1-1-2007

A method for estimating intensity of a Poisson process

Sandhya Gunti
University of Nevada, Las Vegas

A METHOD FOR ESTIMATING INTENSITY OF A POISSON PROCESS

by

Sandhya Gunti

Bachelor of Engineering
Osmania University, Andhra Pradesh, India
April 2003

A thesis submitted in partial fulfillment
of the requirements for the

Master of Science Degree in Mathematical Sciences
Department of Mathematical Sciences
College of Sciences

Graduate College
University of Nevada, Las Vegas
May 2007

UNLV
UNIVERSITY OF NEVADA LAS VEGAS

Thesis Approval
The Graduate College
University of Nevada, Las Vegas

March 2, 2007

The Thesis prepared by

Sandhya Gunti

Entitled

A Method For Estimating Intensity Of A Poisson Process

is approved in partial fulfillment of the requirements for the degree of

Master of Science in Mathematical Sciences

Examination Committee Chair

Dean of the Graduate College

Examination Committee Member

Examination Committee Member

Graduate College Faculty Representative

1017-53

ABSTRACT

A Method for Estimating Intensity of a Poisson Process

by

Sandhya Gunti

Dr. Chih-Hsiang Ho, Examination Committee Chair
Professor of Mathematical Sciences
University of Nevada, Las Vegas

Motivated by its vast applications, we investigate ways to estimate the intensity of a Poisson process. Much of the work on modeling and analysis of repairable systems is based on the assumption of a special type of nonhomogeneous Poisson process (NHPP) known as Weibull process or Power-law process. In this thesis, we link the traditional homogeneous and nonhomogeneous Poisson processes to the classical time series via a sequence of the empirical recurrence rates (ERR), calculated at equally spaced intervals of time. We consider a computationally simple algorithm to calculate the total area and also the area for the last ten recurrence rates under the ERR curve. We conclude that the mean function of an NHPP can be estimated from the ERR values. In addition, we argue by simulation, that the algorithm can be implemented to forecast NHPP observations with various forms of intensity function. A correction factor is defined based on the overall trend of the targeted point process.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to dedicate this thesis to my parents, my husband, Goverdhan Gajjala, and friends whose love, support and understanding have always motivated me to strive for excellence.

I would like to sincerely and wholeheartedly thank Dr. Chih-Hsiang Ho for his guidance and kindness throughout this work. His patience as an advisor, boundless energy while teaching, promptness while reviewing all my writing, and passion for research are to be commended and worth emulating. I owe most of this work to him.

My deeply indebtedness also give to respectable committee members, Dr. Ananda, Dr. Cho and Dr. Qian, for their positive inputs and mentoring during my graduate studies.

CHAPTER 1

INTRODUCTION

Reliability is the ability of a system or component to perform its required functions under stated conditions for a specified period of time and it plays a key role in developing quality products and in enhancing competitiveness. Quality is a snapshot at the start of life and reliability is a motion picture of the day-by-day operation. Time zero defects are manufacturing mistakes that escaped final test. The additional defects that appear over time are "reliability defects" or reliability fallout. Much of the theory of reliability deals with nonrepairable systems or devices and it emphasizes the study of life time models. The main distinction between nonrepairable and repairable systems is that the former can fail only once, and Weibull distribution serves as a life time model for it where as the later is one which can be repaired and can be placed back in service. A repairable system is often modeled by a point process.

1.1 Basic Theory of Point Process

A point process is a stochastic model that describes the occurrences of events in time. These occurrences are thought as points on the time axis. In general the times between failures are neither independent nor identically distributed. Let $X(t)$ be a random variable that denotes the number of failures in the interval $(0,t]$ and X is called the counting random variable.

The probability that a unit survives beyond time t_0 is called the reliability at time t_0, and the reliability function, is defined as

$$R(t_0) = P[T > t_0] = 1 - F(t_0),$$

where $F(t)$ gives the probability that a randomly selected unit will fail by time t.

In biomedical applications the term "survival function" is also used. The probability density function (pdf) is defined to be the derivative of the cdf, provided the derivative exits. That is,

$$f(x) = \frac{d}{dx} F(x) = -\frac{d}{dx} S(x)$$

Life testing model can be characterized in terms of a number of different concepts. The hazard function (HF) is defined by

$$h(t) = \frac{f(t)}{1 - F(t)}$$

In actuarial science $h(t)$ is known as the "force of mortality," and in extreme-value theory $h(t)$ is called the "intensity function." This concept is often referred to as the "failure rate."

1.1.1 Mean function of a point process

The mean function of a point process is defined to be the expectation

$$m(t) = E(X(t))$$

Thus $m(t)$ is the expected number of failures through time t, and so it must be a non-decreasing function. Because $X(t)$ is a non-decreasing step function, it is clear that the mean function must be non-decreasing.

1.1.2 Rate of occurrence of failures (ROCOF)

When $m(t)$ is differentiable we define ROCOF as

$$\rho(t) = \frac{d}{dt} m(t)$$

This can be interpreted as the instantaneous rate of change in the expected number of failures.

1.1.3 Intensity function

Let $X(t)$ denote the number of occurrences in the interval $[0,t]$, and $P_n(t) = P$ [n occurrences in an interval $(0,t)$]. The intensity function of a point process (Rigdon and Basu, 2000) is

$$\lambda(t) = \lim_{h \to 0} \frac{P[X(t+h) - X(t) \geq 1]}{h}$$

Roughly speaking, the intensity function is the probability of failure in a small interval divided by the length of the interval. Thus, there will be many failures over intervals on which $\lambda(t)$ is large, and fewer failures over intervals on which $\lambda(t)$ is small. It is instructive to compare the definitions of the hazard function and intensity function. The hazard function is the limit of a conditional probability that the one and only one failure will occur in a small interval, divided by the length of the interval. This probability is conditioned on survival to the beginning of the interval. The intensity function is the unconditional probability of a failure in a small interval divided by the length of the interval (Rigdon and Basu, 2000). Clearly the intensity and ROCOF functions are the same provided that simultaneous failures cannot occur, i.e., when the mean function $m(t)$ is not discontinuous.

1.2 Poisson Process

A point process $X(t)$ is said to be a Poisson process if

1. $X(0) = 0$

2. $\{X(t), t \geq 0\}$ has independent increments.

3. $P[X(t+h) - X(t) = 1] = \lambda(t)h + o(h)$, where the function, λ is called the intensity function of the Poisson Process.

4. $P[X(t+h) - X(t) \geq 2] = o(h)$

1.2.1 Homogeneous Poisson Process (HPP)

The HPP is a Poisson process with constant intensity function. In terms of a repairable system, this implies that the system is neither improving nor wearing out with age, but rather is maintaining a constant intensity of failure.

1.2.2 Nonhomogeneous Poisson Process (NHPP)

In this thesis we focus on arrival (counting) processes, and more particularly, arrival processes that can be classified as nonstationary point processes. For such processes we are able to observe each arrival time exactly, and in general the arrival intensity (rate) changes over time.

Then $P(X(t) = n) = \dfrac{e^{-m(t)}(m(t))^n}{n!}$ $, n \geq 0$, where $m(t) = \int_0^t \lambda(s)ds$.

Under certain assumptions a nonstationary arrival process can be represented as NHPP (Cinlar, 1975). Using NHPPs, we can accurately represent a large class of arrival processes encountered in practice.

1.2.3 Weibull process

The special type of NHPP known as a Weibull process which is also known in the literature as a Power Law process. The name Weibull process derives primarily from the resemblance of the intensity function of the process to the hazard function of a Weibull distribution. In particular the intensity function has the form

$$\lambda(t) = \left(\frac{\beta}{\theta}\right)\left(\frac{t}{\theta}\right)^{\beta-1}$$

The mean value function of a Weibull function of a Weibull process has the form

$$m(t) = E[X(t)] = \left(t/\theta\right)^{\beta}$$

with scale parameter $\theta > 0$ and shape parameter $\beta > 0$.

