AN EFFICIENT STOCHASTIC COMPUTATIONAL FRAMEWORK
FOR ESTIMATING AND UPDATING
EARTHQUAKE FATALITIES

by

Davis Engler
A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Master of Science (Applied Mathematics and Statistics).

Golden, Colorado
Date ______________________

Signed: ______________________  Davis Engler

Signed: ______________________  Dr. Mahadevan Ganesh
Thesis Advisor

Signed: ______________________  Kishor Jaiswal, Ph.D., P.E.
Thesis Advisor

Golden, Colorado
Date ______________________

Signed: ______________________  Dr. Gregory Fasshauer
Professor and Head
Department of Applied Mathematics and Statistics
ABSTRACT

Following an earthquake it is crucial to estimate the true scope of impact in terms of fatalities and damage as quickly as possible for efficient post-earthquake response and recovery. The USGS Prompt Assessment of Global Earthquakes for Response (PAGER) model is the state-of-the-art for rapid loss estimation following any major earthquake anywhere in the world. In particular, the PAGER model can provide estimates of the total loss within about 20 minutes after the event based on estimates of population, shaking intensity, and the vulnerability of the population to such shaking in the affected region. The latter is modeled as a fatality rate given as a function of shaking intensity based on past earthquake data.

The current PAGER model does not explicitly account for uncertainties in the population, shaking intensity, or fatality rates independently. However, PAGER’s ability to calibrate these fatality rate models using historical earthquake data entails implicitly capturing some of these uncertainties. Historically, the PAGER estimates are only updated if improved estimates of shaking are available through the USGS ShakeMap product. In addition, due to the uncertainties in the mean fatality rates, estimated population, and shaking intensities over a broad region, there is a tendency to overestimate or underestimate these losses if the estimates of shaking, population, or fatality rates are inconsistent with the ground truth. The primary goal of PAGER model is to provide the most precise and accurate estimates as soon as possible. The current interest is to upgrade the PAGER-based loss estimates through efficient spatial and temporal data-driven computational modeling, for quickly obtaining and communicating the true scope of potential loss.

The main aim of this thesis is centered on developing a stochastic computational mathematical framework for efficient incorporation of ground truth data on reported earthquake fatalities to improve PAGER’s overall estimate. The stochastic modeling, to obtain more accurate and precise estimates of total loss, is developed by (i) incorporating additional un-
certainties when estimating the total loss; (ii) updating the initial PAGER estimates using reported losses over time. The computational framework in this work is developed using a combination of forward uncertainty propagation, and inverse Bayesian approaches for both stationary and temporal fatality data. For the latter, we incorporated techniques inspired from a traditional Kalman filtering process, which are used to update the total loss, subject to partial loss data over time provided by authoritative agencies. Using real earthquake data, we demonstrate the efficiency of the proposed computational framework and its effect in terms of improving PAGER loss alerts over time. The proposed framework allows us to propagate uncertainties over time and has the ability to ingest partial loss data over time to improve PAGER’s overall forecast for the total loss.
TABLE OF CONTENTS

ABSTRACT ................................................................................................. iii
LIST OF FIGURES ...................................................................................... viii
ACKNOWLEDGMENTS ............................................................................... xiii
DEDICATION ............................................................................................... xiv
CHAPTER 1 INTRODUCTION .................................................................... 1
CHAPTER 2 ESTIMATING TOTAL LOSS AFTER AN EARTHQUAKE .......... 7
  2.1 Overview of Probability and Statistics .............................................. 7
    2.1.1 Random Variables and Probability Space ............................... 7
    2.1.2 Probability Distributions and Densities ............................... 8
    2.1.3 Expectation and Statistical Moments .................................... 9
    2.1.4 Dependence and Conditionals of Multiple Events or Random Variables . 10
    2.1.5 The Law of Total Probability and Bayes Theorem .............. 11
    2.1.6 Normal and Lognormal Random Variables ....................... 12
    2.1.7 Monte Carlo Simulation ..................................................... 13
  2.2 Shaking Intensity Measures ............................................................. 15
    2.2.1 Measures of Ground Motion and Types of Data .................. 16
    2.2.2 Predicting Macroseismic Intensity Using Data .................... 20
  2.3 Spatial Distribution of Population ..................................................... 22
  2.4 PAGER Fatality Rate Model ............................................................ 25
CHAPTER 3 POTENTIAL ENHANCEMENT TO THE CURRENT PAGER LOSS
  MODEL ................................................................................................... 33
3.1 The Current PAGER Loss Model with Illustration ........................................ 34
  3.1.1 Description of current PAGER model ...................................................... 34
  3.1.2 Illustration of current PAGER model ...................................................... 39
3.2 Accounting for Additional Uncertainties in the PAGER Model .................. 41
  3.2.1 Uncertainty in the estimated population and fatality rate ....................... 42
    3.2.1.1 Expectation Method: ............................................................... 43
    3.2.1.2 Monte Carlo Method: ............................................................. 48
  3.2.2 Accounting for uncertainty in shaking intensity ..................................... 52
    3.2.2.1 Approximating the Distribution of $\nu(S_i)$ ................................ 52
    3.2.2.2 Approximating the Distribution of $\epsilon_\nu(S_i)$ ......................... 55
3.3 Generating a Subregional Loss Ensemble for an Arbitrary Subregion $i$ ...... 59
  3.3.1 Drawing realizations of $\epsilon_{\nu_i}$ .................................................. 60
    3.3.1.1 Calculating $f_{\epsilon_{nu_i}}(\epsilon)$ for any $\epsilon$ ................................. 60
    3.3.1.2 Metropolis Hastings Markov Chain Monte Carlo Method ................. 61
  3.3.2 Example and Results for an Arbitrary $E_i$ ............................................ 63
3.4 Aggregating Subregional Realizations to Obtain Total Loss $E$ Estimate .... 66

CHAPTER 4 BAYESIAN UPDATING OF THE TOTAL LOSS DISTRIBUTION ............ 76
4.1 Overview of Updating Methods ................................................................. 77
4.2 Method 1: $\mu$ and $\sigma$ are both known ............................................... 79
4.3 Method 2: $\sigma$ is known and $\mu$ is unknown ......................................... 85
4.4 Method 3: $\mu$ is known and $\sigma$ is unknown ......................................... 90
4.5 Intro to Primary Thesis Goal: Updating Given Only Partial Observations of Loss ................................................................. 97
CHAPTER 5 HOW REPORTED LOSSES EVOLVE OVER TIME AND BASICS
OF KALMAN FILTERING .............................................. 100

5.1 Modeling Reported Losses Using an Exponential CDF .................. 102
  5.1.1 Nonlinear Regression of Reported Loss Data ....................... 103
  5.1.2 Modeling Loss Projection ......................................... 108
  5.1.3 Fitting 20 Training Earthquakes ................................. 109
  5.1.4 Estimating $\alpha$ from an Earthquake’s Moment Magnitude .... 112

5.2 Kalman Filtering .................................................................. 120
  5.2.1 The Kalman filter loss updating model ............................. 120

5.3 The Kalman Filter Model for Updating Total Loss .................... 125
  5.3.1 Formulation of Model for Normal Assumptions .................. 125
  5.3.2 Illustration of Kalman Filter Loss Updating Model .......... 128

CHAPTER 6 THE RECURSIVE BAYES LOSS UPDATING MODEL .......... 133

6.1 Incorporating $\alpha$ Uncertainty in the Kalman Filter Model .......... 134

6.2 Using the Generalized Kalman Filter Model on the Test Earthquakes 141
  6.2.1 April 2015 magnitude 7.8 earthquake in Gorkha, Nepal ................. 143
  6.2.2 April 2009 magnitude 6.3 earthquake in L’Aquila, Italy ................. 151
  6.2.3 September 2017 magnitude 7.1 earthquake in Mexico City, Mexico ... 155

6.3 Analysis of Quadrature Size in Numerical Integration ............... 160

CHAPTER 7 CONCLUSIONS, IMPROVEMENTS, AND FUTURE WORK ..... 167

REFERENCES CITED ......................................................... 176
LIST OF FIGURES

Figure 2.1 ShakeMap of the 1906 San Francisco earthquake . . . . . . . . . . . . . . 23
Figure 2.2 Fatality rates for three different countries . . . . . . . . . . . . . . . . . . 30
Figure 2.3 Fatality residuals plot for all Indonesia earthquakes in EXPO-CAT . . . . 30
Figure 2.4 USGS onePAGER . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32
Figure 3.1 A histogram of the total loss distribution for the 2011 Van earthquake. . . 41
Figure 3.2 Histogram of log($E$) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 49
Figure 3.3 Statistics of log($E$) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 49
Figure 3.4 Plots of Statistics for Different Sample Size $M$. . . . . . . . . . . . . . . . 50
Figure 3.5 Standard Error of Mean of log($E$). . . . . . . . . . . . . . . . . . . . . . 50
Figure 3.6 Table of fatality distribution results . . . . . . . . . . . . . . . . . . . . . 51
Figure 3.7 Sample statistics and Beta distribution statistics for $\nu(S_i)$. . . . . . . 53
Figure 3.8 Histogram of fatality rate . . . . . . . . . . . . . . . . . . . . . . . . . . . 54
Figure 3.9 Density functions for sample PDF . . . . . . . . . . . . . . . . . . . . . . . 54
Figure 3.10 Histogram of fatality rate error standard deviation . . . . . . . . . . . . . 56
Figure 3.11 Density functions for sample pdf of $\tau_\nu(S_i)$ . . . . . . . . . . . . . . 56
Figure 3.12 Plot of the pdf of $\epsilon_\nu(S_i)$ . . . . . . . . . . . . . . . . . . . . . . . . 58
Figure 3.13 Squared error in approximations to $f_{\epsilon_\nu(S_i)}$ . . . . . . . . . . . . 58
Figure 3.14 Statistics of the different approximations to $f_{\epsilon_\nu(S_i)}$ along with $L^2$-error. . 59
Figure 3.15 Table of the first four sample statistics of chosen $E_i$ . . . . . . . . . . . . . 63
Figure 3.16 Histograms of $E_i$ and log($E_i$) . . . . . . . . . . . . . . . . . . . . . . . 64
Figure 4.11  Method 2b: Prior and posterior cdfs of log($E$) ........................................ 93

Figure 4.12  Method 3: Posterior pdf of log($E$) for different observed losses ............... 94

Figure 4.13  M3: Posterior statistics of log($E$) for different observed losses ............... 95

Figure 4.14  Observed loss after different times following the 2015 Gorkha earthquake. ................................................................. 98

Figure 4.15  Observed loss after different times following the 2009 L'Aquila earthquake. ................................................................. 98

Figure 4.16  Observed loss after different times ......................................................... 99

Figure 5.1  Comparison of nonlinear regression for different norms ......................... 107

Figure 5.2  Observed loss data gathered from Zhao et al. and "earthquake report.com". Plotted over the data is the best fit exponential cdf subject to the L2G norm along with its standard deviation curves. ........................................ 110

Figure 5.3  Observed loss data gathered from and . Plotted over the data is the best fit exponential cdf subject to the L2G norm along with its standard deviation curves. ................................................................. 111

Figure 5.4  Linear regression fit .................................................................................. 115

Figure 5.5  Weighted linear regression fit .................................................................. 117

Figure 5.6  Plot of the log residuals from all 20 earthquakes nonlinear regression fits 119

Figure 5.7  Diagram showing the proposed Kalman filter model ............................... 124

Figure 5.8  Plot of the nonlinear regression model fit for the Gorkha earthquake . 128

Figure 5.9  Plots showing the updated total loss estimates for Gorkha using the basic KF model and best fit decay rate $\alpha$ .................. 129

Figure 5.10  Plot of the range of the best fit $\alpha$, KF basic total losses given different PAGER estimates for the Gorkha earthquake. ......................................................... 130

Figure 5.11  Plots showing the updated total loss estimates for Gorkha using the basic KF model and $M_w$ regression decay rate $\alpha$ .................. 131

Figure 5.12  Plot of the range of the $M_w$ regression $\alpha$, KF basic total losses given different PAGER estimates for the Gorkha earthquake. ......................... 131
Figure 6.1  Plot of the $M_w$ regression predicted total loss for the Gorkha earthquake assuming exact PAGER estimate . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 144

Figure 6.2  Plot of the $M_w$ regression predicted total loss of the Gorkha earthquake for different PAGER estimates . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 145

Figure 6.3  Plot of the range of $M_w$ regression predicted total loss for the Gorkha earthquake for different PAGER estimates . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 145

Figure 6.4  Plots of the skew and kurtosis of the $M_w$ regression predicted total loss for the Gorkha earthquake for different PAGER estimates . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 146

Figure 6.5  Plot of the $M_w$ regression predicted decay rate $A$ for the Gorkha earthquake assuming exact PAGER estimate . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 147

Figure 6.6  Plot of the $M_w$ regression predicted decay rate $A$ of the Gorkha earthquake for different PAGER estimates . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 148

Figure 6.7  PAGER total loss histograms of the $M_w$ regression updated total loss for the Gorkha earthquake assuming exact different initial loss estimates . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 149

Figure 6.8  Plot of the NLR best fit predicted total loss for the Gorkha earthquake assuming exact PAGER estimate . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 150

Figure 6.9  Plot of the $M_w$ regression predicted total loss for the L’Aquila earthquake assuming exact PAGER estimate . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 151

Figure 6.10 Plot of the range of $M_w$ regression predicted total loss for the L’Aquila earthquake for different PAGER estimates . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 152

Figure 6.11 Plot of the $M_w$ regression predicted decay rate $A$ for the L’Aquila earthquake assuming exact PAGER estimate . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 153

Figure 6.12 PAGER total loss histograms of the $M_w$ regression updated total loss for the L’Aquila earthquake assuming exact different initial loss estimates . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 154

Figure 6.13 Plot of the NLR best fit predicted total loss for the L’Aquila earthquake assuming exact PAGER estimate . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 155

Figure 6.14 Plot of the $M_w$ regression predicted total loss for the Mexico City earthquake assuming exact PAGER estimate . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 156
Figure 6.15  Plot of the range of $Mw$ regression predicted total loss for the Mexico City earthquake for different PAGER estimates ......... 156

Figure 6.16  Plot of the $Mw$ regression predicted decay rate $A$ for the Mexico City earthquake assuming exact PAGER estimate ............... 157

Figure 6.17  PAGER total loss histograms of the $Mw$ regression updated total loss for the Mexico City earthquake assuming exact different initial loss estimates ................................................. 158

Figure 6.18  Plot of the NLR best fit predicted total loss for the Mexico City earthquake assuming exact PAGER estimate ................. 159

Figure 6.19  Plots of the error convergence of the updated total loss pdf .......... 162

Figure 6.20  Plots of the error convergence of the updated total loss cdf .......... 162

Figure 6.21  Plots of the error convergence of the updated decay rate $A$ pdf ...... 162

Figure 6.22  Plots of the estimated order of convergence (EOC) of the updated total loss pdf .................................................. 164

Figure 6.23  Plots of the estimated order of convergence (EOC) of the updated decay rate $A$ cdf .................................................. 164

Figure 6.24  Plots of the estimated order of convergence (EOC) of the updated total loss pdf .................................................. 164

Figure 6.25  Plot showing the time (in seconds) it took to perform the Kalman updating filter .................................................. 165
ACKNOWLEDGMENTS

I would like to thank the academy for granting me this prestigious thesis. This project would never have succeeded without the expert tutelage from my Colorado School of Mines thesis advisor Mahadevan Ganesh, and my USGS supervisor Kishor Jaiswal, who have been instrumental in helping me build the skills, tools, and knowledge necessary to produce this work. Additionally, I would like to acknowledge Hae Young Noh out of Carnegie Mellon University, for all of her help, direction, and particularly her knowledge of Kalman filters that made the bulk of this work possible, especially showing me how to best structure my thesis, as well as thank Luis Tenorio and Karin Leiderman for their roles in my thesis committee and input for this work. I would also like to acknowledge to the USGS Geologic Hazards Science Center as a whole for allowing me to focus my work around my research internship, particularly the entire PAGER team, including David Wald, Kristin Marano, and of course Kishor Jaiswal. Additionally, I would like to acknowledge the Colorado School of Mines for expanding my knowledge and directing my career path. Finally, none of this would be possible without the love and support of my family, particularly my wife Joanna Engler, who has been instrumental in helping me keep my sanity and stay focused for the last decade.
To my inspiring wife Joanna.
CHAPTER 1
INTRODUCTION

Shortly after an earthquake occurs, it is important to be able to gauge the amount of overall damage that it will cause, both in casualties and economic loss. These loss estimates can be used to better inform the proper authorities of the potential damage so that they can accordingly plan for aid and recovery. The amount of earthquake losses can vary drastically depending on both the magnitude of energy and how radiates from the earthquake’s epicenter as well as the particular population exposed to the earthquake’s energy. Furthermore, the amount of potential damage for an exposed population depends on how vulnerable that population is to different amounts of shaking in terms of the quality and structure of its built environment. This facilitates the necessity to better understand an earthquake’s possible impact on the affected built environment and its population in order to obtain accurate estimates of the total fatalities and economic loss. Among the different loss estimation models that are used today, none is more well-known and more used than the Prompt Assessment of Global Earthquakes for Response (PAGER) framework, which is a product the United States Geological Survey (USGS) Geologic Hazards Science Center (GHSC) located in Golden, Colorado. Developed in 2010 by Kishor Jaiswal and others at the USGS GHSC [3], PAGER is responsible for quickly estimating the amount of economic loss and fatalities after all major earthquakes worldwide. Its current focus is to enhance these estimates through data-driven, stochastic computational models. The main aim of this thesis is to develop the fatality estimation portion of this proposed PAGER framework by accounting for additional uncertainties and updating total loss estimates given reported fatality data.

The USGS GHSC is responsible for monitoring earthquakes worldwide as well as estimating and analyzing the effect the earthquake will have on the surrounding environment and its population. Except for the specific fault characteristics, the majority of these estimates
are provided within 20 minutes to an hour after an earthquake occurs [4]. The USGS GHSC has developed a number of different products which are designed to characterize the ground motion of the area surrounding an earthquake’s epicenter as well as predict how vulnerable the surrounding built environment is to this ground motion, which will correspond to how vulnerable the exposed population is to shaking related deaths. PAGER uses this information as well as population distributional data and country specific vulnerability or fatality rate model’s to obtain an estimation of the total loss. Because of the variability associated with the different model’s and data, the number of total fatalities is treated as a random variable with a distribution spread across different magnitudes of loss [3]. That is, there is a relatively large uncertainty in the true value of total loss (generally about 1/2-1 order of magnitude) for any given earthquake. By accounting for possible uncertainties in the data and models used to estimate loss, as well as using reported loss data during and after the earthquake to update the total loss distribution, this work will demonstrate that more accurate and precise estimations of total fatalities can be obtained.

There are several components which make up the PAGER loss model. The first of these is the ground motion or shaking intensity of the area surrounding the earthquake. Depending on the depth of the rupture, the fault characteristics, the ground composition (hard or soft) and terrain (mountains, basins, bodies of water) earthquakes with the same magnitudes can have dramatically different effects on their surrounding environment, particularly the built environment. These effects are summarized as ground motions or shaking intensities, and give a measure of the amount of acceleration, perceived shaking, or possible extent of damage for any area around the earthquake. These shaking intensity for a particular location is estimated as a function of its distance to the epicenter and the aforementioned ground characteristics, and nearby shaking observations and measurements.

The USGS GHSC has a product known as ShakeMap which is responsible for predicting the ground motion or shaking intensity on in a region surrounding an earthquake. In other words, ShakeMap provides a description of how the energy propagates and dissipates from an
earthquake’s source in order to determine the areas which are most vulnerable to this ground motion energy. For years, ShakeMap has been the standard for ground motion prediction for three main reasons:

1. Using waveform propagation and energy dispersion analysis to predict shaking at any location.

2. Enhance the prediction based on measured ground motion from nearby stations.

3. Further enhance the predictions by taking in surveys of people affected by the earthquake via the USGS’s Did You Feel It? product [4].

Another component in the PAGER loss system is the estimated population distribution of the area surrounding the earthquake, which is given by LandScan, and represents a 24-hour average population for the given region.

The final component estimates a population’s vulnerability to different shaking levels by what is called the fatality rate given shaking intensity model. The vulnerability of a particular population to different shaking intensities corresponds to the fragility of the built environment to these accelerations. This gives an idea of how likely someone who is exposed to a particular intensity is to die, given the country that they live and the history of earthquakes in that country. This model, also developed by PAGER, uses the available historical earthquakes in a particular country to help predict the fatality rate at different shaking intensities for future earthquakes in that country. As mentioned, the estimated total loss distribution for any given earthquake can be spread across multiple orders of magnitude.

While 1–2 orders of magnitude uncertainty may seem large, it is actually quite reasonable due to the high variability in the extent of damage any given earthquake can cause. This uncertainty comes from the limited number of past earthquakes worldwide used to train the PAGER total loss model, as well as the variability in the population exposed and shaking intensities provided by ShakeMap. Furthermore, the fatality rate model, which describes how vulnerable a population is to different shaking intensities, varies greatly between different
countries. This accounts for the fact that different countries have different building stocks, architectural codes, building materials, and population densities, each of which affects the fragility of the corresponding built environment. Furthermore, these country specific fatality rate estimates are only averages, and the true fatality rate can vary significantly for different earthquakes and regions of that country, and can even vary locally for different areas affected by the same earthquake. The current PAGER loss model only accounts for uncertainty in each country specific, fatality rate model’s ability to to hind-cast the total loss for past earthquakes in that country. The current model does this without taking into account the variability of the shaking intensities, population distribution, or deviations from the average fatality rate model for different events and locations. This work first seeks to efficiently incorporate these different variabilities into the current PAGER model in order to obtain more robust estimations of the total number of fatalities. An overview of the work presented herein is as follows.

Chapter 2 begins by laying the foundation of the different probabilistic and statistical terminology, concepts, tools, and notations that will be used throughout the rest of this work. The chapter will then provide background information regarding the different components PAGER uses to estimate loss. While some of the specific information regarding the ShakeMap PAGER fatality models provided in this chapter are not explicitly needed to understand the analyses performed in the remaining chapters, they are nonetheless included in order to provide a better idea of how the different data-driven, physics-based, and empirical models work, as well as to gain more familiarity with the concepts, terminology, and notations that will be used throughout this work.

Chapter 3 seeks an efficient method for quantifying the change in the total loss distribution arising from incorporating the previously mentioned uncertainties in the shaking intensities, the population distribution, and in the PAGER fatality rate model. The chapter begins by explaining the current PAGER method and total loss distribution, before systematically incorporating these additional uncertainties. This method is then tested for a
specific earthquake to demonstrate how it works. Rather than providing an in-depth analysis of this model in regards to multiple earthquakes, including past earthquakes used to train the PAGER fatality rate model, this chapter mainly serves to illustrate the effects of including multiple levels of uncertainty in a given problem. As such it serves more to illustrate some the ideas, concepts, and mathematics that will be used in the remaining chapters than to provide an operational, fully realized framework, leaving such analyses for future work.

Chapter 4 seeks a method for updating the PAGER loss distribution given an actual reported total loss for a given earthquake. This chapter investigates how the initial PAGER loss distribution is affected when incorporating total loss estimates in a Bayesian setting. Particularly, this chapter illustrates how Bayesian updating can be achieved for problems in which multiple layers of uncertainty exist, particularly when uncertainty in the prior PAGER distributional parameters is included. Similarly to Chapter 3, this chapter is used more to describe the concepts and mathematical techniques that are used in the final two chapters rather than providing a robust analysis of the proposed model regarding numerous earthquakes and actual total loss data, leaving such analyses for future work. While the work in this and the previous chapter are of interest and useful in understanding the reliability of the PAGER total loss distribution for different earthquakes, this thesis aims for the higher hanging fruits in terms of mathematical complexity, focusing the majority of the work on the more complicated problem of updating the total loss distribution given observations of partial loss over time.

Chapter 5 deals with the much more difficult task of updating the total loss distribution given reported loss shortly after the earthquake. Because these observations are not of the total loss, but rather partial or incomplete observations of it, the estimated loss distribution cannot be updated in the usual Bayesian manner. This chapter investigates a generalization of a particular updating model called the Kalman filter, which combines the initial PAGER loss distribution with reported losses over time in a Bayesian setting. In particular, relating partial losses over time to the total loss involves determining a model which describes how
these losses accumulate over time. After describing the candidate loss accumulation model used throughout this chapter and the next, this chapter highlights the complete loss model before illustrating the (generalized) Kalman filter process on a more simplified, basic model. This basic Kalman filter total loss model is used to illustrate how Kalman filtering works, as well as to understand the kind of results obtained through the updating process in regards to real loss data from the 2015 Gorkha earthquake in Nepal. It should be noted that while traditional Kalman filtering involves linear combinations of normal random variables, the generalized Kalman filter model used herein can be used for any distributional assumptions.

Chapter 6 develops an enhanced version of the Kalman filter set up in Chapter 5, which incorporates the uncertainty in the decay rate governing the candidate reported loss accumulation equation. After describing the theory behind the updating and laying out a procedure for accomplishing it, this chapter looks at three test earthquakes again to see how it performs given the reported loss data over time. This chapter also investigates the convergence of the different numerically integrated distributions obtained for increasing numbers of quadrature points.

Chapter 7 concludes this work by discussing the conclusions, limitations, challenges, current work, and future work related to the work presented in Chapters 3-6. Particularly, it discusses limitations in the models due to distributional and model assumptions for each problem, and discusses the different possible further studies that can and should be conducted in order to have a more realistic, well-grounded, robust framework for improving total loss estimates.

This thesis can be summarized as a rigorous look into developing forward and inverse models for estimated total loss after an earthquake. In order to understand how these forward and inverse models will work, it is necessary to understand how the current PAGER model works, all of which is described in the following chapter.
CHAPTER 2
ESTIMATING TOTAL LOSS AFTER AN EARTHQUAKE

In order to understand how the PAGER loss estimation model works, it is first necessary to describe the different components that the PAGER system uses to estimate loss. Before looking at these different components however, an overview of the terminology, notation, and basic description of the probability and statistical theory used throughout this work will be given. After this detour the components that make up PAGER will be described in further detail.

