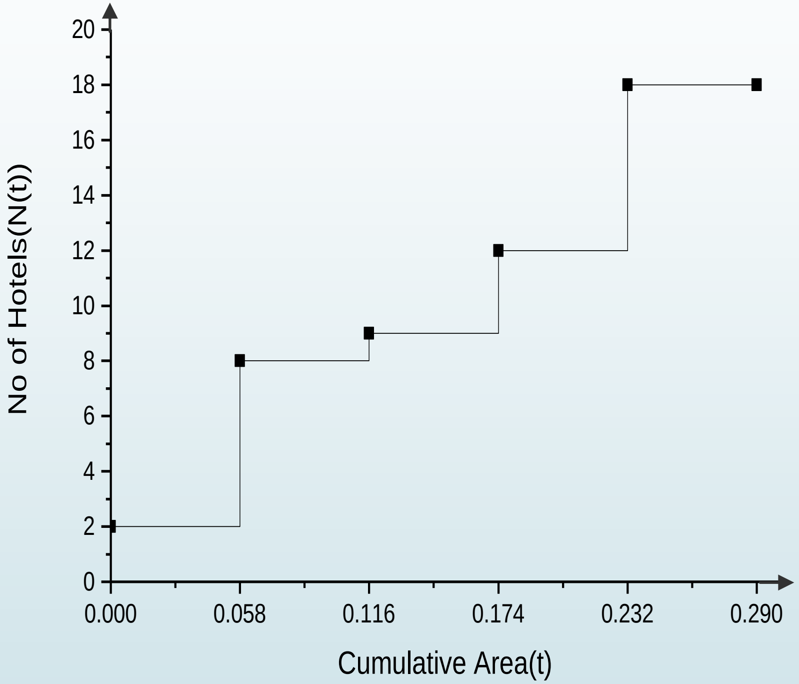
USED FORMULAE

- Mean = $\mu = x \,.\, P(X = x)$
- Variance = $Var(X) = (x - \mu)^2 \,.P(X = x)$
- Standard Deviation = $SD(X) = \sqrt{Var(X)}$
- Coefficient of Variation = $\frac{\text{Standard Deviation}}{\text{Mean}}$

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

COMPARISON OF POINT DATASETS





(SAMPLE PATHS FOR BOTH THE LOCATIONS)

- Cumulative Areas are used as the regeneration points.

LOCATION – 1

Mean = 1.512

Variance = 5.816

Standard Deviation = 2.412

Coefficient of Variation = 1.595

Standard Deviation > Mean

NN Index as Difference = -0.0226 (clustered)