1.3 Objective of the Thesis

Estimating the intensity of the failures is the important task which is the main objective of the thesis. To begin with, we try to establish the relation between HPP and NHPP, and the statistical inferences of Power law Process as shown in chapter 2. From chapter 3, we try to get the mean function from the area under the ERR plot by defining a correction factor. From the results in chapter 4, it will be proved that the correction factor can be used improve predicting the mean function of the coming failures and it will be applied to the mining accidents data in the chapter 5. The algorithms used for simulation are given in chapter 7.

CHAPTER 2

POWER-LAW PROCESS

The Power law process has the intensity function of the form

$$\lambda(t) = \left(\frac{\beta}{\theta}\right)\left(\frac{t}{\theta}\right)^{\beta-1}$$

The β parameter affects how the system deteriorates or improves over time. If $\beta > 1$ then the intensity function $\lambda(t)$ is increasing, and the failures tend to occur more frequently. If $\beta < 1$, then $\lambda(t)$ is decreasing, and the system is improving. Finally if $\beta = 1$, then the power law process reduces to the simpler HPP with intensity $1/\theta$. There are several reasons why the power law process is widely used model for repairable systems. In the coming sections we link the traditional HPP with Weibull process and finally prove that Weibull process is a special case of NHPP.

2.1 Analysis of NHPP

2.1.1 Joint pdf of HPP with intensity $\lambda = 1$

Theorem 2.1: Let the random variables $Z_1 < Z_2 < ... < Z_n$ are distributed as the first n successive times of an HPP with intensity $\lambda = 1$

If we let $Z_0 = 0$, it follows from basic properties of an HPP that the differences $Z_i - Z_{i-1}$ for $i = 1, ..., n$ are independent exponential random variables with mean $1/\lambda = 1$.

It follows that the joint probability density function of $Z_1, Z_2, ..., Z_n$ is,

$$f(Z_1, Z_2, ..., Z_n) = \exp(-z_n), \text{ if } 0 < Z_1 < Z_2 < ... < Z_n < \infty.$$

Proof: Let $Z_1, Z_2, ..., Z_n$ denote the failure times in an HPP, and let $X_i = Z_i - Z_{i-1}$ (where $Z_0 = 0$) be the times between failures. Deriving the joint pdf of $Z_1, Z_2, ..., Z_n$ by using the following relationship

$$f(z_1, z_2, z_3, ..., z_n) = f_1(z_1) f_2(z_2 \mid z_1) f_3(z_3 \mid z_1, z_2) ... f_n(z_n \mid z_1, z_2, ..., z_{n-1})$$

where $0 < z_1 < z_2 < ... < z_n$. The survival function for Z_1 is

$$\begin{aligned} S(z_1) &= P(Z_1 > z_1) \\ &= P(N(0, z_1] = 0) \\ &= \exp\left(-\int_0^{z_1} \lambda dx\right) \\ &= \exp(-\lambda z_1), \quad z_1 > 0 \end{aligned}$$

The pdf of Z_1 is thus

$$\begin{aligned} f_1(z_1) &= -S'_1(z_1) \\ &= \lambda \exp(-\lambda z_1), \quad z_1 > 0 \end{aligned}$$

The conditional survival function of Z_2 given $Z_1 = z_1$ is

$$\begin{aligned} S_2(z_2 \mid z_1) &= P(Z_2 > z_2 \mid z_1) \\ &= P(N(z_1, z_2] = 0) \\ &= \exp\left(-\int_{z_1}^{z_2} \lambda dx\right) \\ &= \exp[\lambda(z_2 - z_1)], \quad z_2 > z_1 > 0 \end{aligned}$$

The conditional pdf of Z_2 given $Z_1 = z_1$ is thus

$$f_2(z_2 \mid z_1) = -\frac{d}{dz_2} S_2(z_2 \mid z_1)$$
$$= \lambda \exp[-\lambda(z_2 - z_1)], \quad z_2 > z_1 > 0$$

Because of the independent increments property, the conditional pdf of Z_3 given $Z_1 = z_1$ and $Z_2 = z_2$ is independent of z_1. The conditional survival function for Z_3 is therefore

$$S_3(z_3 \mid z_1, z_2) = P(Z_3 > z_3 \mid Z_1 = z_1, Z_2 = z_2)$$
$$= P(Z_3 > z_3 \mid Z_2 = z_2)$$
$$= P(N(z_2, z_3] = 0)$$
$$= \exp\left(-\int_{z_2}^{z_3} \lambda dx\right)$$
$$= \exp[-\lambda(z_3 - z_2)], \; z_3 > z_2 > 0$$

In general, we have

$$f_n(z_n \mid z_1, z_2, ..., z_{n-1}) = \lambda \exp[-\lambda(z_n - z_{n-1})], \quad z_n > z_{n-1} > 0$$

The joint pdf of $Z_1, Z_2, ..., Z_n$ is thus

$$f(z_1, z_2, z_3 .... z_n) = [\lambda \exp(-\lambda z_1)]\{\lambda \exp[-\lambda(z_2 - z_1)]\}\{\lambda \exp[-\lambda(z_3 - z_2)]\}$$
$$\times ... \times \{\lambda \exp[-\lambda(z_n - z_{n-1})]\}$$
$$= \lambda^n \exp(-\lambda z_n), \qquad z_n > z_{n-1} > ... > z_2 > z_1 > 0$$

In our case $\lambda = 1$, so

$$f(Z_1, Z_2, ..., Z_n) = \exp(-z_n).$$

2.1.2 Joint distribution of Weibull process using HPP

Theorem 2.2: Suppose $m(t)$ is continuous. If $Z_i = m(T_i)$ for $i = 1,...,n,$ then the random variables $Z_1 < Z_2 < ... < Z_n$ are distributed as the first n successive times of an HPP with intensity $\lambda = 1$

In the case of a Weibull process, $m(t) = \left(t/\theta\right)^{\beta}$, which means that $T_i = Z_i^{1/\beta}\theta$.

This transformation yields the joint probability density function of $T_1, T_2, ..., T_n$, namely

$$f(t_1, t_2, ...t_n) = \left(\frac{\beta}{\theta}\right)^n \prod_{i=1}^{n} \left(\frac{t_i}{\theta}\right)^{\beta-1} e^{-\left[t_n/\theta\right]^{\beta}} \text{ for } 0 < t_1 < t_2 < ...t_n < \infty.$$

Proof: Given $m(t) = \left(t/\theta\right)^{\beta}$ is continuous

Let $T_i = u(Z_i)$ & $Z_i = m(T_i)$

Therefore $Z_i = \left(T_i/\theta\right)^{\beta}$

$$\Rightarrow T_i = Z_i^{1/\beta}\theta$$

Let us derive the joint using distribution using Jacobian under transformation

Therefore the joint pdf of $T_1, T_2, ..., T_n = f_Z(m(t_n)).|J| ...$ **(1)**

Where $J = \begin{vmatrix} \frac{\partial z_1}{\partial t_1} & \frac{\partial z_1}{\partial t_2} \cdots & \frac{\partial z_1}{\partial t_n} \\ \frac{\partial z_2}{\partial t_1} & \ddots & \vdots \\ \vdots & \cdots & \frac{\partial z_n}{\partial t_n} \end{vmatrix}$ is the Jacobian

When $i = j$, $\frac{\partial z_i}{\partial t_i} = \frac{\partial\left(t_i/\theta\right)^{\beta}}{\partial t_i} = \beta\left(\frac{t_i}{\theta}\right)^{\beta-1} \cdot \frac{1}{\theta}$

$$i \neq j, \quad \frac{\partial z_i}{\partial t_j} = \frac{\partial \left({}^{t_i}\!/_{\theta} \right)^{\beta}}{\partial t_j} = 0;$$