2.1 Overview of Probability and Statistics

A quantity of interest is called deterministic if it is completely predictable in terms of its independent variables. In probability and statistical theory an event refers to the realization of an equation or function with a particular set of independent variable values. Events that are not deterministic are known as random events, meaning that their outcome is not fully predictable. The information and notations that will be used are those from [5] and [6].

2.1.1 Random Variables and Probability Space

Consider an indeterministic experiment whose outcomes are not completely predictable. Let \( \omega \) refer to a single outcome of the experiment and introduce the abstract space \( \Omega \) called the sample space which contains every possible outcome \( \omega \) that the experiment can produce. In order to deal with random experiments mathematically, different numerical values can be assigned to the different possible \( \omega \)'s that might arise. This mapping of the possible outcomes in \( \Omega \) to the real line leads to what is called a random variable. A random variable \( X = X(\omega) \) is such that given an outcome \( \omega \) from an experiment with sample space \( \Omega \), it will produce
an output in what is known as a *state space*, denoted by $S$, which is the set containing all possible values of a random variable. This thesis follows the convention of denoting random variables by capital letters and specific realizations of these variables in lower case letters.

A collection of different outcomes is known as an event, and the collection of all possible events including unions, intersections, and complements of these events is known as a *σ-algebra*, which we denote as $\mathcal{F}$. While this paper does not delve into the nuances of σ-algebras and measure theory, and assumes that all random variables herein are measureable with respect to their corresponding σ-algebras and state spaces. The concept of probability gives the likelihood that a particular event will occur. This is accomplished by attaching a weight, or number less than one, to each event such that the probability of no event occurring is zero, the probability of all events occurring is one, and the sum of the weights of all disjoint subsets of $\Omega$ is exactly one. The probability an event $A$ will occur as $P(A)$, where $A \subset \Omega$.

For any particular experiment, the probability measure $P$ assigns weights to the possible outcomes such that the conditions stated earlier in this paragraph are satisfied.

A random experiment can be fully characterized by what is known as a *probability space* given by the triplet $(\Omega, \mathcal{F}, P)$. A probability space contains all information needed to know everything about the possible results of an experiment [5].

### 2.1.2 Probability Distributions and Densities

Given a particular value $x$ of a random variable $X(\omega)$ the *cumulative distribution function* or *cdf* is defined as $F_X(x) = P(X \leq x) = P(\omega \in \Omega : X(\omega) \leq x)$, i.e. it gives the probability of a random variable $X$ having a value less than or equal to $x \in \mathbb{R}$. Given a set of values $B$ in the state space $S$ the probability that the random variable $X$ will have values in $B$ is $P_X(B) = P(X \in B) = P(\omega \in \Omega : X(\omega) \in B)$.

Depending on the experiment, a random variable can have a discrete or countably infinite number of possible values leading to what are called *discrete* or *continuous random variables*, respectively. For discrete random variables, the distribution will have jumps, whereas the
distribution for continuous random variables will be continuous. Continuous distributions often have what are called probability density functions or pdfs $f_X : \Omega \to \mathbb{R}^+ \cup \{0\}$ such that $F_X(x) = \int_{-\infty}^{x} f_X(y)dy$, $x \in \mathbb{R}$, where the subscript $X$ denotes the random variable which the pdf or cdf describes.

### 2.1.3 Expectation and Statistical Moments

Understanding how the possible values of a random variable $X$ are distributed across the state space is of primary interest when analyzing random variables. A statistic is any descriptor related to a random variable which summarizes some aspect of its behavior. One of the most well known ways of summarizing a random variable is found by determining its statistical moments. One of the principle statistics of interest in regarding a random variable is known as an expectation value, also known as the mean. The expectation value of a random variable is known as the first statistical moment and is defined as

$$
\mu_X = \mathbb{E}[X] = \int_{\mathcal{S}} x f_X(x)dx,
$$

and gives the average value expected from the experiment. Another statistic of interest for random variables as a measure of how the possible values of the variable are spread about the average. This spread is called the standard deviation about the mean and is denoted $\sigma$, although more often this spread is determined via the square of the standard deviation. This quantity, known as the variance, is defined as

$$
\sigma_X^2 = \text{Var}(X) = \int_{\mathcal{S}} (x - \mu_X)^2 f_X(x)dx.
$$

In general, the expectation of an arbitrary function $g$ of the random variable $X$ is

$$
\mathbb{E}[g(X)] = \int_{\mathcal{S}} g(x) f_X(x)dx.
$$
2.1.4 Dependence and Conditionals of Multiple Events or Random Variables

Suppose that there are two different random events $A$ and $B$ from the same experiment, both of which are measurable subsets of $\Omega$. Then $A$ and $B$ are defined as independent if and only if

$$P(A \cap B) = P(A)P(B).$$

Two random variables $X$ and $Y$ are called independent variables if and only if

$$P(X \in B_1, Y \in B_2) = P(X \in B_1)P(Y \in B_2)$$

for all measurable subsets $B_1$ and $B_2$ of the state space $S$. Alternatively two random variables $X$ and $Y$ are independent if and only if

$$F_{XY}(x, y) = F_X(x)F_Y(y)$$

or, if $X$ and $Y$ have density functions $f_X$ and $f_Y$ respectively, then they are independent if and only if

$$f_{XY}(x, y) = f_X(x)f_Y(y),$$

where $F_{XY}$ and $f_{xy}$ are called the joint probability distribution and joint probability density functions, respectively. In general, for $N \in \mathbb{N}$ mutually independent random variables $X_1, ..., X_N$, the joint cdf of $X_1, ..., X_N$ is

$$F_{X_1,...,X_N}(x_1, ..., x_N) = F_{X_1}(x_1)F_{X_2}(x_2)...F_{X_N}(x_N)$$

and

$$f_{X_1,...,x_N}(x_1, ..., x_N) = f_{X_1}(x_1)f_{X_2}(x_2)...f_{X_N}(x_N).$$

Another important concept in probability regarding multiple events or random variables is the concept of conditionals. These describe how the probability or distribution of an event or random variable changes depending on if another event has occurred. Given two events $A$ and $B$ with $P(B) \neq 0$ the conditional probability of $A$ given that $B$ has occurred is found by the equation
\[ P(A|B) = \frac{P(A \cap B)}{P(B)}. \] \hspace{1cm} (2.4)

Furthermore, for a random variable \( X \) the \textit{conditional distribution} of \( X \) given the occurrence of event \( B \) is defined as

\[ F_X(x|B) = P(X \leq x|B) = \frac{P(X \leq x, B)}{P(B)}. \]

### 2.1.5 The Law of Total Probability and Bayes Theorem

One of the more useful theorems in probability theory is known as the law of total probability. This says that if a set of \( B_1, \ldots, B_M \) disjoint events covers a sample space \( S \), then the probability of any event \( A \subset S \) occurring is given by the sum of the probabilities of the joint events \( P(A \cap B_i), \quad i = 1, \ldots, M \). The probability of the joint event \( A \cap B \) is found using Equation 2.4.

The fundamental theorem that will be continually used throughout this work is known as the Bayes theorem. This theorem gives the probability of one event \( A \) given another event \( B \), using the fact that \( P(A \cap B) = P(B|A)P(A) \) in Equation 2.4, as

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \] \hspace{1cm} (2.5)

In practice these theorems will be used in terms of the cdfs and pdfs of continuous random variables. For random variables \( X \in \Omega_X \) and \( Y \in \Omega_Y \) with density functions \( f_X : \Omega_X \to \mathbb{R} \) and \( f_Y : \Omega_Y \to \mathbb{R} \), respectively, and conditional density \( f_{X|Y} : \Omega_X \to \mathbb{R} \), then the law of total probability implies

\[ f_X(x) = \int_{\Omega_Y} f_{X|Y}(x|y)dy = \int_{\Omega_Y} f_{X|Y}(x|y)f_Y(y)dy. \] \hspace{1cm} (2.6)

The Bayes equation implies that (using Equation 2.6)

\[ f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)} = \frac{f_{X|Y}(x|y)f_Y(y)}{\int_{\Omega_Y} f_{X|Y}(x|z)f_Y(z)dz}, \] \hspace{1cm} (2.7)
where the denominator of Equation 2.7 is a normalization constant which ensures that
\[ \int_{\Omega_y} f_{Y|X}(y|x)\,dy = 1. \]

### 2.1.6 Normal and Lognormal Random Variables

Throughout this work, the principle random variables used to demonstrate the different techniques and methods are called *normal or lognormal random variables* and their respective *cumulative normal or lognormal distribution functions*. A normally distributed random variable \( X \) with mean \( \mu \) and standard deviation \( \sigma \) has the following probability density function (pdf):

\[
f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right),
\]

(2.8)

where \( x \in \mathbb{R} \) is a realization of \( X \in \Omega_X \). Now consider a random variable \( Y \in \Omega_Y \) such that \( Y = e^X \). Then \( Y \) is called a *lognormally distributed* random variable with parameters \( \mu \) and \( \sigma \), which are respectively the mean and standard deviation of \( X \), and one writes \( Y \sim LN(\mu, \sigma) \). The pdf for a lognormal variable \( Y \) is

\[
f_Y(y) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(\log(y) - \mu\right)^2}{2\sigma^2}\right),
\]

(2.9)

where \( y > 0 \) is a realization of \( Y \). This pdf is found by replacing \( x = \log(y) \) in Equation 2.8 and normalizing the function over all \( y \in \Omega_Y \).

Consider a random variable \( Z \) with pdf \( f_Z(z) \), where \( z \) is a particular realization of \( Z \). Then the cumulative distribution function (cdf) of \( Z \) is the function \( F_Z(z) \) such that

\[
F_Z(z) = \int_{-\infty}^{z} f_Z(z)\,dz,
\]

(2.10)

where again \( F_Z(z) \) is defined as the probability that \( Z < z \). A normal random variable with zero mean and variance 1 is denoted \( Z \sim N(0,1) \) is said to have a *standard* normal distribution. There is a special name for the cdf of a standard normal distribution, called the *error function* and denoted \( \Phi \), which is defined as
\[ F_Z(z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{2\pi} e^{-z^2/2}. \] 

(2.11)

Suppose \( X \sim N(\mu, \sigma) \), then \( X \) can be written in terms of a standard normal variable \( Z \sim N(0, 1) \) as \( X = \sigma Z + \mu \), which implies that \( Z = \frac{X-\mu}{\sigma} \). Then

\[
F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\sigma z + \mu - \mu)^2}{2\sigma^2}\right) (\sigma d\sigma z) \\
= \int_{-\infty}^{x} \frac{1}{2\pi} \exp(-z^2/2) dz = \Phi(z) \\
= \Phi\left(\frac{x-\mu}{\sigma}\right). 
\]

Now suppose that \( Y = e^X \) and thus \( X = \log Y \), then

\[
F_Y(y) = \int_{-\infty}^{y} \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log(y) - \mu)^2}{2\sigma^2}\right) dy \\
= \int_{-\infty}^{x} \frac{1}{e^x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log(e^x) - \mu)^2}{2\sigma^2}\right) (e^x dx) \\
= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) dx = \Phi\left(\frac{x-\mu}{\sigma}\right) \tag{2.12} \\
= \Phi\left(\frac{\log(y) - \mu}{\sigma}\right),
\]

which gives a means of describing any normal or lognormal random variable in terms of the error function \( \Phi \).

2.1.7 Monte Carlo Simulation

One of the simplest and robust simulation methods is known as Monte Carlo simulation. Named after the city renowned for its gambling, this method uses the Strong Law of Large Numbers to estimate the expectation of a random variable distribution by generating numerous realizations of the variable and taking the sample average of these realizations. For a random variable \( M \) in which \( N \) realizations \( m_1, ..., m_N \) are generated, the sample mean,
denoted \( \overline{m} \), is defined to be
\[
\overline{m} = \frac{1}{N} \sum_{i=1}^{N} m_i,
\] (2.13)
and the estimation of the variance of \( M \) is denoted \( s_M^2 \) and is given by
\[
s_M^2 = \frac{1}{N-1} \sum_{i=1}^{N} (m_i - \overline{m})^2.
\] (2.14)

The Strong Law of Large Numbers says that
\[
\lim_{N \to \infty} \overline{m} = \frac{1}{N} (m_1 + m_2 + ... + m_N) = \mu_M,
\] (2.15)
and \( s_M^2 = \frac{\sigma_M^2}{N} \to 0 \) as \( N \to \infty \) with probability 1. This means that no matter the distribution, one can use \( \overline{m} \) to approximate the true expectation \( \mu_M \) within a desired precision by simply increasing the number of realizations [7]. The rate at which the statistical mean \( \overline{m} \) converges to the actual mean is \( 1/\sqrt{N} \). That is, in order to get one decimal of precision closer to the actual mean, the sample size must be increased a hundred fold. This rate of convergence comes from the Central Limit Theorem which says that for \( M_1, M_2, ..., M_N \) independent random variables with the same underlying distribution and with mean \( \mu_M \) and variance \( \sigma_M^2 \), the distribution function of the variable \( U_N = \sqrt{N} \frac{s_N - \mu_M}{\sigma} \) will be that of a normal distribution with expectation 0 and variance 1 [5]. The variable \( U_N \) is a rescaling of the statistical mean \( s_N \) such that the mean and variance will be 0 and 1, respectively.

Monte Carlo simulation is robust in that the convergence rate is independent of the number of random variables or equivalently the dimension of the sample space \( \Omega_M \). The drawback of this method is that the rate of convergence is relatively slow compared to the amount of data needed for increased precision. This concludes the basic probability and statistical overview. The different components used by the PAGER system to estimate total loss following earthquakes will now be discussed.
2.2 Shaking Intensity Measures

Immediately following an earthquake, multiple characteristics of the event must be gathered and computed in a short amount of time. The USGS GHSC is responsible for gathering and analyzing this relevant information about the earthquake and the area in which it occurred in order to summarize the event and its potential impact on the affected area and its built environment. One of the most fundamentally important products that the USGS GHSC produces is called ShakeMap. ShakeMap is a program which estimates the extent of ground motion in a region affected by an earthquake. Understanding how ShakeMap works and the information it provides is crucial for understanding how PAGER estimates total loss.

Developed circa 1999 [4], ShakeMap was designed to provide a more comprehensive and useful metric for describing the potential damage that can be caused by an earthquake. While the magnitude is useful for describing the ground motion at the earthquake’s source, it provides no information on the degree of shaking at other affected locations. As seismic waves radiate from an earthquake’s epicenter, they are affected from both dispersion (increased wavelengths) and attenuation (decreased amplitude) as they interact with the different ground compositions in the surrounding area. This is because waves travel at different velocities through different mediums, for instance the amount of energy radiating from an earthquake can extend to a farther distance and with greater amplitude in solid rock compared to in a loose-soiled basin. Also, due to heterogeneities in the earth and surface elevation, these seismic waves can reflect and interfere with one another, effectively amplifying or dampening the ground motion in different locations.

Because the surface and ground composition vary greatly around the world and even locally for a particular event, understanding how the energy propagates after an earthquake and in which locations the shaking is greatest is a difficult task. This section is designed is to explain how ShakeMap works, assess the kinds of information it incorporates, and highlight its uses. The principle work behind ShakeMap comes from Worden and Wald [4], Boore
et al. [8], Wald et al. [9], Wald et al. [10], and Worden et al. [11]. After describing the different measures of ground motion that ShakeMap incorporates, this section then describes the types ground motion data that it uses, and finally describes how these are combined with the earthquake’s characteristics (epicentral location, magnitude, etc.) and the particular ground composition surrounding the earthquake, in order to obtain estimates of shaking for different locations of an affected region. Again while the specifics regarding the inner workings of ShakeMap are not necessary to understand the methods described in this work, they are included to provide a more concrete understanding of what these shaking intensities are and how they are estimated.

2.2.1 Measures of Ground Motion and Types of Data

Depending on the magnitude and epicenter of a given earthquake, the ShakeMap program decides upon a reasonably large area surrounding the earthquake, generally 100-200 km, and divides this region into smaller subregions 1 km x 1 km [4]. For each subregion, ShakeMap is tasked with determining the estimated peak ground motion for that subregion. These estimates are based on equations describing how the seismic waves attenuate with distance through different mediums as well as on measurements and observations of the ground motion at nearby locations. Along with these expected peak ground motions, ShakeMap also outputs the uncertainty of these values, taking into account the various uncertainties involved in the wave propagation formulas, data uncertainty, fault characterization, and the added uncertainty in transforming data from one measure of ground motion to another.

There are a variety of different ways to quantify the ground motion at any particular location or subregion. These can be divided into two categories: strong ground motion measures and macroseismic intensity measures. Strong ground motion measures are those for which characteristics of the seismic waves are measured and analyzed through instruments located at various seismic stations. The most used of these are the peak ground acceleration (PGA), The peak ground velocity (PGV), and the peak spectral acceleration at .3s, 1s, and
3s (PGA.3, PGA1, and PGA3, respectively). The PGA measures the maximum absolute acceleration of the ground at any particular location, and is generally interpreted in terms of percent gravity. The PGV measures the maximum absolute velocity of the ground at a particular location in the time following an earthquake. Finally the PSA is a measure of the acceleration felt by a damped, harmonic oscillator with period .3s, 1s, or 3s [4]. This gives an indication of the maximum response that can be felt from different structures of different heights corresponding to different modes of oscillation. These are called strong ground motion measures because they describe the actual kinematics of the ground subject to the seismic waves.

Macroseismic intensity on the other hand, is more of a qualitative, categorical measurement of shaking. Rather than describing the response felt by the ground, macroseismic intensities describe the response felt by the built environment or its population as it encounters seismic waves. The principle macroseismic intensity measure used is the Modified Mercalli Intensity (MMI) scale. This is a scale ranging from 1 to 10 describing the shaking in terms of the response felt and type of damage observed. Rather than being a specific measure of acceleration or velocity, it describes how the shaking has affected parts of the built environment. That is, it can be used to describe any shifting of objects around the house, cracks in walls, damage to a building’s foundation, chimney, or facade. MMI can also be used to describe damage experienced by underground pipes, structures, sewers, and electrical systems. The MMI is both used to qualitatively describe the shaking intensity, (not felt, weak, moderate, strong, violent, and extreme), as well as the amount of damage caused (no damage, light, moderate, heavy, very heavy). An MMI of 1 corresponds to the lowest shaking intensity (not felt, no damage) and 10 corresponds to the highest intensity (extreme shaking, very heavy damage). However, because studies have shown that it is difficult for people to consistently distinguish between MMI 9 and 10, ShakeMap and the USGS products that use it use a maximum MMI of 9, combining the previous MMI 9 and 10 levels into one.
The macroseismic intensity measure is the principle measure outputted by ShakeMap because it describes the expected response of the built environment, although the USGS also produces ShakeMaps describing the other measures of ground motion (PGA, PGV, etc.) [4]. One of the caveats that makes ShakeMap so versatile is that it is able to predict ground motion for a large region given sparse instrumental measurements in that region. It is able to accomplish this because of three reasons. The first is that it uses well-developed equations describing how seismic energy propagates as it radiates from an earthquake’s source. These equations describe how the wave attenuates in a constant medium (bedrock is the default medium of calculation). Because the ground composition is generally not bedrock, after all necessary contributions to shaking are incorporated, the shaking estimate is then dampened or amplified depending on the soil conditions. This allows for the measurements or estimations of shaking in one location to influence the estimations of shaking in nearby locations [4].

These ground motion prediction equations, or GMPE’s, are described in Boore et al. [8], and Boore and Atkinson [12], along with the relative uncertainty in using them. Essentially, the GMPE’s predict the ground motion at one location by performing a numerical waveform analysis on the seismic waves, taking into account the path and site heterogeneities in the ground composition. Where GMPE’s describe strong ground motion measures (PGA, PGV, etc.) , there are also well developed intensity prediction equations (IPE’s) which perform the same task except with respect to macroseismic intensity measures [4].

The second reason that ShakeMap can do so well despite often having only sparse instrumental ground motion measurements is that it also incorporates observations of macroseismic intensity felt by those who experienced shaking following an earthquake. In order to survey and collect these observations in the time immediately proceeding the earthquake, the USGS GHSC developed the Did you feel it? (DYFI) system. This worldwide web-based system surveys any participating people or organizations by asking a series of questions regarding the extent of shaking felt as a result of a particular event. Anyone signed up with DYFI
receives a survey following an earthquake, which has numerous questions regarding the shaking felt, as well as the observed response of objects around the house, paintings, damage to walls, ceilings, foundation, plumbing, electrical, to this shaking. For each report filed, the DYFI estimates an MMI experienced for that location. Because the shaking is desired for larger regions than one household, and in order to reduce any bias in individual reports, the MMI is averaged over the DYFI reports in a given area. In order to account for the fact that the shaking reported in a particular subregion is a summary of numerous DYFI reports, a standard deviation of .5 MMI level is given to each DYFI subregion [4]. Analysis with DYFI data and instrumental data for numerous fatal and nonfatal earthquakes worldwide indicates that they are generally in agreement. This means that the ability to use this DYFI information, which is often readily available in large numbers shortly after an earthquake, helps produce accurate shaking estimates in areas with sparse amounts of instrumental strong ground motion data.

The final reason why ShakeMap is so versatile is its ability to use any available ground motion data (strong from stations or macroseismic from DYFI) to predict any type of ground motion (strong or macroseismic) at any location in an affected region. It does this via the use of equations relating one measure of ground motion to another. These equations are known as the ground motion/intensity conversion equations (GMICE’s). These GMICE’s allow for measurement of one type of ground motion (strong or macroseismic) to be converted into an observation of the other type of ground motion (macroseismic or strong, respectively), along with some uncertainty due to the variability of the particular GMICE being used [8, 12].

In summary there are three kinds of data to account for when predicting the ground motion of any subregion: the predicted ground motion (PGM) from the GMPE’s, the influence of nearby strong motion data (SM) in the form of PGA, PGV, or PSA, and finally the influence of nearby macroseismic intensity observations (IM) from DYFI. When ShakeMap produces a grid of predictions of a certain measure of ground motion, say macroseismic intensity, the ground motion must be calculated using an intensity prediction equation (IPE),
and any strong motion observations need to be converted to macroseismic intensities via the appropriate GMICE’s.

Because the PAGER loss system models the fatality rate in terms of the macroseismic intensity (MMI), this will be the principle measure used throughout this thesis. Henceforth, any mention of shaking intensity will be in terms of macroseismic intensity, and any shaking measurements are assumed to have been converted into macroseismic intensity via the appropriate GMICE. In order to understand how ShakeMap takes the different kinds of data and produces a grid of shaking estimates, this work first describes how the predicted ground motion is calculated for a particular earthquake as described in [11].

2.2.2 Predicting Macroseismic Intensity Using Data

Suppose that following an earthquake, the affected area is subdivided into \( N \) subregions with the purpose of predicting the peak shaking intensity \( Y_i \) in each subregion \( i, \quad i = 1, \ldots, N \).

As mentioned, each different kind of data influences the overall predicted ground motion for a given subregion. The contribution of each type of data is given as a weight corresponding to degree of influence it has on the overall predicted ground motion for subregion \( i \). Assuming that the macroseismic intensity is the ground motion measure of interest, there is a predicted ground motion for each location found via an IPE, which takes the physics of the wave propagation into account. Call this contribution \( Y_{GMPE,i} \) and its associated uncertainty \( \sigma_{GMPE,i} \). This uncertainty and the uncertainty in the other data types will be used to determine the corresponding weights to be applied in the prediction of \( Y_i \).

Prediction of shaking at location \( i \) also depends on the native data (intensity measures in this case). Suppose there are \( m \) of these macroseismic observations, that is, there are observed intensity measures in \( m \) subregions, each of which is averaged over all individual reports in that particular subregion from the different DYFI reports. In the ShakeMap model each observed intensity measure, call it \( Y_{IM,j}, j = 1, \ldots, m \) will influence the prediction of shaking at subregion \( i \), the weight being inversely proportional to the variance of that
intensity measure $\sigma^2_{IM,j}$. In order to use the observation of shaking at subregion $j$ to predict a value for subregion $i$ ($i \neq j$), the contribution is scaled to account for the relative distance to the epicenter by multiplying the observed value by the ratio $Y_{GMPE,j}/Y_{GMPE,i}$ where $Y_{GMPE,j}$ and $Y_{GMPE,i}$ are the predicted ground motions for subregions $i$ and $j$ obtained using the appropriate IPE. The final product is multiplied by a relative amplification factor, which represents the ratio of the site amplification at each site due to their corresponding ground compositions. This gives a prediction for the shaking at location $i$ due to the observed intensity at location $j$, call it $Y_{IM,i,j}$. Propagating the uncertainties in the site amplification factor and the two GMPE calculations, one can also obtain the variance $\sigma^2_{IM,i,j}$ for the point $i$ due to the observation at subregion $j$. The variance is calculated as an exponential cdf function $\sigma_{GMPE,j}(1 - \exp(-\sqrt{3r_\Delta}))$, where $r_\Delta$ is the distance between points $i$ and $j$ [11]. It should be noted that when calculating the distance between two subregions, one actually calculates the areal center of mass as the center point of each subregion and then uses the distances between the two centers in to obtain $r_\Delta$. This variance is assumed to be 0 at distance zero, and increases with distance from the observed point.

