Highway Safety Analytics and Modeling

Dominique Lord
Xiao Qin
Srinivas R. Geedipally

ELSEVIER

Highway Safety Analytics and Modeling

by

Dominique Lord, Texas A&M University
Xiao Qin, University of Wisconsin-Milwaukee
Srinivas R. Geedipally, Texas A&M Transportation Institute

To be published by Elsevier

The primary purpose of this textbook is to provide information for practitioners, engineers, scientists and researchers who are interested in analyzing safety data in order to make engineering- or policy-based decisions. This book provides the latest tools and methods documented in the literature for analyzing crash data, some of which have in fact been developed or introduced by the authors. The textbook covers all aspects of the decision-making process, from collecting and assembling data to making decisions based on the results of the analyses. Several examples and case studies are provided to help understand models and methods commonly used for analyzing crash data. Where warranted, helpful hints and suggestions are provided by the authors in the text to support the analysis and interpretation of crash data.

(With the exception of Chapter 1, the word counts for all chapters are between 12,000 and 15,000 words)

CHAPTER 1 – INTRODUCTION

1.1 MOTIVATION

1.2 IMPORTANT FEATURES OF THIS TEXTBOOK

1.3 ORGANIZATION OF TEXTBOOK

1.3.1 Part I: THEORY AND BACKBROUND

1.3.2 Part II: HIGHWAY SAFETY ANALYSES

1.3.3 Part III: ALTERNATIVE SAFETY ANALYSES

1.3.4 Appendices

1.4 FUTURE CHALLENGES AND OPPORTUNITIES

1.5 REFERENCES

Part I: THEORY AND BACKBROUND

CHAPTER 2 – FUNDAMENTALS AND DATA COLLECTION

2.1 INTRODUCTION

2.2 CRASH PROCESS: DRIVERS, ROADWAYS, AND VEHICLES

2.3 CRASH PROCESS: ANALYTICAL FRAMEWORK

2.4 SOURCES OF DATA AND DATA COLLECTION PROCEDURES

2.4.1 Traditional Data

2.4.1.1 Crash Data

2.4.1.2 Roadway Data

2.4.1.3 Traffic Flow Data

2.4.1.4 Supplemental Data

2.4.1.5 Other Safety-Related Data and Relevant Databases

2.4.2 Naturalistic Driving Data

2.4.3 Disruptive Technological and Crowdsourcing Data

2.4.4 Data Issues

2.5 ASSEMBLING DATA

2.6 4-STAGE MODELING FRAMEWORK

2.6.1 Determine Modeling Objective Matrix

2.6.2 Establish Appropriate Process to Develop Models

2.6.3 Determine Inferential Goals

2.6.4 Select Computation Techniques and Tools

2.7 METHODS FOR EVALUATING MODEL PERFORMANCE

2.7.1 Likelihood-Based Methods

2.7.2 Error-Based Methods

2.8 HEURISTIC METHODS FOR MODEL SELECTION

2.9 REFERENCES

CHAPTER 3 - CRASH–FREQUENCY MODELING

3.1 INTRODUCTION

3.2 BASIC NOMENCLATURE

3.3 APPLICATIONS OF CRASH-FREQUENCY MODELS

4.3.1 Understanding Relationships

4.3.2 Screening Variables

3.3.3 Sensitivity of Variables

3.3.4 Prediction

3.3.5 Causal Relationships

3.4 SOURCES OF DISPERSION

3.4.1 Over-Dispersion

3.4.2 Under-Dispersion

3.5 BASIC COUNT MODELS

3.5.1 Poisson Model

3.5.2 Negative Binomial Model

3.5.3 Poisson-Lognormal Model

3.5.4 Other Poisson-Mixture Models

3.6 GENERALIZED COUNT MODELS FOR UNDER-DISPERSION

3.6.1 Conway-Maxwell-Poisson Model

3.6.2 Other Generalized Models

3.7 FINITE MIXTURE AND MULTIVARIATE MODELS

3.7.1 Finite Mixture Models

3.7.2 Multivariate Models

3.8 MULTIVARIABLE MODELS

3.8.1 Negative Binomial-Lindley Model

3.8.2 Other Multivariable Models

3.9 MODELS FOR BETTER CAPTURING UNOBSERVED HETEROGENEITY

3.9.1 Random-Effects/Multilevel Model

3.9.2 Random-Parameters Model

3.10 SEMI- AND NON-PARAMETRIC MODELS

3.10.1 Semi-Parametric Models

3.10.2 Dirichlet Process Models

3.10.3 Non-Parametric Models

3.11 MODEL SELECTION

3.12 REFERENCES