Therefore

$$J = \begin{vmatrix} \left(\beta\left(\frac{t_1}{\theta}\right)^{\beta-1}\cdot\frac{1}{\theta}\right) & 0 & 0 & \cdots & \cdots \\ 0 & \left(\beta\left(\frac{t_2}{\theta}\right)^{\beta-1}\cdot\frac{1}{\theta}\right) & 0 & & \\ 0 & & \ddots & & \\ \vdots & & & \ddots & 0 \\ \vdots & & & 0 & \left(\beta\left(\frac{t_n}{\theta}\right)^{\beta-1}\cdot\frac{1}{\theta}\right) \end{vmatrix}$$

$$= \beta\left(\frac{t_1}{\theta}\right)^{\beta-1}\frac{1}{\theta}.\beta\left(\frac{t_2}{\theta}\right)^{\beta-1}\frac{1}{\theta}\ldots\beta\left(\frac{t_n}{\theta}\right)^{\beta-1}\cdot\frac{1}{\theta}$$

$$= \left(\frac{\beta}{\theta}\right)^{n}\prod_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta-1}$$

Therefore from (1)

$$f_{t_1,t_2,\ldots,t_n} = f_Z(m(t_n)).\left(\frac{\beta}{\theta}\right)^{n}\prod_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta-1}$$

From theorem 2.1, $f(z_1, z_2, \ldots, z_n) = \exp^{-(z_n)}$

$$m(t_n) = \left({}^{t_n}\!/_{\theta} \right)^{\beta}$$

Substituting these in the equation for joint pdf:

$$f_{t_1,t_2\ldots,t_n} = f_{z_1,z_2,\ldots,z_n}\left(\left({}^{t_n}\!/_{\theta} \right)^{\beta}\right).\left(\frac{\beta}{\theta}\right)^{n}\prod_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta-1}$$

$$= \left(\frac{\beta}{\theta}\right)^{n}\prod_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta-1} e^{-\left[{}^{t_n}\!/_{\theta}\right]^{\beta}} \ldots \qquad \textbf{(2)}$$

This is the joint distribution of a Weibull process with intensity $\lambda(t) = \frac{\beta}{\theta}\left(\frac{t_i}{\theta}\right)^{\beta-1}$. The following theorem shows that the joint distribution of Weibull process is indeed an NHPP.

2.1.3 Relating Weibull process with NHPP:

Theorem 2.3: If the joint distribution of Weibull process is

$f(t_1,t_2,..t_n) = \left(\frac{\beta}{\theta}\right)^n \prod_{i=1}^{n}\left(\frac{t_i}{\theta}\right)^{\beta-1} e^{-\left[t_n/\theta\right]^\beta}$ for $0 < t_1 < t_2 < ..t_n < \infty$, then it represents a NHPP with intensity function $\lambda(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1}$

Proof: The joint pdf formula of the failure times $T_1, T_2, ..., T_n$ from an NHPP is derived by using the formula

$f(t_1,t_2,t_3,...,t_n) = f_1(t_1)f_2(t_2 \mid t_1)f_3(t_3 \mid t_1,t_2)...f_n(t_n \mid t_1,t_2,...,t_{n-1})$, where $\lambda(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1}$ is the intensity of the NHPP as stated in the theorem.

For $0 < t_1 < t_2 < ... < t_n$, the survival function for T_1 is

$$S(t_1) = P(T_1 > t_1)$$

$$= P(N(0,t_1] = 0)$$

$$= \exp\left(-\int_0^{t_1} \lambda(x)dx\right),\ t_1 > 0$$

The pdf of T_1 is thus

$$f_1(t_1) = -s'_1(t_1)$$

$$= -\frac{d}{dt_1}\exp\left(-\int_0^{t_1} \lambda(x)dx\right)$$

$$= \lambda(t_1)\exp\left(-\int_0^{t_1} \lambda(x)dx\right)$$

The conditional survival function of T_2 given $T_1 = t_1$ is

$$S_2(t_2 \mid t_1) = P(T_2 > t_2 \mid t_1)$$

$$= P(N(t_1, t_2] = 0)$$

$$= \exp\left(-\int_{t_1}^{t_2} \lambda(x)dx\right), \quad t_2 > t_1 > 0$$

The conditional pdf of T_2 given $T_1 = t_1$ is thus

$$f_2(t_2 \mid t_1) = -\frac{d}{dt_2} S_2(t_2 \mid t_1)$$

$$= \lambda(t_2)\exp\left(-\int_{t_1}^{t_2} \lambda(x)dx\right), t_2 > t_1 > 0$$

We conclude that survival function of T_k given $T_1 = t_1, T_2 = t_2, ..., T_{k-2} = t_{k-2}, T_{k-1} = t_{k-1}$ is

$$S_k(t_k \mid t_1, t_2, ..., t_{k-1}) = S(t_k \mid t_{k-1})$$

$$= P(T_k > t_k \mid T_{k-1} = t_{k-1})$$

$$= P(N(t_{k-1}, t_k] = 0)$$

$$= \exp\left(-\int_{t_{k-1}}^{t_k} \lambda(x)dx\right), t_k > t_{k-1}$$

Thus

$$f(t_k \mid t_{k-1}) = \lambda(t_k)\exp\left(-\int_{t_{k-1}}^{t_k} \lambda(x)dx\right), t_k > t_{k-1}.$$

The joint pdf of $T_1, T_2, ..., T_n$ is therefore

$$f(t_1, t_2, t_3, ..., t_n) = \left[\lambda(t_1)\exp\left(-\int_0^{t_1} \lambda(x)dx\right)\right]\left[\lambda(t_2)\exp\left(-\int_{t_1}^{t_2} \lambda(x)dx\right)\right] \times ...$$

$$\times\left[\lambda(t_n)\exp\left(-\int_{t_{n-1}}^{t_n} \lambda(x)dx\right)\right]$$

$$= \left(\prod_{i=1}^{n} \lambda(t_i) \right) \exp\left(- \int_0^{t_n} \lambda(x) dx \right), t_n > t_{n-1} > \ldots > t_2 > t_1 > 0$$

Therefore, for $\lambda(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1}$

$$f(t_1, t_2, \ldots, t_n) = \left(\prod_{i=1}^{n} \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1} \right) \exp\left(- \int_0^{t_n} \frac{\beta}{\theta}\left(\frac{x}{\theta}\right)^{\beta-1} dx \right)$$

$$= \left(\frac{\beta}{\theta}\right)^n \prod_{i=1}^{n} \left(\frac{t_i}{\theta}\right)^{\beta-1} e^{-\left[t_n/\theta\right]^{\beta}}$$

Therefore the joint distribution of Weibull process represents an NHPP with intensity function $\lambda(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1}$

2.1.4 Relating the theorems to the simulation

In order to generate NHPP, first we generate an exponential distribution with mean $1/\lambda = 1$, where λ is the mean of HPP. From theorem 2.1 it is evident that HPP is formed by the sum of variables from exponential distribution. Then by using the transformation stated in theorem 2.2 we generate the Power Law process for $\beta = 0.5$, 1, 2, and for increasing and decreasing step intensities to be described later. The NHPP generated is used for further analysis. Stated below are basic inferences of NHPP which can be used to judge a given data.