In addition to ground motion predictions and native data contributions, the overall prediction of ground motion for a subregion $i$ is also determined from the converted intensity observations obtained from say $K$ strong motion observations (PGA, PGV, etc.) made at different locations. Following the same process as with the native observations, the converted observation at each subregion $k$, $k = 1, \ldots, K$ predicts an intensity $Y_{SM,i,k}$ and has variance $\sigma^2_{SM,i,k}$ which is composed of the uncertainty in conversion of the strong motion data, which increases with distance from the observed location similarly to the native data. The final mean predicted shaking intensity for subregion $i$ is then calculated by taking a weighted average of the contributions from the observed data and predicted ground motion data from all other subregions, assigning a weight $w_{data} = \frac{1}{\sigma^2_{data}}$, where $data$ can be either $GMPE,i$, $IM,i,j$, or $SM,i,j$, depending on the data being used. Each weight is multiplied by its corresponding prediction and a ground motion estimate is obtained by summing over all of
these weighted predictions and dividing by the total sum of the various weights. Thus, the predicted ground motion for subregion $i$ is given as

$$Y_i = \frac{Y_{GMP_{E,i}}}{\sigma_{GMP_{E,i}}^2} + \frac{\sum_{j=1}^{M} Y_{IM_{i,j}}}{\sigma_{IM_{i,j}}^2} + \frac{\sum_{k=1}^{K} Y_{SM_{i,k}}}{\sigma_{SM_{i,k}}^2}, \quad i = 1, \ldots, N,$$

(2.16)

When determining each predicted ground motion, the variance of this prediction is set equal to the quotient of Equation 2.16, i.e. it is the sum of the weights of all ground motion data, or the sum of the inverse of the data variances:

$$\sigma_i^2 = \frac{1}{\frac{1}{\sigma_{GMP_{E,i}}^2} + \sum_{j=1}^{N} \frac{1}{\sigma_{IM_{i,j}}^2} + \sum_{k=1}^{M} \frac{1}{\sigma_{SM_{i,k}}^2}}.$$

(2.17)

After computing the grids of predicted ground motions for all relevant ground motion measures, the USGS publishes a map for each measure, which can be found under the specific earthquake’s USGS event catalog. ShakeMap also produces a map of the uncertainty in shaking. In addition to these maps, files can be downloaded which contain the shaking estimates and their uncertainties for each grid point as grid files. Below (Figure 2.1) is a ShakeMap generated for the 1906 San Francisco earthquake, illustrating how even events over a century ago can be reasonably approximated based on reports, observations, and ground motion prediction equations.

Again, while the preceding description of how ShakeMap produces shaking estimates is not necessary to understand the analysis contained in this thesis, it does provide a more detailed picture of what these shaking intensities are and how they depend on the various types of data. This will be of use later when deciding what type of distribution can be used to reasonably estimate these predicted shaking intensities.

### 2.3 Spatial Distribution of Population

The USGS PAGER loss system depends on accurate population distribution information for any region in which an earthquake occurs. This is because in order to better predict the
Figure 2.1: ShakeMap of the 1906 San Francisco earthquake, found at earthquake.usgs.gov
total economic loss and fatalities after an earthquake, the PAGER model needs to know the amount of people exposed to the earthquake. Currently the PAGER loss system relies on LandScan for this information. This section is more brief compared to the ShakeMap and PAGER fatality rate model sections because much of the areal population data given by LandScan is proprietary and there is limited or no access to the uncertainties in the data or the exact means with which it is calculated [13].

According to their website, LandScan uses a conglomeration of different types of data and methods to predict their areal population distribution. Data includes census data, historical and geographic data, remote sensed data, and imagery analysis. For any given locale it uses a combination of different region based population distribution models in order to best mitigate the difference in applying certain models to certain locations [13].

The typical resolution for LandScan is 1 km x 1 km, which is the same resolution ShakeMap generally uses [13]. In order for PAGER to use the LandScan data, it must first reconcile the ShakeMap grid and LandScan grid to ensure that the two grids describe the same subregions. The PAGER system then uses these grids to get an idea of what populations are being exposed to which shaking intensities, which it then uses to estimate the total loss. As noted, there is limited to no access to the uncertainties in this population data, and much less its spatial correlation structure. Because of this, this thesis will simply assume an uncorrelated spatial population distribution, with a certain coefficient of variation (CoV) of .3 assumed for each subregional population (CoV = Standard Deviation / Mean).

In order to see how the PAGER system uses the ShakeMap and LandScan data to predict the number of fatalities in each subregion and hence in the entire region, it is necessary to understand how PAGER determines the estimated fatality rate as a function of shaking intensity for an earthquake, based on the country or region in which the earthquake occurs. The next section describes how PAGER developed its fatality rate model.
2.4 PAGER Fatality Rate Model

The PAGER loss system relies on a model which describes the fatality rate for populations exposed to different shaking intensities. This function of fatality rate in terms of shaking was found to depend on the country in which the earthquake occurred under the assumption that building stock and population demographics are relatively similar across that country. As mentioned the fatality rate describes the vulnerability of a population to different shaking intensities, which is a consequence of the fragility of the particular built environment to these different intensities. This assumption has proven to work very well for the PAGER loss system [3].

Estimating the total loss for a particular event can be achieved by determining the population exposed to the different shaking intensities, as well as the fatality rate given shaking intensity model. This is done by overlaying the population grid with the ShakeMap shaking grid to get a population with a corresponding shaking intensity for each subregion. The model sorts all of these subregions in terms of shaking intensity, or MMI for short, and sums across all of the populations exposed to these different MMI values to get the population exposed to each MMI. For each intensity, the exposed population is multiplied by the estimated fatality rate at that MMI to get the estimated loss due to each shaking intensity. Finally, all estimated losses for the different intensities are added up to get an estimation of total loss.

PAGER uses an empirical model to estimate the fatality rate given MMI for a future earthquake by hind-casting all previous earthquakes for that country, described in detail in Jaiswal and Wald [3]. As Jaiswal and Wald [3] describe in their formulation, determining how vulnerable a specific population is to different shaking intensities is difficult to quantify. In order to gain a full understanding regarding the vulnerability of the population, a complete model is needed which describes the fragility of a specific built environment to the specific shaking it experiences. In practice however, information about a specific area’s building stock or even general statistics regarding the built environment is rarely obtainable and
computationally expensive to employ. Because of this, an empirical approach to fatality rate estimation was proposed [3].

The model works by determining a fatality rate given MMI function that best predicts the past earthquakes for a given country. This is achieved by using a nonlinear regression of parameters related to the fatality rate model proposed by Jaiswal and Wald [3]. Specifically, they used a cumulative lognormal distribution function to describe the fatality rate as a function of MMI, described earlier in this chapter in Equation 2.12. This function was chosen because a linear change in the independent variable (MMI) is expected to lead to an exponential change in the dependent variable (fatality rate).

Because fatality rate data is scarce following an earthquake, the model seeks to choose the parameters which best predict the total loss of the previous earthquakes. It does this by searching over different values of the lognormal cdf parameters. For each set of parameters, it calculates the estimated loss, and then determines the parameter values which minimize the error of these estimated losses in comparison with the true observed losses, with respect to a special norm known as the L2G norm [3], explained shortly.

In order to perform this regression for a given country, the empirical loss model requires two pieces of information regarding recorded past earthquakes for that country. The first of these is the estimated population exposed to different shaking levels for each past earthquake in a country. This is found via a USGS catalog known as the exposure catalog (EXPO-CAT) [14] and gives the estimated population exposure to different shaking intensities for a large number of global earthquakes. This is achieved by using an estimated population distribution model for the at the time of each past earthquake, and combining it with an estimated shaking grid for that earthquake, which is provided by ShakeMap [3].

Secondly, the empirical fatality rate model needs to know the number of total shaking related deaths following each of these earthquakes. The USGS has been collecting information about all major earthquakes for numerous countries worldwide since around 1900. This information is summarized in the USGS PAGER catalog (PAGER-CAT) [15]. It should be
noted that the PAGER system only estimates the total loss due to shaking, and not from secondary hazards such as fire, tsunami, ground liquefaction, or subsequent aftershocks, and the PAGER-CAT has numerous fields which help discern the fatalities caused from different hazards for all all earthquakes in its catalogue.

When trying to model the fatality rate of a certain country, the empirical model looks at all of the past earthquakes in that country which appear in both PAGER-CAT and EXPO-CAT, and tries to predict the losses for those earthquakes [3]. Many countries do not have a large number of past earthquakes, or no major ones at all. In order to describe the fatality rate for these countries, the PAGER system groups up similar countries according to population demographics and geological similarities into different regions [3].

While the PAGER empirical model accounts for the uncertainty in the estimated parameters due to the regression, which accounts for variability in the fatality rate model due to hind-casting past earthquakes, it does not account for uncertainty in the shaking, population, or observed loss totals from those past earthquakes. Because of the exponential relationship between the fatality rate and shaking intensity assumed in the PAGER model, the cumulative lognormal function was chosen by looking at the residual plot of estimated losses in the log-log scale for multiple countries [3].

The PAGER model proposed in Jaiswal and Wald [3] calculates the fatality rate $\nu(S_j)$ for shaking intensity $S_j$, where $j = 1, ..., 9$ corresponds to half intensity increments starting at intensity 5 ($j = 1$) and ending at intensity 9 ($j = 9$) using the equation

$$\nu(S_j) = \Phi \left( \frac{\log(S_j/\theta)}{\beta} \right),$$  

(2.18)

where $\Phi$ is the cumulative lognormal distribution function with parameters $\theta$ and $\beta$ corresponding to the mean and standard deviation of a Gaussian function of $\log(S_j)$, log referring to the natural logarithm both here and in the rest of this thesis, unless otherwise specified.

Given a country with $M$ previous earthquakes appearing in the USGS earthquake catalogs, the observed and estimated losses due to shaking are determined, denoted $O_i$ and $E_i$
respectively, where \( i, \ i = 1, \ldots, M \). As mentioned earlier, the estimated deaths from an earthquake are given by the sum of the estimated population \( P_i(S_j) \) at shaking intensity \( S_j \) multiplied by the fatality rate of that intensity \( \nu(S_j) \), or

\[
E_i = \sum_{j=1}^{9} P_i(S_j) \nu(S_j) .
\] (2.19)

In order to compare the observed and estimated losses for each earthquake, [3] proposed using two different residual norms. The first is the traditional L2 norm which defines the root sum of squares error of the estimations from the \( M \) different earthquakes. Call this error \( \zeta_{L2} \) such that

\[
\zeta_{L2} \propto \sqrt{\sum_{i=1}^{M} (E_i - O_i)^2} .
\] (2.20)

This norm works well for determining \( \theta \) and \( \beta \) such that the residuals in predicting the high fatality earthquakes dominate the error, although performing regression to minimize this norm will not factor in the residuals of low or zero fatality earthquakes nearly as much [3].

In order to give more precedence to the low fatality residuals, [3] described the logarithmic norm, which they referred to as the G norm, in order to obtain another error, call it \( \zeta_G \), which is the root sum of squares of the difference \( \log(E_i) - \log(O_i) = \log(E_i/O_i) \), or

\[
\zeta_G \propto \sqrt{\sum_{i=1}^{M} (\log(E_i/O_i))^2} .
\] (2.21)

This norm does very well at predicting the low fatality earthquakes, although it does a worse job of predicting the high fatality earthquakes [3]. While it is important that the fatality rate model does well at predicting the low fatality earthquakes, which are much more frequent, it is equally important if not more so that the model does well at predicting the high fatality earthquakes. Because of this Jaiswal and Wald [3] proposed using what they called the L2G norm, which combines the two norms by summing them after taking the logarithm of the \( \zeta_{L2} \) norm error, or

\[
\zeta_{L2G} = \log(\zeta_{L2}) + \zeta_G .
\] (2.22)
In their paper Jaiswal and Wald [3] show that this norm does a much better job at predicting both low and high fatality earthquakes, compared to either the $L_2$ or $G$ norm alone. Finally, in order to quantify the uncertainty in each estimated loss value, they proposed using the $G$ norm to describe the standard deviation of the residuals of the log difference of observed and estimated losses, which they call $\zeta$ or

$$
\zeta = \sqrt{\frac{1}{M - 2} \sum_{i=1}^{M} \log^2(E_i/O_i)}.
$$

(2.23)

It should be noted that in order to avoid taking log(0) for earthquakes in which zero deaths were observed or estimated, a value of .5 is added to each $E_i$ and $O_i$ values. Thus for each country, [3] used a nonlinear solver known as the Nelder-Mead optimization to determine the optimal lognormal cdf parameters. By minimizing the $L_2G$ norm over possible values of $\theta$ and $\beta$, a fatality rate model was be determined for each country which best predicts the total losses for all past earthquakes for that country. In addition, Equation 2.23 can be used to describe the average uncertainty of the estimated losses[3], giving a distribution for the PAGER estimated total loss.

In Figure 2.2 the fatality rate given MMI curves for three different countries can be seen. They are Indonesia, Iran, which represents the highest fatality rates and therefore the largest vulnerability and fragility of the built environment, and the U.S.(California specifically), which represents the lowest fatality rates in the world relative to shaking.

Note the logarithmic scale of fatality rate in the plot, which reinforces the exponential increase in fatality rate as the shaking intensity increases. Figure 2.3 shows a plot of the residuals of the estimated earthquakes and their spread for Indonesia’s past earthquakes.

In Figure 2.3 the residual spread is centered along the center-line ($\log(E) = \log(O)$) in the log-log plot, and the spread looks mostly constant as the observed or estimated fatalities grow. The spread of residuals is also within an order of magnitude of the center-line, which is the PAGER goal when determining any of its country or regional models. Included below is a sample PAGER output (Figure 2.4) showing a summary of the information that it compiled.

29
Figure 2.2: Fatality rates for three different countries, Indonesia, Iran, and US (California)(from Jaiswal and Wald [3])

Figure 2.3: Fatality residuals plot for all Indonesia earthquakes in EXPO-CAT(from Jaiswal and Wald [3]).
and estimated for the 8.8 magnitude earthquake off the shore of Maule, Chile in February of 2010.

Known as a onePAGER, it represents the typical output product that PAGER sends to all of its subscribers and summarizes the predicted loss (as well as economic loss) distributions as well as population exposed to different shaking intensities. The total loss distribution is represented by the histogram in the top left of the onePAGER, separated into the different probabilities of the possible alert levels. PAGER forms this histogram by subdividing the predicted loss into different orders of magnitude (0-1, 1-10, 10-100,...) and displays the probability of the total loss being in each of these bins. The range of loss with the highest probability determines the alert level that is issued. Estimating 0-1 deaths is considered a ’Green Alert’, and will generally not require very much help outside of the municipalities affected. Estimating 1-100 deaths is considered a ’Yellow Alert’, and indicates that the recommended response is still locally concentrated between municipalities, regions, or states. A PAGER estimated 100-1000 deaths signifies a ’Orange Alert’, indicating a recommended response on the national level. Finally, an estimated loss of >1000 deaths indicates a ’Red Alert’ for which international response is recommended [3]. The overall alert level for an earthquake listed in the onePAGER is determined to be the highest alert level between the fatality and economic loss histograms.

With these components and some basic knowledge of random variables and probability definitions, the current PAGER loss system can now be fully understood. After examining how the information contained in this chapter is used to estimate loss, an enhanced PAGER model will be introduced to incorporate additional uncertainties in the PAGER components.
Figure 2.4: USGS onePAGER summarizing the PAGER estimated loss and exposure after the 2/27/2010 earthquake in Maule, Chile (from earthquakes.USGS.gov).
CHAPTER 3
POTENTIAL ENHANCEMENT TO THE CURRENT PAGER LOSS MODEL

Even in situations where the true loss is in the tail of the loss distribution, it is not necessarily due to the PAGER model being wrong. For instance, regardless of how well the PAGER model estimates the average fatality rate for an earthquake, if a region with a large population experiences a higher shaking intensity than was estimated, or the population for that region was underestimated, or its building stock was significantly lower than the local average, then the PAGER estimation could be several orders of magnitude different than it would have been had more accurate estimates of these quantities been available.

In absence of the actual values of these quantities, which are often unobtainable, better estimates can be obtained by accounting for the variability of the shaking, population, and fatality rate estimates. In order to understand how these additional uncertainties will affect the total loss distribution, further understanding of the current PAGER loss system [3] is needed. Furthermore, the primary interest of this chapter is to determine how close the resulting probability distribution of loss is to a lognormal distribution. The reason for this is because in subsequent chapters the updating will be done assuming a lognormally distributed loss $E$, and if this enhanced loss distribution can be well approximated with a lognormal distribution, then all updating methods in the remainder of this thesis can be equally applicable to either the current loss distribution or the distribution obtained by incorporating these additional uncertainties. Furthermore, this chapter seeks a means of obtaining these enhanced estimates relatively quickly, finding ways of approximating different distributions involved in computing these enhanced loss estimates.
3.1 The Current PAGER Loss Model with Illustration

3.1.1 Description of current PAGER model

Suppose an estimation is needed for the total number of deaths in a certain region after an earthquake. This is determined by dividing the region in question into say \( N \) different subregions (generally 1km x 1km) and estimating the total number of fatalities in each subregion. The number of fatalities for each of these can be estimated by multiplying the population and estimated fatality rate for each subregion. The total estimated loss is then found by taking the sum of all of these estimates across the area of interest. That is, for \( N \) subregions, the total loss \( E \) is found as the sum of the different estimated losses \( E_1, \ldots, E_N \), or

\[
E = \sum_{i=1}^{N} E_i = \sum_{i=1}^{N} \text{pop}_i \nu(s_i),
\]

where \( s_i, \text{pop}_i \) and \( \nu(s_i) \) are respectively the shaking intensity (in MMI), population, and fatality rate given shaking intensity \( s_i \) for subregion \( i = 1, \ldots, N \). The fatality rate given MMI model used by the PAGER system [3] is a cumulative lognormal distribution function with parameters \( \theta \) and \( \beta \) and as mentioned are determined based on the country in which the earthquake is occurring [3]. The true fatality rate \( \nu(s_i) \) is not precisely known for any subregion \( i \) and so the estimated average fatality rate \( \nu_{PAGER}(s_i) \) for that particular country is used instead.

In [3] the PAGER fatality rate model parameters were determined by minimizing the residuals found from trying to predict the total fatalities of past earthquakes for that specific country. As mentioned in the previous chapter, [3] quantified the uncertainty of this model by minimizing the total residuals through the use of the special norm called the L2G norm. After applying this special norm to the minimized residuals, they obtained an error term \( \zeta \) which describes the standard deviation of the residuals found from using the optimal fatality rate parameters. This error term quantifies how well each country model’s parameters predicted the total losses for that particular country’s past earthquakes.
The shaking intensity (and its uncertainty) for each subregion is provided by ShakeMap. The estimated populations are given by LandScan in the form of a 24-hour average, although the uncertainties in these population estimates are not generally available, and the precise knowledge of population at the time of the earthquake is also not available. Because this population is just an average, it is already natural to try and account for the possible variability, which is mostly caused by the time of day of the earthquake. For instance, the population of a city would be expected to be higher than average during the day and lower than average at night, and vice versa for the surrounding suburban and rural areas. This implies that the population of a region is spatially correlated. Rather than modeling these different effects, which is beyond the scope of this work, this thesis assumes some constant coefficient of variation for the population data and does not take the spatial correlation or heteroscedacity of these variables into account.

The PAGER loss system assumes that the total loss $E$ is lognormally distributed with parameters $\mu$ and $\sigma$. This means that its logarithm is a normally distributed random variable with mean $\mu$ and standard deviation $\sigma$. The PAGER model sets $\mu = \log(\sum_i pop_i \nu(s_i))$ and variance $\sigma^2 = \zeta^2$ where $pop_i$, $s_i$, and $\nu(s_i)$ are the average estimations for each subregion $i$.

In order to demonstrate how the PAGER system works, the 2011 Vans Turkey earthquake will be used as an example. According to PAGER, the country model for Turkey has parameters $\theta = 11.0674$ and $\beta = .1063$, and has a model standard deviation of $\zeta = 1.52$. The fatality rate given MMI for Turkey as modeled in Jaiswal and Wald [3] follows Equation 2.18. Recall $\zeta$ arises from the uncertainty in the fatality rate model although it itself is not a fatality rate. Rather, $\zeta$ describes the uncertainty in the total loss model PAGER uses to estimate $E$. Because of the logarithmic feature of the fatality rate model, the norm used to calculate $\zeta$, and the linear correlation and regularity of residuals in the log-log plot, it is instead used by Jaiswal and Wald [3] as the standard deviation of the logarithm of total loss $E$. In order to quantify this uncertainty in total loss, the lognormally distributed $\eta \sim LN(0, \zeta^2)$ is introduced such that
\[ E = \eta \sum_{i}^{N} \text{pop}_i \nu_{\text{PAGER}}(s_i) = \eta E_{\text{PAGER}}. \] (3.2)

This lognormal error factor \( \eta \) was chosen because the country model residuals were determined to be reasonably lognormal, although for some countries in which only a handful of past earthquakes exist, the log-residuals would be better estimated with a student's t distribution, which has wider tails and allows for higher probability of outlying residuals, although this is not currently done by PAGER and is beyond the scope of this paper. Because \( E \) is a function of the random variable \( \eta \), it is also a random variable, recalling that in the current PAGER model does not account for the variability in the variables that make up \( E_{\text{PAGER}} \).

In this light, the distribution of total loss is the distribution of the error factor \( \eta \) scaled by an amount \( E_{\text{PAGER}} \), which is calculated as the sum in Equation 3.2.

A simple example can be used to help one understand how \( E \) is distributed. Let \( X \sim N(\mu, \sigma^2) \), \( Y = e^X \), and \( Z = aY, \ a \in \mathbb{R} \).

Then \( Z = ae^X = e^{\log(a)}e^X = e^{(X+\log(a))} \) and thus \( \log(Z) = X + \log(a) \sim N(\mu + \log(a), \sigma^2) \) and therefore \( Z \sim LN(\mu + \log(a), \sigma^2) \).

Using this, \( E = \eta E_{\text{P}} \implies \log(E) = \log(\eta) + \log(E_{\text{PAGER}}) \), and since \( \eta \sim LN(0, \zeta^2) \), \( \log(\eta) \sim N(0, \zeta^2) \), and thus \( \log(E) \sim N(\log(E_{\text{PAGER}}), \zeta^2) \implies E \sim LN(\log(E_{\text{PAGER}}), \zeta^2) \).

Because \( E \) is lognormally distributed, its parameters can be verified by first determining its mean \( \mu_E \) and standard deviation \( \sigma_E \) and using the following relations:

\[ \sigma^2_{\log E} = \log \left( 1 + \frac{\sigma_E^2}{\mu_E^2} \right), \] (3.3)

\[ \mu_{\log E} = \log(\mu_E) - \sigma^2_{\log E}. \] (3.4)

Using the properties of taking expectations of products and sums of random variables, one obtains
\[ \mu_E = \mathbb{E}[\eta] \mathbb{E} \left[ \sum_{i=1}^{N} \text{pop}_i \nu(s_i) \right] = \mu_\eta \sum_{i=1}^{N} \text{pop}_i \nu(s_i) \]

\[ = \exp \left( \mu_{\log \eta} + \frac{\sigma_{\log \eta}^2}{2} \right) \sum_{i=1}^{N} \text{pop}_i \nu(s_i) \]

\[ = \exp(0 + \zeta^2 / 2) \sum_{i=1}^{N} \text{pop}_i \nu(s_i) \]

\[ = e^{\zeta^2 / 2} \sum_{i=1}^{N} \text{pop}_i \nu(s_i), \]

and

\[ \sigma^2_E = \mathbb{E}[E^2] - \mu^2_E = \mathbb{E}[\eta^2] \mathbb{E} \left[ \left( \sum_{i=1}^{N} \text{pop}_i \nu(s_i) \right)^2 \right] - e^{\zeta^2} \left( \sum_{i=1}^{N} \text{pop}_i \nu(s_i) \right)^2 \]

\[ = (\mathbb{E}[\eta^2] - e^{\zeta^2}) \left( \sum_{i=1}^{N} \text{pop}_i \nu(s_i) \right)^2 \]

\[ = \sigma^2_\eta \left( \sum_{i=1}^{N} \text{pop}_i \nu(s_i) \right)^2. \]

Then, using the fact that \( \sigma^2_\eta = \left( e^{\sigma_{\log \eta}^2} - 1 \right) e^{2\mu_{\log \eta} + \sigma_{\log \eta}^2} = \left( e^{\zeta^2} - 1 \right) e^{\zeta^2} \) for lognormally distributed \( \eta \), this implies

\[ \sigma^2_{\log \, E} = \log \left( 1 + \frac{\sigma^2_E}{\mu^2_E} \right) \]

\[ = \log \left( \frac{e^{\sigma_{\log \eta}^2} - 1}{1} e^{2\mu_{\log \eta} + \sigma_{\log \eta}^2} \left( \sum_{i=1}^{N} \text{pop}_i \nu(s_i) \right)^2 \right) \]

\[ = \log \left( 1 + e^{\zeta^2} - 1 \right) \]

\[ = \zeta^2, \]

and
\[
\mu_{\log E} = \log(\mu_E) - \sigma_{\log E}^2/2
= \log(e^{\xi^2/2} + \log\left(\sum_{i=1}^{N} \text{pop}_i \nu(s_i)\right)) - \zeta^2
= \log\left(\sum_{i=1}^{N} \text{pop}_i \nu(s_i)\right),
\]
and thus \(E \sim LN\left(\log\left(\sum_{i=1}^{N} \text{pop}_i \nu(s_i)\right), \zeta^2\right)\), as it was defined.

It is important to understand that the value \(e^\mu = E^{\text{PAGE}}\) is not the mean of \(E\), but is actually its median. This is due to the properties of a lognormal distribution, where the median of \(E\), referred to as \(GM_E\), equals the exponential of \(\mu\), and the fact that the median by definition describes the value for which there is an equal probability (.5) of \(E\) being either less than or greater than its median. The actual mean of a lognormal random variable is \(\mu_E = e^{\mu + \sigma^2/2}\). The reason that the median of \(E\) is used instead of its mean as the loss estimator is because the median of a lognormal random variable is known as the geometric mean of that variable and is an unbiased estimator. The mean of a lognormal variable on the other hand, is extremely sensitive to outliers making it a biased estimator. The geometric mean is similar to the more traditional arithmetic mean in that it describes the point around which the distribution is spread in the logarithmic sense. Thus when dealing with lognormal random variables one is often interested in using the median as the primary estimator. In practice, one generally deals with everything in terms of the logarithm of \(E\) so that all calculations are done in respect to normal random variables.

