12-2011

Development of a simulation tool for analysis of freeway crashes due to cell phone usage

Sourabh Sriom
University of Nevada, Las Vegas

DEVELOPMENT OF A SIMULATION TOOL FOR ANALYSIS OF FREEWAY CRASHES DUE TO CELL PHONE USAGE.

By

Sourabh Sriom

A thesis submitted in partial fulfillment

of the requirements for the

Master of Science in Electrical and Computer Engineering

Department of Electrical and Computer Engineering

School of Graduate Studies

University of Nevada Las Vegas

December 2011

THE GRADUATE COLLEGE

We recommend the thesis prepared under our supervision by

Sourabh Sriom

entitled

Development of a Simulation Tool for Analysis of Freeway Crashes Due to Cell Phone Usage

be accepted in partial fulfillment of the requirements for the degree of

Master of Science in Electrical and Computer Engineering
Department of Electrical and Computer Engineering

Puskin Kachroo, Committee Chair

Rama Venkat, Committee Member

Brendan Morris, Committee Member

Alexander Paz, Graduate College Representative

Ronald Smith, Ph. D., Vice President for Research and Graduate Studies
and Dean of the Graduate College

December 2011

Abstract

Research shows that drivers are susceptible to numerous distractions while driving which can be held responsible as the reason of a crash. Usage of cell phones, i.e. talking or texting while driving is considered one of the prominent distractions which causes a crash. This dissertation aims to study the number of crashes occurring on the freeways and their relation with the drivers of these cars using cellphones while they are driving. Since crashes occur relatively less frequently, the study of crashes is done using what is called the "rare event" theory. Java based simulations are done to model a six lane freeway. The cars travelling on the freeway are assigned probabilities of having and using a cell phone. This in turn leads to a probability of getting distracted and changing their normal behaviour which may or may not cause a crash. A Life table is constructed using the data observed for 24 hours and using the techniques described in the thesis, it is analysed to see which rare event probabilistic model it fits into.

Dedication

To my Grandfather Mr.Shambhoo Narayan Rai

Acknowledgements

I would like to express my sincere thanks and gratitude towards Dr.Pushkin Kachroo for inspiring, motivating and guiding me throughout my masters degree and the dissertation. I am also grateful to Dr.Alexander Paz, Dr.Rama Venkat and Dr.Brendan Morris for their insightful remarks.

I would like to take this opportunity to show my appreciation towards the love of my parents Mr.Hari Shankar Prasad and Mrs. Madhu Lata and be beloved grand parents, Mr.Shambhoo Narayan Rai and Mrs. Devkala Devi.

I am also grateful to my friends Pratik Verma, Atul Sancheti, Romesh Khaddar, Shaurya Agarwal and Anuj Nayyar for all their support and help during the period of my work.

Contents

List of Tables

Chapter 1

Introduction

1.1 Background

Research shows that drivers are susceptible to numerous distractions while driving which can be held responsible as the reason of a crash. A 2001 study conducted by the AAA foundation, which studied the nationwide crashes occurring in the USA for the period of 1995-1999 indicated that 8.3% of the total crashes could be attributed to the driver being distracted. The most common sources of distraction cited in these crashes were an outside person,object, or event (29.4% of crash-involved drivers); adjusting the radio, cassette, or CD (11.4%),passengers (10.9%), distraction due to some sort of cellphone usage was found out to be 1.5%.

In another naturalistic study conducted in 2001 for the AAA Foundation, the driving of 70 volunteers was recorded using in-vehicle cameras to identify potential sources of distraction (Stutts, Feaganes,Rodgman, et al., 2003). The researchers reviewed 3 hours of data per subject and found that the potential distractions in

which the greatest proportion of drivers engaged were manipulating the vehicle controls(100% of drivers), conversing with passengers (77% of drivers), eating or drinking (71% of drivers), and grooming (46% of drivers). Thirty-four percent of drivers used cell phones during the video samples that were reviewed, and did so for 3.8% of their total observed driving time.

Ten years after the report we have seen a significant increase in the number of cellphone users in the country and subsequently the number of drivers who use their cellphones while driving. An International Telecommunication Union survey in January 2008 reports that 9 out of every 10 adults i.e. 89% of all adults in the country are active cell phone users. Another research carried out by AT&T in collaboration with Harvard Centre for Risk Analysis found out that 80-90% of these cell phone users admit to using their cell phones at least once while they are driving. The study also found out that 37% of these drivers spend from 30 seconds to 2 minutes on call while driving.

Looking objectively at Distracted driving, it is any non-driving activity a person engages in that has the potential to distract him or her from the primary task of driving and increase the risk of crashing. Over the years research and numerous surveys have showed that drivers are more susceptible to crashes if they are distracted while driving. Distractions as it may be noted can be classified into three types.

1. **Visual** : Taking your eyes of the road.

2. **Manual** : Taking your hands of the wheel.

3. **Cognitive** :Taking your mind off what you are doing, which in this case is driving.

While all distractions can endanger drivers safety, Cell phone usage is the most alarming because it involves three types of distraction and hence the most dangerous type of distraction. Considering a vehicle travelling on a freeway with a speed of 65 miles per hour, a vehicle travels 88 feet in one second and thus even a small distraction can end up causing a lot of damage.

1.2 Motivation

The aim of this thesis is to study the relations between the trend of cell phone usage in the state of Nevada, using data collected from agencies such as FAST and the number of crashes that occur everyday on freeways. This thesis aims to explore these crashes from a *rare event* or *life-time data* point of view and see whether we can mathematically model the timing of the crashes assuming certain probability distributions for the usage of cell phones, vehicle arrival time on freeways, call waiting duration and the actual call duration while driving.

1.3 Mode of study

Besides consulting various surveys that the government bodies have carried out over the years, this study will include simulation of a 4 lane freeway using the Java 2D API. The vehicles will then be populated Exponentially distributed using a preset arrival rate. Each vehicle will then have a particular probability of having a cell phone and subsequently using it. The probability of a driver getting distracted is then determined and also considering the age various actions while the driver is distracted

are set. All these events might ultimately culminate in a vehicle being involved in a crash. The timings of this crash will now be noted and using chi-square distribution it will be studied which of the probable *rare events* distribution best fits the obtained data.

Chapter 2

Literature survey

Research shows that approximately 40,000 people die every year in United States in different kind of road accidents due to various reasons. Here are some of the various statistics to draw attention towards serious issues and gaps related to day to day Transportation. System (NASS) General Estimates System (GES) shows that: In 2008, there were a total of 34,017 fatal crashes in which 37,261 individuals were killed. In 2008, 5,870 people were killed in crashes involving driver distraction (16% of total fatalities). The under-20 age group had the highest proportion of distracted drivers involved in fatal crashes (16%). The age group with the next greatest proportion of distracted drivers was the 20-to-29-year-old age group (12%). Motorcyclists and drivers of light trucks had the greatest percentage of total drivers reported as distracted at the time of the fatal crashes (12%). An estimated 21 percent of 1,630,000 injury crashes were reported to have involved distracted driving.

Road accidents occur due to various reasons (for example shortcoming in vehicle side, shortcoming in road structure, shortcoming from drivers side and many more).

Various studies based on crash data and accident analysis shows that a majority of the incidents are related to shortcoming from Drivers side. When we talk about the shortcomings in Driving, we are majorly concerned about the attentiveness of driver while driving. Drivers inattention leads to various accident and crashes. Due to inattentive state, Driver recognises a trouble situation little late and gets lesser time to react to a particular situation than he/she would have in attentive state. This decrement in reaction time leads to the uncertain response to the trouble situation, which sometimes results in crash or a near crash situation.

Inattentive state can be narrowed to Distractions while driving. Distractions are some particular situations (Outside or inside the vehicle) which can drag drivers attention away from driving. Distractions can be further classified broadly into two categories:-

1. Distractions which can cause driver taking his/her eye off from the road (For example:- Texting, calling, dialling, changing CD, adjusting AC etc)

2. Distractions which do NOT cause driver taking his/her eye off from road but mental attention is dragged from driving. (For example:- Listening, singing, talking, getting emotional, Listening/talking to co passenger etc)

A major percentage of accidents are due to Distracted Driving. Below are some statistics:-

- In 2008, slightly more than almost 20 percent of all crashes in the year involved some type of distraction. (National Highway Traffic Safety Administration - NHTSA).

- Nearly 6,000 people died in 2008 in crashes involving a distracted driver, and more than half a million were injured. (NHTSA)

- The younger, inexperienced drivers under 20 years old have the highest proportion of distraction-related fatal crashes.

- Drivers who use hand-held devices are four times as likely to get into crashes serious enough to injure themselves. (Source: Insurance Institute for Highway Safety)

- Using a cell phone use while driving, whether its hand-held or hands-free, delays a driver's reactions as much as having a blood alcohol concentration at the legal limit of .08 percent. (Source: University of Utah)

- The proportion of drivers reportedly distracted at the time of the fatal crashes has increased from 8 percent in 2004 to 11 percent in 2008(GES).

There various studies/research have been performed from various aspects of Distracted Driving. This work certainly helps us in understanding the problem in depth, getting acquainted with the drivers mind perspective while driving and his/her reaction to a particular situation. Research done by various organisations and Institutions focus on different areas which are classified as below:-

1. Types of Distraction and Quantisation of Impact due to a particular type of Distraction.

2. Conclusions based upon crash data analysis and GES Data.

3. Study of behavioural response of Driver in real time scenario.

4. Study of behavioural response of Driver in a driving simulation setup.

5. Physiological data collection at the time of Distraction.

2.1 Types of Distraction and Quantisation of Impact due to a particular type of Distraction

2.1.1 Virginia Tech Transportation Institute

A study has been performed in Virginia Tech Transportation Institute focusing on the risk of crash or near miss associated with the various types of distraction (for example:- Texting, Dialling etc). This work also based upon the type of vehicle involved in driving (for example Car, Truck) because risk associated may vary from one type vehicle to other type of vehicle. During this study various drivers were monitored more than 6 M of miles via camera and data was collected. It mainly focus upon the eye glace analysis bases upon various operations involved in cell phone use.