2.2 Statistical Inferences

2.2.1 Maximum Likelihood Estimators (MLE) of NHPP

Suppose that a repairable system is observed until n failures occur, so we observe the failure times $0 < t_1 < t_2 < ... < t_n$, so the joint pdf of a failure truncated NHPP as

$$f(t_1, t_2, ..., t_n) == \frac{\beta^n}{\theta^{n\beta}} \left(\prod_i^n t_i \right)^{\beta - 1} \exp\left[-\left(\frac{t_n}{\theta} \right)^{\beta} \right]$$

To get the maximum likelihood estimator's, we take the logarithm of this joint density and set the first partial derivatives (with respect to θ and β) equal to zero. The log-likelihood function is

$$l(\theta, \beta | t) = n \ln \beta - n\beta \ln \theta + (\beta - 1) \sum_{i=1}^{n} \ln t_i - \left(\frac{t_n}{\theta} \right)^{\beta} \quad \text{and}$$

$$0 = \frac{\partial l}{\partial \theta} = -\frac{n\beta}{\theta} + \frac{\beta}{\theta} \left(\frac{t_n}{\theta} \right)^{\beta}$$

$$0 = \frac{\partial l}{\partial \beta} = \frac{n}{\beta} - n \ln \theta + \sum_{i=1}^{n} \ln t_i - \left(\frac{t_n}{\theta} \right)^{\beta} \ln \left(\frac{t_n}{\theta} \right)$$

The first equation simplifies to $0 = -n + \left(\frac{t_n}{\theta} \right)^{\beta}$

which can be solved for θ (in terms of β) to obtain

$$\hat{\theta} = t_n / n^{1/\hat{\beta}}$$

Substituting back into the first equation yields

$$0 = \frac{\partial l}{\partial \beta} = \frac{n}{\beta} - n \ln \frac{t_n}{n^{1/\beta}} + \sum_{i=1}^{n} \ln t_i - \left(\frac{t_n n^{1/\beta}}{t_n} \right)^{\beta} \ln \left(\frac{t_n n^{1/\beta}}{t_n} \right)$$

Solving for β yields

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n-1} \ln\left(t_n/t_i\right)}.$$

2.2.2 Deriving the test statistic for NHPP

Theorem 2.4: $\dfrac{2n\beta_0}{\hat{\beta}} = 2n\beta_0 \left(\dfrac{n}{\sum_{i=1}^{n-1} \ln(t_n/t_i)} \right)^{-1} = 2\beta_0 \sum_{i=1}^{n-1} \ln(t_n/t_i)$ has a chi-squared distribution with $2n-2$ degrees of freedom (e.g., Crow, 1974, 1982; Rigdon and Basu, 2000).

Proof: We can write the expression $2n\beta / \hat{\beta}$ as

$$\frac{2n\beta}{\hat{\beta}} = 2\beta \sum_{i=1}^{n} \log(t_n / t_i)$$

We know that conditioned on $T_n = t_n$ the random variables $T_1 < T_2 < \ldots < T_{n-1}$ are distributed as n-1 order statistics from the distribution with cdf

$$G(y) = \begin{cases} 0, y \le 0 \\ m(y)/m(t_n), 0 < y < t_n \\ 1, y \ge t_n. \end{cases}$$

The proof of this theorem is provided in Statistical methods (Rigdon Steven. E., Basu Asit. P., 2000)

For the power law process, we have $m(t) = \left(t/\theta \right)^{\beta}$, so

$$\frac{m(y)}{m(t_n)} = \frac{(y/\theta)^{\beta}}{(t_n/\theta)^{\beta}} = \left(\frac{y}{t_n} \right)^{\beta}.$$

In this case we have

$$G(y)=\begin{cases}0, y\le 0\\ (y/t_n)^{\beta}, 0<y<t_n\\ 1, y\ge t_n.\end{cases}$$

Letting Y be a random variable with cumulative distributive function G, we have on the one hand

$$\left(\frac{y}{t_n}\right)^{\beta}=G(y)=P(Y\le y)$$

for 0<y<t_n. On the other hand,

$$P(Y\le y)=P(Y/t_n\le y/t_n)=P((Y/t_n)^{\beta}\le (y/t_n)^{\beta}).$$

This implies that

$$P((Y/t_n)^{\beta}\le (y/t_n)^{\beta}) = (y/t_n)^{\beta}$$

for $0<y<t_n$ or $0<y/t_n<1$. This means that the random variable $(Y/t_n)^{\beta}$ has a uniform distribution over the interval (0, 1). Therefore the quantities $(T_i/t_n)^{\beta}$, i = 1, 2, ..., n-1 are distributed as n-1 order statistics from a uniform distribution on the interval (0,1). The sum

$$\sum_{i=1}^{n}-\log(t_i/t_n)^{\beta}=-\beta\sum_{i=1}^{n}\log(t_i/t_n)$$

is thus distributed as the sum of n-1 exponential random variables, each with mean 1. The proof of this statement is provided in Statistical Methods (Lemma 30, Rigdon Steven. E., Basu Asit. P., 2000). The sum of n-1 exponential random variables has a gamma distribution with parameters $k=n-1$ and θ = 1. Finally twice a gamma distribution with parameters n-1 and yields a chi-square distribution with 2(n-1) degrees of freedom.

Thus

$$-2\beta\sum_{i=1}^{n}\log(t_i/t_n)=\frac{2n\beta}{\hat{\beta}} \text{ has a } x^2(2(n-1)) \text{ distribution.}$$

Thus, a size α test of $H_0:\beta=\beta_0$ against $H_a:\beta\neq\beta_0$ is to reject H_0 if

$$2n\beta_0/\hat{\beta}\leq\chi^2_{\alpha/2}(2n-2) \text{ or } 2n\beta_0/\hat{\beta}\geq\chi^2_{1-\alpha/2}(2n-2),$$

where $\chi^2_{\alpha/2}(2n-2)$ is the $100\alpha/2$ percentile of chi-squared distribution with $2n-2$ degrees of freedom.

The above inferences about NHPP can be used to determine whether a particular data is a power law process and its further analysis can be done.

CHAPTER 3

EMPIRICAL RECURRENCE RATES TIME SERIES

A nonhomogeneous Poisson process is often suggested as an appropriate model when a system whose rate varies over time. If the process is waning or developing, the rate λ should be a monotonically decreasing or increasing function of t.

3.1 Relationship between Mean and the Intensity Functions of NHPP

Theorem 3.1: The nonhomogeneous Poisson process (NHPP) has a mean value function denoted by $m(t|\Theta)$, where Θ is a vector of parameters. The intensity function $\lambda(t|\Theta)$ is described as follows:

$$\lambda(t|\Theta) = \frac{d}{dt} m(t|\Theta)$$

Therefore the mean value function of a power law process is

$$m(t|\theta,\beta) = (t/\theta)^{\beta} = \int_0^t \lambda(s)ds$$

Proof: Let us consider the random variable $X = N(a,b]$ has a Poisson distribution with mean m, then its moment generating function is

$$M_t(s) = E(e^{sX}) = \exp[m_t(e^s - 1)]$$

Let $a < b$. Then by Statistical methods (Theorem 15, Rigdon Steven. E., Basu Asit. P., 2000)

$$N(a) \sim POI\left(\int_0^a \lambda(x)dx\right)$$

and $N(b) \sim POI\left(\int_0^b \lambda(x)dx\right)$. The moment generating functions of $N(a)$ and $N(b)$ are therefore

$$M_{N(a)}(s) = \exp\left[\left(\int_0^a \lambda(x)dx\right)\left(e^s - 1\right)\right]$$

and $$M_{N(b)}(s) = \exp\left[\left(\int_0^b \lambda(x)dx\right)\left(e^s - 1\right)\right]$$

By the independent increments property, the random variables $N(a)$ and $N(a,b]$ are independent. Using the result that the moment generating function of the sum of independent random variables is the product of their moment generating functions (Bain and Engelhardt, 1992), we have

$$M_{N(b)}(s) = M_{N(a)+N(a,b]}(s)$$

$$= M_{N(a)}(s) + M_{N(a,b]}(s)$$

Thus

$$\exp\left[\left(\int_0^b \lambda(x)dx\right)\left(e^s - 1\right)\right] = \exp\left[\left(\int_0^a \lambda(x)dx\right)\left(e^s - 1\right)\right] M_{N(a,b]}(s)$$

Solving for $M_{N(a,b]}(s)$ yields

$$M_{N(a,b]}(s) = \frac{\exp\left[\left(\int_0^b \lambda(x)dx\right)\left(e^s - 1\right)\right]}{\exp\left[\left(\int_0^a \lambda(x)dx\right)\left(e^s - 1\right)\right]}$$

$$= \exp\left[\left(\int_0^b \lambda(x)dx - \int_0^a \lambda(x)dx\right)\left(e^s - 1\right)\right]$$

$$=\exp\left[\left(\int_a^b \lambda(x)dx\right)\left(e^s-1\right)\right]$$

This is the moment generating function for a Poisson random variable with mean

$$m=\int_a^b \lambda(x)dx$$

So for a $N(0,t]$ the mean function is

$$m(t|\theta,\beta)=\int_0^t \lambda(s)ds$$

Recall that the intensity function λ is defined as probability of failure in small interval divided by the length of the interval, which motivates the following developments. First, we generate the Empirical Recurrence Rates of the NHPP and then calculate the area under it based on the number of failures per a particular span of time divided by that time span.