The variance of a lognormal random variable \(E\) is \(\sigma_{E}^2 := (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}\) and from the definition of \(\mu_E\), the lognormal distribution parameter \(\sigma\) is be defined as

\[
\sigma = \sqrt{\log\left(1 + \frac{\sigma_{E}^2}{\mu_{E}^2}\right)}.
\]  

(3.5)

In order to understand how the lognormal distribution is spread about the geometric mean, the geometric standard deviation \(GSD_E = e^\sigma\) is formed. This value can be used
to construct confidence intervals for lognormal random variables in a manner analogous to determining confidence intervals for normal random variables. Suppose that the 95% confidence interval of lognormal $E$ is desired. This is equivalent to finding the 95% confidence interval for the normally distributed $\log(E)$, which is $[\mu - 1.96\sigma, \mu + 1.96\sigma]$. Because the logarithm is a nondecreasing and convex upward operator, there exists a bijection (one-to-one and onto) between confidence bounds of normal and lognormal random variables. This means that the confidence intervals of $\log(E)$ can be exponentiated in order to obtain the confidence intervals of $E$. Given this information, the 95% confidence interval for $E \sim LN(\mu, \sigma^2)$ is

$$CI_{E}^{95} = [e^{\mu - 1.96\sigma}, e^{\mu + 1.96\sigma}] = [e^{\mu}/e^{1.96\sigma}, e^{\mu}e^{1.96\sigma}] = [GM_E * GSD_{E}^{-1.96}, GM_E * GSD_{E}^{1.96}]$$.

### 3.1.2 Illustration of current PAGER model

The example scenario throughout this chapter and the next is the 2011 Van earthquake in Turkey. The PAGER fatality rate model for Turkey has parameters $\theta = 11.0674$ and $\beta = .1063$ and has uncertainty $\zeta = 1.52$. Using the grid files of population and shaking intensity for each cell provided by LandScan and Shakemap, respectively, along with the Jaiswal and Wald [3], the mean and standard deviation of the logarithm of loss are

$$\mu = \mu_{\log E} = \log \left( \sum_{i=1}^{N} \text{pop}_i \nu(s_i) \right) = \log(570.0121) = 6.3457,$$

$$\sigma = \sigma_{\log E} = \zeta = 1.52.$$

As mentioned, the primary estimators for lognormal random variables are the geometric mean $GM_E = e^{\mu} = 570.0121$ and geometric standard deviation $GSD(E) = e^{\sigma} = 4.5722$. These can be used to form confidence intervals for lognormal random variables. If an interval of width
$2n\sigma$ is taken, where $n = \mathbb{R}$ around the mean of the log of $E$, $\mu$, then one obtains the following interval in terms of log $E$: $[\mu - n\sigma, \mu + n\sigma]$. Thus there is a corresponding confidence level that $E$ lays in the interval $[\exp(\mu - n\sigma), \exp(\mu + n\sigma)] = [570.0121 \times 4.5722^{-n}, 570.0121 \times 4.5722^n]$.

Therefore the 95% confidence interval for $E$, corresponding to $n \approx 1.96$ is 

$$CI_E^{(95)} = [28.98, 11213].$$

Finally, using the cumulative distribution function of $E$ the histogram of the total loss distribution can be formed in the format that PAGER uses to decide which alert level is predicted for each earthquake. Rather than broadcast the median loss, or even the entire loss density, PAGER chooses to publish the results in a histogram whose bins represent one order of magnitude in terms of powers of 10. As mentioned, the histogram bin with the largest probability is used to determine the overall alert level for that earthquake. Recall that the alert level goes from green ($0 - 1$ deaths) to yellow ($1 - 100$ deaths) to orange ($100 - 1000$ deaths) to red ($> 1000$ deaths), where each color indicates the likely amount of support needed for that particular earthquake, specifically local, regional, national, international respectively. PAGER does not broadcast its median estimated value because this value misrepresents the accuracy of the PAGER model. For instance, suppose that PAGER underestimates the total loss by a factor of 2. While this could be seen as an inaccurate estimate, the large spread in the total loss distribution actually makes this median estimate quite accurate.

With $GM_E \approx 570$, one can see (Figure 3.1) an 'orange' alert for this earthquake.

As mentioned in the introduction, one of the primary goals of this thesis is to improve upon the existing PAGER model by incorporating additional uncertainties in the population, shaking intensity, and fatality rates which the loss estimation model uses. Now that a layout of the current PAGER model has been given, an improved model will now be discussed which takes these uncertainties into account. Accounting for these should help explain the additional variability in the total estimated loss beyond the current PAGER model’s capabilities.
Figure 3.1: A histogram of the total loss distribution for the 2011 Van earthquake.

3.2 Accounting for Additional Uncertainties in the PAGER Model

The aforementioned additional uncertainties and their effects will be incorporated in order of simplicity. The simplest adaptation of the current PAGER model is to only assume variability in the population and fatality rate, and not in the shaking intensity. Then the estimated loss $E_i$ in a subregion $i$ can be seen as a random variable comprised of the product of two random variables.

First, this work seeks to demonstrate that while the resulting distribution will be approximately lognormal, determining the best parameters cannot be done by simply determining the mean and variance of $E$ and using these to estimate the lognormal parameters. As mentioned, estimations of the mean of a lognormal random variable are generally biased and tend to overestimate the expected median loss. The reason for this is because after incorporating these uncertainties, the resulting distribution is close to lognormal, except it has a much heavier left tail than right compared to an actual lognormal random variable. This means that using the mean of a lognormal random variable to approximate this distribution will give more weight to the larger losses than the actual distribution, which can lead to much larger estimations of the median value.

The benefit of the method of moments matching is that expectations and variances of sums of random variables are quite easy, particularly when they are uncorrelated. In fact, no
distributional assumptions need to be made in this case because estimation of the mean and variance of $E$ involves only knowing the mean and variances of the individual populations and fatality rates. However, the only way to avoid the bias obtained from estimating the mean of a lognormal random variable is to instead estimate the median of total loss, which is not biased. Fortunately, the median of a lognormal variable is the exponential of the mean of its logarithm, and so one can estimate the lognormal parameters of $E$ by matching the moments of the logarithm of $E$ instead. However, the logarithm of a sum is not equal to the sum of logarithms, which makes the method of moments much more difficult to perform. Furthermore, because PAGER publishes a histogram instead of just the median value, one needs to form the entire distribution of loss $E$, which cannot be obtained from just knowing its mean and variance.

3.2.1 Uncertainty in the estimated population and fatality rate

For mathematical simplicity suppose that there is a population error factor $\epsilon_{p_i} \sim LN(0, \tau_{p_i}^2)$ and fatality rate error factor $\epsilon_{\nu}(s_i) \sim LN(0, \tau_{\nu}(s_i))$ for each cell, which represents the intraevent uncertainty in fatality rate due to differences in buildings, ground composition, and a number of other variables at different locations.

Suppose the random variables $Pop_i = \epsilon_{p_i} pop_i$ and $FR_i = \epsilon_{\nu}(s_i) \nu(s_i)$ are introduced, where for now consider $s_i$ to be deterministic. Then $E_i$ is a random variable for each cell, and since $\epsilon_{p_i}$ and $\epsilon_{\nu_i}$ are lognormally distributed, the same technique used to show that $E \sim LN(\log(E_{PAGER}), \zeta^2)$ can be used to show that $Pop_i \sim LN(\log(pop_i), \tau_{p_i}^2)$ and $FR_i \sim LN(\log(\nu(s_i)), \tau_{\nu_i}^2)$, and the total loss is now

$$E = \eta \sum_i E_i = \eta \sum_i Pop_i FR_i.$$ 

The distribution of $\eta$ will be the same as it was without assuming other uncertainties, i.e.,

$$\eta \sim LN(0, \zeta).$$

The distribution of the total loss estimate for an arbitrary cell $i$ can be determined using the fact that the product of lognormal random variables is a lognormal random
variable. From the definitions of $\text{Pop}_i$ and $\text{FR}_i$, $E_i = \text{Pop}_i \cdot \text{FR}_i \implies \log(\text{Pop}_i) = \log(\text{Pop}_i) + \log(\text{FR}_i) = \log(\epsilon_p) + \log(\epsilon_{\nu}(s_i)) + \log(\text{pop}_i \nu(s_i))$. Now $\log(\epsilon_p) \sim N(0, \tau_p)$ and $\log(\epsilon_{\nu}(s_i)) \sim N(0, \tau_{\nu}(s_i))$ and thus since $\log(\text{pop}_i \nu(s_i))$ is deterministic

$$\log(E_i) \sim N \left( \log(\text{pop}_i \nu(s_i)), \sqrt{\tau_p^2 + \tau_{\nu}^2(s_i)} \right)$$

and thus

$$E_i \sim LN \left( \log(\text{pop}_i \nu(s_i)), \sqrt{\tau_p^2 + \tau_{\nu}^2(s_i)} \right).$$

Because $E = \eta \sum_i E_i$, $\log(E) = \log(\eta) + \log(\sum_i E_i)$. However, the sum of lognormal random variables is not a lognormal random variable, and therefore the logarithm of their sum cannot be a normal random variable. This means that $\log(E)$ cannot be a normally distributed and therefore $E$ is no longer lognormally distributed. However, it has been shown that in some cases the sum of $N$ lognormal random variables can be approximated by a lognormal distribution, although not in general, see the Schwartz and Yeh [16] or Fenton and Wilkinson [17] approximations. This section will first estimate the distribution of $E$ using a method of moments, which involves the propagation of error for variance and expectations of sums of random variables. Alternatively, another approach will be demonstrated which uses the Monte Carlo method to generate independent realizations of the population and fatality rate distributions. This will allow for a histogram to be formed and for approximations of the distribution of $E$ and more specifically $\log(E)$ to be made which do not suffer from the bias obtained using the method of moments of $E$ using the propagation of expectations and variance method.

### 3.2.1.1 Expectation Method:

This method shows a straightforward way to calculate $\mu_E$ and $\sigma_E$ given the model $E = \eta \sum_{i=1}^N \text{Pop}_i \cdot \text{FR}_i$ and describes how to approximate the lognormal parameters $\mu$ and $\sigma$ of $E$ from them. First $\mu_E$ is computed:
\[ \mu_E = \mathbb{E}[E] = \mathbb{E}[\eta] \cdot \mathbb{E} \left[ \sum_{i=1}^{N} E_i \right] + \text{cov} \left( \eta, \sum_{i=1}^{N} E_i \right) \]

\[ = \mathbb{E}[\eta] \sum_{i=1}^{N} \mathbb{E}[E_i] + \sum_{i=1}^{N} \text{cov}(\eta, E_i) \]

\[ = e^{\zeta^2/2} \sum_{i=1}^{N} \mathbb{E}[\text{Pop}_i FR_i] + \sum_{i=1}^{N} e^{\zeta^2/2} \text{cov}(\text{Pop}_i, FR_i) + \text{cov}(\eta, \text{Pop}_i FR_i) \]

where \( \text{cov}(\cdot, \cdot) \) is the covariance between two different random variables. If \( \eta \) is assumed to be independent of each \( E_i, i = 1, ..., N \), and \( \text{Pop}_i \) is independent of \( FR_i, i = 1, ..., N \), then, using the fact that \( \mathbb{E}[\text{Pop}_i] = \exp(\log(\text{Pop}_i) + \tau_p^2/2) \) and \( \mathbb{E}[FR_i] = \exp(\log(\nu(s_i)) + \tau_\nu^2/2) \), one obtains

\[ \mu_E = e^{\zeta^2/2} \sum_{i=1}^{N} \text{Pop}_i \nu(s_i) \exp \left( \frac{\tau_p^2 + \tau_\nu^2(s_i)}{2} \right) . \]

The variance of \( E, \sigma_E^2 \), is calculated as follows (assuming independence of \( \eta \) and all \( E_i \), as well as independence of \( \text{Pop}_i \) and \( FR_i \) for all \( i = 1, ..., N \):

\[ \sigma_E^2 = \mathbb{E}[E^2] - \mu_E^2 \]

\[ = \mathbb{E} \left[ \eta^2 \left( \sum_{i} \text{Pop}_i FR_i \right)^2 \right] - e^{\zeta^2} \left( \sum_{i=1}^{N} \text{Pop}_i \nu(s_i) \exp \left( \frac{\tau_p^2 + \tau_\nu^2(s_i)}{2} \right) \right)^2 \]

\[ = \mathbb{E}[\eta^2] \left( \sum_{i=1}^{N} \mathbb{E}[\text{Pop}_i^2] \mathbb{E}[FR_i^2] + \sum_{j \neq i} \mathbb{E}[\text{Pop}_i \text{Pop}_j] \mathbb{E}[FR_i FR_j] \right) \]

\[ - e^{\zeta^2} \left( \sum_{i=1}^{N} \text{Pop}_i^2 \nu^2(s_i) \exp(\tau_p^2 + \tau_\nu^2(s_i)) \right) \]

\[ + \sum_{j \neq i} \text{Pop}_i \text{Pop}_j \nu(s_i) \nu(s_j) \exp \left( \frac{\tau_p^2 + \tau_\nu^2(s_i)}{2} \right) \exp \left( \frac{\tau_p^2 + \tau_\nu^2(s_j)}{2} \right) \]
If one assumes no spatial correlation in population or fatality rate, then \( \text{cov}(\text{Pop}_i, \text{Pop}_j) = 0 \) and \( \text{cov}(F_{Ri}, F_{Rj}) = 0 \), for all \( i, j = 1, \ldots, N \), then the variance of \( E \) is

\[
\sigma_E^2 = \sum_{i=1}^{N} \mathbb{E}[\eta^2] \mathbb{E}[\text{Pop}_i^2] \mathbb{E}[F_{R_i}^2] - e^{\sigma^2 \text{Pop}_i \nu^2} \exp(\tau_{p_i}^2 + \tau_{\nu}^2(s_i)) \\
+ \sum_{i=1}^{N} \sum_{j \neq i} (\mathbb{E}[\eta^2] - e^{\sigma^2 \text{Pop}_i \nu^2} \nu(s_i) \nu(s_j) \exp\left(\frac{\tau_{p_i}^2 + \tau_{\nu}^2(s_i)}{2}\right) \exp\left(\frac{\tau_{p_j}^2 + \tau_{\nu}^2(s_j)}{2}\right))
\]

\[
= \sum_{i=1}^{N} \left(\mathbb{E}[\eta^2] \mathbb{E}[\text{Pop}_i^2] \mathbb{E}[F_{R_i}^2] - e^{\sigma^2 \text{Pop}_i \nu^2} \exp(\tau_{p_i}^2 + \tau_{\nu}^2(s_i))
\right) - \sigma^2 \text{Pop}_i \nu^2(s_i) \exp(\tau_{p_i}^2 + \tau_{\nu}^2(s_i))
+ \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma^2 \text{Pop}_i \text{Pop}_j \nu(s_i) \nu(s_j) \exp\left(\frac{\tau_{p_i}^2 + \tau_{p_j}^2 + \tau_{\nu}^2(s_i) + \tau_{\nu}^2(s_j)}{2}\right)
\]

\[
= \sum_{i=1}^{N} \mathbb{E}[\eta^2] \left(\mathbb{E}[\text{Pop}_i^2] \mathbb{E}[F_{R_i}^2] - \text{Pop}_i \nu^2(s_i) \exp(\tau_{p_i}^2 + \tau_{\nu}^2(s_i))\right)
+ \sigma^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \text{Pop}_i \text{Pop}_j \nu(s_i) \nu(s_j) \exp\left(\frac{\tau_{p_i}^2 + \tau_{p_j}^2 + \tau_{\nu}^2(s_i) + \tau_{\nu}^2(s_j)}{2}\right)
\]

\[
= (\sigma^2 + \mu^2) \sum_{i=1}^{N} \left(\sigma^2_{\text{Pop}_i} \sigma^2_{F_{R_i}} + \sigma^2_{\text{Pop}_i} \mu^2_{F_{R_i}} + \mu^2_{\text{Pop}_i} \sigma^2_{F_{R_i}} + \sigma^2_{\text{Pop}_i} \mu^2_{F_{R_i}}\right) + \sigma^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \mu_{\text{Pop}_i} \mu_{\text{Pop}_j} \mu_{F_{R_i}} \mu_{F_{R_j}},
\]

with
\[ \mu_\eta = e^{\zeta^2/2}, \quad \sigma_\eta^2 = (e^{\zeta^2} - 1)e^{\zeta^2}, \]
\[ \mu_{Pop_i} = \text{pop}_i e^{\tau_i^2/2}, \quad \sigma_{Pop_i}^2 = (e^{\tau_i^2} - 1)\text{pop}_i^2 e^{\tau_i^2}, \]
\[ \mu_{FR_i} = \nu(s_i) e^{\tau_i^2(s_i)/2}, \quad \sigma_{FR_i}^2 = (e^{\tau_i^2(s_i)} - 1)\nu^2(s_i) e^{\tau_i^2(s_i)}. \]

Finally, using \( \sigma_E^2 \) and \( \mu_E \) to determine the lognormal parameters \( \sigma \) and \( \mu \) of \( E \) (the standard deviation and mean of \( \log(E) \), respectively), one obtains the following:

\[
\sigma^2 = \log \left( 1 + \frac{(\sigma_\eta^2 + \mu_\eta^2) \sum_{i=1}^{N} \left( \sigma_{Pop_i}^2 \sigma_{FR_i}^2 + \sigma_{Pop_i}^2 \mu_{FR_i}^2 + \mu_{Pop_i}^2 \sigma_{FR_i}^2 \right) \mu_\eta^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \mu_{Pop_i} \mu_{Pop_j} \mu_{FR_i} \mu_{FR_j}}{\mu_\eta^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \mu_{Pop_i} \mu_{Pop_j} \mu_{FR_i} \mu_{FR_j}} \right)
\]
\[
= \log \left( 1 + \frac{\sigma_\eta^2 \sum_{i=1}^{N} \left( \sigma_{Pop_i}^2 \sigma_{FR_i}^2 + \sigma_{Pop_i}^2 \mu_{FR_i}^2 + \mu_{Pop_i}^2 \sigma_{FR_i}^2 \right) \mu_\eta^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \mu_{Pop_i} \mu_{Pop_j} \mu_{FR_i} \mu_{FR_j}}{\mu_\eta^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \mu_{Pop_i} \mu_{Pop_j} \mu_{FR_i} \mu_{FR_j}} \right)
\]
\[
= \log \left( e^{\zeta^2} + D \right),
\]

where

\[ D = \frac{(\sigma_\eta^2 + \mu_\eta^2) \sum_{i=1}^{N} \left( \sigma_{Pop_i}^2 \sigma_{FR_i}^2 + \sigma_{Pop_i}^2 \mu_{FR_i}^2 + \mu_{Pop_i}^2 \sigma_{FR_i}^2 \right) \mu_\eta^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \mu_{Pop_i} \mu_{Pop_j} \mu_{FR_i} \mu_{FR_j}}{\mu_\eta^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \mu_{Pop_i} \mu_{Pop_j} \mu_{FR_i} \mu_{FR_j}} \]

is a positive constant which implies that \( \sigma > \zeta \). Additionally, the parameter \( \mu \) can be calculated:

\[ \mu = \log(\mu_E) - \sigma^2/2 = \log \left( e^{\sigma^2/2} \sum_{i=1}^{N} \mu_{Pop_i} \mu_{FR_i} \right) - \log(e^{\zeta^2} + D)/2 \]
\[ = \log \left( \sum_{i=1}^{N} \mu_{Pop_i} \mu_{FR_i} \right). \]

Thus the approximate distribution of total loss \( E \) is

\[ E \sim LN \left( \log \left( \sum_{i=1}^{N} \mu_{Pop_i} \mu_{FR_i} \right), \log \left( e^{\zeta^2} + D \right) \right). \]
Confidence intervals of this loss can then be estimated. Suppose a width of $2n \in \mathbb{R}$ geometric standard deviations around the geometric mean, the corresponding confidence interval is then

$$CI_E^{(\Phi^{-1}(n))} = \left[ (e^{\zeta^2} + D)^{-1} \sum_{i=1}^{N} \mu_{Pop_i} \mu_{FR_i}, \quad (e^{\zeta^2} + D) \sum_{i=1}^{N} \mu_{Pop_i} \mu_{FR_i} \right],$$

where $\Phi^{-1} : \mathbb{R} \to [0, 1]$ is the inverse error function such that $\Phi^{-1}(z) = p$, where $p = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ is the probability that a standard normally distributed $Z < z \in \mathbb{R}$ described in the second chapter.

The 2011 Van earthquake in Turkey is used to illustrate this method. Recall that the fatality rate model for Turkey has parameters $\theta = 11.0674$ and $\beta = .1063$ and has uncertainty $\zeta = 1.52$ which is the standard deviation of the logarithm of the fatality rate model error $\eta$. In this example, the population coefficient of variation (CoV) is assumed to be .30 for every cell. The intraevent fatality rate CoV is also chosen for each cell, which depends on the shaking intensity for that cell. For the purposes of demonstration, this works assumes that $CoV_{FR_i}(s_i) = as_i + b$. Based off of basic assumptions provided by Kishor Jaiswal from the USGS, this relationship was chosen assuming that smaller shaking intensities have larger variance in fatality rate. Choosing $CoV_{FR_i}(4) = .7$ and $CoV_{FR_i}(9) = .3$ implies that $a = -.08$ and $b = 1.02$. Again this is purely for demonstration purposes so that the effect of a linear relationship between the shaking intensity and the fatality rate CoV can be seen when describing the distribution of the fatality rate $FR_i$ for a subregion $i$. That is, subregions with different expected shaking values will have different fatality rate variances. Under these assumptions,

$$\mu = \log \left( \sum_{i=1}^{N} \mu_{Pop_i} \mu_{FR_i} \right) = 6.4543 \implies GM(E) = e^{\mu} = 635.4113,$$

$$\sigma = \log \left( e^{\zeta^2} + D \right) = 1.5215 \implies GSD(E) = e^{\sigma} = 4.5790.$$

Finally, the 95% confidence interval for $E$ is
\[ C I_{E}^{(95)} = \left[ e^\mu e^{-1.96 \log(e^{\kappa^2}+D)}, e^\mu e^{1.96 \log(e^{\kappa^2}+D)} \right] \]
\[ = [6.8006, 59369]. \]

3.2.1.2 Monte Carlo Method:

Another way of examining how these additional uncertainties affect the distribution of total loss \( E \) would be to perform a Monte Carlo simulation. This method would seek to generate an ensemble of values for \( \eta \) and each \( E_i \), and then form a histogram from the different realizations of \( E \) calculated from this ensemble. That is, given \( M \) i.i.d samples for each \( E_i \) and \( \eta \), \( M \) i.i.d. samples of \( E \) can be formed by summing the \( E_i \) values and multiplying them by the ensemble \( \eta \) values. As the sample size \( M \) increases, the law of large numbers says that the sample statistics of \( E \) will approach its true statistics, and the histogram of the ensemble of realizations of \( E \) will more closely resemble the probability density of \( E \).

In order analyze how the statistics of \( \log(E) \) converge as the number of realizations grows for the 2011 Van earthquake example, suppose that successively larger ensembles of \( E \) are generated and analyzed. Because the primary interest of this chapter is to see how closely \( E \) can be approximated with a lognormal distribution, the statistics of the ensembles of \( \log(E) \) will be compared to the known statistics of a normal distribution. Below (Figure 3.2) a histogram of \( \log(E) \) is shown which was formed using \( M = 1,024,000 \) samples.

This histogram of \( \log(E) \) looks very close to a normal distribution, and looking at the statistics as a function of sample size \( M \), the 3rd and 4th central statistical moments, skewness and kurtosis of \( \log(E) \), respectively, are close to the skewness of 0 and kurtosis of 3, which are the skewness and kurtosis of a normal random variable. Below (Figure 3.3) is a
Figure 3.2: Histogram of log($E$) for the 2011 Van Earthquake.

table showing these sample statistics for each sample size $M = 1000, 2000, ...1,024,000$.

<table>
<thead>
<tr>
<th>$M$</th>
<th>Mean log($E$)</th>
<th>Std log($E$)</th>
<th>Skewness log($E$)</th>
<th>Kurtosis log($E$)</th>
<th>time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>6.4379488</td>
<td>1.4636028</td>
<td>-0.1273948</td>
<td>2.9459878</td>
<td>14.0177833</td>
</tr>
<tr>
<td>2000</td>
<td>6.4842538</td>
<td>1.4975828</td>
<td>-0.0858158</td>
<td>3.1264268</td>
<td>14.6347683</td>
</tr>
<tr>
<td>4000</td>
<td>6.4883788</td>
<td>1.5259608</td>
<td>-0.0608878</td>
<td>3.0880628</td>
<td>28.6473373</td>
</tr>
<tr>
<td>8000</td>
<td>6.4813768</td>
<td>1.5367748</td>
<td>-0.0407798</td>
<td>3.0535778</td>
<td>58.0383233</td>
</tr>
<tr>
<td>16000</td>
<td>6.4826568</td>
<td>1.5239238</td>
<td>-0.0209948</td>
<td>3.0238908</td>
<td>161.8961573</td>
</tr>
<tr>
<td>32000</td>
<td>6.4670208</td>
<td>1.5183208</td>
<td>-0.0127658</td>
<td>3.0445008</td>
<td>231.4307733</td>
</tr>
<tr>
<td>64000</td>
<td>6.4677158</td>
<td>1.5159348</td>
<td>0.0151628</td>
<td>3.0237788</td>
<td>707.6293963</td>
</tr>
<tr>
<td>128000</td>
<td>6.4588498</td>
<td>1.5161858</td>
<td>0.0075258</td>
<td>3.0055178</td>
<td>894.1835123</td>
</tr>
<tr>
<td>256000</td>
<td>6.4535248</td>
<td>1.5186978</td>
<td>0.0025448</td>
<td>2.9896188</td>
<td>1813.2234543</td>
</tr>
<tr>
<td>512000</td>
<td>6.4529268</td>
<td>1.5206018</td>
<td>-0.0010508</td>
<td>2.9936158</td>
<td>3559.5032163</td>
</tr>
<tr>
<td>1024000</td>
<td>6.4524438</td>
<td>1.5195498</td>
<td>-0.0029728</td>
<td>2.9969568</td>
<td>6539.7330593</td>
</tr>
</tbody>
</table>

Figure 3.3: Statistics of log($E$) Using Monte Carlo Simulation.