2.1.2 University of South Carolina (Highway Safety Research Centre) For: - AAA Foundation for Traffic safety

This work is mainly concerned with the The Role of Driver Distraction in Traffic Crashes. This paper also considers CDS data analysis. It considers various factors related to environment, vehicle and crash into account. It also quantizes the other factors having potential affect on Driver distraction.

2.1.3 SURVEY AND EVALUATION RESEARCH LABORATORY CENTER FOR PUBLIC POLICY VIRGINIA COMMONWEALTH UNIVERSITY

To ascertain the frequency of specific behaviours that may constitute driver inattention, a survey was developed for officers to complete at crash scenes that involved driver inattention. These factors were thought to control for concerns such as type of road network (e.g., urban vs. rural vs. interstate), population density, different reporting agencies, and for areas in which combinations of agencies may investigate crashes provide a framework to create prior expectations of how many crashes occur in each county or independent city.

2.2 Conclusions based upon Crash data analysis and GES Data

2.2.1 National Highway Traffic safety Administration (NHTSA)

An examination of Drivers Distraction as recorded in NHTSA Database was done by National Center for statistical Analysis. The research note provides fatality, injury, on-scene crash investigation, and survey data associated with distracted driving and to summarise recent data from NHTSA and other DOT modes pertaining to distracted-driving crashes.

2.2.2 Highway Safety Information System (HSIS)

Study was related to Development of a speeding-related Crash Typology. Analysis was Done based Fatal Analysis Reporting System (FARS). the study used 2005 data from the two major national crash databases: the National Automotive Sampling Systems General Estimates System (GES) and FARS.(9,1) GES data are derived from a nationally representative sample of police-reported motor vehicle crashes of all types, from minor to fatal. Approximately 60,000 police accident reports are included each year. Sample weights are assigned to each crash based on a sampling protocol.

2.2.3 Data from Wisconsin motor vehicle accident report form (MV4000)

This survey intent was to provide an introduction to the issue, or problem, of motorist cell phone use and to prompt research and queries. The issue of cell phone use by motorists is complex enough to warrant further review of many variables, including driver demand and/or need for cell phones, driver education, other driver distractions, varying driver abilities, physiological factors of cell phone use, cell phone reporting procedures, and a review of empirical data.

2.3 Study of behavioural response of Driver in real time scenario

2.3.1 Federal Highway Administration (FHWA)

The researchers also developed a brief take-home survey. The Federal Highway Administration (FHWA) examined safety areas: driver behaviour at intersections, the development of tools and procedures for intersection design, and human factors literature reviews for Safety R&D program areas, including Intersections, Pedestrians and Bicyclists, Speed Management, and Visibility. The work was divided into different scenarios like 1) Red-light running)2) Left turn in traffic 3) Left turn at stop sign etc and survey was conducted on different site locations.

2.3.2 Research, Development, and Technology Turner-Fairbank Highway Research Center

Twenty drivers participated in the experiment and were divided into groups based on age. Ten subjects were between the ages of 18 and 25 (younger drivers), and 10 subjects were between the ages of 65 and 75 (older drivers). Within each age group, five subjects were male and five were female. Driver behaviour was investigated on-road using an instrumented 1995 Oldsmobile Aurora four door sedan. The primary apparatuses used in the study were:

1. the automobile

2. cameras and sensors

3. an IVIS display

4. software and hardware interfaces for information portrayal and data collection

5. the information portrayed and the portrayal format.

2.3.3 Monash University Accident Research Centre Monash University

Visual clutter and external-to-vehicle driver distraction. Participants undertook the experiment in groups of approximately ten. Drivers filled out a brief questionnaire on their age, sex, licence type, and driving experience. Participants viewed a series of photographs of various road scenes for another experiment being conducted concurrently, and then took part in the focus group discussion. Participants were shown a photograph of a crowded city and asked a series of questions about which objects contributed to the amount of clutter in the scene; what could distract them from the driving task; and how easy they thought it would be to find a street sign. The discussion was recorded on audiotape for later transcription by the experimenter.

2.4 Study of behavioural response of Driver in a driving simulation setup

2.4.1 Strayers research group at the University of Utah

submitted test subjects to different levels of distractions while driving in a simulator. The researchers were able to conclude that cellular phone conversations while driving

caused the subjects to react slower to stimuli and perform tasks with considerably reduced precision. Specifically, while engaged in cell phone conversations the subjects were twice as likely to miss simulated traffic signals compared to when they were not distracted.

2.4.2 Insurance Corporation of British Columbia

Complex method of identifying specific cell phone users/non-users through in-field observations, and linking these people with their driving records was used. This method presents some obvious limitations or uncertainty about the user classification; however, the results corresponded well with other identifying methods. The driving records of the cell phone users had higher counts of moving violation citations over the previous four years, to include speeding, alcohol, and failure to use seat belts, aggressive driving violations, and non-moving violations.

2.4.3 Massachusetts Institute of Technology, Cambridge

Age Related Changes in Cognitive Response Style in the Driving Task:- In this research, different patterns of physiological response were observed between younger and older drivers during a simulated cellular telephone conversation. Results of the study indicate that while younger drivers showed heart rate acceleration during the cell phone task, older adult drivers, as a group, showed no change. Questioning the apparent lack of reaction in the older drivers, and considering the literature on sensory intake and rejection which suggests that heart rate can increase or decrease depending on how individuals attend to cognitive processing demands.

2.5 Physiological data collection at the time of Distraction

2.5.1 Brain Research Center, University System of Taiwan, Hsinchu, Taiwan

Department of Electrical and Control Engineering, National Chiao-Tung University, Hsinchu, Taiwan. Department of Computer Science and Information Engineering, National Cheng-Kung University, Tainan, Taiwan.

This study was to investigate Electroencephalography (EEG) response to distraction during driving. To study human cognition under driving task, Virtual Reality (VR) based driving simulation to simulate events were used including unexpected car deviations and mathematics questions (math) in real driving. For further assessing effects of the stimulus onset asynchrony (SOA) between the deviation onset and math presented on the EEG dynamics, five cases with different SOA were designed. Results showed that increases of theta band (5 7.8 Hz) and beta band (12.2 17 Hz) power were observed in the frontal cortex. Results demonstrated that reaction time and multiple cortical EEG sources responded to the driving deviations and math occurrences differentially in the stimulus onset asynchrony. Results also suggested that the theta band power increase in frontal area could be used as the distracted indexes for early detecting drivers inattention in the future.

2.5.2 GM Corp., Wayne State University Medical School, and Henry Ford Hospital

This study used Functional Magnetic Resonance Imaging (MRI) and Magnetoencephalography (MEG) to locate essential brain activated structures and their corresponding dynamics. The authors suggest that there are situations where behavioural indicators will show that the mind is on the road, but in reality, it is not. With this understanding, the authors set out to uncover the exact neural mechanisms that are associated with distracted behaviours while driving

2.5.3 Carnegie Mellon University

study by using Functional Magnetic Resonance Imaging (MRI) to investigate the impact of concurrent auditory language comprehension on the brain activity when simultaneously exposed to a simulated driving experience. Participants operated a driving simulator, either undisturbed or while listening to statements they had to identify as true or false. This auditory language comprehension was designed to mimic talking on a cell phone. The participants brain activity was monitored during the simulations and was compared against the MRI scans of the undisturbed drivers brain.

2.5.4 Department of Psychology, University of Minnesota

University of Groningen, the Netherlands

Psychological Measures Of Driver Distraction and Workload While Intoxicated. This study compares driver performance while conversing on a hands-free cell phone

to conditions of operating common in-vehicle controls and alcohol intoxication. In addition, the study examined the combined effects of being distracted and being intoxicated given that there may be a higher risk of a crash if the driver engages in a combination of risk factors. During simulated traffic scenarios, resource allocation was assessed through an event-related potential (ERP) novelty oddball paradigm. Intoxicated drivers were less attentive to all stimuli and drivers engaged in secondary tasks had weaker responses to unexpected novel sounds in brain regions associated with evaluative processing. Drivers conversing on the cell phone and in-vehicle tasks while sober had lower accuracy during the target tone task than intoxicated drivers not completing any secondary task.

2.5.5 Distracted Driving Nation Safety Council

This provides a understanding of the full impact of driving while engaging in cell phone conversations on both handheld and hands-free phones. It explains how cognitively complex it is to talk on the phone and drive a vehicle at the same time, and why this drains the brains resources. This report also talks about the multitasking myths related to human brain.

2.5.6 IBM T.J. Watson Research Center

Physiological User Interfaces . This paper presents an overview of some research projects that demonstrate how physiological signals including GSR, EMG, EKG and respiration can be used as an alternative user interface, especially in new areas of computing such as mobile and The system used for collecting the signals is introduced

first, then three scenarios for using a physiological user interface are presented: a method for guiding music selection; an algorithm for automatic control of a digital camera and a system for detecting driver stress.

2.5.7 Pennsylvania State University, University Park

The objective of the project is to assess the extent of distraction caused by billboards by analysing the drivers eye movement in the presence of billboards. The analysis is to be carried out in a simulator, as any distraction caused on the road would be dangerous. The eye tracking system tracks the eye movement of the driver as he drives and the scene camera arranged in the car gives a view of the road ahead of the driver. The data collected from the scene camera and the eye tracking system can later be combined. The combined data gives a video recorded by the scene camera. The video gives an idea of the safety effects of billboards on drivers.

2.6 Gaussian probability distribution

In probability theory, the normal (or Gaussian) distribution is a continuous probability distribution that is often used as a first approximation to describe real-valued random variables that tend to cluster around a single mean value. The graph of the associated probability density function is "bell"-shaped, and is known as the *Gaussian function or bell curve*:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \tag{2.1}$$

where parameter μ is the mean (location of the peak) and σ^2 is the variance (the measure of the width of the distribution). The distribution with $\mu = 0$ and $\sigma^2 = 1$ is

called the **standard normal**.

de Moivre [1] developed the normal distribution as an approximation to the binomial distribution, and it was subsequently used by Laplace [2] in 1774 to study measurement errors and by Gauss in 1809 [3] in the analysis of astronomical data.