CHAPTER 4 – CRASH-SEVERITY MODELING

4.1 INTRODUCTION

4.2 CHARACTERISTICS OF CRASH INJURY SEVERITY DATA AND METHODOLOGICAL CHALLENGES

4.2.1 Ordinal Nature of Crash Injury Severity Data

4.2.2 Unobserved Heterogeneity

4.2.3 Omitted Variable Bias

4.2.4 Imbalanced Data between Injury Severity Levels

4.3 RANDOM UTILITY MODEL

4.4 MODELING CRASH SEVERITY AS AN UNORDERED DISCRETE OUTCOME

4.4.1 Multinomial Logit Model

4.4.2 Nested Logit Model

4.4.3 Mixed Logit Model

4.5 MODELING CRASH SEVERITY AS AN ORDERED DISCRETE OUTCOME

4.5.1 Ordinal Probit/Logistic Model

4.5.2 Generalized Ordered Logistic and Proportional Odds Model

4.5.3 Sequential Logistic/Probit Regression Model

4.6 MODEL INTERPRETATION

4.7 REFERENCES

Part II: HIGHWAY SAFETY ANALYSES

CHAPTER 5 – EXPLORATORY ANALYSES OF SAFETY DATA

5.1 INTRODUCTION

5.2 QUANTITATIVE TECHNIQUES

5.2.1 Measures of Central Tendency

5.2.2 Measures of Variability

5.2.3 Measures of Association

5.2.4 Confidence Intervals

5.2.5 Hypothesis Testing

5.3 GRAPHICAL TECHNIQUES

5.3.1 Box-And-Whisker Plot

5.3.2 Histogram

5.3.3 Bar Graphs

5.3.4 Error Bars

5.3.5 Pie charts

5.3.6 Scatterplots

5.3.7 Bubble Chart

5.3.8 Radar/Web Plot

5.3.9 Heatmap

5.3.10 Contour Plot

5.3.11 Population Pyramid

5.4 REFERENCES

CHAPTER 6 – CROSS-SECTIONAL AND PANEL STUDIES IN SAFETY

6.1 INTRODUCTION

6.2 TYPES OF DATA

6.2.1 Time Series Data

6.2.2 Cross-Sectional Data

6.2.3 Panel Data

6.3 DATA AND MODELING ISSUES

6.3.1 Over-dispersion and Under-dispersion

6.3.2 Low Sample Mean and Small Sample Size

6.3.3 Under-Reporting

6.3.4 Omitted Variables Bias

6.3.5 Endogenous Variables

6.3.6 Unobserved Heterogeneity

6.4 DATA AGGREGATION

6.5 APPLICATION OF CRASH-FREQUENCY AND CRASH-SEVERITY MODELS

6.5.1 Functional Form

6.5.1.1 Flow-Only Models

6.5.1.2 Flow-only Models with CMFs

6.5.1.3 Model with Covariates

6.5.2 Variable Selection

6.5.3 Crash Variance and Confidence Intervals

6.5.4 Sample Size Determination

6.5.5 Outlier Analysis

6.5.6 Model Transferability

6.6 OTHER STUDY TYPES

6.6.1 Cohort Studies

6.6.2 Case-Control Studies

6.6.3 Randomized trials

6.7 REFERENCES

CHAPTER 7 – BEFORE–AFTER STUDIES IN HIGHWAY SAFETY

7.1 INTRODUCTION

7.2 CRITICAL ISSUES WITH BEFORE-AFTER STUDIES

7.2.1 Regression-to-the-Mean

7.2.2 Site Selection Bias

7.3 BASIC METHODS

7.3.1 Simple Before-After Study

7.3.2 Before-After Study Using Control Groups

7.4 BAYESIAN METHODS

7.4.1 Empirical Bayes Method

7.4.2 Bayes Method

7.5 ADJUSTING FOR SITE SELECTION BIAS

7.6 PROPENSITY SCORE MATCHING METHOD

7.7 BEFORE-AFTER STUDY USING SURVIVAL ANALYSIS

7.8 SAMPLE SIZE CALCULATIONS

7.8.1 Factors Influencing Sample Size Calculations

7.8.2 Sample Size Estimation using Known Crash Counts for Both Time Periods

7.8.3 Sample Size Based on the Variance and Ratio r_d (Before Period)

7.9 REFERENCES

CHAPTER 8 – IDENTIFICATION OF HAZARDOUS SITES

8.1 INTRODUCTION

8.2 OBSERVED CRASH METHODS

8.2.1 Crash Frequency Method

8.2.2 Crash Rate Method

8.2.3 Rate Quality Control Method

8.2.4 Equivalent Property Damage Only (EPDO) Method

8.2.5 Severity Index Method

8.2.6 Composite Safety Score

8.3 PREDICTED CRASH METHODS

8.3.1 Potential for Improvement using Predicted Crashes

8.3.2 Level of Service of Safety

8.4 BAYESIAN METHODS

8.4.1 Empirical Bayes Method

8.4.2 Bayes Method

8.5 COMBINED CRITERIA

8.6 GEOSTATISTICAL METHODS

8.6.1 Clustering Methods

8.6.1.1 K-means clustering

8.6.1.2 Ripley's K-function

8.6.1.3 Nearest neighborhood clustering

8.6.1.4 Moran’s I Index

8.6.1.5 Getis-Ord General G^* (d)

8.6.2 Kernel Density Estimation