Next, in Figure 3.4 the different statistics are plotted as a function of the Monte Carlo sample size $M$ in order to see how they converge as $M$ increases.

Finally, a plot of the standard mean errors is shown below Figure 3.5 in order to check whether they are converging at the theoretical convergence rate of $1/\sqrt{M}$ that the Monte Carlo method guarantees.

The results indicate that in this particular case one is justified in approximating $E$ with a lognormal distribution, and therefore, one can simply calculate the mean $\mu_E$ and standard deviation $\sigma_E$ of $E$ using the expectation method demonstrated earlier in this section. The
Figure 3.4: Plots of Statistics for Different Sample Size \( M \).

Figure 3.5: Standard Error of Mean of \( \log(E) \).
method of moments can then be used estimate the mean $\mu$ and standard deviation $\sigma$ of the logarithm of $E$ using the following relationships for lognormal distributions:

$$
\sigma = \sqrt{\log \left( 1 + \frac{\sigma^2}{\mu^2} \right)},
$$

$$
\mu = \mu E e^{-\sigma^2/2}.
$$

It should be reiterated that these results only correspond to the case where only the uncertainty in population and intraevent fatality rates are accounted for, and does not account for any uncertainty in the shaking intensity, which will be investigated shortly. Figure 3.6 shows the parameter values of $E$ without these additional uncertainties compared to the parameter values obtained using both the Monte Carlo simulation as well as the method of moments approximation.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$GM(E) = e^\mu$</th>
<th>$GSD(E) = e^\sigma$</th>
<th>95% Conf. Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current PAGER Model Results</td>
<td>6.3457</td>
<td>1.52</td>
<td>570.0121</td>
<td>4.572</td>
<td>[28.98, 11213]</td>
</tr>
<tr>
<td>Monte Carlo Approximation Results</td>
<td>6.4522</td>
<td>1.5212</td>
<td>634.0958</td>
<td>4.578</td>
<td>[32.16, 12503]</td>
</tr>
<tr>
<td>(M =1.024*10^6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lognormal Approximation Results</td>
<td>6.4543</td>
<td>1.5215</td>
<td>635.4113</td>
<td>4.579</td>
<td>[32.21, 12536]</td>
</tr>
</tbody>
</table>

Figure 3.6: Table of fatality distribution results obtained using the different methods.

The results indicate that one can achieve the same order of precision when performing the Monte Carlo approximation up to $1.024 \times 10^6$ realizations compared to just assuming that $E$ is lognormally distributed and determining its parameters via the method of moments and the propagation of expectations method. One can also see that incorporating intraevent uncertainties in population and fatality rate has increased the geometric mean and geometric standard deviation, leading to a 95% confidence interval that has a larger range than if these additional uncertainties were not included.

It should be of note that the assumption that the intraevent population uncertainty follows a lognormal distribution was purely for mathematical convenience. In reality this uncertainty should be more similar to a normal distribution, and so one would have to see...
if the same results hold when changing this assumption about the population uncertainty. While this section showed that reliable results can be achieved without using the expensive Monte Carlo approach, this will not be the case in the following sections. Upon factoring in the shaking intensity uncertainty, the expectation method will no longer give accurate estimations, and the more robust Monte Carlo method must be used. Before seeing how the total loss distribution will be changed when incorporating uncertainties in the shaking intensities, the distributions of the fatality rate equation $\nu(s)$ and intraevent fatality rate error $\epsilon(\nu(s))$ will first be investigated.

### 3.2.2 Accounting for uncertainty in shaking intensity

Suppose that the shaking intensity for each subregion is uncertain and follows a normal distribution with the mean and standard deviation given by ShakeMap, that is, $S_i \sim N(\mu_{s_i}, \sigma_{s_i})$ for all cells $i$, where $S_i$ is the random variable version of $s_i$. Then $\nu(S_i)$ becomes a random variable as well, with

$$\nu(S_i) = \Phi\left(\frac{\log\left(\frac{S_i}{\theta}\right)}{\beta}\right).$$

This section will first investigate the affect that these uncertain shaking intensities have on the fatality rate and its error $\epsilon(\nu(S_i))$. After this, the distribution for the subregional losses $E_i$ will be approximated using Monte Carlo simulation. Finally, given a means of generating realizations for each $E_i$, the distribution of total loss $E = \sum_i E_i$ will then be approximated from the realizations of each $E_i$.

#### 3.2.2.1 Approximating the Distribution of $\nu(S_i)$

In order to understand how the variable $\nu(S_i)$ is distributed given the assumed normally distributed $S_i$, Monte Carlo simulation is again used. The usual procedure is followed in which an $M$ sized ensemble $s_{i,j}$, $j = 1, ..., M$, of $S_i$ is generated for a general subregion $i$. 

52
For each realization $s_{i,j}$, the fatality rate $\nu(s_{i,j})$ is then calculated giving an $M$ sized sample of the fatality rate $\nu(S_i)$ for cell $i$. This can be used to approximate the statistics of $\nu(S_i)$ as well as its distribution through a histogram. For Monte Carlo methods, determining the skewness and kurtosis of a distribution accurately (1-3 digits of precision) generally requires upwards of 1-10 million samples. Given some arbitrary cell $i$ from the 2011 Van earthquake data, the distribution of $\nu(S_i)$ is approximated in order to demonstrate this Monte Carlo method.

Suppose for cell $i$, $\mu_{s_i} = 8.09$, and $\sigma_{s_i} = 1.0$. Further suppose that 10 million samples are drawn from the distribution $N(\mu_{s_i}, \sigma_{s_i})$ and the fatality rate is calculated for each of them. Below is a table summarizing the first four statistics for both $\nu(S_i)$ and $\log(\nu(S_i))$. Looking at the statistics for $\log(\nu(S_i))$ one can see that it is definitely not normally distributed because of its negative skewness and kurtosis $>3$. The histogram of $\log(\nu(S_i))$ in Figure 3.7, shows that $\log(\nu(S_i))$ is a left skewed distribution with a heavier left tail than for a typical normal distribution.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Sample Mean</th>
<th>Sample Std</th>
<th>Sample Skewness</th>
<th>Sample Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu(S_i)$</td>
<td>0.0062</td>
<td>0.0132</td>
<td>5.6771</td>
<td>57.009</td>
</tr>
<tr>
<td>$\log(\nu(S_i))$</td>
<td>-6.7206</td>
<td>2.241</td>
<td>-0.8013</td>
<td>4.1096</td>
</tr>
</tbody>
</table>

Figure 3.7: Sample statistics and Beta distribution statistics for $\nu(S_i)$.

Next the histogram of the ensemble of fatality rates can be seen (Figure 3.8). Finally the approximate probability density of $\nu(S_i)$ is computed and plotted (Figure 3.9) from this histogram.

The results indicate that $\nu(S_i)$ is not lognormal, although it does have some logarithmic behavior in its left tail, however the upper tail exhibits decay closer to a normal random variable as opposed to a lognormal one. This kind of distribution is present for any value of $\mu_{S_i}$ and $\sigma_{S_i}$ due to the fact that it involves the logarithm of a normal random variable. Given an idea of how the fatality rate is distributed given uncertainty in shaking, the fol-
Figure 3.8: Histogram of fatality rate as a function of random $S_i$.

Figure 3.9: Density functions for sample PDF of $\nu(S_i)$ versus beta density function.
lowing subsection looks at how it affects the distribution of the intraevent fatality rate error \( \epsilon_\nu(S_i) \).

### 3.2.2.2 Approximating the Distribution of \( \epsilon_\nu(S_i) \)

The uncertainty in \( S_i \) will also have an affect on the intraevent fatality rate error because the standard deviation of its logarithm, \( \tau_\nu(S_i) \), is a function of \( S_i \). Thus \( \tau_\nu(S_i) \) is itself a random variable with some distribution. Assuming that \( S_i \) is normally distributed, the distribution of \( \tau_\nu(S_i) \) and its affect on the distribution of \( \epsilon_\nu(S_i) \) can be investigated.

Again, this work assumes that the CoV of \( \epsilon_\nu(s) \) is linearly dependent on the shaking intensity \( s \), or \( CoV_{\epsilon_\nu(s)} = a_\nu s + b_\nu \), where \( a_\nu \) and \( b_\nu \) are constants. Under the assumption that \( \epsilon_\nu(s) \sim LN(0, \tau^2_\nu(s)) \), the properties of lognormal random variables can be used to determine \( \tau_\nu(s) \):

\[
\tau_\nu(s) = \sqrt{\log \left( 1 + CoV_{\epsilon_\nu(s)}^2 \right)} = \sqrt{\log \left( 1 + (a_\nu s + b_\nu)^2 \right)}.
\]

Following a method similar to that which was done for the random variable \( \nu(S_i) \), a Monte Carlo simulation is performed by drawing \( M \) realizations \( s_{i,j}, \quad j = 1, ..., M \) of \( S_i \) and then calculating \( \tau_\nu(s_{i,j}) \) and inferring an approximation to its density by looking at the histogram in Figure 3.10. The histogram is generated from \( M = 10 \) million samples, again using the sample cell from before, i.e. \( S_i \sim N(8.09, 1) \).

Using this histogram a plot showing the approximate pdf of \( \tau_\nu(S_i) \) can be seen in Figure 3.11, along with the normal and lognormal distributions found by fitting the sample mean and standard deviation using the method of moments.

These plots indicate that the distribution for \( \tau_\nu(S_i) \) closely resembles a normal distribution, and after repeating this analysis multiple times, this work obtained an average sample skewness = \(-.0662\) and a sample kurtosis= \(3.001\). Considering that a normal distribution has skew 0 and kurtosis 3, perhaps one can be justified approximating \( \tau_\nu(S_i) \) as a normally distributed random variable without it dramatically affecting the overall total loss.
Figure 3.10: Histogram of fatality rate error standard deviation $\tau_\nu$ as a function of random $S_i$.

Figure 3.11: Density functions for sample pdf of $\tau_\nu(S_i)$ versus fitted normal and lognormal density functions.
estimation. With a sample mean and variance being \( \mu \tau \nu (S_i) = .3603 \) and \( \sigma \tau \nu (S_i) = .0413 \), respectively, then the standard deviation of the intraevent fatality rate error for cell \( i \) would be \( \tau \nu (S_i) \sim N(.3603, .0413) \) under this normality assumption.

In order to understand how the uncertainty in \( S_i \) affects the density of \( \epsilon \nu (S_i) \) involves marginalizing the joint distribution of shaking intensity and fatality rate with respect to the shaking intensity. Using the definition of conditional densities and the law of total probability, the distribution of intraevent fatality rate error can be obtained:

\[
 f_{\epsilon \nu i}(\epsilon) = \int_{-\infty}^{\infty} f_{\epsilon \nu, S_i}(\epsilon, s) ds \\
= \int_{-\infty}^{\infty} f_{\epsilon \nu | S_i}(\epsilon | S_i = s) f_{S_i}(s) ds.
\]  

(3.6)

(3.7)

Assuming that \( S_i \sim N(\mu_{S_i}, \sigma_{S_i}) \), \( f_{S_i}(s) \) is simply the pdf for a normal random variable with mean \( \mu_{S_i} \) and variance \( \sigma_{S_i}^2 \). Conditional on \( S_i = s \), \( f_{\epsilon \nu | S_i}(\epsilon | S_i = s) \) is the pdf for a lognormal random variable with parameter \( \mu_{\log \epsilon \nu} = 0 \) and \( \sigma_{\log \epsilon \nu} = \tau \nu (s) \). Essentially, this is equivalent to saying that \( f_{\epsilon \nu i}(\epsilon) = \mathbb{E}_{S_i} [f_{\epsilon \nu (S_i)}(\epsilon | S_i)] \), where \( \mathbb{E}_{S_i} \) refers to an expectation taken over the probability space to which \( S_i \) belongs.

Below, Figure 3.12 shows a plot of \( \epsilon \nu (S_i) \) values can be seen given the normal shaking distribution for an arbitrary cell \( i \), where again \( S_i \sim N(8.09, 1) \), the density of \( \epsilon \nu (S_i) \) is calculated using numerical integration with respect to \( S_i \) using a mesh size of 10000.

Figure 3.12 also shows two other curves which try to approximate the actual density curve of \( \epsilon \nu (S_i) \). The first of these (black curve) is a lognormal approximation to the resulting distribution using a method of moments on the logarithm of of \( \epsilon \nu (S_i) \) while the second (red curve) is the distribution of the intraevent fatality rate uncertainty without including in shaking, or \( \epsilon \nu (\mu_{S_i}) \). There are some differences between these curves, which can be seen in Figure 3.13, which shows the squared error in these two approximations compared to the density of \( \epsilon \nu (S_i) \) obtained using the Monte Carlo simulation.
Probability Density Functions of $\epsilon_{\nu_i}$ for $S_i \sim N(8.09, 1.0)$ and $\epsilon_{\nu_i}|S_i \sim LN(0, \sqrt{\log(1 + (aS_i + b)^2)})$

where $a = -0.08$ and $b = 1.02$

Figure 3.12: Plot of the pdf of $\epsilon_{\nu}(S_i)$ along with some approximations.

Squared Error of pdf of $\epsilon_{\nu_i}$ and Approximate pdfs

for $S_i \sim N(8.09, 1.0)$ and $\epsilon_{\nu_i}|S_i \sim LN(0, \sqrt{\log(1 + (aS_i + b)^2)})$

where $a = -0.08$, and $b = 1.02$

Figure 3.13: Squared error in approximations to $f_{\epsilon_{\nu_i}(S_i)}$. 
A table showing the statistics of these two approximations and as well as for the numerically obtained $f_{\nu(S_i)}$ is shown in Figure 3.14.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Mean</th>
<th>Std</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>$\mu$ Parameter</th>
<th>$\sigma$ Parameter</th>
<th>$L_2$norm (Actual - Approx.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual</td>
<td>1.0699</td>
<td>0.4134</td>
<td>1.5560</td>
<td>8.7282</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LN Approx</td>
<td>1.0699</td>
<td>0.4134</td>
<td>1.2170</td>
<td>5.7441</td>
<td>-2.06E-03</td>
<td>0.3731</td>
<td>0.05968</td>
</tr>
<tr>
<td>LN No Shake Unc.</td>
<td>1.0672</td>
<td>0.3979</td>
<td>1.1702</td>
<td>5.5299</td>
<td>3.96E-08</td>
<td>0.3607</td>
<td>0.03862</td>
</tr>
</tbody>
</table>

Figure 3.14: Statistics of the different approximations to $f_{\nu(S_i)}$ along with $L^2$-error.

The data indicates that the distribution of $\epsilon_{\nu}(S_i)$ has too large of a right tail to be lognormally distributed, and one wonders if the approximate distributions will be valid enough to capture the distribution of loss for a cell $i$. The following analysis should determine whether the difference in these distributions have any noticeable effects on the distribution of $E_i$ for an arbitrary cell $i$. With all additional uncertainties included, the equation for total loss becomes

$$E = \eta \sum_i Pop_i FR_i = \eta \sum_i Pop_i \nu(S_i) \epsilon_{\nu_i},$$

(3.8)

where $\eta \sim LN(0, \zeta)$, and $Pop_i \sim LN(\log(pop_i), \tau_{p_i})$. The end goal of this chapter is to find an approximate lognormal distribution for $E$. The basic brute force manner of doing this would be to follow the Monte Carlo method used previously in this chapter. Again, this can be done by generating an ensemble of say $M$ realizations of $\eta$ along with $M$ realizations of each $E_i$ for every $i = 1, ..., N$. One can then use Equation 3.8 to obtain $M$ realizations of $E$. This work will first investigate how to draw realizations of each $E_i$ efficiently, before looking at how these can be used to find an approximate distribution of $E$ given the included uncertainties.

### 3.3 Generating a Subregional Loss Ensemble for an Arbitrary Subregion $i$

Recall that $E_i = Pop_i \nu(S_i) \epsilon_{\nu_i}$ for each $i = 1, ..., N$. If $M$ independent realizations of $Pop_i$, $\nu(S_i)$, and $\epsilon_{\nu_i}$, are drawn and multiplied together, this will give $M$ realizations of
Even though the exact distribution of $\nu(S_i)$ is not known, one can nonetheless generate realizations from it using the Monte Carlo method described earlier in this section. The exact distribution of $\epsilon_{\nu_i}$ is also unknown, although there is a functional form Equation 3.7 which can be approximated using numerical integration. However, drawing realizations from the distribution of $\epsilon_{\nu_i}$ using this numerical approximation is difficult and requires a special method known as Markov chain sampling.

### 3.3.1 Drawing realizations of $\epsilon_{\nu_i}$

In order to draw realizations from the density function of $\epsilon_{\nu_i}$ by approximating Equation 3.7, this thesis uses a sampling method known as the Markov Chain Monte Carlo (MCMC) approach. In order to use the MCMC algorithm to generate these realizations, one first requires a method that will calculate $f_{\epsilon_{\nu_i}}(\epsilon)$ for any $\epsilon$. Below is a method outlining how to calculate (in a Monte Carlo manner) the value of $f_{\epsilon_{\nu_i}}$ for any $\epsilon$.

#### 3.3.1.1 Calculating $f_{\epsilon_{\nu_i}}(\epsilon)$ for any $\epsilon$

1. Choose $\epsilon$ for which to evaluate the pdf of $\epsilon_{\nu_i}$, call it $\epsilon^*$.  

2. Discretize the space $S_i$ belongs to, $(-\infty, \infty)$, into $K_S$ quadrature points $s_1, ..., s_{K_S}$ and corresponding quadrature weights $w_1, ..., w_{K_S}$.  

3. For each $k = 1, ..., K_S$, evaluate $f_{S_i}(s_k)$ and $f_{\epsilon_{\nu_i}|S_i}(\epsilon^*, s_k)$  

4. For each $k = 1, ..., K_S$ multiply the quadrature weight $w_k$ with the quantity $f_{\epsilon_{\nu_i}|S_i}(\epsilon^*, s_k) f_{S_i}(s_k)$.  

5. Sum all of the products from the previous step in order to obtain a numerical approximation to $f_{\epsilon_{\nu_i}}(\epsilon^*)$.

Essentially, this means that
\[ f_{\epsilon_{\nu}}(\epsilon^*) \approx \sum_{k=1}^{K_S} w_k f_{\epsilon_{\nu}|S_i}(\epsilon^*, s_k) f_{S_i}(s_k). \] (3.9)

Given this method for evaluating the pdf of \( \epsilon_{\nu} \) for any \( \epsilon \), the Metropolis Hastings Markov Chain Monte Carlo (MHMCMC) method for generating realizations of \( \epsilon_{\nu} \) will now be described.

### 3.3.1.2 Metropolis Hastings Markov Chain Monte Carlo Method

1. Pick a starting point \( \epsilon_0 \) guaranteed to be in the allowable range of \( \epsilon_{\nu} \).

2. Pick a number of MCMC points to generate, called \( M_{MHMCMC} \) (note that this needs to be much larger than the number of realizations needed, explained later).

3. Choose what is called a proposal density, which is generally a symmetric bivariate density, such as the two-dimensional Gaussian distribution. Call the proposal density \( g(\cdot, \cdot) \). This distribution chooses how far away from the current \( \epsilon \), which is \( \epsilon_0 \), one looks for the next \( \epsilon \), call it \( \epsilon' \).

4. Start \( \ell = 0 \) and set \( \epsilon_{\ell} = \epsilon_0 \).

5. Draw a candidate \( \epsilon' \) from the distribution \( g(\epsilon, \epsilon_{\ell}) \).

6. Calculate what is known as the acceptance ratio
\[
\alpha = \frac{f_{\epsilon_{\nu}}(\epsilon')}{f_{\epsilon_{\nu}}(\epsilon_{\ell})}
\]
by using the preceding algorithm for calculating values of \( f_{\epsilon_{\nu}} \).

7. Accept \( \epsilon' \) with probability \( \alpha \). That is, draw a number \( d \) from the distribution \( U(0, 1) \) and either accept \( \epsilon' \) if \( \alpha > d \) or reject it if \( \alpha \leq d \).

8. If one accepts \( \epsilon' \), set \( \epsilon_{\ell+1} = \epsilon' \), and if it is rejected set \( \epsilon_{\ell+1} = \epsilon_{\ell} \).

9. Repeat steps 5-8 until \( M_{MHMCMC} \) points have been generated.
The next step is to look at what is known as the autocorrelation (which describes how correlated nearby MHMCMC samples are) between any two MHMCMC points $\epsilon_\ell$ and $\epsilon_\ell^*$, where $\ell^* = \ell + n$, $n = 1, 2, \ldots$, for all $\ell = 0, 1, \ldots$. The goal is to find an $n = d$ for which the autocorrelation between $\epsilon_\ell$ and $\epsilon_d$ is zero for each $\ell$. This implies that every $d^{th}$ MHMCMC point is independent. Thus, by choosing every $d^{th}$ point generated, one will have an independent and identically distributed set of realizations of $\epsilon_\nu_i$. Another characteristic of the MHMCMC algorithm is that it often requires generating a certain number of points, called the burn-in, before the generated points begin exploring the density of the target distribution. In practice the number of points equal to the burn-in are thrown away, and then every $d^{th}$ point is selected to form an ensemble of realizations of $\epsilon_\nu_i$. Both of these features are the reason why one needs to generate many more MHMCMC points than $M$. Specifically, $M_{MHMCMC} \geq Md + (Burn - in)$.

This means of generating intraevent fatality rate error realizations is then used to generate realizations of $E_i$ for an example subregion. The realizations of $E_i$ are drawn by generating realizations of $Pop_i$, $\nu(S_i)$, and $\epsilon_\nu(S_i)$ and multiplying them together. When generating the realizations of $\epsilon_\nu(S_i)$ for each $i$, three different distributions will be used and their corresponding effects on the distribution of $E_i$ will be compared. The first of these is the actual distribution of $\epsilon_\nu(S_i)$, which requires the MHMCMC algorithm just mentioned to produce realizations. The second distribution is that found by performing the method of moments to fit a lognormal distribution to the probability density function

$$f_{\epsilon_\nu(S_i)} = \int_{-\infty}^{\infty} f_{\epsilon_\nu|S_i}(\epsilon|S_i = s) f_{S_i}(s)ds.$$  

The third approximate distribution of $\epsilon_\nu(S_i)$ will be that obtained by assuming no shaking uncertainty in the intraevent fatality rate, so that $\epsilon_\nu(S_i) \sim LN(0, \tau_\epsilon(\mu_{S_i}))$. If the results indicate that there is a negligibly small difference between the estimated distributions of $E_i$ using these three different methods, then this means the least expensive approximation of $\epsilon_\nu(S_i)$ should be able to be used without drastically affecting the estimation of $E$. This
is highly desirable because the MHMCMC algorithm takes a fairly long time and requires optimizing its parameters (burn-in and autocorrelation lag) for each subregion \( i \), so any way to avoid generating realizations in this manner is desirable.

For the example subregion \( i \) used throughout this chapter, assume that the population for it is distributed as \( \text{Pop}_i \sim N(\mu_{\text{Pop}_i}, \sigma_{\text{Pop}_i}) \), with \( \mu_{\text{Pop}_i} = 832 \) and \( \sigma_{\text{Pop}_i} = 1.0 * \mu_{\text{Pop}_i} \). Again assume a mean shaking intensity for this cell of \( \mu_{S_i} = 8.09 \), and a standard deviation \( \sigma_{S_i} = 1.0 \).

### 3.3.2 Example and Results for an Arbitrary \( E_i \)

After drawing successively larger ensembles of sample size \( M \) of the random variable \( E_i \), the following results are obtained. First, Figure 3.15 shows the first four sample statistics of \( \log(E_i) \) obtained from each set of \( M \) realizations.

![Table of the first four sample statistics](image)

This table illustrates that the first four sample moments are converging as \( M \) increases, although slowly, particularly for the higher moments. It is also clear that the distribution of \( \log(E_i) \) is not normally distributed, as it has a sizeable negative skew and kurtosis more than twice that of a normal distribution. Next, (Figure 3.16) histograms of both \( E_i \) and \( \log(E_i) \) from the sample with \( M = 8,192,000 \) realizations can be seen.
This histogram shows a definite logarithmic behavior for $E_i$, as its logarithm is close to being normally distributed, albeit with a negative skew and larger kurtosis. Using the histogram for $\log(E_i)$ above, an approximate probability density and empirical cumulative distribution function for $\log(E_i)$ can be formed. These approximations are both plotted in Figure 3.17 along with a normal approximation of the density of $\log(E_i)$ used by matching the first two sample moments of $\log(E_i)$.

Again, one can see that the MC distribution differs significantly from the normal approximation. These results can be compared with the other approximate distributions mentioned for $\epsilon_{\nu}(S_i)$, which were the lognormal distribution fitting the moments obtained from the MC approximation of the distribution of $\epsilon_{\nu}(S_i)$, and the lognormal distribution obtained assuming no uncertainty in shaking. Figure 3.18 shows a table summarizing the first four statistics of $\log(E_i)$ from each of these methods as well as the mean, median, and mode of $E_i$ obtained from each method.

The results from this table indicate that while there are subtle differences between the results given by the different methods, they are relatively close. In order to demonstrate that these differences do little to affect the distribution of $E_i$, the approximate cdfs of $E_i$
Figure 3.17: Approximate PDFs and CDFs of log($E_i$) using MC method and normal approximation.