3.2 The Empirical Recurrence Rates (ERR)

Let $t_1,....,t_n$ of the NHPP be the n ordered failures during an observation period, $(0,T]$, where we recommend $T=h\left\{\left[\frac{t_n}{h}\right]+1\right\}$, and where $\left[\frac{t_n}{h}\right]$ is the largest integer less than or equal to $\frac{t_n}{h}$, from the first occurrence to the last occurrence. The time-step h can be varied accordingly, and the suggested time-step is the sample mean of the NHPP which is used in the current work. Then a discrete time series $\{y_l\}$ is generated sequentially at equidistant time intervals $h,2h,...,lh,...,Nh\ (=T)$. If 0 is adopted as the time-origin and h as the time–step, then we regard y_l as the observation at time $t=lh$.

Therefore, we propose a time series of the empirical recurrence rates as follows:

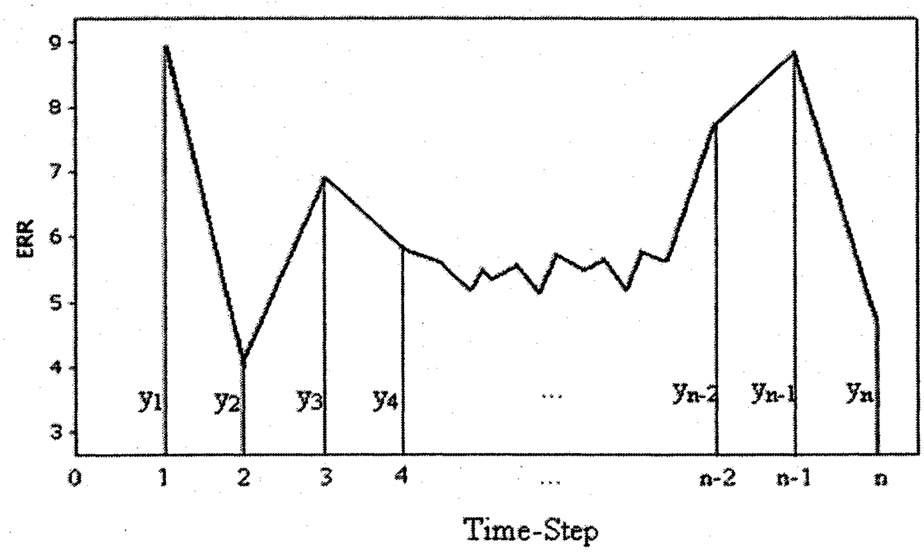
$$y_l = n_l / lh = \text{Total number of failures in}\,(0, lh)/lh,$$

where $l = 1, 2, \ldots, N$. Note that y_l evolves over time and it is simply the MLE for the mean rate of a simple Poisson process observed in $(0, lh)$. The time plots of the empirical recurrence rates can be obtained which are used to calculate the area as given below.

3.3 Area under ERR Plots

The formula for area under the ERR curve is derived by dividing the ERR plot into trapezoids as shown in fig. 3.1 and adding up the area of trapezoids.

Figure 3.1 Example of ERR Plot with Time Step h=1 for a Random Data Set



Area under ERR curve = sum of the areas of the trapezoids

$$= \frac{1}{2}h(y_1 + y_2) + \frac{1}{2}h(y_2 + y_3) + \ldots + \frac{1}{2}h(y_{n-2} + y_{n-1}) + \frac{1}{2}h(y_{n-1} + y_n)$$

$$=\frac{1}{2}h(y_1+2y_2+\ldots+2y_{n-1}+y_n)$$

$$=h\left(\frac{1}{2}y_1+y_2+\ldots+y_{n-1}+\frac{1}{2}y_n\right)$$

The above formula is used to calculate the total area under the ERR curve, which should be close to $m(t)$, the total number of events in $(0,T]$ for a HPP. A computationally simple algorithm designed to predict the mean function of an NHPP will be presented in chapter 7.

3.4 Methodology

The NHPP is generated as stated in 2.1.4 and the total area under the ERR curve is calculated after generating the ERR's as stated in the previous two sections. In order to predict the mean function of next few observations, which is the main goal of the thesis, we need to define a correction factor. The purpose of correction factor is as described below.

When we try to predict the mean function of the last ten ERR's, we find the area for the last nine time intervals and observe that it is not equal to the mean number of the last nine time-steps. The mean number is nothing but the number of events present in that nine time steps which is the value of mean function at a particular time period. So a correction factor is needed to be derived which could give the correct mean function for the last nine time-steps when we multiply the area with this factor. So based on the total area and true mean function of NHPP a correction factor is calculated using the formula defined below.

$$C.F = \frac{\text{True mean function of NHPP}}{\text{Total area under ERR curve}}$$

The true mean function is given from the Power law Process by the formula

$$\text{True mean function} = m(T|\theta) = (T/\theta)^{\beta} \text{ where } T = h\left\{\left[\frac{t_n}{h}\right]+1\right\}$$

And the formula for predicting mean function in the required time-steps is given by

Predicted mean function of 'n' time-steps= C.F $\times$ Area under the ERR curve for 'n' time-steps

Finally, the predicted mean function attained after using this correction is proved to be close to the actual mean number. So this correction factor can be used for predicting the mean function of the coming failures.

The simulation is done for decreasing $(\beta = 0.5)$, constant $(\beta = 1)$, increasing $(\beta = 2)$ intensities, and also for decreasing and increasing step intensities of NHPP. The events at which jumps occur for the step intensities are taken to be at one-third, one-half and two-thirds into the process (with respect to n, the total number of failures observed). So we consider τ, the jump points as 33, 51, 66. In case of the step intensity since β is not specified, the formula for the transformation as stated in theorem 2.2 can not be used to get NHPP from the given HPP, and also the true mean function cannot be calculated as stated above. So the other way of calculating them is as shown below:

For increasing step intensity $\lambda(t) = 1$, for $0 \le t \le \tau$ and $\lambda(t) = 3$, for $\tau < t \le \infty$. And from theorem 3.1 and theorem 2.2 $Z_i = m(T_i) = \int_0^t \lambda(s)ds$

So we generate NHPP as follows, after solving the above equation for T_i:

$$T_i = Z_i, \qquad \text{for } 0 \le t \le \tau$$

$$T_i = \frac{Z_i + 2\tau}{3}, \text{ for } \tau < t \le \infty$$

True mean function for increasing intensity is

$$m(t) = 3T - 2\tau \text{ where } T = h\left\{\left[\frac{t_n}{h}\right] + 1\right\}$$

For decreasing step intensity $\lambda(t) = 3$, for $0 \le t \le \tau$ and $\lambda(t) = 1$, for $\tau < t \le \infty$

So we generate NHPP as follows:

$$T_i = Z_i / 3, \quad \text{for } 0 \le t \le \tau$$

$$T_i = Z_i - 2\tau, \text{ for } \tau < t \le \infty$$

And since for a decreasing step intensity the failures when $0 \le t \le \tau$ will be thrice that of $\tau < t \le \infty$, we take the jump points at $\tau = 33/3, 51/3, 66/3$.

True mean function for decreasing intensity is

$$m(t) = T + 2\tau \text{ where } T = h\left\{\left[\frac{t_n}{h}\right] + 1\right\}$$

The average values for 10000 iterations are considered for better results. And the results as described above are attained, are shown in the coming chapter.