The normal distribution is considered the most prominent probability distribution in statistics[1]. There are several reasons for this: First, the normal distribution is very tractable analytically, that is, a large number of results involving this distribution can be derived in explicit form. Second, the normal distribution arises as the outcome of the central limit theorem, which states that under mild conditions the sum of a large number of random variables is distributed approximately normally. Finally, the "bell" shape of the normal distribution makes it a convenient choice for modelling a large variety of random variables encountered in practice.For this reason, the normal distribution is commonly encountered in practice, and is used throughout statistics, natural sciences, and social sciences[2] as a simple model for complex phenomena. For example, the observational error in an experiment is usually assumed to follow a normal distribution, and the propagation of uncertainty is computed using this assumption.

2.6.1 Characteristics of the Probability Density Function

The probability density function (pdf) of a random variable describes the relative frequencies of different values for that random variable. The pdf of the normal distribution is given by the formula explained in detail in the previous subsection:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \tag{2.2}$$

This is a proper function only when the variance σ^2 is not equal to zero. In that case this is a continuous smooth function, defined on the entire real line, and which is called the "Gaussian function" and it has the following properties.

- Function $p(x)$ is uni modal and symmetric around the point $x = \mu$, which is at the same time the mode, the median and the mean of the distribution[4].
- The inflection points of the curve occur one standard deviation away from the mean (i.e., at $x = \mu\sigma$ and $x = \mu + \sigma$).
- Function $p(x)$ is log-concave.
- The standard normal density $\phi(x)$ is an eigenfunction of the Fourier transform.
- The function is super smooth of order 2, implying that it is infinitely differentiable.[5]
- The first derivative of $\phi(x)$ is $\phi\prime(x) = x\phi(x)$; the second derivative is $\phi\prime\prime(x) = (x^2 1)\phi(x)$. More generally, the nth derivative is given by $\phi(n)(x) = (1)^n H_n(x)\phi(x)$, where H_n is the Hermite polynomial of order n.[4]

2.7 Poisson Process

A Poisson process, named after the French mathematician Simon-Denis Poisson (17811840), is a stochastic process in which events occur continuously and independently of one another (the word event used here is not an instance of the concept of event frequently used in probability theory). Examples that are well-modeled as Poisson processes include the radioactive decay of atoms, telephone calls arriving at a switchboard, page

view requests to a website, and rainfall. A Poisson process is usually described as a function of time, although it need not be.

The Poisson process is a collection $N(t) : t \succeq 0$ of random variables, where N(t) is the number of events that have occurred up to $timet$ (starting from $time0$). The number of events between $timea$ and $timeb$ is given as $N(b)N(a)$ and has a Poisson distribution. Each realisation of the process $N(t)$ is a non-negative integer-valued step function that is non-decreasing, but for intuitive purposes it is usually easier to think of it as a point pattern on $[0, \infty)$ (the points in time where the step function jumps, i.e. the points in time where an event occurs).

There are several types of Poisson processes which are listed as below :

- Homogenous
- Non-Homogenous
- Spatial
- Space-time

In our simulation we will be using the *Homogenous Poisson Process* to generate cars in a single lane.

2.7.1 Homogenous Poisson Process

The homogeneous Poisson process is one of the most well-known *Lvy processes*. This process is characterised by a rate parameter λ, also known as *intensity*, such that the number of events in time interval $(, t + \tau]$ follows a Poisson distribution with

associated parameter $\lambda\tau$. This relation is given as

$$P[(N(t+\tau) - N(t)) = k] = (\lambda\tau)^k e^{-\lambda\tau}/k! \tag{2.3}$$

where $N(t+\tau)N(t) = k$ is the number of events in time interval $(t, t+\tau]$.

Just as a Poisson random variable is characterised by its scalar parameter λ, a homogeneous Poisson process is characterised by its rate parameter λ, which is the expected number of "events" or "arrivals" that occur per unit time.

$N(t)$ is a sample homogeneous Poisson process, not to be confused with a density or distribution function.

2.7.2 Properties of Poisson Process

Poisson processes have an exponentially-distributed inter-arrival time property, to illustrate this consider a homogeneous Poisson process $N(t)$ with rate parameter λ, and let T_k be the time of the kth arrival, for $k = 1, 2, 3, ...$. Clearly the number of arrivals before some fixed time t is less than k if and only if the waiting time until the kth arrival is more than t. In symbols, the event $[N(t) < k]$ occurs if and only if the event $[T_k > t]$ occurs. Consequently the probabilities of these events are the same:

$$P(T_k > t) = P(N(t) < k) \tag{2.4}$$

In particular, consider the waiting time until the first arrival. Clearly that time is more than t if and only if the number of arrivals before time t is 0. Combining this latter property with the above probability distribution for the number of homogeneous Poisson process events in a fixed interval gives:

$$P(T_1 > t) = P(N(t) = 0) = P[(N(t) - N(0)) = 0] = e^{-\lambda t} \tag{2.5}$$

Consequently, the waiting time until the first arrival T_1 has an exponential distribution, and is thus memory less. One can similarly show that the other inter arrival times $T_k T_{k1}$ share the same distribution. Hence, they are independent, $identically-distributed(i.i.d.)$ random variables with parameter $\lambda > 0$; and expected value $1/\lambda$.

2.8 Exponential Probability Distribution

In probability theory and statistics, the exponential distribution (also known as negative exponential distribution) is a family of continuous probability distributions. It describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate, as has been proved in the previous section.

2.8.1 Probability Density Function of the Exponential distribution

The probability distribution function(*pdf*) of the exponential distribution is as described by the equation below:

$$p(x;\lambda) = \lambda e^{-\lambda x} H(x) \tag{2.6}$$

Here $\lambda > 0$ is the parameter of the distribution, often called the rate parameter and $H(x)$ is the unit-step function. The distribution is supported on the interval $[0, \infty)$. If a random variable X has this distribution, we write $X\ Exp(\lambda)$.

2.9 Linear Least square curve fitting

linear least squares is an approach to fitting a mathematical or statistical model to data in cases where the idealised value provided by the model for any data point is expressed linearly in terms of the unknown parameters of the model. The resulting fitted model can be used to summarise the data, to predict unobserved values from the same system, and to understand the mechanisms that may underlie the system.

Mathematically, linear least squares is the problem of approximately solving an overdetermined system of linear equations, where the best approximation is defined as that which minimises the sum of squared differences between the data values and their corresponding modeled values. The approach is called "linear" least squares since the solution depends linearly on the data. Linear least squares problems are convex and have a closed-form solution that is unique, provided that the number of data-points used for fitting equals or exceeds the number of unknown parameters, except in special degenerate situations. In contrast, non-linear least squares problems generally must be solved by an iterative procedure, and the problems can be non-convex with multiple optima for the objective function.

2.9.1 A general linear least squares solution

Consider an overdetermined system of m linear equations in n unknown coefficients, $\beta_1, \beta_2, , \beta_n$, with $m > n$

$$\sum_{j=1}^{n} X_{ij}\beta_j = y_i, (i = 1, 2, ..., m) \tag{2.7}$$

This can be written in matrix form as :

$$\mathbf{X}\beta{=}\mathbf{y} \tag{2.8}$$

where, $\mathbf{X} =$

$$\begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{pmatrix}$$

, $\beta =$

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}$$

, $\mathbf{y} =$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

Such a system usually has no solution, so the goal is instead to find the coefficients β which fit the equations best, in the sense of solving the quadratic minimisation problem:

$$\hat{\beta} = argmin_{\beta} S(\beta), \tag{2.9}$$

where the objective function S is given by:

$$\tag{2.10}$$

A justification for choosing this criterion is given in properties below. This minimisation problem has a unique solution, provided that the n columns of the matrix X are linearly independent, given by solving the normal equations:

$$(X^T X)\hat{\beta} = X^T y \tag{2.11}$$

2.9.2 Derivation of the normal equations

Define the ith **residual** to be:

$$r_i = y_i - \sum_{j=1}^{n} X_{ij}\beta_j \tag{2.12}$$

Then S can be rewritten as:

$$S = \sum_{i=1}^{m} r_i^2 \tag{2.13}$$

S is minimised when its gradient vector is zero. (This follows by definition: if the gradient vector is not zero, there is a direction in which we can move to minimise it further - see maxima and minima.) The elements of the gradient vector are the partial derivatives of S with respect to the parameters:

$$\partial S/\partial \beta_j = 2\sum_{i=1}^{m} r_i \partial r_i/\partial \beta_j (j = 1, 2, ..., n) \tag{2.14}$$

The derivatives are:

$$\partial r_i/\partial \beta_j = -X_{ij} \tag{2.15}$$

Substitution of the expressions for the residuals and the derivatives into the gradient equations gives:

$$\partial S/\partial \beta_j = 2\sum_{i=1}^{m}(y_i - \sum_{k=1}^{n} X_{ik}\beta_k)(-X_{ij}), (j = 1, 2, ..., n) \tag{2.16}$$

Thus if $\hat{\beta}$ minimises S, we have:

$$2\sum_{i=1}^{m}(y - \sum_{k=1}^{n} X_{ik}\hat{\beta}_k)(-X_{ij}) = 0, (j = 1, 2, ..., n) \tag{2.17}$$

The solution of this equation gives us the vector $\hat{\beta}$ of the optimal parameter values.

Chapter 3

Tools Used

3.1 NetBeans

NetBeans IDE is an open-source integrated development environment. NetBeans IDE supports development of all Java application types (Java SE including JavaFX, (Java ME, web, EJB and mobile applications) out of the box. Among other features are an Ant-based project system, Maven support, refactorings, version control (supporting CVS, Subversion, Mercurial and Clearcase).