<table>
<thead>
<tr>
<th>ev(Si ) Distribution</th>
<th>Mean(log(E)$_i$)</th>
<th>Std(log(E)$_i$)</th>
<th>Skew(log(E)$_i$)</th>
<th>Kurt(log(E)$_i$)</th>
<th>Mean(E)$_i$</th>
<th>Median(E)$_i$</th>
<th>Mode(E)$_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC Method</td>
<td>-0.61464</td>
<td>4.3334</td>
<td>-1.4828</td>
<td>7.0535</td>
<td>18.7858</td>
<td>1.3126</td>
<td>7.7718</td>
</tr>
<tr>
<td>LN Approx</td>
<td>-0.61274</td>
<td>4.3324</td>
<td>-1.4814</td>
<td>7.0457</td>
<td>18.8425</td>
<td>1.3143</td>
<td>6.1780</td>
</tr>
<tr>
<td>LN (No Shake Unc.)</td>
<td>-0.61284</td>
<td>4.3279</td>
<td>-1.4727</td>
<td>6.9479</td>
<td>18.7780</td>
<td>1.3116</td>
<td>7.1571</td>
</tr>
</tbody>
</table>

Figure 3.18: Table of statistics of log($E_i$) and $E_i$ using the three different approximate distributions of $\epsilon_\nu(S_i)$.
using each method are plotted together in Figure 3.19, along with their corresponding mean, median, and modes.

![Figure 3.19: CDF of $E_i$ for different approximate distributions of $\epsilon(\nu(S_i))$.](image)

One can see the difference between the methods for this particular subregional loss $E_i$ is very small. This means that reasonably, one can be justified in using the easiest approximate distribution of $\epsilon(\nu(S_i))$, which assumes no shaking uncertainty. This avoids having to perform the Markov chain Monte Carlo method for each and every subregion $i = 1, ..., N$, which would take a substantial amount of time. Instead, one can use the simpler distribution for $\epsilon(\nu(S_i))$ and generate realizations from the distribution $N(0, \tau(\mu_S))$, exponentiating them to get approximate realizations of $\epsilon(\nu(S_i))$. Henceforth when generating realizations of total loss $E$, this approximate distribution of the intraevent fatality rate error will be used.

### 3.4 Aggregating Subregional Realizations to Obtain Total Loss $E$ Estimate

Now that an efficient method has been found to generate say $M$ realizations of each subregional loss estimate $E_i$, $i = 1, ..., N$, this method can be used to generate $M$ realizations of $E$ by summing over all $i = 1, ..., N$ for each set of realizations. As an illustrative example an assumed CoV of .1 for the population in each cell will be used. The same functional form for
the intraevent standard deviation parameter \( \tau(s) = \sqrt{\log(1 + (as + b)^2)} \) will be used, where recall that \( a \) and \( b \) are picked such that \( \text{CoV}_{\epsilon(S_i)}(s = 4) = .7 \) and \( \text{CoV}_{\epsilon(S_i)}(s = 9) = .3 \).

Again the 2011 Van earthquake is used as an example and numerous samples of different sizes are drawn. Figure 3.20 shows a table summarizing the first four statistics of \( \log(E) \) for various sample sizes \( M \).

<table>
<thead>
<tr>
<th>( M )</th>
<th>Mean ( \log(E) )</th>
<th>Std ( \log(E) )</th>
<th>Skewness ( \log(E) )</th>
<th>Kurtosis ( \log(E) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>6.1493508</td>
<td>2.8660498</td>
<td>-0.3527658</td>
<td>2.9788348</td>
</tr>
<tr>
<td>2000</td>
<td>6.2056248</td>
<td>2.9223058</td>
<td>-0.4061818</td>
<td>3.0956328</td>
</tr>
<tr>
<td>4000</td>
<td>6.2172168</td>
<td>2.8615368</td>
<td>-0.4232078</td>
<td>3.1999928</td>
</tr>
<tr>
<td>8000</td>
<td>6.1864558</td>
<td>2.8650548</td>
<td>-0.4214678</td>
<td>3.273968</td>
</tr>
<tr>
<td>16000</td>
<td>6.1666768</td>
<td>2.8849628</td>
<td>-0.4293298</td>
<td>3.3343858</td>
</tr>
<tr>
<td>32000</td>
<td>6.1518628</td>
<td>2.8886088</td>
<td>-0.4212968</td>
<td>3.4471628</td>
</tr>
<tr>
<td>64000</td>
<td>6.1555168</td>
<td>2.8995188</td>
<td>-0.4178268</td>
<td>3.4662998</td>
</tr>
<tr>
<td>128000</td>
<td>6.1551068</td>
<td>2.9142108</td>
<td>-0.4338298</td>
<td>3.5108358</td>
</tr>
<tr>
<td>256000</td>
<td>6.1512338</td>
<td>2.9128788</td>
<td>-0.4249238</td>
<td>3.4658588</td>
</tr>
<tr>
<td>512000</td>
<td>6.1521508</td>
<td>2.9120678</td>
<td>-0.4246468</td>
<td>3.4518458</td>
</tr>
<tr>
<td>1024000</td>
<td>6.1520818</td>
<td>2.9091598</td>
<td>-0.4271298</td>
<td>3.4605138</td>
</tr>
<tr>
<td>2048000</td>
<td>6.1575978</td>
<td>2.9084478</td>
<td>-0.4255828</td>
<td>3.4655178</td>
</tr>
<tr>
<td>4096000</td>
<td>6.1589328</td>
<td>2.9103968</td>
<td>-0.4239948</td>
<td>3.4614618</td>
</tr>
<tr>
<td>8192000</td>
<td>6.1573638</td>
<td>2.9099478</td>
<td>-0.4244668</td>
<td>3.4626978</td>
</tr>
</tbody>
</table>

Figure 3.20: Table of sample statistics of \( \log(E) \) for various sample sizes \( M \).

The statistics do seem to be converging, although quite slowly, as about 8 million realizations are needed just to get 2 significant digits accuracy in each statistic. One can immediately see that histogram is not normal, as it has a negative skew as well as a kurtosis which is greater than that of a normal distribution. However, even with the different skew and kurtosis, the PAGER histogram formed from this exact distribution versus one formed by assuming a normal distribution for \( \log(E) \) would be only slight, and would not generally affect the PAGER alert level.

Figure 3.21 shows the histograms formed for \( E \) and \( \log(E) \). One can see that it closely resembles a normal distribution, albeit for the negative skew and larger left tail.
Next, in Figure 3.22 a few different approximations to the distribution of $\log(E)$ are shown. The blue curve represents the approximate distribution of $\log(E)$ formed from the previous histogram. This curve is the Monte Carlo approximation using a sample size of 8,704,000 realizations. The red curve represents the distribution of total loss using the current PAGER method without the additional uncertainties. The green line represents a normal distribution fitted via the method of moments (MOM) with respect to $E$, while the black curve represents a normal distribution fitted again with the method of moments, but this time with respect to $\log(E)$. Finally, the cyan curve represents a normal distribution fitted such that the exponential of its mean equals the median value of $E$. Like the black curve, this curve is also conditioned by the standard deviation of the $\log(E)$.

Figure 3.22: Approximate distributions of $\log(E)$.
The results indicate that incorporating additional uncertainties has led to an increased spread in the distribution of \( \log(E) \). In addition, the green curve representing the MOM with respect to \( E \) has a much larger mean value and thus would lead to a larger expected total loss than the other approximations. It turns out that using this method will overestimate the total loss so much that the alert level changes for this earthquake, motivating the use of the Monte Carlo method as the more accurate of the two when the shaking intensity uncertainty is included.

The next plots (Figure 3.23) show these different distributional approximations in terms of \( E \) instead of \( \log(E) \). Various statistics for each curve are also included, which illustrate how the three approximations were made using the statistics from the MC approximation.

![Graphs showing different distributional approximations for E.](image)

**Figure 3.23**: Candidate approximate pdfs of \( E \).

In addition to the two best approximate pdfs from the previous plot, the following plot Figure 3.24 includes two other candidate distribution fits for the pdf of the MC approximation of \( \log(E) \). Included in the plot are the \( L^2 \)-norms for each approximation.
The black and cyan (top left and top right respectively) of Figure 3.24 represent the earlier fits using the MOM with respect to \( \log(E) \) and that obtained by matching the median of \( E \), respectively. Also included is a Weibull distribution fit (bottom left) determined to best approximate the mean and std of \( \log(E) \) simultaneously, and the bottom right distribution is a gamma distribution (shifted and reflected) fitted using the MOM with respect to \( \log(E) \), which obtained the smallest \( L^2 \)-error. Other distributions were tried, including lognormal, log-logistic, and inverse-gamma, although their norms and errors were worse than the four pictured above. Below in Figure 3.25 we include a plot of the \( L^2 \)-error between our approximate densities and the MC approximation.

Clearly the gamma (green) fit is performing the best, both at describing the distribution in its tails and in its higher probability density region. The Weibull (red) distribution also does well in the interior of the density, but performs the worst in the tails. Meanwhile, the method of moments approximation (black) of \( \log(E) \) with a normal distribution out-performs the median matching method (cyan) in the tails, but under-performs comparatively in the higher probability density regions. Because in the end PAGER would only publish the
Figure 3.25: $L^2$-error of approximations to log($E$) between the Monte Carlo approximate pdf and the four best approximate parametric pdfs

PAGER histogram of total loss, it will be of interest to see how these errors change the final PAGER histograms depending on the approximation used. If it can be shown that there is a negligible difference in the total loss histograms, then using the normal approximation would be of primary interest because it is the simplest and easiest to implement in a Bayesian scheme in addition to it being the same distribution as the current PAGER total loss scheme. In order to see how these different probability distributions compare, the next plot (Figure 3.26) shows the approximate cumulative distribution functions for $E$ and lists their corresponding median values.

Looking at the cdfs one can see that they all share the same major features across the different orders of magnitude of loss, however they all perform differently in the middle and in the tails. The results indicate that entirely different median losses can arise from the different methods, with some being larger or smaller than the Monte Carlo median. Fortunately, this large discrepancy is of little consequence because PAGER does not publish...
its median expected loss. The first three PAGER histograms below (Figure 3.27) show the current PAGER model for this earthquake along with the Monte Carlo approximation as well as the MOM with respect to $E$ approximation, which greatly overpredicted the median loss.

The results indicate that a noticeable increase in the variance of total loss occurs when incorporating these additional uncertainties. As noted earlier, the MOM approximation (right)
over predicts the loss, which ends up changing the alert level for this earthquake. Below in Figure 3.28 the PAGER histograms are shown regarding the different approximations of $E$ that were present in Figure 3.26.

The results indicate that the closest approximations to $\log(E)$ have very little difference between them when viewed as PAGER histograms. This justifies the use any of them to approximate the total loss, although the lognormal approximations will be the easiest to manipulate and analyze when performing Bayesian updating in the following chapter. Because the current PAGER model loss is given as lognormal distribution, having a lognormal approximation to this enhanced loss model means that without loss of generality, all loss distributions henceforth can be seen as either coming from the current PAGER model or this enhanced model.

While the formulation of a better forward model is important, this thesis ultimately seeks to develop a model which changes and adapts according to reported losses following an earthquake. For instance, suppose that after observing a final loss count that differs greatly from the predicted loss, one wishes to update the original PAGER loss distribution with this information in order to better explain the variability of loss for that particular earthquake, which can be useful in training the fatality rate model for future earthquakes. However, this thesis is primarily interested in updating the total loss distribution long before the observed loss reaches its final value, and ideally within 1-2 days, after only a fraction of the loss has been observed.

In the fourth chapter, after spending some time gaining familiarity with the Bayesian framework, this work looks at updating the total loss distribution given noisy observations of total loss. The fourth chapter helps one to understand how updating the different parameters of the total loss distribution affects the process with which it incorporates observed total loss. Recall that the motivation behind this is that when the PAGER alert level is wrong for a particular earthquake, it does not mean that the PAGER model is wrong. Regardless, the primary interest of the PAGER team is to get the correct alert level issued as quickly
Figure 3.28: PAGER histograms for different approximate distributions of $E$. 
as possible, which the final two chapters illustrate can be done by using the reported losses over time. This proves much more difficult because one is now forced to find or develop a time evolving model which describes the observation of loss over time.
CHAPTER 4
BAYESIAN UPDATING OF THE TOTAL LOSS DISTRIBUTION

The last chapter showed that an incorrect PAGER estimation could be caused by multiple factors, such as incorrect estimates of shaking and fatality rates, particularly for high population areas. In addition, the last chapter showed that by accounting for these additional uncertainties one can better understand if the PAGER estimate was statistically reasonable or whether it is evidence that the model is inconsistent in regards to a particular event.

This chapter is designed to illustrate how the total loss distribution can be updated given the actual reported total loss a Bayesian setting. Furthermore, it investigates how the updated total loss distributed depends on different levels of uncertainty in the reported loss data as well as in the loss distribution parameters $\mu$ and $\sigma$. Ideally an updating method is desired which takes the uncertainty of the reported total loss into account, as well as one that sufficiently updates the total loss distribution when the observed fatality count is inconsistent with the PAGER model’s prediction.

As mentioned earlier, the primary purpose of this chapter and the last was to get a hands on look at the different models and frameworks associated with estimating loss, in order to build more familiarity with some of the concepts, notations, and analysis used in the fifth and sixth chapters. As such this chapter does not delve too deeply into the analysis of the updating procedure presented with regards to multiple earthquakes. Similarly to Chapter 3, this chapter applies its model for only the 2011 Van earthquake in Turkey, and leaves analysis of different earthquakes for future work. This chapter merely serves as an illustration of how a lognormal loss distribution changes when taking uncertain reported total losses into account, specifically with regards to changes in the error in the data, the prior distributions, and the initial PAGER loss estimates. The degree of updating explained in this chapter can be summarized by two different methods, one of which is an enhancement of the other.
4.1 Overview of Updating Methods

The first of these updating methods uses the PAGER total loss distribution as a prior and forms a likelihood from the error distribution of an observed total loss, call it \( M^* \), which will be explained in further detail. Essentially this method seeks to reconcile the data with the PAGER estimate by finding the range of deaths where they both agree.

The second method expands on this method by further assuming uncertainty in the parameters of the PAGER model by assuming they also have some prior distributions. This method represents a multi-level updating scenario, which first estimates the true \( E^* \) from an uncertain \( M^* \) and then uses this \( E^* \) to update the prior parameters. In practice, one simply forms a new prior which is marginalized over the distributions of the parameters before proceeding to update in the traditional Bayesian manner. This method has different executions depending on whether one wishes to update the loss distribution mean parameter \( \mu \), standard deviation parameter \( \sigma \), or both of them together. Essentially these different submethods differ in the degree of uncertainty assumed in the PAGER loss parameters.

For mathematical and computational ease the lognormal approximation of total loss assumed by PAGER will be used, whether it is the original PAGER estimate or the enhanced estimate following the method introduced in the previous chapter. Furthermore, this chapter will be taking advantage of the conjugate prior relationships for the normal likelihood distribution depending on its parameters. This allows one to find closed form marginal densities of \( E \) by updating the prior parameters for most of the updating cases. For portions where closed form solutions are not available, numerical integration will be used to obtain the posterior and marginal predictive distributions.

Suppose that there is a candidate lognormal distribution of total loss \( E \), i.e., \( E \sim LN(\mu, \sigma) \), where \( \mu \) and \( \sigma \) are determined either from the current PAGER loss estimate or the enhanced version with additional uncertainties incorporated. Further suppose that a noisy measurement \( M \) of total loss is made such that
\[ M = E \mathcal{E}, \quad (4.1) \]

where \( \mathcal{E} \sim LN(\mu_\mathcal{E}, \sigma_\mathcal{E}) \) represents a residual error factor. The reason for the multiplicative error is because analysis will be performed on the logarithm of Equation 4.1, in which case it becomes a sum of normal random variables, which itself is a normal random variable.

This problem can be recast in this logarithmic domain by introducing the random variables \( L_E = \log(E) \), \( L_\mathcal{E} = \log(\mathcal{E}) \), and \( L_M = \log(M) = L_E + L_\mathcal{E} \). Then \( L_E \sim N(\mu, \sigma) \) and \( L_\mathcal{E} \sim N(\mu_\mathcal{E}, \sigma_\mathcal{E}) \). In the end, given a noisy observation \( M^* \) of \( M \) one wishes to determine the distribution of \( E \) given \( M^* \), or \( E | (M = M^*) \). Before this can be accomplished, one must first determine the conditional density of \( L_E \) given \( m^* = \log(M^*) \), or \( f_{L_E | L_M}(\cdot | m^*) \). Now, for any realization \( e \) of \( L_E \), Bayes’ theorem states

\[
f_{L_E | L_M}(e | m^*) = \frac{f_{L_M | L_E}(m^* | e) f_{L_E}(e)}{C}, \quad (4.2)
\]

where \( C \) is a normalization constant ensuring the distribution integrates to 1. Given a fixed realization \( e \) of \( L_E \), the distribution of \( L_M | [L_E = e] \) is simply the distribution in the log error factor \( L_\mathcal{E} \) shifted by an amount \( e \), or

\[
f_{L_M | L_E}(m^* | e) = f_{L_\mathcal{E}}(m^* - e).
\]

Because \( L_\mathcal{E} \sim N(\mu_\mathcal{E}, \sigma_\mathcal{E}) \) by assumption, this means that \( L_M | (L_E = e) \sim N(e + \mu_\mathcal{E}, \sigma_\mathcal{E}) \). Depending on the amount of confidence in the initial PAGER loss distribution’s parameters, there are different levels of updating that can be performed. The simplest updating method assumes that these parameters have no uncertainty. The second of the updating methods allows for uncertainty in one or both of these parameters, and investigates how the updating results change depending on the degrees of uncertainty in these parameters. In order to understand how the basic updating method works, the first method will now be illustrated.
4.2 Method 1: \( \mu \) and \( \sigma \) are both known

This case is relatively straightforward as one can input the candidate normal distribution for \( L_E \) into Equation 4.2, giving

\[
f_{L_E|L_M}(e|m^*) = \frac{f_{L_E}(m^*-e)f_{L_E}(e)}{C}. \tag{4.3}
\]

Because the likelihood distribution \( L_M|L_E = e \) is a normal distribution, which is conjugate to the normal distribution in regards to its mean, the posterior is a normal distribution (since \( L_E \) is also normal). The new posterior distribution will have the updated parameters \( \mu' \) and \( \sigma' \), i.e. \( L_E|L_M = m^* \sim N(\mu', \sigma') \). Consulting a list of conjugate priors written by Fink [18], formulas for the posterior parameters in terms of the measurement \( m^* \) and prior parameters can be obtained by the method of completing the square in the exponents of the normal pdfs.

For Equation 4.3 the conjugate normal pair of distributions implies that

\[
L_E|L_M = m^* \sim N(\mu', \sigma'),
\]

where

\[
\mu' = \frac{1}{\sigma^2 + \frac{1}{\sigma_E^2}} \left( \frac{\mu}{\sigma^2} + \frac{m^* + \mu_E}{\sigma^2} \right), \tag{4.4}
\]

\[
\sigma' = \frac{1}{\sigma^2 + \frac{1}{\sigma_E^2}},
\]

and thus

\[
E[M = M^*] \sim LN(\mu', \sigma').
\]

Example of updating for the 2011 Van, Turkey magnitude 7.1 earthquake

The Van earthquake will again be used to demonstrate these updating techniques. Initially, these methods will look at a hypothetical scenario where 10,000 fatalities were reported. This is much larger than the true value, but it will serve to help demonstrate how the different methods differ by exaggerating the process. After looking at this specific ex-
ample, analysis of the updating results will be shown for an entire range of observed losses in order to better understand what to expect when large discrepancies occur between the PAGER estimate and actual reported losses. These methods assume some error in the reported total loss and initially investigates the process assuming a coefficient of variation of 1 for this measurement error, and look at the posterior solutions for different magnitudes of this CoV. This value and the others used are not at all realistic, however they again are chosen to be large enough so that the updating can be easily seen visually. For these examples, the PAGER $\zeta$ value will be that obtained in chapter 3, which was 2.910. Again this chapter is not concerned with analyzing the specifics regarding this earthquake, and it is reiterated that this chapter is mainly to demonstrate the effects of Bayesian updating in connection to the problem that will be necessary for Chapters 5 and 6.

Figure 4.1 shows plots of the prior and posterior pdfs of $\log(E)$, along with some of the statistics and parameter values with an assumed error factor CoV of 1.

![Figure 4.1: Method 1: Prior and posterior pdfs of $\log(E)$ for an assumed reported loss error factor CoV](image)

The blue curve represents the prior distribution of $\log(E)$, while the red is the posterior update given the noisy reported loss. Clearly a large update has occurred in the direction of the observation, and the variance has decreased significantly, which is to be expected if the
updating process is working correctly. In order to see how this translates to the total loss $E$, a plot showing its prior and posterior cdfs is shown in Figure 4.2.

![Figure 4.2: Method 1: Prior and posterior cdfs of $E$.](image)

This plot illustrates a very significant change in the estimate of $E$ given the reported loss, shifting the median estimation to around 8000. Recall that the PAGER median estimate from Chapter 3 for this earthquake was about 569, indicating that the updated estimate dramatically shifted towards the observation. The reason for this is due to both the mean and variance of the reported loss. Because it was so large compared to the estimate, it was considered statistically improbable by the prior, meaning that the prior isn’t able to contribute nearly as much as it would had the observation been closer. The measurement error appears to control just how much the new posterior mean will shift, as well as the posterior variance. These relationships will become clear in the remaining examples. As mentioned, this updating process will be examined for several hypothetical reported loss values and measurement errors in order to see just how this Bayesian model reacts to different observed data and uncertainty in that data. In order to illustrate this, the Bayesian updating method will be performed on a set of observed $E$ ranging from 1 to 10,000 for different measurement error CoV’s of 1,100, 10,000, and 1,000,000. Recall that the error factor is assumed to be lognormally distributed, so these CoV’s correspond to standard deviation parameters $\sigma_E = 0.8326, 3.0349, 4.2919, \text{ and } 5.2565$, respectively. These respectively translate to roughly 1,
3, 4, and 5 orders of magnitude standard deviation about the observed measurement. Again these uncertainties are unrealistic, however they are shown in order to have a more clear understanding of how the uncertainty in data can affect the posterior loss distributions.

Figure 4.3 contains three dimensional plots showing the posterior probability density of \( \log(E) \) for different values of \( E^{obs} \), where each panel represents a different assumed CoV for the observed data. That is, each individual subplot represents a suite of possible posterior updates (x-axis) given different possible observations values (y-axis) for a particular measurement error CoV. Note that in all plots both the posterior densities and the observations are shown in terms of the logarithms \( \log(E) \) and \( \log(E^{obs}) \), respectively.

![Posterior Densities of \( \log(E) \)](image)

Figure 4.3: Method 1: Posterior pdf of \( \log(E) \) for different observed losses \( E^{obs} \) and different CoVs for the \( E^{obs} \).

One immediate result that can be seen from these plots is that in all four cases the posterior shifts linearly over the different \( E^{obs} \). This is to be expected because the conjugate
pair relationship for this produces the posterior mean is a linear combination between the prior mean and measurement value as seen in Equation 4.4. One can also see that for increasing CoV values of $E_{\text{obs}}$ the posterior distribution has a larger variance. In addition the posterior can be seen shifting towards the observation much more slowly for the large CoV values. This is to be expected because the larger uncertainty in the data makes the posterior more heavily weighted towards the prior distribution. Equation 4.4 reflects this relationship because the measurement mean is divided by its variance, and because increasing the CoV while keeping the mean constant leads to an increased variance. This makes the contribution from the measurement in the mean and variance calculation less significant as the measurement error uncertainty increases. Interestingly, because of the nature of Equation 4.4, the opposite effects occur when the prior variance becomes large, producing an updated loss distribution which is very close to that of the measurement distribution. In effect, this means that the weaker the prior precision, the less contribution it will have when producing the posterior, and similarly for the data. In fact, it is the relationship between these two uncertainties that determines how the posterior behaves, and not necessarily their specific variances. Figure 4.4 shows plots of the CDF of $E$ for different $E_{\text{obs}}$ and corresponding CoV’s, similarly to how they were shown in Figure 4.3.

These plots show the scale with which the different observed loss values and their CoV’s affect the posterior distribution of $E$. The results also show that the larger error CoV’s and larger $E_{\text{obs}}$ have estimated medians (.5 on the z-axis) that are much closer to the prior estimated median loss, and there is less variation in the CDF of $E$ from this perspective.

Next, Figure 4.5 shows how the posterior updated mean $\mu$ and standard deviation $\sigma$ of log($E$) change for different observed loss values and different observed loss CoV’s. For a fixed CoV, the posterior mean (left plot) increases linearly with increasing log($E_{\text{obs}}$), and at a faster rate as the measurement error decreases. This is essentially the relationship that was observed in the four 3D plots in Figure 4.3, except it is much more pronounced in this case. Clearly the variance is not changing with respect to $E_{\text{obs}}$, but it does with increased CoV,
Figure 4.4: Method 1: Posterior cdf of total loss $E$ for different observed losses $E^{obs}$ and different observation $CoVs$

Figure 4.5: Method 1: Posterior statistics of $\log(E)$ for different observed losses $E^{obs}$ and different observation $CoVs$
although at a decreasing rate as the observed error increases. This is no surprise, because the updated posterior variance of a conjugate normal pair is independent of both the prior mean and the measurement. These relationships can also be seen in the pdf plot as well as in Equation 4.4.

Method 1 demonstrates how a noisy observed loss affects the posterior estimates in relation to the prior PAGER distribution. By determining the degree the posterior shifts in the direction of the data, one can determine how likely this observed loss was given the prior PAGER distribution. This can be used to determine if the PAGER estimate was inconsistent with the actual total loss. The next sections investigate the updating process assuming the prior parameters $\mu$ and $\sigma$ themselves are random variables with their own uncertainties. These uncertainties arise naturally because the PAGER model only produces estimated values of $\mu$ and $\sigma$, and not their actual values.