CHAPTER 4

RESULTS

Suppose $T_1,...,T_n$ denote the first n successive times of occurrence of an NHPP, and let $m(t)$ denote the mean function of the process. It is well known that if $Z_j = m(T_j) = \int_0^{T_j} \lambda(s)ds$, for $j = 1,...,n$, then the random variables $Z_1 < Z_2 < ... < Z_n$ are distributed as the first n successive occurrence times of an HPP with intensity $\lambda = 1$. Therefore, the standard method of simulating an NHPP with the failure-truncated sampling is to simulate an HPP with $\lambda = 1$, and then to do appropriate time transformation to get an NHPP realization. The area and the correction factor are calculated in the simulation.

The simulations are done for 10000 iterations and the average value of the required values is tabulated as given below:

4.1 Summary for $\beta = 0.5, 1, 2$

For n=100(total number of failures) and N=10^4 (number of iterations)

Table 4.1 Results for NHPP with $\beta = 0.5, 1, 2$

β	*0.5*	*1*	*2*
Average total area	179.9033	98.74559	50.46421
Average area for last 9 ERR's	9.076592	8.910172	8.591306
Average $m(t) = (T/\theta)^{\beta}$ (From Power law Process)	100.5130	101.9594	101.9217
Average $m(t) = (T/\hat{\theta})^{\hat{\beta}}$ (Expected mean function)	100.5080	101.0235	102.0526
Average # of T_i's in the last 9 time steps (mean number $m(92h < t < Th)$)	5.021	8.9176	16.2397
Correction factor	.055867	1.02316	2.022276
Predicted mean function	5.07081	9.11653	17.37815

4.2 Summary for Increasing Step Intensity

For n=100(total number of failures) and N=10^4 (number of iterations)

Table 4.2 Results for NHPP with $\lambda(t)=1$, for $0 \le t \le \tau$ & $\lambda(t)=3$, for $\tau < t \le \infty$

τ	*33*	*51*	*66*
Average total area	65.62367	71.658	78.87832
Average area for last 9 ERR's	8.679504	8.560146	7.463256
Average $m(T)$ (True mean function)	101.1060	101.5846	101.4423
Average # of T_i's in the last 9 time steps (mean number $m(92h < t < Th)$)	14.4372	17.2139	19.4196
Correction factor	1.540694	1.417652	1.28606
Predicted mean function	13.37245	9.77367	9.59819

4.3 Summary for Decreasing Step Intensity

For n=100(total number of failures) and N=10^4 (number of iterations)

Table 4.3 Results for NHPP with $\lambda(t)=3$, for $0 \le t \le \tau$ & $\lambda(t)=1$, for $\tau < t \le \infty$

τ	*33/3*	*51/3*	*66/3*
Average total area	140.1084	143.0810	137.7996
Average area for last 9 ERR's	8.985952	9.027335	9.062618
Average $m(t)$ (True mean function)	100.7025	100.5044	100.5679
Average # of T_i's in the last 9 time steps (mean number $m(92h < t < Th)$)	7.1816	6.2517	5.4808
Correction factor	0.717913	0.70243	0.729812
Predicted mean function	6.45113	6.34107	6.614007

4.4 Description of the Simulation Results

From the simulation, we get the total area under ERR curve and the true mean function. The correction factor is also calculated based on the average values of the above two, after 10^4 iterations. We can justify the correction factors from table 4.1 as follows. The correction factor for $\beta = 0.5$ is less when compared to $\beta = 1$ & 2 because for $\beta = 0.5$, the intensity function is decreasing, the number of events present initially will be more when compared to that at the end. So there is an over estimate of number of events at the end of given time period. In order to normalize it, the correction factor should be low so when we get the actual number of events by multiplying the respective area with the correction factor, we get a less number and there is no over estimation. And in case of $\beta = 2$, since the intensity is increasing, initially there will be less events so there could be under estimation of the number of events and hence the correction factor is high to balance it. And for $\beta = 1$, since there is no under or over estimation the correction factor is almost equal to 1.

The averages of the area for the last nine time-steps and the actual mean number (i.e. the number of failures in the last nine time steps) are calculated. The average area for last nine time-steps is multiplied with the correction factor in the simulation itself to give the predicted mean function. And from the results, it is evident that the predicted mean function for the last nine time-steps is approximately equal to the actual mean number in table 4.1 and in table 4.3 whereas there is slight variation in table 4.2 because the intensity is increasing step, so when $\tau < t \leq \infty$, the number of failures will be thrice than the failures before the jump. So the actual mean number will be slightly more than the predicted mean function whereas this effect is compensated in decreasing case by taking $\tau = 33/3, 51/3, 66/3$ even before generating the NHPP.

So the results are quite satisfactory to state that this correction factor could be used to predict the mean function, and the intensity of that fore-coming failures could be estimated by using the formula $\lambda(t|\Theta) = \frac{d}{dt}\mu(t|\Theta)$ and the goal of this thesis is achieved.

CHAPTER 5

APPLICATIONS

The control of industrial accidents generally requires, from time to time, new safety equipment, safety regulations, improved machinery, etc.; hence, one may expect that the number of accidents occurring would tend to decrease with time. Because of serious injuries or, perhaps, deaths that may occur as a result of an industrial accidents, it is usually important to know whether or not the safety action are resulting in a significant decrease in the number of accidents. The nonhomogeneous Poisson process with Weibull intensity function may possibly be useful in measuring this decrease (Crow, 1974). The technique described in chapter 3 can be well applied to the mining data to predict the number of failures in the coming years. So we will able to find the intensity of the accidents using this prediction.

5.1 Mining Data

The data in Table 5.1 (Maguire et al, 1952, Table 1) represent days between explosions in mines in Great Britain involving more than 10 men killed. The data cover the period from December 6, 1875 to May 29, 1951. Using the inferences in chapter 2 it can be easily determined that the data represents an NHPP with decreasing trend.

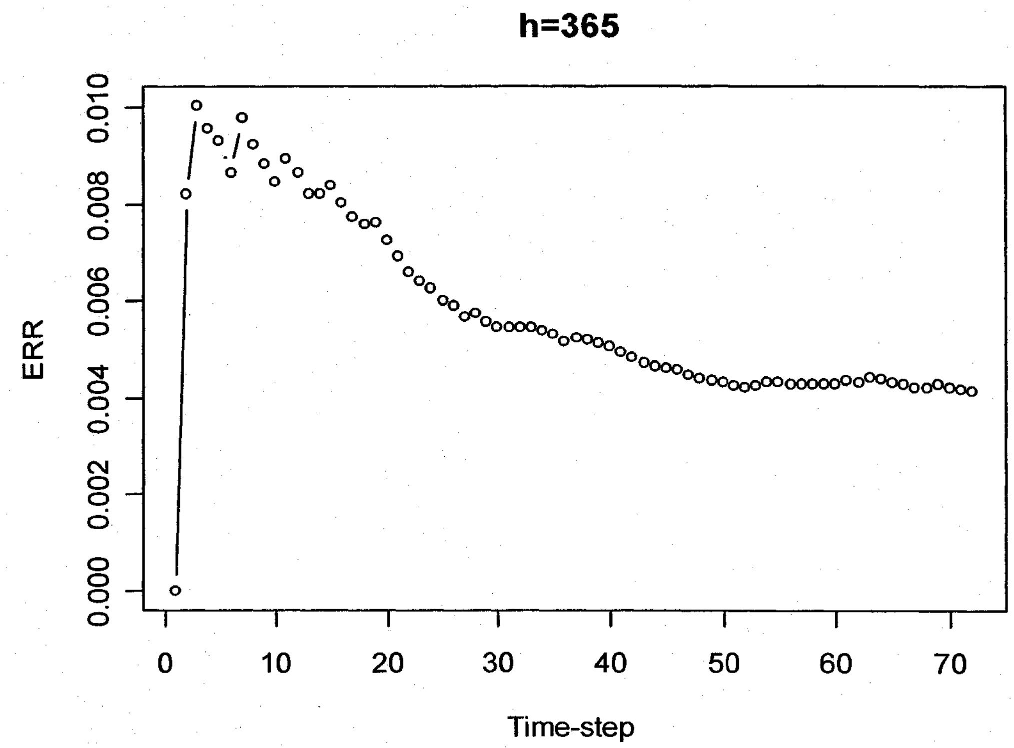
Table 5.1 Time Intervals in Days between Explosions in Mines, Involving 10 Men Killed, from 6 December 1875 to 29 May 1951