8.7 CRASH CONCENTRATION LOCATION METHODS

8.7.1 Sliding Window Method

8.7.2 Peak Searching Method

8.7.3 Continuous Risk Profile (CRP)

8.8 PROACTIVE METHODS

8.9 EVALUATING SCREENING METHODS

8.10 REFERENCES

CHAPTER 9 – MODELS FOR SPATIAL DATA

9.1 INTRODUCTION

9.2 SPATIAL DATA AND DATA MODELS

9.3 MEASUREMENTS OF SPATIAL ASSOCIATION

9.3.1 Global Statistics for Spatial Association

9.3.1.1 Getis-Ord General $G_i^*(d)$

9.3.1.2 Moran's I

9.3.2 Local Indicators of Spatial Association

9.3.2.1 Local $G_i^*(d)$

9.3.2.2 Local Moran's I_i

9.4 SPATIAL WEIGHTS AND DISTYANCE DECAY MODELS

9.5 POINT DATA ANALYSIS

9.5.1 First- and Second-order Process

9.5.2 Kernel Density Estimation

9.5.3 Ripley's K-function

9.5.4 Cross-K Function

9.5.5 Spatial Regression Analysis

9.5.6 Spatial Econometrics Methods

9.5.6.1 Spatial Autoregressive Model

9.5.6.2 Spatial Error Model

9.6.2 Generalized Linear Model with Spatial Correlation

9.6.2.1 Generalized Linear Mixed Model (GLMM)

9.6.2.2 Hierarchical Bayesian Model

9.6.3 Modeling Local Relationships in Crash Data

9.7 REFERENCES

CHAPTER 10 – CAPACITY, MOBILITY, AND SAFETY

10.1 INTRODUCTION

10.2 MODELING SPACE BETWEEN VEHICLES

10.3 SAFETY AS A FUNCTION OF TRAFFIC FLOW

10.4 CHARACTERIZING CRASHES BY REAL-TIME TRAFFIC

10.5 PREDICTING IMMINENT CRASH LIKELIHOOD

10.6 REAL-TIME PREDICTIVE ANALYSIS OF CRASHES

10.6.1 Binary Logistic Regression Model

10.6.2 Conditional Logistic Regression Model

10.6.3 A Note about Binary Logit and Conditional Logistic Regression Models

10.7 USING TRAFFIC SIMULATION TO PREDICT CRASHES

10.7.1 Cell Transmission Model (CTM)

10.7.2 Fundamental Diagram (FD) Calibration

10.7.3 CTM Simulation Algorithm

10.7.4 Crash Modeling

10.7.5 Crash Prediction

10.8 REFERENCES

Part III: ALTERNATIVE SAFETY ANALYSES

CHAPTER 11 – SURROGATE SAFETY MEASURES

11.1 INTRODUCTION

11.2 AN HISTORICAL PERSPECTIVE

11.3 TRAFFIC CONFLICTS TECHNIQUE

11.4 FIELD SURVEY OF TRAFFIC CONFLICTS

11.5 PROXIMAL SURROGATE SAFETY MEASURES

11.5.1 Collision Course

11.5.2 Time- and Distance-based Proximal Surrogate Safety Measures

11.5.2.1 Time-to-Collision Family

11.5.2.2 Encroachment Time Family

11.5.2.3 Proportion of Stopping Distance (PSD)

11.5.2.4 Other Indicators

11.6 THEORETICAL DEVELOPMENT OF SAFETY SURROGATE MEASURES

11.6.1 Block Maxima (BM) using the Generalized Extreme Value (GEV) Distribution

11.6.2 Peak Over Threshold (POT) Using the Generalized Pareto (GP) Distribution

11.6.3 Block Maxima (BM) or Peak Over Threshold (POT)

11.7 SAFETY SURROGATE MEASURES FROM TRAFFIC MICROSIMULATION MODELS

11.8 SAFETY SURROGATE MEASURES FROM VIDEO AND EMERGING DATA SOURCES

11.9 REFERENCES

CHAPTER 12 – DATA MINING AND MACHINE LEARNING TECHNIQUES

12.1. INTRODUCTION

12.2. ASSOCIATION RULES

12.3. CLUSTERING ANALYSIS (CA)

12.3.1. K-means Clustering (KC)

12.3.2. Latent Class Cluster (LCC)

12.4. DECISION TREE MODEL

12.4.1. The CART Model

12.4.2. Random Forests (RF)

12.4.3. Gradient Boosted Trees (GBT)

12.5. BAYESIAN NETWORKS

12.6. NEURAL NETWORK

12.6.1. Multilayer Perceptron (MLP) Neural Network

12.6.2. Convolutional Neural Networks (CNN)

12.6.3. Long Short-Term Memory - Recurrent Neural Networks (LSTM-RNN)

12.6.4 Bayesian Neural Networks (BNN)

12.7. SUPPORT VECTOR MACHINES (SVM)

12.8. SENSITIVITY ANALYSIS

12.9 REFERENCES

APPENDICES

Appendix A: Negative Binomial Regression Models and Estimation Methods
Appendix B: Summary of Crash-Frequency and Crash-Severity Models in Highway Safety
Appendix C: Computing Codes
Appendix D: List of Exercise Datasets