4.3 Method 2: $\sigma$ is known and $\mu$ is unknown

The incorporation of possible uncertainties in the prior loss distribution parameters involves an enhancement of the first method in which the prior distribution is marginalized with respect to these uncertainties. In the second method, the parameter $\mu$ is treated as a random variable, referred to as $\Theta_1$, and the standard deviation $\sigma$ is considered to be constant and known. Recall that the Bayes theorem gives

$$f_{L_E|L_M}(e|m^*) = \frac{f_{L_M|L_E}(m^*|e)f_{L_E}(e)}{C}.$$  

Given a fixed realization $e$ of $L_E$, this method has the same likelihood distribution as in the first method

$$f_{L_M|L_E}(m^*|e) = f_{L_E}(m^* - e).$$

All that remains is to quantify how the uncertainty in $\mu$ will change the prior distribution $f_{L_E}(e)$ and hence the posterior distribution. Using the law of total probability,
\[ f_{L_E}(e) = \int_{\Theta_1} f_{L_E|\Theta_1}(e|\theta_1) f_{\Theta_1}(\theta_1). \]

For a fixed realization \( \theta_1 \) of \( \Theta_1 \), \( L_E(\Theta_1 = \theta_1) \sim N(\theta_1, \sigma) \). As mentioned before, the conjugate prior of a normal distribution with unknown mean and known standard deviation is a normal distribution. Therefore, for mathematical simplicity, assume that \( \Theta_1 \sim N(\mu_{\Theta_1}, \sigma_{\Theta_1}) \), implying there is a conjugate pair underneath the integral in the previous equation. It turns out that the predictive prior distribution \( f_{L_E}(e) \) will also be a normal distribution, which can be seen as a convolution of two normal distributions, which produces a normal distribution. Let \( \theta_1 = e - z \), and let \( Z \sim N(0, \mu_{\Theta_1}) \), then the integral above can be rewritten as

\[ f_{L_E}(e) = \int_{-\infty}^{\infty} f_Z(z) f_{\Theta_1}(e - z) \, dz. \]

Now the convolution of two normal random variables \( X \sim N(\mu_X, \sigma_X) \) and \( Y \sim N(\mu_Y, \sigma_Y) \) will be the random variable \( Z \sim N\left(\mu_X + \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2 + 2\rho \sigma_X \sigma_Y}\right) \), where \( \rho \sigma_X \sigma_Y = \text{cov}(X, Y) \). Given a realization \( \theta_1 \) of \( \Theta_1 \), \( L_E(\Theta_1 = \theta_1) \) is independent of \( \Theta_1 \) since \( f_{L_E|\Theta_1}(e|\theta_1) = f_{L_E|\Theta_1}(e|\theta_1) \), and thus \( \text{cov}(Z, \Theta_1) = 0 \) and therefore

\[ L_E \sim N\left(\mu_{\Theta_1}, \sqrt{\sigma^2 + \sigma_{\Theta_1}^2}\right). \]

The normal conjugate pair relationship implies that the posterior \( L_E|\Theta_1 = m^* \) will also be normally distributed, with updated parameters calculated similarly to the first updating method by using Equation 4.4. In this case, \( \sigma^2 \) is replaced with \( \sigma^2 + \sigma^2_{\Theta_1} \) and \( \mu \) with \( \mu_{\Theta_1} \). Calling the updated parameters \( \mu' \) and \( \sigma' \) gives

\[ L_E|\Theta_1 = m^* \sim N(\mu', \sigma'), \]

where

\[ \mu' = \frac{1}{\frac{1}{\sigma^2 + \sigma_{\Theta_1}^2} + \frac{1}{\sigma'^2}} \left( \frac{\mu_{\Theta_1}}{\sigma^2 + \sigma_{\Theta_1}^2} + \frac{m^* + \mu_L}{\sigma_{\Theta_1}^2} \right), \]

\[ \sigma' = \frac{1}{\frac{1}{\sigma^2 + \sigma_{\Theta_1}^2} + \frac{1}{\sigma'^2}}, \]

86
and thus

\[ E|(M = M^*) \sim LN(\mu', \sigma'), \]

where \( \mu' \) and \( \sigma' \) have different values than they did in Method 1.

**Example of updating for the 2011 Van earthquake**

The Van earthquake example is again used to demonstrate this updating method. Again assume a hypothetical scenario in which 10,000 deaths were observed. The first application again assumes a measurement error CoV of 1, and will assume a CoV of 1 for \( \Theta_1 \), the mean parameter, signifying a significant amount of uncertainty in this parameter. Figure 4.6 shows the prior and posterior pdfs of \( \log(E) \) for this method, along with some of the statistics and parameter values.

![Figure 4.6: Method 2: Prior and posterior pdfs of log(E) for the 2011 Van earthquake assuming a measurement error CoV of 1, a CoV of 1 for the mean parameter, and assuming an observed total loss of 10,000, chosen only for demonstrative purposes.](image)

This plot again shows the prior distribution shifting significantly in the direction of the reported loss observation. This effect is even more pronounced than in the previous method, because the uncertainty in the mean parameter has dramatically increased the spread in the prior distribution. This is expected because as mentioned in the last method, an increase in the prior variance will give it less contribution when producing the posterior, resulting
in the posterior distribution being much closer to the measurement distribution. However, even with a very large prior distribution, the updated posterior still has a smaller standard deviation than the measurement alone, implying that the prior information has led to an improved estimate, even with its large uncertainty. Translating everything back in terms of the total loss $E$, Figure 4.7 shows plots of the prior and posterior cdfs for this example.

![Prior and Posterior CDFs](image)

Figure 4.7: Method 2: Prior and posterior cdfs of $E$

The above results show that the update posterior loss distribution has a median loss value of about 9600. Thus the relatively low confidence in the prior mean value has given even more weight to the data, or equivalently less weight to the prior, than with the previous method. This also makes sense, because having additional uncertainty in the prior should be giving it less influence in the updating process.

Recall that the relationship between the measurement error and the prior uncertainty affected the rate at which the posterior shifted toward the data in the first method. Specifically, with larger error CoVs the posterior shifted at a slower rate for different observed losses. Because of the uncertainty in the prior parameters, there is an increased prior variance, meaning
that the posterior will shift more towards the observed data than in the previous method, even for large measurement CoV’s. Figure 4.8 illustrates the four panel 3D plot of different posterior updates given different observed total losses (y-axis), and different measurement error CoVs (different subplots).

Compared to the first method one can see that all four of these plots have a much larger shift in posterior mean as $E^{obs}$ increases then their respective method 1 plots. One can also see that the additional uncertainty in the mean leads to a much larger posterior variance than for the previous method as a result of having a larger spread in the prior distribution. Figure 4.9 shows the posterior statistics of log($E$) for method 2.

Looking at the mean parameter plot (left), one can see the surface is much more flat than in the previous method which had more of a saddle feature. This confirms the earlier observation that the mean is changing rather constantly as the error CoV is increased. The
Figure 4.9: Method 2: Posterior statistics of log(\(E\)) for different observed losses \(E^{obs}\) and different observation CoVs.

variance plot doesn’t show too much difference, except that it allows a much larger change in variance as the error CoV is increased than with the first method.

These results indicate that having uncertainty in the prior mean for a normal conjugate pair leads to normal posterior distributions with increased variance. Similarly to the first method, the relative size of the measurement error, the prior standard deviation \(\sigma\), and the standard deviation of the mean parameter \(\Theta_1\) determines how much the posterior shifts in the direction of the observed loss. This method is rather straightforward due to the normality assumptions in the prior, likelihood, and distribution of \(\Theta_1\). Now a different method is explained to describe what happens when when \(\mu\) is assumed to be constant and known, and \(\sigma\) is treated as uncertain.

4.4 Method 3: \(\mu\) is known and \(\sigma\) is unknown

In this scenario, the prior standard deviation \(\sigma\) will be treated as a random variable, or rather \(\sigma^2\), referred to as \(\Theta_2\). The prior variance \(\sigma^2\) is chosen instead of \(\sigma\) because in terms of an uncertain variance the normal likelihood has an \textit{inverse gamma distribution} as its conjugate prior. That is, if the variance parameter \(\Theta_2\) has an inverse gamma distribution,
then the posterior will also be an inverse gamma distribution. Again recall that the Bayes theorem gives

\[ f_{L_E|m^*} = \frac{f_{LM|m^*} f_{L_E}}{C} \]

where the likelihood distribution again is unchanged and is simply the distribution of the measurement error shifted by an amount equal to the observed log loss value. As before, all that remains is to quantify how the uncertainty in \( \sigma \) will change the prior density \( f_{L_E} \) before calculating the posterior loss distribution. By the law of total probability,

\[ f_{L_E} = \int_{\Theta_2} f_{L_E|\Theta_2}(e|\theta_2) f_{\Theta_2}(\theta_2), \]

and for a fixed realization \( \theta_2 \) of \( \Theta_2 \), \( L_E|\Theta_2 = \theta_2 \sim N(\theta_2, \sigma) \). For mathematical simplicity assume that \( \Theta_2 \sim \Gamma^{-1}(\alpha, \beta) \), where \( \alpha > 0 \) and \( \beta > 0 \) are known as hyperparameters of \( \Theta_2 \). Using a similar convolution argument that was used in the previous method, it turns out [19] that the marginal distribution of the logarithm of total loss \( f_{L_E} \) in the preceding equation follows a *scaled- non-central student t’s distribution* with \( 2\alpha \) degrees of freedom. This distribution is a 3-parameter generalization of the student’s t distribution allowing for nonzero mean and scalability, similarly to how the \( \beta \) parameter functions in the inverse-gamma distribution. The location parameter is simply \( \mu \), which is constant in this method, and the scale parameter is \( \alpha/\beta \) giving

\[ L_E \sim t_{2\alpha}(\mu, \alpha/\beta), \]

which has the probability density

\[ f_{L_E}(e) = \frac{\Gamma(\alpha + 1/2)}{\Gamma(\alpha)} \frac{1}{\sqrt{2\pi\beta}} \left( 1 + \frac{(e - \mu)^2}{2\beta} \right)^{-(\alpha+1/2)}, \]

where \( \Gamma \) is the gamma function whereby \( \Gamma(z) = \int_0^\infty x^{z-1}e^{-x} \).

This distribution is not a conjugate prior for the likelihood measurement distribution, and so the posterior distribution will have to be evaluated numerically. Plugging this marginal prior distribution into the Bayes equation gives
\[
f_{L_{E}|L_{M}}(e|m^*) = \frac{f_{L_{M}|L_{E}}(m^*|e)f_{L_{E}}(e)}{C} = \frac{1}{C} \left( \frac{1}{\sqrt{2\pi\sigma_{\xi}^2}} e^{-\frac{(e-m^*+\mu_{\xi})^2}{2\sigma_{\xi}^2}} \left( \frac{\Gamma(\alpha+1/2)}{\Gamma(\alpha)} \frac{1}{\sqrt{2\pi\beta}} \left( 1 + \frac{(e-\mu)^2}{2\beta} \right)^{-\alpha+1/2} \right) \right)
\]

where \( C \) can be calculated by numerically integrating the above equation over all positive \( e \in L_{E} \). The posterior distribution should be similar to a normal distribution, except with a wider tails, which results from having a t-distributed prior instead of a normal distribution as in the previous method.

**Example of updating for the 2011 Van earthquake**

The Van example is once again used to test this updating method. Again assume a hypothetical scenario in which 10,000 deaths were observed. This method again assumes a measurement error CoV of 1 first before investigating the affects of different error uncertainties, and will assume a CoV of 1 for \( \sigma^2 \), signifying a large amount of uncertainty in this parameter.

Figure 4.10 shows the prior and posterior pdfs of \( \log(E) \), along with some of the statistics and parameter values with an assumed error factor CoV of 1 for this uncertain variance updating method.

The plot shows a similar updated posterior as the previous methods, however visually it is difficult to ascertain the effects of the uncertain prior variance on this posterior. However, because the prior was a student-t distribution, it should be similar to the method 1 prior, except with fatter tails as a result of the t-distributed prior. Notice that while the posterior density (red) appears symmetric and normal, it actually has a larger right tail than left as a result of the large tails in the prior distribution. Thus just because this prior looks more certain than the preceding method, the larger tails mean that outlying observations are more probable than for the normally distributed prior. This highlights the main difference
between updating with respect to a mean or variance. When updating with respect to the mean, it shifts the mean and variance to better explain the data, while updating with respect to the variance instead increases the tails to better explain the data. In order to see how this translates to the total loss $E$, the prior and posterior cdfs are shown in Figure 4.11.

This plot shows that a strong update has occurred in the direction of the observed loss, leading to a median loss value of about 7400. While this isn’t as significant as in either of
the previous methods, it should be noted that rather than shifting the mean alone, the tails have also changed to reflect the data. Figure 4.12 shows how this updating changes given different possible observed losses and measurement error CoV’s.

Figure 4.12: Method 3: Posterior pdf of log($E$) for different observed losses $E^{obs}$ and different observation CoV’s.

It appears that incorporating uncertainty in the prior variance $\Theta_2$ seems to have the opposite effect as the previous method. Whereas with $\mu$ unknown, the posterior dramatically shifted in the direction of the observation, in this method it tends to shift at a much smaller rate than even the first method. Moreover, it appears that the density has its largest peak values when the observed loss is close to the prior predicted loss. Furthermore, instead of seeing the dramatic increase in the spread as the error CoV is increased, these results show that instead the tails become fatter. Additionally a clear skew in the direction of the $E^{obs}$ can be seen, with it resembling a student’s t distribution when the observed loss is close to
the prior median loss. As the error CoV increases, the posterior shifts towards the data at a slower rate, while the skew increases. Figure 4.13 shows the posterior statistics for the different observed losses and measurement CoVs for this updating method.

Figure 4.13: M3: Posterior statistics of $\log(E)$ for different observed losses $E^{\text{obs}}$ and different observation CoVs.

The mean (left) can be seen behaving much like it did in Method 1, although it tends toward the prior mean more dramatically than in that method as the error CoV increases. Additionally, it still appears to be changing linearly with change in observed loss for a fixed CoV. The posterior variance plot (right), shows similar features as it did for the first method, except that it is no longer constant with change in $E^{\text{obs}}$. One can see that when the observed loss is near the prior predicted loss, the spread appears to decrease, which can be seen as an increase in the posterior pdf peak value at in this region. This is because the posterior distribution is behaving like a student’s t distribution in this region, which seeks to make outliers more probable by increasing the weight in the tails rather than increasing the variance as in the second method. This method illustrates that without the normal conjugate pairs seen in the previous two methods, the posterior distribution is no longer
normally distributed. Rather, it tends to skew in the direction of the measurement.

Together, each of these methods represent different confidences in the PAGER loss distribution and measurement error. Naturally, one could treat both prior parameters as uncertain, which would lead an updated distribution having combinations of the characteristics from the first two methods. In this light, everything can be seen in terms of just one parameter, which will update the mean and variance simultaneously, giving different preferences between methods 1 and 2 depending on the relative error in the prior parameters.

As mentioned, the main goal of this chapter was to give an in-depth look at how Bayesian updating works, particularly when dealing with multiple levels of uncertainty. The main motivation behind this is to build the groundwork for the next two chapters, which will use some of the same notations and strategies when updating the partial reported losses over time. While the conjugate pair relationships were useful to obtain close form solutions for most of the previous results, they were only chosen for mathematical simplicity, a more robust updating method would allow for non-conjugate relationships, meaning that posterior calculations would have to be done numerically.

While the particular results from this chapter do not give any additional information regarding the Van earthquake given the actual observation, they did highlight how the uncertainty in the observation compared to that of the prior, as well as how the statistical likelihood of the data given the prior, affect the posterior updates. Future work around this topic would look at understanding how well the PAGER estimates work given the actual observations for many different earthquakes, but requires proper formulation of the typical observed total loss uncertainty, as well as estimates of the uncertainty in the PAGER prior parameters and their possible distributions through some other means or expert elicitation.
4.5 Intro to Primary Thesis Goal: Updating Given Only Partial Observations of Loss

As mentioned, the process of updating the distribution of loss is much more difficult when one is not observing the random variable $E$ directly, but rather partial observation of it over time. That is, at any given time the number of people observed is a random variable, and after an infinite (or practically large) amount of time it will converge to the true total loss $E$. The main difficulty is that records of the growth of observed loss over time are not readily available or are hard to come by. Furthermore, not all of the data can be obtained from reputable sources, and there is often significant uncertainty in these observed loss reports.

In addition to the availability of said data, the rate at which these observed fatalities occur can vary dramatically between different countries, and even for different earthquakes in the same country. For instance, one would expect the rate of observed loss to be much faster for developed countries than undeveloped countries, due to the greater resources available for response and recovery. In order to understand how reported losses accumulate over time, reported loss data for 20 different earthquakes will be examined and fit with a candidate reported loss accumulation model. These are called the training earthquakes in this work, because they are used to predict the reported loss accumulation rates for any earthquake. Three additional earthquakes were chosen in order to test the following partial loss updating models. The reported loss data for these test earthquakes is shown in the following pages.

The first observed loss test data is for the 2015 Gorkha earthquake [20] compiled from Nepal police reports and NEOC death toll counts in CATDAT in Figure 4.14. For ease in seeing the behavior of this data, each reported loss data set is shown in the linear, semilog, and loglog scales.

Looking at the semilog and loglog plots, one can see there is definitely some sort of logarithmic behavior to this growth rate. In addition, this data set has a relatively smooth growth behaviour.
Figure 4.14: Observed loss after different times following the 2015 Gorkha earthquake.

The second test earthquake is the 2009 L’Aquila earthquake in Italy in Figure 4.15, and its reported loss data was compiled by Max Wyss from the International Center for Earth Modelling [21].

Figure 4.15: Observed loss after different times following the 2009 L’Aquila earthquake.

Interestingly these plots show the same features as the Gorkha data, illustrated by the 'S' shape in the middle semilog plots.
The final test earthquake is the 2017 Mexico City earthquake (Figure 4.16). This data was compiled by earthquake-report.com which gathers data from various news sources in the hours and days after the earthquake occurred [22].

![Figure 4.16: Observed loss after different times following the 2017 Mexico City earthquake.](image)

Again, one can see similar results, although we do not have enough data points at the tails to tell whether this earthquake also followed the 'S' trend in the semilog plot. The next two chapters describe a loss updating framework which will use the 20 training earthquakes, a candidate reported loss accumulation model, and the initial PAGER loss distribution in order to obtain updated total loss predictions given actual reported losses over time.
CHAPTER 5
HOW REPORTED LOSSES EVOLVE OVER TIME AND BASICS OF KALMAN FILTERING

Recall that the USGS PAGER team is responsible for determining and communicating the potential scope of loss quickly so that its users can appropriately plan aid and recovery. Regardless of whether an incorrect PAGER alert level estimate was statistically reasonable or not, it is of more importance to be able to obtain the correct alert level as quickly as possible. As mentioned, PAGER currently only updates its estimates when the ShakeMap gets updated. As such, the PAGER estimate alone is not capable of being updated without any other information, and so a different means of estimating total loss is needed. One piece of possibly useful data could be the accumulating losses reported over time following an earthquake.

If a reasonable relationship between the reported losses over time and some of the earthquake characteristics could be found, then this model could be used to estimate the total loss given the reported losses, which represent partial observations of the total loss. However, deadly earthquakes are rare, and the rate at which these losses are reported could not only depend on the magnitude of the earthquake, but also the maximum shaking felt by the affected population as well as the country or region in which it occurred, and possibly other characteristics of the event. Because of this, using reported loss data from past deadly earthquakes to estimate loss for a future earthquake will have large uncertainties, similarly to how the uncertainty in the PAGER fatality rate model for different countries arose from it being trained from past earthquakes. In addition, because the loss estimation needs to be done as fast as possible, say within one day or so, these updated total loss estimates can only be made with a handful of reported losses, leading to larger uncertainties in the final loss given only a handful of observations. Rather than discarding the PAGER estimate in
favor of a different, perhaps equally uncertain model, a method which incorporates both of them should be able to give more accurate and precise estimates than either method alone, even with only a few data points within the first day or so after the event.

The purpose of this chapter and the next is to formulate a model which will update the PAGER estimated losses given real time reported loss data. Such data is scarce for any given earthquake, and the earthquakes the reported data has large uncertainties, particularly in the early times. For some earthquakes, the local government or police will keep a tally of the reported loss and its time, however most reported loss data has to be obtained by investigating different media reports following the incident to get an idea of the number of fatalities at the time the report was written. Such is the case for the 2017 Mexico City earthquake data [22] used in this thesis, which, along with several other earthquakes’ loss data, was obtained from earthquake-report.com. Several other earthquakes’ loss data used in training the reported loss model comes from a paper which was trying fit these earthquakes’ loss data using an exponential cumulative distribution function, see Zhao et al. [1]. This exponential cdf used in Zhao’s work showed reasonable fits to the curve for earthquakes with more than a hundred people. Earthquakes with much less than a hundred generally do not grow at this exponentially decaying rate, and estimating their total losses given partial observations will not be studied herein.

However, deadly earthquakes are rather rare, and ones with more than a hundred deaths are even more so, with maybe one or two occurring each year. Because of this, there is a limited number of earthquakes with the reported loss data for earthquakes with more than a hundred fatalities, and of these, some of them caused tsunamis which contributed to the majority of the total losses. PAGER does not currently have a model which estimates tsunami loss, and so does not have the capability of accurately estimating losses for earthquakes which cause them. Removing any earthquakes with tsunami losses as well as those which caused less than a hundred fatalities, this thesis was able to compile reported loss data for 23 different earthquakes. Twenty of these are used to train the reported loss model, while
the remaining three will be used to test the total loss updating model given the reported losses over time. While this work makes certain distributional and model assumptions, it is important to note that most of these are for mathematical or demonstrative purposes, and further investigation is required to develop a more realistic, readily implementable version of the updated loss model presented herein.

It should be noted that observed loss does not begin at the same moment as the earthquake, but generally after a few hours following, depending on the time of day and efficiency with which recovery and support can be assembled and begin observing loss. In order to account for this, the time at which the reported losses begin is set to some positive number, and all data are shifted accordingly when training the reported loss model and in the Kalman filter updating process. The main motivation behind this is that without it the first observation often falls well below where it is predicted to be because the model thinks the reported loss process begins at the same time as the earthquake.

5.1 Modeling Reported Losses Using an Exponential CDF

The exponential cdf reported loss model has been used by in multiple studies to model the rate of reported loss following earthquakes and tsunamis [1, 23]. The model describes a process whose derivative follows an exponential decay at some rate $\alpha$. This cdf model gives the number of observed loss $N(t)$ at time $t$ to be

$$N(t) = N_f(1 - e^{-\alpha t}),$$

where $N_\infty$ is the total loss and $\alpha$ is an exponential decay rate. Note that $N(t)/N_\infty = 1 - e^{-\alpha t}$ is an exponential cdf with parameter $\alpha$, hence the reported loss model’s name. In order to determine the optimum parameters for Equation 5.1 given a set of reported loss data for a particular earthquake, a nonlinear regression of these parameters given the data can be performed. Rather than setting $N_\infty$ to be equal to the actual total reported loss, it is left uncertain in this regression process. The reason for this is because forcing the regression to
match the final loss limits how well the model can fit the early loss data, which can provide inaccurate estimations of reported losses for these times.

5.1.1 Nonlinear Regression of Reported Loss Data

Supposing that reported loss data is available for a past earthquake, specific parameters of the reported loss model can be found which minimize some error measure. Suppose that $n$ observed partial losses are observed at subsequent times $t_1, ..., t_n$ following an earthquake. Let the functions $f_i$ and $g_i$ be such that

$$f_i(N_\infty, \alpha) = N_\infty (1 - e^{-\alpha t_i}) - N_i,$$
$$g_i(N_\infty, \alpha) = \log(N_\infty) + \log(1 - e^{-\alpha t_i}) - \log(N_i), \quad i = 1, ..., n$$

then let $F$ be a function such that

$$F(N_\infty, \alpha) = \frac{1}{2} \sum_{i=1}^{n} g_i^2(N_\infty, \alpha) + \frac{1}{2} \log \left( \sum_{i=1}^{n} f_i^2(N_\infty, \alpha) \right).$$

The goal of this nonlinear regression is to determine values of $N_\infty$ and $\alpha$ that minimize $F(N_\infty, \alpha)$. In order to minimize this, this thesis used the built-in MatLab function fminsearch.m, which minimizes an objective function given an initial guess value by using the Nelder-Mead downhill simplex method, the default nonlinear solver used by fminsearch.m. Because it is the early growth that will dictate how the loss prediction works for early times, the $G$-norm to quantify the residuals of our estimations using this nonlinear regression, as it will be used later to estimate the variability of $\log(N(t))$ around the exponential cdf model for different times $t$. After determining $N_\infty$ and $\alpha$ from the above minimization process, call them $N_\infty^*$ and $\alpha^*$, a residual variance of can be found by taking the $G$-norm of the residuals found using $N_\infty^*$ and $\alpha^*$ in Equation 5.1.