LOCATION – 2

Mean = 0.467

Variance = 0.184

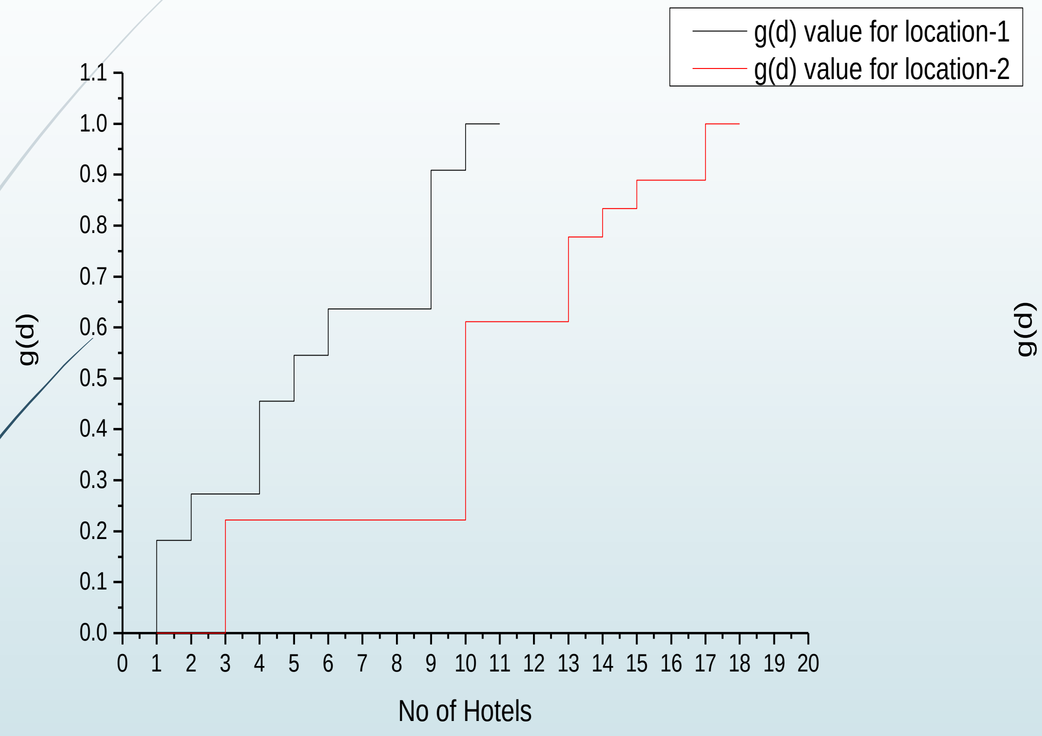
Standard Deviation = 0.429

Coefficient of Variation = 0.919

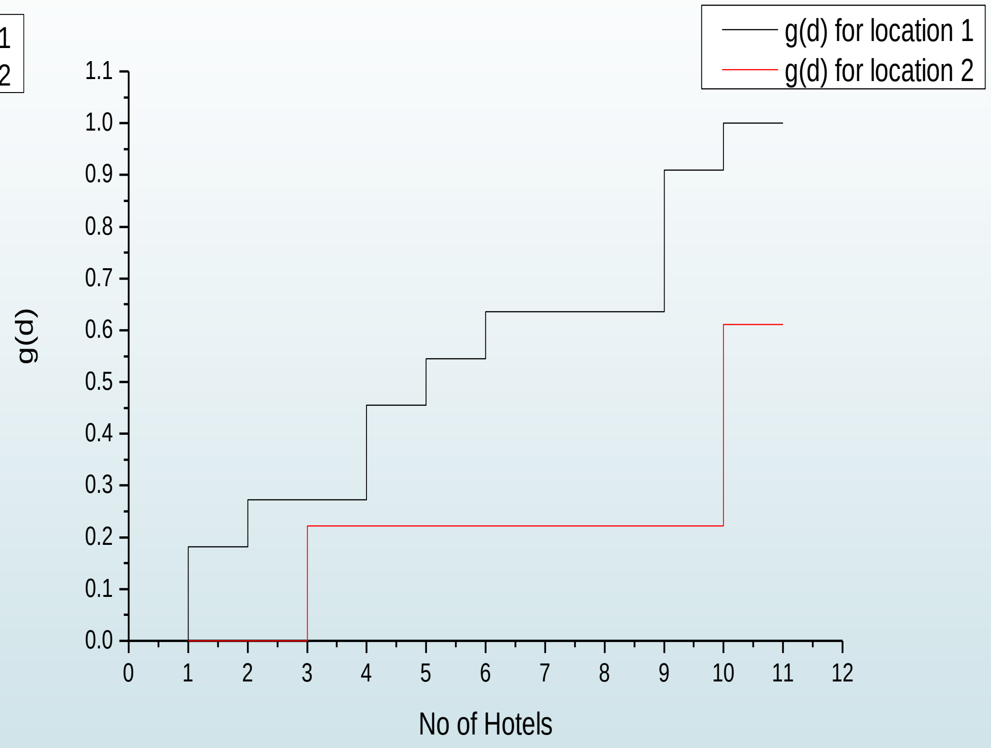
Standard Deviation < Mean

NN Index as Difference = -0.0052 (clustered)

g – Function Comparison



Comparison of g(d) values



Comparison of g(d) values of first 11 points

CONCLUSION

- Location – 2 has a better distribution of hotels.
- Not concentrated in only one place which helps people find them easily.
- In real life these methods can be applicable in analysing the locations of shopping malls and other buildings, distribution of natural resources in a particular area, to identify individual stands of specific types of trees, the location of diseased trees in forests etc.
- We are currently working on a paper showing the application of Renewal Theory on Point Pattern Analysis. We are just mentioning it as have not got all the results.

Thank You …