378	59	54	498	217	156
36	61	217	49	120	47
15	1	113	131	275	129
31	13	32	182	20	1630
215	189	23	255	66	29
11	345	151	195	291	217
137	20	361	224	4	7
4	81	312	566	369	18
15	286	354	390	338	1357
72	114	58	72	336	
96	108	275	228	19	
124	188	78	271	329	
50	233	17	208	330	
120	28	1205	517	312	
203	22	644	1613	171	
176	61	467	54	145	
55	78	871	326	75	
93	99	48	1312	364	
59	326	123	348	37	
315	275	457	745	19	

5.2 ERR-Plotting of Mining Data

ERR-plots for the observation period, (0, T], are produced respectively for the mining data (Table 5.1). In this application we use h = 365. Because the sample total of the 109 successive mine accidents is 26,263, in order to equally divide the total period into 72 years, we take additional 17 days as the final observation but we don't consider that an accident took place on 17th day. In this the first 62 years are used to predict the mean function of the next 10 years. So now the total number of days of observation is $T = 26280$. The ERR-plot is displayed in Figure 5.1.

Fig. 5.1 ERR Plot of Mining Data for Time-Step $h = 365$



5.3 Predicting the Mean Function

The area under the ERR curve of mining data is calculated for 62 time-steps (years) and the true mean function in the 62 years is found to be 98(i.e. the number of accidents). The correction factor is determined based on the above two which is given in table 5.2. To get the maximum likelihood estimators of β and θ over 62 years, we take additional 13 days as the 99^{th} observation, so, there are $T = 22630$ total days in the data which is a multiple of 365. Now $\hat{\beta}$ and $\hat{\theta}$ are determined to find the expected mean function $\hat{m}(t_{62}) = (T/\hat{\theta})^{\hat{\beta}}$, and it is found to be the same as the true mean number 98. The results are given in table 5.2

Table 5.2 Results for the First 62 years

Total area for 62 time steps	135.7342
Correction factor	0.7219992
$\hat{\beta}$	0.7221503
$\hat{\theta}$	39.01416
$\hat{m}(t_{62})$	98.9999

The actual mean number is determined for each year, after 62 years and their cumulative numbers are also calculated. Now the area for each time-step under the ERR curve is calculated along with the cumulative areas after 62 years which is multiplied with the correction factor to predict the mean function in the last ten time steps.

The expected mean function is also determined for the last ten time steps using the Power Law process formula $\hat{m}(t_{62+k}) = \left(\frac{22630+365k}{\hat{\theta}}\right)^{\hat{\beta}}, k=1,2,...,10$. The cumulative mean function is calculated by the formula $\hat{m}(t_{(62<h<62+k)}) = \hat{m}(t_{62+k}) - \hat{m}(t_{62})$.

And this when compared to the predicted mean function is proved from table 5.3 to be approximately equal. Although the actual mean numbers vary during the first few years they are approximately equal to the predicted values in the final years. Hence, this correction factor could be used to predict the mean function in the coming years, i.e., the number of accidents in the coming years.

5.4 Summary of the Simulation Results for Mining Data:

Table 5.3 Results for the Last Ten years, with C.F = 0.7219992

No. of time steps	*No of failures in each step*	*Cum No (actaul mean number m(t))*	*Cum area of time steps*	*Cum $\hat{m}(t)$ Power Law process*	*CF * cum area of each step (predicted $\hat{m}(t_{(62<h<62+k)})$)*
1	4	4	1.599846	1.150545	1.155088
2	1	5	3.214058	2.296026	2.320547
3	0	5	4.811053	3.436545	3.473576
4	0	5	6.383664	4.572199	4.609000
5	0	5	7.932623	5.703082	5.727348
6	2	7	9.473339	6.829285	6.839743
7	3	10	11.028006	7.950895	7.9262212
8	0	10	12.582044	9.067997	9.084225
9	0	10	14.114036	10.180674	10.190322
10	1	11	15.631543	11.289005	11.285961

CHAPTER 6

CONCLUSIONS AND FUTRUE WORK

6.1 Conclusions

The main goal of this thesis is to estimate the mean function of a monotonic point process with a discrete time series. The relation between HPP and NHPP is well established by proving the theorem, which is used to generate NHPP in the simulation. The bridge between Poisson process and time series is easily demonstrated which can be very helpful to get recurrence rates from any kind of Poisson process. The method to predict the mean function of NHPP using area under ERR curves and the correction factor proves to be useful as the correction factor given can be extended to the fore-coming failures. The correction factor is well justified for three different recurrence rates. The proposed algorithm to generate NHPP, to find the actual mean number, to get the area under ERR curve and finally to calculate the correction factor after finding the true mean function using the formula from Power law Process, shows tremendous potential to forecast NHPP with various forms of intensity function and recurrence rates. Applying the above technique to forecast the intensity of the accidents in the mining data is proved to be successful from the results, even though there is slight variation in the predicted and actual mean numbers for the first few time-steps, the mean function from the Power law Process is approximately equal to the predicted one.

6.2 Future Work

From the results it is noticeable that the predicted mean function for increasing step intensity differs from the actual mean number. The reason for that could be the selection of jump points. So, this can be rectified in the future by properly analyzing the jump points of the step, both for increasing and decreasing step intensity.

Finally for predicting the intensity of the NHPP, the predicted mean function could be used and safety measures could be followed to avoid failures such as accidents in the mining data.

CHAPTER 7

R-PROGRAM

Given are the programs to find the total area, area for last 10 ERR's under ERR curve, number of time intervals in the last ten time sets, observed mean function and the correction factor for the NHPP:

7.1 Program for $\beta = 0.5, 1, 2$

```
beta = 0.5    # modify to 1, 2
for(r in 1:10000){
g = 0
theta =  1
x = rexp(100,1)
for(k in 1:100){
z[k] = sum(x[1:k])
t[k] = (th*((z[k])^(1/beta)))}
betahat = 100/log(prod((t[100])/t[1:99]))
thetahat = ((t[100])/(100^(1/betahat)))
h = (t[100])/100    #mean
T = ((t[100]/h)+1)
for(j in 1:T){
w[j] = (j*h)
for(i in 1:100){
if(t[i] < w[j]){
g = g+1}}
y[j] = (g/w[j])
g = 0}
d = 0
for(i in 1:100){
if(t[i] < w[92]){
d = d+1}}
Numoft = (100-d)
```

```
soy = sum(y[92:101])
areasmall[r] = ((soy-(y[92]/2)-(y[101]/2))*h)
f = cumsum(y)
arealarge[r] = ((f[101]-(y[1]/2)-(y[101]/2))*h)
expez[r] = ((T*h/tethahat)^(betahat))
truez[r] = ((T*h/theta)^beta)}
totalareasmall = cumsum(areasmall)
avgarealast10ERR = totalareasmall[10000]/10000
totalnumoft = cumsum(numoft)
avgnumoft = (totalnumoft[10000]/10000)
totalarealarge = cumsum(arealarge)
avgtotalarea = totalarealarge[10000]/10000
totalexpez = cumsum(expez)
avgexpez = totalexpez[10000]/10000
totaltruez = cumsum(truez)
avgtruez = totaltruez[10000]/10000
correctionfactor = avgtruez * avgarealast10ERR / avgtotalarea
```