3.1.1 Modularity

All the functions of the IDE are provided by modules. Each module provides a well defined function, such as support for the Java language, editing, or support for the CVS versioning system, and SVN. NetBeans contains all the modules needed for Java development in a single download, allowing the user to start working immediately. Modules also allow NetBeans to be extended. New features, such as support for

other programming languages, can be added by installing additional modules. For instance, Sun Studio, Sun Java Studio Enterprise, and Sun Java Studio Creator from Sun Microsystems are all based on the NetBeans IDE.

3.2 Java 2D API

The Java 2D API is a set of classes for advanced 2D graphics and imaging, encompassing line art, text, and images in a single comprehensive model. The API provides extensive support for image compositing and alpha channel images, a set of classes to provide accurate color space definition and conversion, and a rich set of display-oriented imaging operators.

The following are the basic *objects* that are necessary for every Java 2D drawing operation :

3.2.1 Shapes

A shape in Java 2D is a boundary which defines an inside and an outside. Pixels inside the shape are affected by the drawing operation, those outside are not.

Trying to fill a straight line segment will result in no pixels being affected, as such a shape does not contain any pixels itself. Instead, a thin rectangle must be used so that the shape contains some pixels.

3.2.2 Paints

A paint generates the colors to be used for each pixel of the fill operation. The simplest paint is java.awt.Color, which generates the same color for all pixels. More

complicated paints may produce gradients, images, or indeed any combination of colors. Filling a circular shape using the color yellow results in a solid yellow circle, while filling the same circular shape using a paint that generates an image produces a circular cutout of the image.

3.2.3 Composites

During any drawing operation, there is a source (the pixels being produced by the paint) and a destination (the pixels already onscreen). Normally, the source pixels simply overwrite the destination pixels, but the composite allows this behaviour to be changed.

The composite, given the source and destination pixels, produces the final result that ultimately ends up onscreen. The most common composite is java.awt.AlphaComposite, which can treat the pixels being drawn as partially transparent, so that the destination pixels show through to some degree.

3.2.4 Filling

To fill a shape, the first step is to identify which pixels fall inside the shape. These pixels will be affected by the fill operation. Pixels that are partially inside and partially outside the shape may be affected to a lesser degree if anti-aliasing is enabled.

The paint is then asked to generate a color for each of the pixels to be painted. In the common case of a solid-color fill, each pixel will be set to the same color.

The composite takes the pixels generated by the paint and combines them with the pixels already onscreen to produce the final result.

3.3 Java Swing

Swing is the primary Java GUI widget toolkit. It is part of Oracle's Java Foundation Classes (JFC) an API for providing a graphical user interface (GUI) for Java programs.

Swing was developed to provide a more sophisticated set of GUI components than the earlier Abstract Window Toolkit. Swing provides a native look and feel that emulates the look and feel of several platforms, and also supports a pluggable look and feel that allows applications to have a look and feel unrelated to the underlying platform. It has more powerful and flexible components than AWT. In addition to familiar components such as buttons, check box and labels, Swing provides several advanced components such as tabbed panel, scroll panes, trees, tables and lists.

Unlike AWT components, Swing components are not implemented by platform-specific code. Instead they are written entirely in Java and therefore are platform-independent. The term "lightweight" is used to describe such an element.

Chapter 4

Lifetime Distributions Analysis

4.1 Introduction

The statistical analysis of what is variously referred to as lifetime [6], survival time or failure time data is an important topic in the fields of engineering and biomedical sciences. It finds it's applications ranging from investigations into endurance of manufactured items to research involving human disease. In this particular Case we will look into modelling the survival time of vehicles travelling on freeways and using cell phones in the process. Owing to a substantial amount of research in this area in the recent times there are now many methods to dealing with lifetime data. In the following section we will have a brief review of the concepts involved in the analysis of lifetime data.

4.2 Basic concepts of lifetime distributions

4.2.1 Continuous Models

Let us begin by considering the case of a single lifetime variable T, where T is a non-negative random variable representing the lifetimes of individuals in some population. T is considered to be continuous in most cases and hence we shall discuss this first. All functions discussed here are defined over the interval $[0, \infty)$. Let $f(t)$ denote the probability density function($p.d.f$) of T and let the cumulative distribution function be $F(t)$ which will then be:

$$F(t) = Pr(T \preceq t) = \int_0^t f(x)\,\mathrm{d}x. \tag{4.1}$$

The probability of an individual surviving till time t is given by the *survivor function* $S(t)$ given by :

$$S(t) = Pr(T \succeq t) = \int_0^\infty f(x)\,\mathrm{d}x. \tag{4.2}$$

Sometimes the survivor function may also be referred to as *reliability function.* $S(t)$ is a monotone decreasing function with $S(0) = 1$ and $S(\infty) = lim_{t\to\infty} S(t) = 0$

Another important term worth defining in the study of lifetime data is the *Hazard function* $h(t)$ which specifies the instantaneous rate of *death* at a given time t, given that the individual survives up till t. Thus $h(t)$ is the probability of death in the $[t, t + \Delta t]$, given survival till time t. The Hazard function then can be defined by the following equation:

$$h(t) = f(t)/S(t) \tag{4.3}$$

Thus, continuous lifetime data can be studied using the functions described above.

4.2.2 Discrete Models

In some cases, as is the topic of this thesis, when lifetimes are grouped and it refers to an integral number of cycles e.g. number of deaths measured every month or year, it is beneficial to treat T as a discrete random variable. Assuming that T can take on values $t_1, t_2, ...$, with $0 \preceq t_1 < t_2 < ...$, and let the probability function $(p.f)$ be:

$$p(t_j) = Pr(T = t_j)j = 1, 2, ... \tag{4.4}$$

The survivor function is then:

$$S(t) = Pr(T \succeq t) = \sum_{t_j \succeq t}^{\infty} p(t_j) \tag{4.5}$$

Just like the continuous case, the survivor function is monotonically decreasing left-continuous function, with $S(0) = 1$ and $S(\infty) = 0$.

The Hazard function is now defined as:

$$h(t_j) = Pr(T = t_j | T \succeq t_j) = p(t_j)/S(t_j)j = 1, 2, ... \tag{4.6}$$

Since $p(t_j = S(t_j) - S(t_{j+1})$, this implies that:

$$h(t_j) = 1 - S(t_{j+1})/S(t_j) \tag{4.7}$$

For the discrete case, we define a *H(t)* which is defined as:

$$H(t) = -logS(t) \tag{4.8}$$

4.3 Some important models

4.3.1 The Exponential Distribution

The exponential distribution has been widely used as a model in areas ranging from studies on the lifetimes of manufactured items to research involving survival or remission times in chronic diseases. The distribution is characterised by a constant hazard function.

$$h(t) = \lambda \qquad t \succeq 0 \tag{4.9}$$

where $\lambda > 0$. The p.d.f and the survivor function can be found out to be:

$$f(t) = \lambda e^{\lambda t} \tag{4.10}$$

and

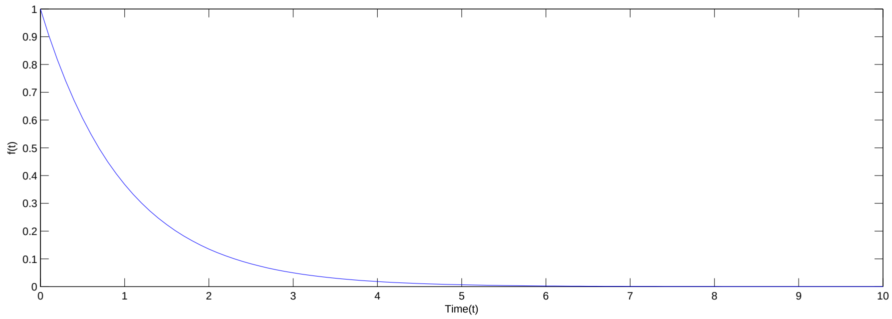
$$S(t) = e^{-\lambda t} \tag{4.11}$$

respectively. The distribution is also often written using the parametrisation $\theta = \lambda^{-1}$, in which case the p.d.f becomes:

$$f(t) = \theta^{-1} e^{-t/\theta} \qquad t \succeq 0 \tag{4.12}$$

The mean and variance of the distribution are θ and θ^2, respectively, and the pth quantile is $t_p = -\theta log(1 - p)$. The distribution where $\theta = 1$ is called the standard exponential distribution.

Figure 4.1: Exponential distribution pdf



4.3.2 The Weibull Distribution

The Weibull distribution is perhaps the most widely used lifetime distribution model. Its application in connection with lifetimes of many types of manufactured items has been widely advocated, and it has been used as a model with diverse types of items such as vacuum tubes and electrical insulation. It is also widely used in biomedical applications.

The Weibull distribution has a hazard function of the form:

$$h(t) = \lambda\beta(\lambda t)^{\beta-1} \tag{4.13}$$

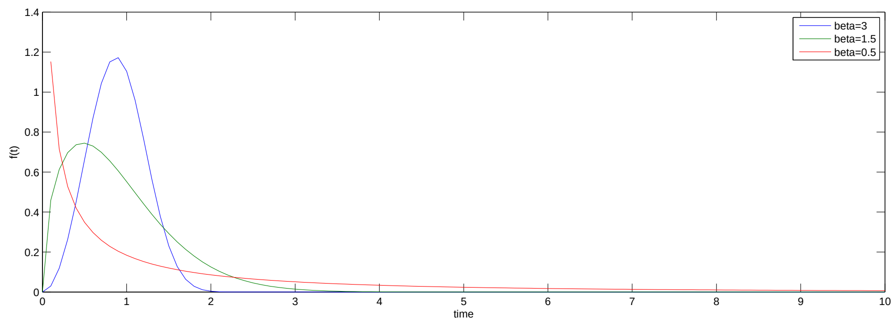
where, $\lambda > 0$ and $\beta > 0$ are parameters. For $\beta = 1$ it becomes the exponential distribution discussed in the previous section. The p.d.f and the survivor function of the distribution are:

$$f(t) = \lambda\beta(\lambda t)^{\beta-1} exp[-(\lambda t)^{beta}] \qquad t > 0 \tag{4.14}$$

and

$$S(t) = exp[-(\lambda t)^{\beta}] \qquad t > 0 \tag{4.15}$$

Figure 4.2: Weibull distribution pdf



The hazard function of the Weibull distribution is monotone increasing if $\beta > 1$ and decreasing if $\beta < 1$, and constant for $\beta = 1$. The model is fairly flexible and has been found to provide a good description of many types of lifetime data. The shape of the Weibull p.d.f depends upon the value of β, which is also called the *shape parameter* sometimes. Typically β varies from application to application.

4.3.3 The Extreme Value Distribution

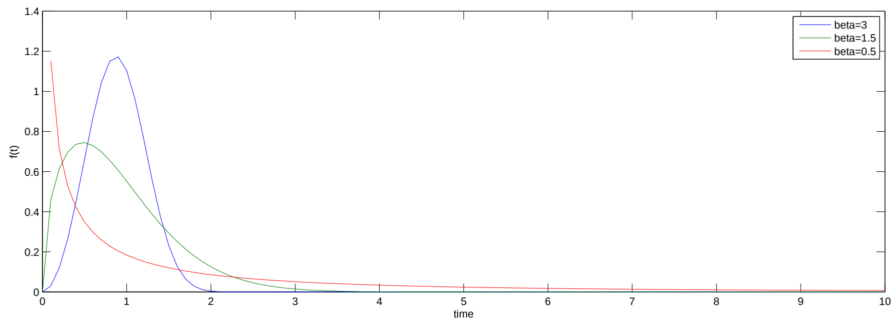
The Extreme value distribution is closely related to the Weibull distribution. This is the so-called first asymptotic distribution of extreme values, hereafter to simply as the extreme value distribution. This distribution is extensively used in a number of areas and is also referred to as *the Gumbel distribution*. The p.d.f. and survivor function of the extreme value distribution are,

$$f(x) = b^{-1}exp[\frac{x-u}{b} - exp(\frac{x-u}{b})] \qquad -\infty < x < \infty \tag{4.16}$$

and

$$S(x) = exp[-exp(\frac{x-u}{b})] \qquad -\infty < x < \infty \tag{4.17}$$

Figure 4.3: Gumbel distribution pdf



where, $b > 0$ and $u(-\infty < u < \infty)$ are parameters. This distribution is directly related to the Weibull distribution by the easily shown fact that if T has a Weibull distribution p.d.f, then $X = log(T)$ has an extreme value distribution with $b = \beta^{-1}$ and $u = -log(\lambda)$

4.4 Standard Life Tables

The life table is primarily a device for portraying the survival experience of a group of individuals, referred to as a *cohort*. The group of individuals is assumed to be a random sample from some population, in which case the life table data also provides an estimates of survival probabilities for the population. The proportion of individuals in the sample surviving to time a_j is an estimate of the probability of surviving to a_j. Let us now look at the process of completing a life table.

The time axis is divided into $k+1$ intervals $I_j = [a_{j-1}, a_j), j = 1, ..., k+1$, with a_0, and $a_{k+1} = \infty$, where T is an upper limit on observation. In the case of the last interval I_{k+1} it is assumed that all the surviving members till that point of time die in this interval. Let us now define the following quantities:

- N_j = Number of individuals *at risk* at time a_{j-1}

- D_j = Number of *deaths* in $I_j = [a_{j-1}, a_j)$

- W_j = Number of *withdrawls*(the number of samples which can no longer be surveyed due to one reason or the other) in $I_j = [a_{j-1}, a_j)$

The number of individuals alive at the start of I_j is N_j, and thus N_1 is n and

$$N_j = N_{j-1} - D_{j-1} - W_{j-1} \qquad j = 2, 3, ..., k+1 \tag{4.18}$$

Assuming that $S(t)$ is the survivor function of the lifetimes of the population under study, we define the following quantities:

- $P_j = S(a_j) = \Pr(\text{an individual survives beyond } I_j)$

- $p_j = \frac{P_j}{P_{j-1}} = \Pr(\text{an individual survives beyond } I_j \text{ — he survives beyond } I_{j-1})$

- $q_j = 1 - p_j = \Pr(\text{an individual dies in } I_j \text{ — he survives beyond } I_{j-1})$

for all these relations j varies from 1,2,...,$k+1$, with P_0 defined to be unity. Also we set $P_{k+1} = 0$ and $q_{k+1} = 1$ and P_j can be written as:

$$P_j = p_1 p_2 ... p_j \qquad j = 1, 2, ..., k+1 \tag{4.19}$$

The concept behind the life table is to employ the equations listed above to estimate the values of P_j and is based on the observation that even when there is censoring/withdrawls, it is generally possible to give sensible estimates of the $p_j's$. For a life table analysis, $q_j's$ and $p_j's$ are estimated which then yield estimates of $P_j's$ and these quantities are then displayed in a format called the life table. Considering

that there might be sensoring of data and assuming that the withdrawn individual is at risk for half the interval, a reliable estimate of q_j is:

$$\hat{q}_j = \frac{D_j}{N_j} - W_j/2 = \frac{D_j}{N_j\prime} \tag{4.20}$$

which in case of no withdrawls becomes:

$$\hat{q}_j = \frac{D_j}{N_j} \tag{4.21}$$

since $p_j = 1 - q_j$ hence an estimate of p_j , $\hat{p}_j$ is $\hat{p}_j = 1 - \hat{q}_j$ and then $\hat{P}_j = \hat{p}_1 ... \hat{p}_j$ can be found and used to complete the life table.

4.4.1 Nonparametric Estimation of the Survivor function

Using the data available to us, we can calculate an empirical distribution function from an empirical survivor function. If there are no withdrawls in a sample of size n, the empirical survivor function(ESF) is defined as:

$$S(\hat{t}) = \frac{Number of observations \succeq t}{n} \qquad t \succeq 0 \tag{4.22}$$

As can be seen, this is a step function which decreases by $\frac{1}{n}$ just after each observed lifetime if all observations are distinct. We now define the *Product limit*(PL) estimate of the survivor function, which is sometimes called the Kaplan-Meier estimate. Suppose that there are observations on n individuals and that there are $k(k \preceq n)$ distinct times $t_1 < t_2 < ... < t_k$ at which deaths occur. The possibility of there being more than one death at t_j is allowed, and we let d_j represent the number of deaths at t_j. In addition to the lifetimes $t_1 < t_2 < ... < t_k$ there are also censoring times L_i for individuals whose lifetimes are not observed. The product-limit estimate of $S(t)$

is now defined as:

$$\hat{S(t)} = \prod_{j:t_j<t} \frac{n_j - d_j}{n_j} \tag{4.23}$$

where n_j is the number of individuals who have survived till time t_{j-1}. Thus, an effective estimate of the hazard function is given by:

$$\hat{H(t)} = -log\hat{S(t)} \tag{4.24}$$

Chapter 5

Implementation of model

5.1 Graphical implementation and animation

To simulate the freeway environment Java 2D API has been used to draw up a six lane freeway with 3 lanes in either direction. The opposite directions have a barrier between them such that cars travelling forward cannot cross over to the opposite direction lanes, even in case of distractions. The cars have been modelled by a single image icon, hence assuming they are all of the same length. The repainting of the board displaying the freeway takes place every 100 milliseconds.

5.2 Implementation of Car following model

To best simulate a freeway environment and give the vehicles a sort of artificial intelligence to decide the speed they ought to be travelling at observing the vehicle ahead of them.Car-following models have been studied extensively since the early 1950s. The earliest work focused on the principle that vehicle separation is governed by safety

considerations by which distance or time headway between vehicles is a function of relative vehicle speeds. A larger number of studies have focused on calibration of parameters (, m, and l) in the GHR model(the model developed by Gazis, Herman and Rothery, 1962) and it variants. The general equation for the GHR model is:

$$a_F(t + \Delta t) = \alpha {V_F}^m (V_L(t) - V_F(t))/(X_L(t) - X_F)^l \tag{5.1}$$

where,

- a_F = acceleration of the following vehicle.
- V_F = speed of the following vehicle.
- V_L = speed of the leading vehicles.
- α , m and l are model parameters.

for $m = 0$ and $l = 2$ the GHR model turns into Greenshields car following model which we will be using in our simulation.

5.3 Implementation of car parameters

The car parameters have been varied generously to observe the impact of changes in results of the different simulations. The following is a list of parameters which differ in simulations:

- Length of each car.
- Minimum distance between cars at the time of generation.
- Average duration call after which a driver is actually distracted.

- Sensitivity coefficient, α of the Greenshields model.

- Reaction time of drivers.

Following is the list of parameters that been fixed constant for all the simulations.

- Average distance between two cars at the time of generation, taken to be 400 ft.

- Freeway speed-limit, which is 70 *miles/hr*.

- Number of call during the 24 hour simulation.

- Average waiting period for each call, which is 8 hours.

- Probability threshold for left or right lane change and of speeding while being distracted, set as 0.99.

5.4 Mathematical model implementation

The simulation is based on certain assumptions mentioned above and mathematical modelling of a lot of features involving cell phone usage while driving a car. Features such as distance between two cars at the time generation, the waiting period for all the three assumed calls, duration of these calls, initial speed of cars, behaviour of the cars while they have been distracted due to cell phone usage have all been modelled using different probability distributions mentioned in chapter 4.

5.4.1 Distance between cars

The distance between two cars at the time of generation has been implemented using exponential random number generation assuming that the arrival of cars on a freeway is a Poisson's process. The average rate of arrival of cars is set as one car per 400 ft. and to prevent cars from being generated too close to each other the minimum distance between two cars at the time of generation has been set as 100 ft.

5.4.2 Call waiting period

Like has been mentioned above, an average person receives 3 calls per day. These 3 calls are to be spread evenly in the 24 hours of driving. Thus, we model the arrival of incoming/outgoing calls as a Poisson's process, with the average call rate taken as 1 call every 8 hours. Since a person can have two consecutive calls, so no minimum separation between 2 calls has been set.