7.2 Program for Increasing Step Intensity

```
tau = 33  # modify to 51, 66
for(r in 1:10000)
{
g = 0
x = rexp(100,1)
for(k in 1:100){
z[k] = sum(x[1:k])
if(z[k] < tau){
t[k] = z[k]}
if(z[k] > tau){
t[k] = (z[k]+(2*tau))/3}}
h = (t[100])/100   #mean
T = ((t[100]/h)+1)
for(j in 1:T){
w[j] = (j*h)
for(i in 1:100){
if(t[i] < w[j]){
g = g+1}}
y[j] = (g/w[j])
g = 0}
d = 0
for(i in 1:100){
if(t[i] < w[92]){
d = d+1}}
```

```
Numoft = (100-d)
soy = sum(y[92:101])
areasmall[r] = ((soy-(y[92]/2)-(y[101]/2))*h)
f = cumsum(y)
arealarge[r] = ((f[101]-(y[1]/2)-(y[101]/2))*h)
capt[r] = (T*h)
truez[r] = (3*capt)-(2*tau)
}
totalareasmall = cumsum(areasmall)
avgarealast10ERR = totalareasmall[10000]/10000
totalnumoft = cumsum(numoft)
avgnumoft = (totalnumoft[10000]/10000)
totalarealarge = cumsum(arealarge)
avgtotalarea = totalarealarge[10000]/10000
totalcapt = cumsum(capt)
avgcapt = totalcapt[10000]/10000
totaltruez = cumsum(truez)
avgtruez = totaltruez[10000]/10000
correctionfactor = avgtruez * avgarealast10ERR / avgtotalarea
```

7.3 Program for Decreasing Step Intensity

```
tau = 33/3   # modify to 51/3, 66/3
for(r in 1:10000){
g = 0
x = rexp(100,1)
for(k in 1:100){
z[k] = sum(x[1:k])
l[k] = (z[k]/3)
if(l[k] < tau){
t[k] = l[k]}
if(l[k] > tau){
t[k] = (z[k]-(2*tau))}}
h = (t[100])/100    #mean
T = ((t[100]/h)+1)
for(j in 1:T){
w[j] = (j*h)
for(i in 1:100){
if(t[i] < w[j]){
g = g+1}}
y[j] = (g/w[j])
g = 0}
d = 0
for(i in 1:100){
```

```
if(t[i] < w[92]){
d = d+1}}
Numoft = (100-d)
soy = sum(y[92:101])
areasmall[r] = ((soy-(y[92]/2)-(y[101]/2))*h)
f = cumsum(y)
arealarge[r] = ((f[101]-(y[1]/2)-(y[101]/2))*h)
capt[r] = (T*h)
truez[r] = (capt)+(2*tau)}
totalareasmall = cumsum(areasmall)
avgarealast10ERR = totalareasmall[10000]/10000
totalnumoft = cumsum(numoft)
avgnumoft = (totalnumoft[10000]/10000)
totalarealarge = cumsum(arealarge)
avgtotalarea = totalarealarge[10000]/10000
totalcapt = cumsum(capt)
avgcapt = totalcapt[10000]/10000
totaltruez = cumsum(truez)
avgtruez = totaltruez[10000]/10000
correctionfactor = avgtruez * avgarealast10ERR / avgtotalarea
```

7.4 Program for Mining Data

```
t =
c(378,36,15,31,215,11,137,4,15,72,96,124,50,120,203,176,55,93,59,315,59,61,1,13,189,
345,20,81,286,114,108,188,233,28,22,61,78,99,326,275,54,217,113,32,23,151,361,312,3
54,58,275,78,17,1205,644,467,871,48,123,457,498,49,131,182,255,195,224,566,390,72,2
28,271,208,517,1613,54,326,1312,348,745,217,120,275,20,66,291,4,369,338,336,19,329,
330,312,171,145,75,364,37,19,156,47,129,1630,29,217,7,18,1357,17)
h = 365
m = cumsum(t)
z =
c(378,36,15,31,215,11,137,4,15,72,96,124,50,120,203,176,55,93,59,315,59,61,1,13,189,
345,20,81,286,114,108,188,233,28,22,61,78,99,326,275,54,217,113,32,23,151,361,312,3
54,58,275,78,17,1205,644,467,871,48,123,457,498,49,131,182,255,195,224,566,390,72,2
28,271,208,517,1613,54,326,1312,348,745,217,120,275,20,66,291,4,369,338,336,19,329,
330,312,171,145,75,364,13)
r = cumsum(z)
betahat = 99/log(prod(r[99]/r[1:98]))
thetahat = (r[99]/(99^(1/betahat)))
for(j in 1:72){
w[j] = (j*h)
for(i in 1:110){
if(m[i]<w[j]){
```

```
g = g+1}}
y[j] = (g/w[j])
g = 0}
for(p in 1:10){
d = 0
for(q in 1:110){
if((w[61+p]<m[q])&(m[q]<w[61+(1+p)])){
d = d+1}}
num[p] = d
soy[p] = sum(y[(61+p):(62+p)])
area[p] = ((soy[p]-(y[61+p]/2)-(y[62+p]/2))*h)
mt[p] = ((22630+(365*p))/thetahat)^betahat}
for(p in 2:10){
cumnum[1] = num[1]
cumnum[p] = num[p]+cumnum[p-1]
cumarea[1] = area[1]
cumarea[p] = area[p]+cumarea[p-1]}
f = sum(y[1:62])
tarea = ((f-y[1]/2-y[62]/2)*h)
fm = cumsum(y)
ttarea = ((fm[72]-y[1]/2-y[72]/2)*h)
cf = 98/tarea
for(p in 1:10){
cfarea[p] = cf*cumarea[p]
mtl[p] = mt[p]-((22630/th)^bh)}
```

REFERENCES

1. Ascher H. (1983). Discussion on statistical Methods in Reliability, by J. F. Lawless. Technometrics v25, p.305-335.
2. Bain L. J. and Engelhardt, M. (1980). Inferences on the Parameters and current System Reliability for a Time Truncated Weibull Process. Technometrics v. 22, No .4, p.305-335.
3. Bain L. J. and Engelhardt, M. (1991). Statistical Analysis of Reliability and Life-Testing Models - Theory and Methods, (2^{nd} ed.), New York: Marcel Dekker
4. Box G.E.P., and Jenkins G.M. (1976). Time Series Analysis Forecasting and Control, Holden Day.
5. Brockwell P. J. and Davis R. A. (2003). Introduction to Time Series and Forecasting. Springer Texts in Statistics.
6. Crow L. H. (1974). Reliability Analysis for Complex Repairable Systems, reliability and biometry, eds. pp. 379-410.
7. Ho C.-H. (1993). Forward and Backward Tests for an Abrupt Change in the Intensity of a Poisson Process: J. Statist. Comput. Simul. v.48, No.2, p. 245-252.
8. Michael E. K., Halim D. and James R. W. (1998). Least squares estimation of Nonhomogeneous Poisson process: Proceedings of 1998 Winter Simulation Conference D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, eds.

9. NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/,7/18/2006.
10. Reliability, Wikipedia contributors, 12/4/2006 http://en.wikipedia.org/w/index.php?title=Reliability&oldid=92020435
11. Rigdon S. E. and Basu A. P. (2000). Statistical Methods for the Reliability of Repairable Systems

VITA

Graduate College
University of Nevada, Las Vegas

Sandhya Gunti

Local Address:
4217, Grove Circle, Apt#1,
Las Vegas, Nevada, 89119

Degrees:
Bachelor of Engineering, Electrical and Electronics Engineering, 2003
Univ. College of Engineering, Osmania University, Hyderabad, India.

Thesis Title:
A Method for Estimating Intensity of a Poisson Process

Thesis Examination Committee:
Chairperson, Dr. Chih-Hsiang Ho, Ph. D.
Committee Member, Dr Malwane Ananda, Ph. D.
Committee Member, Dr Hokwon Cho, Ph. D.
Graduate Faculty Representative, Dr. Shizhi Qian, Ph. D.