5.4.3 Call Duration

AT&T Wireless Communications commissioned the Harvard Center for Risk analysis to conduct an independent, comprehensive risk-benefit analysis of the use of cellular phones while driving. This results of this study for call durations can be summarised in the table below:

Using this survey, we assume the call durations to be a Gaussian Random variable, with mean call duration as 120 seconds and variance as 60 seconds. The same parameters have been used for all the 3 calls expected within the 24 hours of driving.

Table 5.1: Cellular Phone Call Duration While Driving

Call Duration while Driving	Percentage of Correspondents
Less than 30 seconds	23%
30 sec to 2 minutes	37%
2 to 5 minutes	18%
more than 5 minutes	4%
Do not talk while driving	10%
Did not respond	8%

5.4.4 Initial speed of cars

Using the fact that the speed limit for the I-15 freeway going through Las Vegas is 70 *m.p.h.* The initialisation speed of vehicles has been set using Gaussian distribution again. The mean speed for generation of vehicles as 61.5 *m.p.h* and the variance has been set as 6.1 *m.p.h* Since the speeds of the vehicles has a maximum limit the initialisation speed has been set to be below the speed limit of the freeway and a lower limit for speed of generated cars has also been set in trying to make the simulation closer to reality.

5.4.5 General behaviour of cars

For a vast majority of the time while the cars are on the freeway, they are calls free and subsequently distraction free. This is a result of the assumption that the only distraction cause in this simulation can be due to scenario of a driver being on a call. Thus, the Greenshields model of car following has been used to model the

behaviour of the cars during the period when they are not distracted. Also a free-speeding distance has been set which makes sure that the speed of the vehicles can be independent of the vehicles ahead of them when the distance between the two is more than this free-speeding distance.

5.4.6 Call duration for cars to be distracted

The simulation take into account the fact that not every call can distract a driver. Thus, modelling of the distracted behaviour has been done such that it takes a driver 1 hour on an average to get distracted, if this period is modelled as a Poisson's process and generated using exponentially random numbers. The Transportation Research Center at UNLV is also working on developing physiological sensors which would help study the distractions better. The following sensors are currently under development at the center:

- Electrocardiogram(ECG)
- Electroencephalogram(EEG)
- Galvanic Skin Response(GSR)

The center intends to study the above three responses to model the reactions of various drivers while they are distracted during the driving process. The aim is to provide the software with better numerical values to model the time required for entering the distraction phase and consequentially help us to model the phenomena of distraction better.

5.4.7 Distracted behaviour

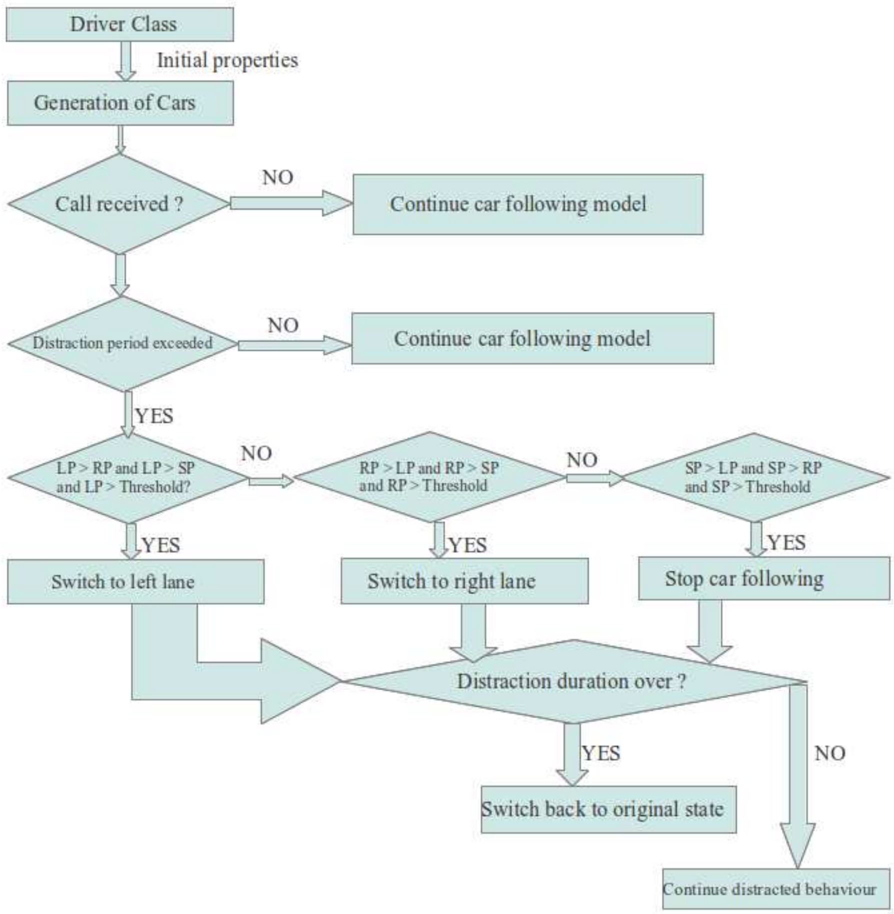
The behaviour of the driver while being distracted has been set to include one of the four things listed below:

- Moving in to the left lane.
- Moving in to the right lane.
- Speeding up.
- Speeding down.

For the left and right lane changes, there are certain restrictions while prevent the vehicles to cross over to the lanes travelling in the opposite direction, assuming that there is a physical barrier between the two directions. A series of uniformly random numbers is generated while the cars are in their *distraction period* and thresholds have been set for all the three possibilities listed above(speeding up and down have been clubbed together into speed changing). Once these thresholds are crossed the vehicles may make any of the state changes listed above. For the changing of speed, which actually means the cars stop moving according to the car following model for the period of distraction. Also once the drivers are distracted, we set a period of 3 seconds after which if the cars have switched to any of the states listed above, switch back to their natural states. The following is a flowchart representing the overall

working of the software.

Figure 5.1: Flowchart illustrating the working of the software



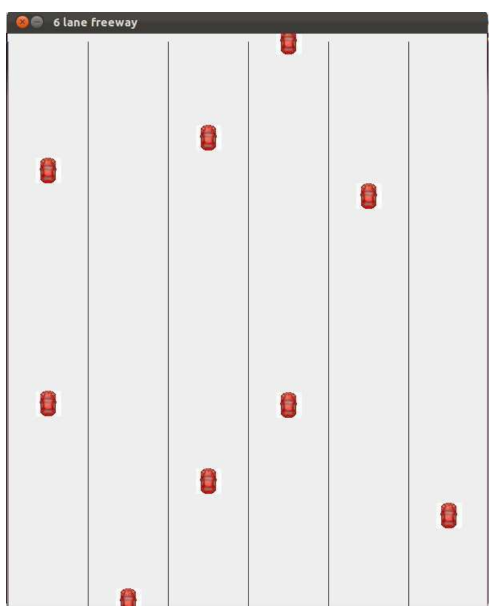
5.4.8 Crash detection and reporting

The car length has been set as 10 ft and once the cars are within this range of each other, a crash is detected and reported. A crash report includes the following details:

- Crash location.
- Indices of the vehicles involved in the crash.
- Direction in which the cars were travelling.
- Initial speeds of these cars.
- Speeds of these cars at the time of crash.
- Time at which the crash occurs.

The following is a screen shot of the simulation through the software:

Figure 5.2: Screen shot of the simulation



Chapter 6

Results

Using the implementation technique described in chapter 6, three different simulations were run and life tables were created using these observations for the product-limit estimate of $S(t)$ and then the estimate of $H(t)$.

6.1 Analysis of first simulation

The following are the life tables for the four different simulations run for a period of 24 hours to observe the number of crashes and hence number of deaths given the pre-set parameters for the cars. We discussed some important models for life time data analysis in chapter 5 and the life table data will be used to see if we can use any of the models discussed. Here we begin with the analysis of the first of the three 24 hour simulations.

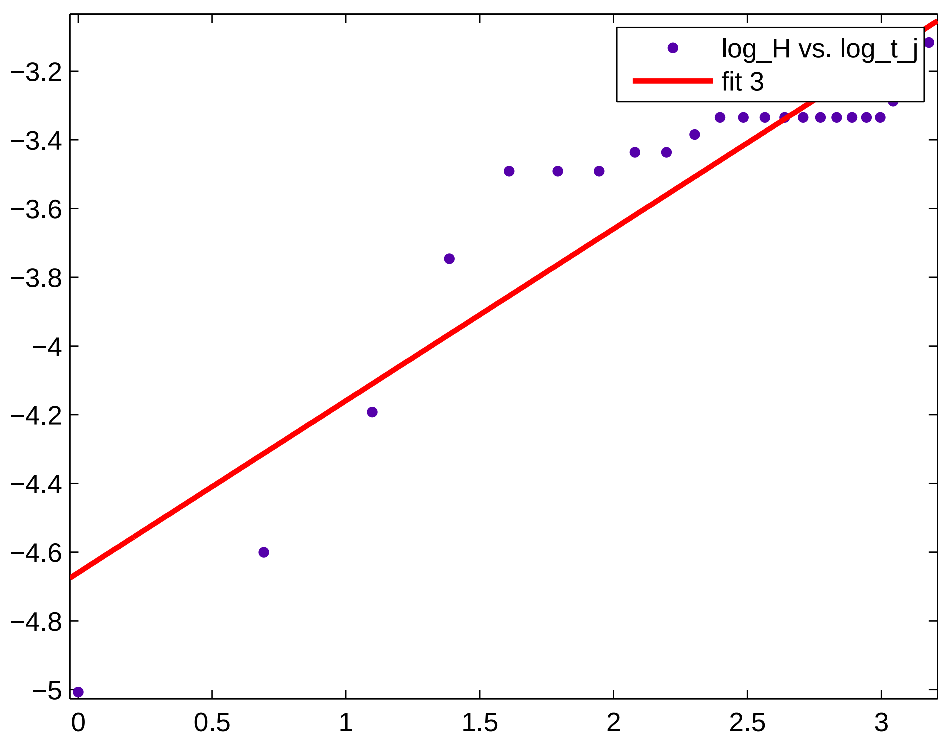
Table 6.1: 1st simulation results

t_j	n_j	d_j	$S(\hat{t_j})$	$H(\hat{t_j})$	$logH(\hat{t_j})$	$logt_j$
1	1200	8	0.993333	0.006689	-5.007292664	0
2	1192	4	0.99	0.01005	-4.600149227	0.693147
3	1188	6	0.985	0.015114	-4.192157777	1.098612
4	1182	10	0.976667	0.02361	-3.746090619	1.386294
5	1172	8	0.97	0.030459	-3.49136695	1.609438
6	1164	0	0.97	0.030459	-3.49136695	1.791759
7	1164	0	0.97	0.030459	-3.49136695	1.94591
8	1164	2	0.968333	0.032179	-3.436444372	2.079442
9	1162	0	0.968333	0.032179	-3.436444372	2.197225
10	1162	2	0.966667	0.033902	-3.384294493	2.302585
11	1160	2	0.965	0.035627	-3.334646515	2.397895
12	1158	0	0.965	0.035627	-3.334646515	2.484907
13	1158	0	0.965	0.035627	-3.334646515	2.564949
14	1158	0	0.965	0.035627	-3.334646515	2.639057
15	1158	0	0.965	0.035627	-3.334646515	2.70805
16	1158	0	0.965	0.035627	-3.334646515	2.772589
17	1158	0	0.965	0.035627	-3.334646515	2.833213
18	1158	0	0.965	0.035627	-3.334646515	2.890372
19	1158	0	0.965	0.035627	-3.334646515	2.944439
20	1158	0	0.965	0.035627	-3.334646515	2.995732
21	1158	2	0.963333	0.037356	-3.287267452	3.044522
22	1156	4	0.96	0.040822	-3.198534261	3.091042
23	1152	0	0.96	0.040822	-3.198534261	3.135494
24	1152	4	0.956667	0.0443	-3.116764758	3.178054

6.1.1 Life table

Table 7.1 can be used to study the relationships between $S(\hat{t}_j)$ and t_j

6.1.2 plot of $logH(\hat{t}_j)$ against $logt_j$



The graph shows that a roughly linear relationship can be found between $logH(\hat{t}_j)$ and $logt_j$. In the next section we will compute the parameters of such a possible curve.

6.1.3 Calculation of slope and intercept

For weibull distribution, there should be a linear relation between $logH(\hat{t}_j)$ against $logt_j$. Using MATLAB's curve fitting tool, we can determine the slope and intercept of such a possible relationship. Assuming that the Linear model Polynomial is of the

form: $f(x) = p1 * x + p2$ we find that the Coefficients (with 95% confidence bounds) are as below:

- p1 = 0.5002 (0.4064, 0.594)
- p2 = -4.66 (-4.887, -4.432)

6.1.4 Goodness of fit

Assuming that the relationship turns out to be linear, we calculate using MATLAB's curve fitting tool the "The goodness of fit"

- SSE: 0.7143
- R-square: 0.8475
- Adjusted R-square: 0.8406
- RMSE: 0.1802

6.1.5 Calculation of β and λ for the distribution

Using the results above we can assume that there is a linear relationship between $logH\hat{(}t_j)$ against $logt_j$. We know that for Weill distribution, this relationship is:

$$log(H(t)) = \beta log\lambda + \beta logt \tag{6.1}$$

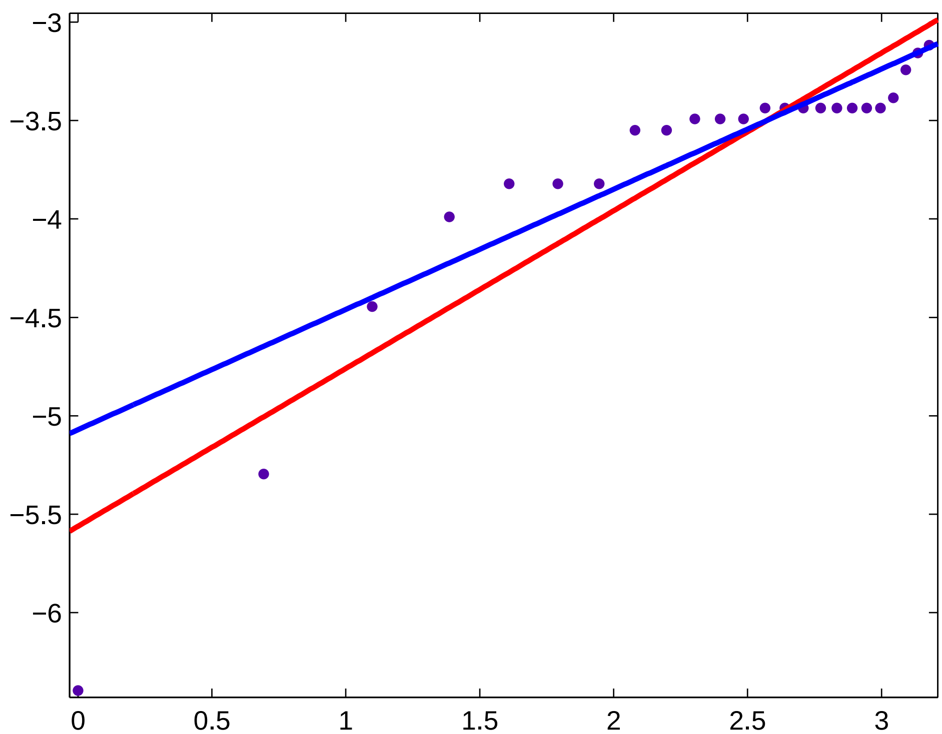
substituting the values obtained from our simulation we get the following values of the weibull parameters:

- $\beta = 0.5002$
- $\lambda = 0.00009$

6.2 Analysis of the second simulation

6.2.1 Life Table

6.2.2 plot of $logH(\hat{t}_j)$ against $logt_j$



The graph shows that a roughly linear relationship can be found between $logH(\hat{t}_j)$ and $logt_j$. In the next section we will compute the parameters of such a possible curve.

6.2.3 Calculation of slope and intercept

For weibull distribution, there should be a linear relation between $logH(\hat{t}_j)$ against $logt_j$. Using MATLAB's curve fitting tool, we can determine the slope and intercept of such a possible relationship. Assuming that the Linear model Polynomial is of the

Table 6.2: Add caption

Time	N_j	D_j	S_j	H_j	log_log_S_j	log_t
1	1200	2	0.998333	0.001668	-6.396095743	0
2	1198	4	0.995	0.005013	-5.295812143	0.693147
3	1194	8	0.988333	0.011735	-4.445157616	1.098612
4	1186	8	0.981667	0.018503	-3.989796912	1.386294
5	1178	4	0.978333	0.021905	-3.821047873	1.609438
6	1174	0	0.978333	0.021905	-3.821047873	1.791759
7	1174	0	0.978333	0.021905	-3.821047873	1.94591
8	1174	8	0.971667	0.028742	-3.549379499	2.079442
9	1166	0	0.971667	0.028742	-3.549379499	2.197225
10	1166	2	0.97	0.030459	-3.49136695	2.302585
11	1164	0	0.97	0.030459	-3.49136695	2.397895
12	1164	0	0.97	0.030459	-3.49136695	2.484907
13	1164	2	0.968333	0.032179	-3.436444372	2.564949
14	1162	0	0.968333	0.032179	-3.436444372	2.639057
15	1162	0	0.968333	0.032179	-3.436444372	2.70805
16	1162	0	0.968333	0.032179	-3.436444372	2.772589
17	1162	0	0.968333	0.032179	-3.436444372	2.833213
18	1162	0	0.968333	0.032179	-3.436444372	2.890372
19	1162	0	0.968333	0.032179	-3.436444372	2.944439
20	1162	0	0.968333	0.032179	-3.436444372	2.995732
21	1162	2	0.966667	0.033902	-3.384294493	3.044522
22	1160	6	0.961667	0.039087	-3.241955403	3.091042
23	1154	4	0.958333	0.04256	-3.156849494	3.135494
24	1150	2	0.956667	0.0443	-3.116764758	3.178054

form: $f(x) = p1 * x + p2$ we find that the Coefficients (with 95% confidence bounds) are as below:

- p1 = 0.8016 (0.6462, 0.957)
- p2 = -5.561 (-5.938, -5.185)

6.2.4 Goodness of fit

Assuming that the relationship turns out to be linear, we calculate using MATLAB's curve fitting tool the "The goodness of fit"

- SSE: 1.96
- R-square: 0.8387
- Adjusted R-square: 0.8314
- RMSE: 0.2985

As we can see the values of the "goodness of fit" parameters are not good. We now use the robust least square analysis to determine a better relationship. We will exclude the outliers from the first curve and then calculate the parameters again. Here we take the first data point as the outlier since it lies a large distance of the estimated line as seen in the plot. Re-evaluating the data now will give the following results:

6.2.5 Calculation of slope and intercept using robust least square analysis

As described above, in this section we re-calculate the slope and intercept excluding the "outliers" again, assuming that the Linear model Polynomial is of the form: $f(x) = p1 * x + p2$ we find that the Coefficients (with 95% confidence bounds) are as below:

- p1 = 0.6108 (0.481, 0.7406)
- p2 = -5.07 (-5.392, -4.749)

6.2.6 Goodness of fit

Assuming that the relationship turns out to be linear, we calculate using MATLAB's curve fitting tool the "The goodness of fit"

- SSE: 1.96
- R-square: 0.8387
- Adjusted R-square: 0.8314
- RMSE: 0.2985

6.2.7 Calculation of β and λ for the distribution

Using the results above we can assume that there is a linear relationship between $logH(\hat{t_j})$ against $logt_j$. We know that for weibull distribution, this relationship is:

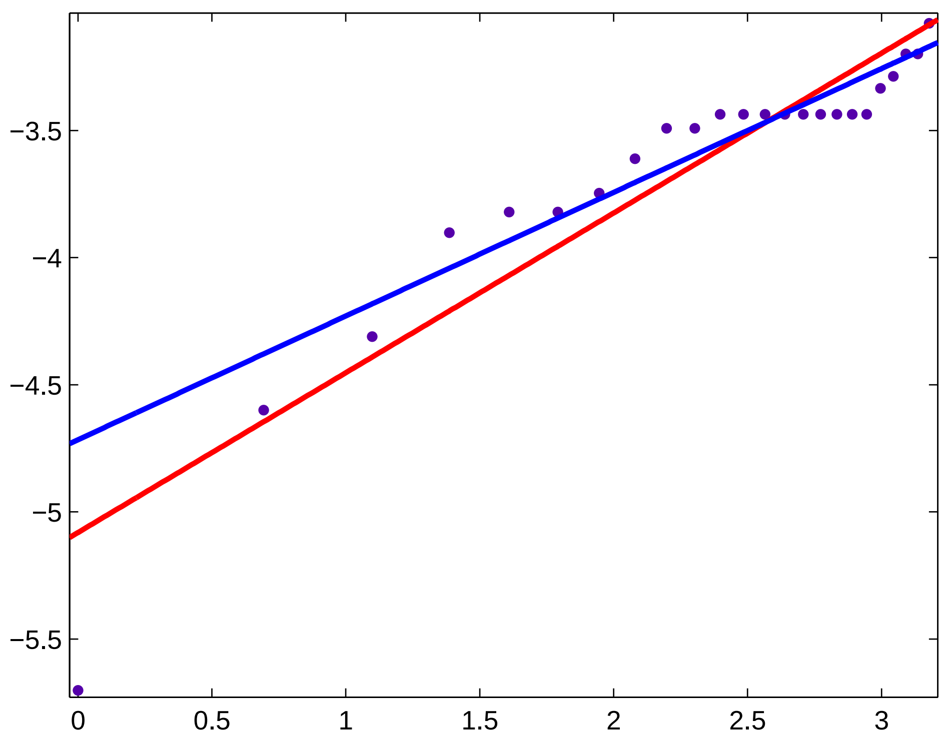
$$log(H(t)) = \beta log\lambda + \beta logt \tag{6.2}$$

substituting the values obtained from our simulation we get the following values of the weibull parameters:

- $\beta = 0.6108$
- $\lambda = 0.00024$

6.3 Analysis of the third simulation

6.3.1 Life Table



The graph shows that a roughly linear relationship can be found between $logH(\hat{t}_j)$ and $logt_j$. In the next section we will compute the parameters of such a possible curve.

Table 6.3: Add caption

Time	N_j	D_j	S_j	H_j	log_log_S_j	log_t
1	1200	4	0.996667	0.003339	-5.702113489	0
2	1196	8	0.99	0.01005	-4.600149227	0.693147
3	1188	4	0.986667	0.013423	-4.310784111	1.098612
4	1184	8	0.98	0.020203	-3.901938658	1.386294
5	1176	2	0.978333	0.021905	-3.821047873	1.609438
6	1174	0	0.978333	0.021905	-3.821047873	1.791759
7	1174	2	0.976667	0.02361	-3.746090619	1.94591
8	1172	4	0.973333	0.027029	-3.610857036	2.079442
9	1168	4	0.97	0.030459	-3.49136695	2.197225
10	1164	0	0.97	0.030459	-3.49136695	2.302585
11	1164	2	0.968333	0.032179	-3.436444372	2.397895
12	1162	0	0.968333	0.032179	-3.436444372	2.484907
13	1162	0	0.968333	0.032179	-3.436444372	2.564949
14	1162	0	0.968333	0.032179	-3.436444372	2.639057
15	1162	0	0.968333	0.032179	-3.436444372	2.70805
16	1162	0	0.968333	0.032179	-3.436444372	2.772589
17	1162	0	0.968333	0.032179	-3.436444372	2.833213
18	1162	0	0.968333	0.032179	-3.436444372	2.890372
19	1162	0	0.968333	0.032179	-3.436444372	2.944439
20	1162	4	0.965	0.035627	-3.334646515	2.995732
21	1158	2	0.963333	0.037356	-3.287267452	3.044522
22	1156	4	0.96	0.040822	-3.198534261	3.091042
23	1152	0	0.96	0.040822	-3.198534261	3.135494
24	1152	6	0.955	0.046044	-3.078159154	3.178054

6.3.2 Calculation of slope and intercept

For weibull distribution, there should be a linear relation between $logH(\hat{t_j})$ against $logt_j$. Using MATLAB's curve fitting tool, we can determine the slope and intercept of such a possible relationship. Assuming that the Linear model Polynomial is of the form: $f(x) = p1 * x + p2$ we find that the Coefficients (with 95% confidence bounds) are as below:

- p1 = 0.6284 (0.5262, 0.7305)
- p2 = -5.081 (-5.329, -4.834)

6.3.3 Goodness of fit

Assuming that the relationship turns out to be linear, we calculate using MATLAB's curve fitting tool the "The goodness of fit"

- SSE: 0.8464
- R-square: 0.8809
- Adjusted R-square: 0.8755
- RMSE: 0.1961

As we can see the values of the "goodness of fit" parameters are not good. We now use the robust least square analysis to determine a better relationship. We will exclude the outliers from the first curve and then calculate the parameters again. Here we take the first data point as the outlier since it lies a large distance of the

estimated line as seen in the plot. Re-evaluating the data now will give the following results:

6.3.4 Calculation of slope and intercept using robust least square analysis

As described above, in this section we re-calculate the slope and intercept excluding the ”outliers” again, assuming that the Linear model Polynomial is of the form: $f(x) = p1 * x + p2$ we find that the Coefficients (with 95% confidence bounds) are as below:

- p1 = 0.4865 (0.4185, 0.5545)
- p2 = -4.716 (-4.884, -4.548)

6.3.5 Goodness of fit

Assuming that the relationship turns out to be linear, we calculate using MATLAB's curve fitting tool the ”The goodness of fit”

- SSE: 0.2341
- R-square: 0.9133
- Adjusted R-square: 0.9092
- RMSE: 0.1056

6.3.6 Calculation of β and λ for the distribution

Using the results above we can assume that there is a linear relationship between $logH(\hat{t_j})$ against $logt_j$. We know that for weibull distribution, this relationship is:

$$log(H(t)) = \beta log\lambda + \beta logt \tag{6.3}$$

substituting the values obtained from our simulation we get the following values of the weibull parameters:

- $\beta = 0.4865$
- $\lambda = 0.00006$

Chapter 7

Conclusions

7.1 Development of the freeway simulator

In this thesis, a java based software was developed , which simulates a six lane freeway, wit three lanes each in either direction. The software gives us the option of choosing the number of cars to be generated, various other characteristics and initial conditions can be alloted probabilistically as described in chapter 5 such as:

- Initial speed.
- Initial position.
- Probability of carrying a cell phone.
- Age of the driver.
- Call waiting periods.
- Call durations.

- Distraction periods.
- Speed of vehicles during simulation(Using Greenshields car following model)
- Behaviour of the cars in the distraction period.

Using this simulation, crashes and crash timings were observed and it was concluded that it is useful to study these crashes using Weibull Distribution model.

7.2 Future work

In the future, more details can be added to the software. E.g. inclusion of ramps and exits which have been excluded in this version. A survey can be done to find out the actual average time taken by a driver on call to be distracted. The center is also working on developing physiological sensors which will help study and model the human responses to different distractions and thus help us assign more accurate values for distraction thresholds. This inclusion would help us model the freeway traffic more effectively.

Bibliography

[1] de Moivre, Abraham , *The Doctrine of Chances. ISBN 0821821032.*. 1738.

[2] la Place, M. de . *"Mmoire sur la probabilit des causes par les vnemens". Mmoires de Mathmatique et de Physique, Presents l'Acadmie Royale des Sciences, par divers Savans & ls dans ses Assembles, Tome Sixime: 621656. Translated by S.M.Stigler in Statistical Science 1 1986: JSTOR 2245476* 1774

[3] Gauss, Carolo Friderico (in Latin). *Theoria motvs corporvm coelestivm in sectionibvs conicis Solem ambientivm [Theory of the motion of the heavenly bodies moving about the Sun in conic sections]* 1809

[4] [4] Patel, Jagdish K.; Read, Campbell B. *Handbook of the normal distribution (2nd ed.). CRC Press. ISBN 0-824-79342-0.* 1996

[5] [5] Fan, Jianqing. "On the optimal rates of convergence for nonparametric deconvolution problems". *The Annals of Statistics 19 (3): 12571272. doi:10.1214/aos/1176348248. JSTOR 2241949.* 1991

[6] [6] J.F.Lawless *Statistical models and methods for lifetime data.* Wiley series in probability and mathematical statistics

VITA

Sourabh Sriom

1861, Arbol Verde Way, Las Vegas, Nevada

Email: sourabh.iitg@gmail.com

Phone no: 702-302-6087

Education:

- Master of Science in Electrical & Computer Engineering from University of Nevada, Las Vegas(2010-11), GPA = 3.93/4. Master's thesis : *Development of a simulation tool for analysis of freeway crashes due to cell phone usage.*

- Bachelor of Technology in Electronics & Communication Engineering from Indian Institute of Technology Guwahati(2005-09), GPA = 7.15/10

Publications:

- Sourabh Sriom, Mohit Khurana, Puneet Lakhanpal and Dr.Amit Kumar Mishra, *"Robust text independent speaker recognition based on hybrid LPC and MFCC algorithm"*, 15th International Conference on Advanced Computing and Com- munication(ADCOM 07). Published by the IEEE Computer Society press.

- Sourabh Sriom, Mukund Jha, Mirium Lujan, CD Martinez and Alberto Sanchis, *"Improved supervised speech recognition system using MLLR speaker adaptation and confidence measurement"*, V Biennial workshop on speech technology, 2008.