This residual variance gives an idea of the spread of the residuals using $N_\infty^*$ and $\alpha^*$ in the reported loss model, indicating an uncertainty in the observed loss at any given time. However, it is also necessary to quantify the uncertainty in the estimated parameters $N_\infty^*$ and
\( \alpha^* \). For nonlinear regression, this involves computing the Hessian of the function \( F(N_\infty, \alpha) \), evaluated at \( N_\infty^* \) and \( \alpha^* \) [24]. For the two parameter system \( F \), its Hessian evaluated at these estimated values is

\[
H = \begin{bmatrix}
\frac{\partial^2 F}{\partial N_\infty^2}(N_\infty^*, \alpha^*) & \frac{\partial^2 F}{\partial N_\infty \partial \alpha}(N_\infty^*, \alpha^*) \\
\frac{\partial^2 F}{\partial \alpha^2}(N_\infty^*, \alpha^*)
\end{bmatrix}.
\]

[24] Suppose that the residuals are independent and identically distributed random variables with variance \( \sigma^2 \), which can be approximated by the square of the residual standard deviation using the \( G \)-norm, then the covariance matrix \( C \) of the estimated parameters \( N_\infty^*, \alpha^* \) is [24]

\[
C = \begin{bmatrix}
\text{Var}(N_\infty^*) & \text{Cov}(N_\infty^*, \alpha^*) \\
\text{Cov}(\alpha^*, N_\infty^*) & \text{Var}(\alpha^*)
\end{bmatrix} = \sigma^2 H^{-1}.
\]

Therefore one can not only obtain the uncertainty in the estimation of \( N_\infty^* \) and \( \alpha^* \), but also their covariance, which is necessary to form confidence bounds around the expected exponential cdf model fit. Before the Hessian arguments for \( F(N_\infty, \alpha) \) can be calculated, the partial derivatives up to order 2 of \( g_i(N_\infty, \alpha) \) and \( f_i(N_\infty, \alpha) \) must be calculated. For \( g_i(N_\infty, \alpha) = \log(N_\infty) + \log(1 - e^{-\alpha t_i}) - \log(N_i) \),

\[
\frac{\partial g_i}{\partial N_\infty} = \frac{1}{N_\infty}, \quad \frac{\partial g_i}{\partial \alpha} = -\frac{1}{N_\infty} \frac{t_i e^{-\alpha t_i}}{1 - e^{-\alpha t_i}},
\]

and for \( f_i(N_\infty, \alpha) = N_\infty (1 - e^{-\alpha t_i}) - N_i \),

\[
\frac{\partial f_i}{\partial N_\infty} = 1 - e^{-\alpha t_i}, \quad \frac{\partial f_i}{\partial \alpha} = t_i N_\infty e^{-\alpha t_i}, \quad \frac{\partial^2 f_i}{\partial \alpha^2} = -t_i^2 N_\infty e^{-\alpha t_i},
\]

The partial derivatives of \( F(N_\infty, \alpha) = \frac{1}{2} \sum_{i=1}^n g_i^2(N_\infty, \alpha) + \frac{1}{2} \log(\sum_{i=1}^n f_i^2(N_\infty, \alpha)) \) up to order 2 are then
\[
\frac{\partial F}{\partial N_\infty} = \sum_{i=1}^{n} g_i(N_\infty, \alpha) \frac{\partial g_i}{\partial N_\infty} + \sum_{i=1}^{n} f_i(N_\infty, \alpha) \frac{\partial f_i}{\partial N_\infty} \frac{n}{\sum_{i=1}^{n} f_i^2(N_\infty, \alpha)},
\]

\[
\frac{\partial F}{\partial \alpha} = \sum_{i=1}^{n} g_i(N_\infty, \alpha) \frac{\partial g_i}{\partial \alpha} + \sum_{i=1}^{n} f_i(N_\infty, \alpha) \frac{\partial f_i}{\partial \alpha} \frac{n}{\sum_{i=1}^{n} f_i^2(N_\infty, \alpha)},
\]

and the second order derivatives are:

\[
\frac{\partial^2 F}{\partial N_\infty^2} = \sum_{i=1}^{n} \left[ \left( \frac{\partial g_i}{\partial N_\infty} \right)^2 + g_i(N_\infty, \alpha) \frac{\partial^2 g_i}{\partial N_\infty^2} \right] + \left( \sum_{i=1}^{n} f_i^2(N_\infty, \alpha) \right) \left( \sum_{i=1}^{n} \left( \frac{\partial f_i}{\partial N_\infty} \right)^2 + f_i(N_\infty, \alpha) \frac{\partial^2 f_i}{\partial N_\infty^2} \right) - 2 \left( \sum_{i=1}^{n} f_i(N_\infty, \alpha) \frac{\partial f_i}{\partial N_\infty} \right)^2 \frac{n}{\sum_{i=1}^{n} f_i^2(N_\infty, \alpha)^2},
\]

\[
\frac{\partial^2 F}{\partial \alpha^2} = \sum_{i=1}^{n} \left[ \left( \frac{\partial g_i}{\partial \alpha} \right)^2 + g_i(N_\infty, \alpha) \frac{\partial^2 g_i}{\partial \alpha^2} \right] + \left( \sum_{i=1}^{n} f_i^2(N_\infty, \alpha) \right) \left( \sum_{i=1}^{n} \left( \frac{\partial f_i}{\partial \alpha} \right)^2 + f_i(N_\infty, \alpha) \frac{\partial^2 f_i}{\partial \alpha^2} \right) - 2 \left( \sum_{i=1}^{n} f_i(N_\infty, \alpha) \frac{\partial f_i}{\partial \alpha} \right)^2 \frac{n}{\sum_{i=1}^{n} f_i^2(N_\infty, \alpha)^2},
\]

\[
\frac{\partial^2 F}{\partial N_\infty \partial \alpha} = \sum_{i=1}^{n} \left[ \frac{\partial g_i}{\partial N_\infty} \frac{\partial g_i}{\partial \alpha} + g_i(N_\infty, \alpha) \frac{\partial^2 g_i}{\partial N_\infty \partial \alpha} \right] + \frac{n}{\sum_{i=1}^{n} f_i^2(N_\infty, \alpha)} \left( \sum_{i=1}^{n} \frac{\partial f_i}{\partial N_\infty} \frac{\partial f_i}{\partial \alpha} + f_i(N_\infty, \alpha) \frac{\partial^2 f_i}{\partial N_\infty \partial \alpha} \right) - \frac{2 \left( \sum_{i=1}^{n} f_i(N_\infty, \alpha) \frac{\partial f_i}{\partial N_\infty} \right) \left( \sum_{i=1}^{n} f_i(N_\infty, \alpha) \frac{\partial f_i}{\partial \alpha} \right)}{\left( \sum_{i=1}^{n} f_i^2(N_\infty, \alpha) \right)^2}.
\]

Thus for any earthquake with reported loss data, the estimated parameter covariance matrix can be determined by first evaluating the partial derivatives of \( f_i \) and \( g_i \) with respect to \( N_\infty \) and \( \alpha \), evaluated at the optimal \( N_\infty^* \) and \( \alpha^* \), and then plugging these into the the
corresponding terms in the second order partial derivatives of \( F(N_\infty, \alpha) \) in the preceding equation. Once each of these partials is computed for \((N_\infty^*, \alpha^*)\), the Hessian matrix \( H \) can be formed, from which the parameter covariance matrix \( C \) of the regression parameters can be determined using

\[
C = \sigma^2 H^{-1},
\]

where again, \( \sigma \) is the standard deviation of the log residuals. For any time \( t_i \) the reported loss model can be used with the estimated parameters \( N_\infty^* \) and \( \alpha^* \) to get a mean expected value of the logarithm of loss at time \( t_i \) by simply letting \( \log(N(t_i)) = \log(N_\infty^*) + \log \left( 1 - e^{-\alpha^* t_i} \right) \).

In order to obtain a confidence band around this expected value, the uncertainty in the prediction of loss at time \( t_i \) must be determined.

Suppose that \( X(t_i) = \left[ \frac{\partial g(t_i)}{\partial N_\infty} \big| (N_\infty^*, \alpha^*) \, , \, \frac{\partial g(t_i)}{\partial \alpha} \big| (N_\infty^*, \alpha^*) \right]^T \), then the propagation of uncertainty formula gives \([24]\)

\[
\text{Var}(\log(N(t_i))) = X(t_i)^T C X(t_i).
\]

Thus an estimate of the loss at any time \( t \) can be obtained (not just the observed loss times), as well as an estimate of the variance in loss at that time. If the residuals are considered to be normally distributed, then one must use a \( t\)-statistic with \( n - 2 \) degrees of freedom to form a confidence bound, because there are \( n \) data points based on a model with 2 parameters. Thus for a \( (1 - \delta)100\% \) confidence level, the corresponding confidence interval around \( \log(N(t_i)) \) is found by :

\[
(1 - \delta)\% \text{C.I.} = \log(N(t_i)) \pm t_{n-2}(\delta/2) \sqrt{X(t_i)^T C X(t_i)},
\]

where \( t_{n-2}(\delta/2) \) is the \( t\)-statistic with \( n - 2 \) degrees of freedom evaluated at \( \delta/2 \) \([24]\). Subsequently, the formulations with respect to the \( L_2 \) and \( G \)-norms can be readily obtained from the \( L_2 G \) formulation. In order to understand the benefits of the \( L_2 G \) norm regression fit one of the training earthquakes will be used as an example.

The 7.7 Mw Pakistan earthquake on September 24, 2013 had a total of 515 reported deaths according to "earthquake report.com" \([2]\). Using the reported loss data from this
website was used to find best fits to the reported loss model with respect to all three norms in order to better understand their differences and the particular choice of the $L2G$ norm for this thesis in Figure 5.1.

![Comparison of nonlinear regression for different norms](image)

Figure 5.1: Comparison of nonlinear regression for different norms

The results indicate that the norms are behaving as predicted. The $L2$-norm regression fit is more strongly influenced by the final losses, while the initial losses are less certain. In contrast, the $G$-norm fit has a larger uncertainty in the final losses, but captures the early reported loss behavior well. As predicted the $L2G$-norm is being near equally influenced from the both initial and final losses, and is most certain in the interior around where the saturation begins taking over around the 70-80% total loss location.

The plots contain information about the fitted decay rate $\alpha$ and the estimated total losses, along with their corresponding uncertainties. The residual standard deviation estimates of $N(t)$ and $\log(N(t))$ are also included in Figure 5.1. As expected, the $L2$-norm fit has the lowest $L2$-norm residual standard deviation, with the $L2G$-norm estimate falling between...
the $G$ and $L2$ estimates. Conversely, the $G$-norm regression estimate performs the best with respect to the $G$-norm, with the $L2G$ estimate falling between it and the $L2$ approximation. In short, it is clear that the $L2G$ regression fit does a good job in both the loss and logarithm of loss domains because it is sensitive to both the early and late behavior when estimated the decay rate and total loss parameters. In order to estimate the decay rate for a future earthquake with no reported loss data, this $L2G$ nonlinear regression will be applied to the training earthquakes.

5.1.2 Modeling Loss Projection

If no reported loss data is available for an earthquake, such as immediately after the event, then the reported loss model parameters $N_\infty$ and decay rate $\alpha$ must be estimated by some other means. Because these parameters must be estimated, the uncertainties in their estimations will affect how the Kalman filter model works. Because the initial PAGER total loss estimate is assumed to be lognormally distributed, it is natural to transform Equation 5.1 by taking the natural logarithm of it. Let $\theta(t) := \log(N(t))$ and $\theta_\infty = \log(N_\infty)$. Because the decay rate and total loss are random variables, the loss at any time $t$ is also a random variable. Following traditional random variable notation, let $\Theta(t)$, $\Theta_\infty$ and $A$ be random variables representing the logarithm of loss at at time $t$, the logarithm of total loss, and the exponential decay rate, respectively. Lower-case values of these variables will be reserved to represent specific realizations of their corresponding random variable quantities. Using this change in notation, the logarithm of Equation 5.1 becomes

$$\Theta(t) = \Theta_\infty + \log \left(1 - e^{-At}\right). \quad (5.2)$$

Based roughly on the work by [1], this work assumes the initial distribution of $A$ to be lognormally distributed. As such, similarly to $\Theta_\infty$ it is more convenient to speak in terms of the logarithm of $A$. Let $\mu_{\Theta_\infty}$ and $\sigma_{\Theta_\infty}^2$ denote the mean and variance of $\Theta_\infty$, respectively,
and let $\mu_{\log(A)}$ and $\sigma^2_{\log(A)}$ denote the mean and variance of $\log(A)$. However, this model assumes no variability in $\Theta(t)$ apart from is estimated parameters. Realistically, one should never expect a discrete, accumulating process to behave exponentially at any given time, even though the mean trend does. This means that at any given time $t$, the true observed loss can be seen as estimated exponential CDF equation multiplied by some residual factor. Just like in Chapter 3, this multiplicative error is chosen for mathematical convenience due to the lognormal PAGER total loss assumption. This error residual factor represents the fact that the model and exponential rate do not exactly reflect reality. Under the logarithmic domain, let $E_{\Theta}(t)$ represent the error in $\Theta(t)$ not explained by the reported loss model at any time $t$. Then Equation 5.2 becomes

$$\Theta(t) = \Theta_\infty + \log(1 - e^{-At}) + E_{\Theta}(t).$$

With this reported loss model in hand, a generalized procedure of updating the total loss given reported losses over time can be constructed through what is known as Kalman filtering.

### 5.1.3 Fitting 20 Training Earthquakes

Below (Figure 5.2 and Figure 5.3) is the data from the 20 training earthquakes with exponential cdfs fitted by minimizing the $L2G$ combined norm, which again was chosen because it finds an exponential rate $\alpha$ that best describes both the early and late reported loss behavior. For each earthquake, the data, the best model fit, and the 67% and 95% confidence curves are plotted, assuming each earthquake’s residuals follow a t-distribution due to the small number of observations for many of the earthquakes studied. Also listed are the residual standard deviation of both the loss ($L2$-norm) and the logarithm of loss ($G$-norm), and the uncertainty in the best fit parameters $N_*^\infty$ and $\alpha^*$. The fitted results indicate that the exponential cdf model does very well at capturing the dominant trend for many of the training earthquakes, while for others sharp changes in the
Figure 5.2: Observed loss data gathered from Zhao et al. [1] and "earthquake report.com" [2]. Plotted over the data is the best fit exponential cdf subject to the L2G norm along with its standard deviation curves.
Figure 5.3: Observed loss data gathered from [1] and [2]. Plotted over the data is the best fit exponential cdf subject to the L2G norm along with its standard deviation curves.
decay rate rate in the beginning, middle, or later times lead to wider uncertainties in the model fit parameters $\alpha$ and $N_\infty$. This is because a fixed rate model is being used whereas some of these earthquakes have different rates at different time scales. These abrupt changes in exponential rate are difficult to model and explain as they could be caused by numerous factors and will not be studied in this work. In addition, many of these earthquakes only have a handful of data points, which makes the uncertainties in the model fits quite uncertain. These larger uncertainties should be taken into account when trying to use this training fit data to estimate future decay rates, which will be seen shortly in the weighted linear least-squares regression analysis illustrated shortly.

Using the best-fit decay rates from the 20 training earthquakes, this work uses them to estimate the decay rate for a future earthquake. This is achieved by finding some relationship between $\alpha$ and some other quantity related to the earthquake or possibly the specific region in which the event occurred. Using linear regression techniques, this thesis investigates whether a relationship between $\alpha$ the moment magnitude ($M_w$) exists for that particular earthquake. There are many other regressors that could be tried, and the use of them will be discussed in the future work portion of the final chapter, however this serves as a starting point to help demonstrate the method. This regression of $\alpha$ in terms of magnitude is used by [1]. Due to current lack of data and understanding about the nature of these rates for different earthquakes and regions, developing better regressor variables is left for future work. However, with the generalized nature of the problem as of now, any estimations for any of the parameters in Equation 5.8 can be substituted.

### 5.1.4 Estimating $\alpha$ from an Earthquake’s Moment Magnitude

A popular approach for determining the correlation between two random variables involves finding some functional relationship between them. In the previous subsection, the exponential model was used to describe the functional relationship between the reported losses at any time, the final loss, and the exponential decay rate. If a functional relationship
between two random variables is not known, generally one can try to determine it by figuring 
out what functional form is needed so that there is a linear correlation between two variables 
after being transformed under this functional form. After this transformation is applied, 
the covariance between the two transformed variables can be found via the method of linear 
least-squares regression.

In Zhao et al. [1], they used the obtained $\alpha$ parameter for their 12 training earthquakes 
and found a reasonably linear correlation between the surface magnitude ($M_s$) and $\log(\alpha)$. 
It should be reiterated that in Zhao et al. [1], fitting of the exponential was done with respect 
to the L2 norm, and their model did not shift the model to begin at some positive time after 
the earthquake. This thesis will be also be finding a linear regression of $\log(\alpha)$ in terms of an 
earthquake’s magnitude, except it is done with respect to the moment magnitude as opposed 
to the surface magnitude. This is because the moment magnitude is used in the ShakeMap 
shaking estimations, and because it better signifies the amount of energy produced by an 
earthquake. The regression between $\log(\alpha)$ and the surface magnitude is demonstrated in 
two different ways. The first follows Zhao et al. [1] by assuming that the obtained $\alpha$ values for 
the 20 earthquakes are certain, and the second incorporates the uncertainty in each of these 
$\alpha$ values obtained from the nonlinear regression using the exponential cdf model. Before 
these different linear regression techniques are applied, the basics of how linear least-squares 
fitting will be described.

Suppose that there are two random variables $X$ and $Y$ for which $n$ realizations $x_1, ..., x_n$ 
and $y_1, ..., y_n$, respectively, have been observed. Here $X$ is some arbitrary quantity not to be 
confused with the observed loss values. This notational abuse is only for demonstrating the 
mathematics behind least-squares linear regression and will not be used elsewhere. Further 
suppose that there is a linear relationship between $X$ and $Y$, in that there exists $\beta_0$ and $\beta_1$ 
for which $Y = \beta_0 + \beta_1 X$, and for each $i$, $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, where $\epsilon_i$ is called a residual. 
Suppose that $\epsilon_1, ..., \epsilon_n$ are i.i.d. with mean zero and variance $\tau^2$. The unbiased and minimum 
variance estimators of $\beta_0$ and $\beta_1$ for a linear relationship such as this are found by minimizing
the least squares norm of the residuals. That is, let \( F(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \). The best linear estimator will be found by minimizing \( F(\beta_0, \beta_1) \), or when

\[
\begin{bmatrix}
0 \\
0
\end{bmatrix} = \begin{bmatrix}
\frac{\partial F}{\partial \beta_0} \\
\frac{\partial F}{\partial \beta_1}
\end{bmatrix} = \begin{bmatrix}
-2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) \\
-2 \sum_{i=1}^{n} x_i(y_i - \beta_0 - \beta_1 x_i)
\end{bmatrix}
\begin{bmatrix}
-2 \sum_{i=1}^{n} y_i + 2n\beta_0 + 2\beta_1 \sum_{i=1}^{n} x_i \\
-2 \sum_{i=1}^{n} x_i y_i + 2\beta_0 \sum_{i=1}^{n} x_i + 2\beta_1 \sum_{i=1}^{n} x_i^2
\end{bmatrix}.
\]

For the top element of the previous equation, one obtains

\[
\hat{\beta}_0 = \frac{1}{n} \sum_{i=1}^{n} y_i - \hat{\beta}_1 \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{y} - \hat{\beta}_1 \bar{x},
\]

where \( \bar{y} \) and \( \bar{x} \) are the sample means of the samples of \( Y \) and \( X \), respectively, and the hats added to the linear parameters signify that they are only estimates of the true \( \beta_0 \) and \( \beta_1 \).

Inserting this relationship into the bottom element from above gives

\[
\sum_{i=1}^{n} x_i y_i = (\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2
\]

\[
= \bar{y} \sum_{i=1}^{n} x_i + \hat{\beta}_1 \left( \sum_{i=1}^{n} x_i^2 - \bar{x} x_i \right)
\]

\[
\Rightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \bar{y} \bar{x}}{\sum_{i=1}^{n} x_i^2 - \bar{x} x_i} = \frac{\sum_{i=1}^{n} x_i y_i - \bar{y} \bar{x}}{\sum_{i=1}^{n} x_i^2 - \bar{x} x_i}.
\]

This reduces to

\[
\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\text{Cov}(X, Y)}{s_X^2},
\]

where \( \text{Cov}(X, Y) \) is the sample covariance of the realizations of \( X \) and \( Y \) and \( s_X^2 \) is the sample variance of the realizations of \( X \). The estimations of the residuals \( \epsilon_1, \ldots, \epsilon_n \) can be found by

\[
\hat{\epsilon}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i.
\]

An estimate of the spread around this linear relationship can also be found by taking the standard deviation of these residual estimates. Now this can be applied to the current case
of expressing \( \log(\alpha) \) linearly in terms of the moment magnitude \( M_w \) for the 20 training earthquakes.

Using linear least-squares regression of \( \log(\alpha) \) on \( M_w \), this relationship is found. Below in Figure 5.4 scatterplot of \( \log(\alpha) \) and \( M_w \) is plotted for these earthquakes along with the fit and single standard deviation lines, including the linear relationship found and the estimated residual standard deviation, which quantifies the uncertainty in \( \log(\alpha) \) using \( M_w \) as a regressor.

![Linear Regression Plot of \( \log(\alpha) \) vs. Moment Magnitude \( (M_w) \)](image)

\[
\log(\alpha) = -0.556 \times M_w + 3.877
\]

\[
\text{residual std} = 1.007
\]

Figure 5.4: Linear regression fit of \( \log(\alpha) \) versus the moment magnitude \( (M_w) \)

Unfortunately there doesn’t seem to be noticeable correlation between the two. In Zhao et al. [1] they found a more reasonable trend using only 12 earthquakes, 9 of which are present in training earthquake data, the remaining 3 being from earthquakes which caused major tsunamis. The small amount of data makes it difficult to ascertain too much about how this data is actually correlated. Generally more regressors can be added to obtain a better regression for \( \log(\alpha) \), but as mentioned that topic is left for future work. However, this method assumes no uncertainty in the \( \alpha \) values themselves, and perhaps giving less importance to values with higher variance will help determine a stronger relationship.
In order to give better preference to those values with smaller variance in the least squares fitting, the traditional linear least-squares approach can be modified using what is known as weighted least squares regression [25].

In this method weights $w_1, \ldots, w_n$ are introduced corresponding to $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$, where again $x$ is some arbitrary quantity not to be confused with the observed loss values. The weighted least-squares method seeks to minimize the function $F_w(\beta_0, \beta_1) = \sum_{i=1}^n w_i (y_i - \beta_0 - \beta_1 x_i)^2$, and chooses $w_i = \sigma_{y_i}^2$, where $\sigma_{y_i}^2$ represents the measurement noise in the observation $y_i$. This ensures that points with larger variance have less weight in the least squares sum and will not contribute as much to the linear fit. According to Mack (2016), if $\overline{y}_w$ denotes the weighted sample mean of our realizations of $Y$, and $\overline{x}_w$ denotes the weighted sample mean of the realizations of $x_1, \ldots, x_n$ of $X$, where

$$\overline{x}_w = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i},$$
$$\overline{y}_w = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i},$$

then the optimal estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ are given by

$$\hat{\beta}_0 = \overline{y}_w - \hat{\beta}_1 \overline{x}_w,$$
$$\hat{\beta}_1 = \frac{\sum_{i=1}^n w_i (y_i - \overline{y}_w)(x_i - \overline{x}_w)}{\sum_{i=1}^n w_i (x_i - \overline{x}_w)^2}.$$

Applying this to the regression of $\log(\alpha)$ in terms of $Mw$ this work obtained the results shown in the following plot, which summarizes the data, its uncertainty, and linear fit along with its uncertainty in Figure 5.5.

These results indicate a stronger trend than was available using the traditional least-squares regression, because the data points farthest from the trend line have some of the largest uncertainties. Because this second regression model accounts for the uncertainty in each earthquakes $\alpha$ and has a stronger relationship, it will be used henceforth when trying to estimate how the reported losses will accumulate for an earthquake before any partial
observations of loss have been made. In addition, results from the 20 training parameter covariance matrices from the 20 training earthquakes showed a strong negative correlation between the decay rate and the final loss. The sample mean of all 20 correlations were \( .74 \) with a standard deviation of \( .15 \). This indicates that larger earthquakes should tend to have smaller decay rates. Physically this makes sense because for large earthquakes causing many fatalities, the local authorities may not be able to proportionally scale their response to the amount of people dying or in danger, limiting the rate at which loss is observed. While this correlation will not be dealt with explicitly, this work assumes some additional uncertainty in the total loss due to estimating the decay rate from this regression which will be referred to as \( R_{\Theta_{\infty}} \sim N(0, \phi_{\Theta}^2) \), signifying that it is an error from the nonlinear regression of the exponential reported loss model. Similarly, there is a regression uncertainty for the decay rate, call it \( R_A \sim N(0, \phi_A^2) \), although this does not appear in the basic model to to the constant \( \alpha \) assumption, although it will be used in the following chapter.

Without any reported loss information, the total loss regression error \( R_{\Theta_{\infty}} \) and process noise \( E_{\Theta}(t) \) are not known. As mentioned when describing the full model, the variances of these distributions can be estimated from the model and process error variances from
the training earthquakes. The sample mean of the uncertainties in \( \log(N_\infty) \) from the 20 earthquakes can be used in order to estimate the model error variance \( \phi_\Theta^2 \approx 0.170 \), although this estimate does carry some uncertainty. This work will not include this uncertainty, although in the full model it mentioned that viewing the model variance itself as a random variable with statistics determined by the sample mean and variance of the model error variances could be applied. In effect, including this uncertainty in the estimated model variance from the 20 training earthquakes will lead to updated distributions with larger tails, similarly to how the updated loss transformed when updating with respect to the uncertain \( \sigma^2 \) in the previous chapter.

The process noise \( E_\Theta(t) \) is more difficult, because in the logarithmic sense it is not homoscedastic and the variance in the residuals appears to decay over time. Because of the small amount of data for many of these earthquakes, determining this noise decay over time for any one earthquake would be difficult and highly uncertain. In order to rectify this, this thesis uses a different approach where it looks at the log loss residuals from the fitted earthquakes all together as can be seen in Figure 5.6. Specifically, looking at the log residuals at their respective times for all earthquakes allowed us to see that \( \Theta \) process noise was decaying over time. During the first day the standard deviation of these log residuals was found to be about .571, and after the first day was about .10. This motivates the use of a decaying process noise, and this paper assumes that the process noise over time drops at an exponential rate of \( 1/t^5 \) (or the process variance decays at a rate of \( 1/t \)), although this is just a starting approximation and set only to visually fit the data in Figure 5.6. This paper chooses a based standard deviation equal to \( \tau_{Base} \approx .571 \), and fixes this to be the variance at a time of 12 hours after the event. That is to say, it sets the standard deviation at a half day to be the residual standard deviation from all points less than a day. The log residuals from all training earthquakes along with the decaying process noise fit are shown below in Figure 5.6.
Figure 5.6: Plot of the log residuals from all 20 earthquakes nonlinear regression fits along with the estimated process noise curve (red) using a decay rate of -.5 and fixing the standard deviation at 12 hours to be .571, equal to the residual std of all data taken at $t < 1$ day.

Note that this fitted curve is simply an empirical visual fit used as a starting approximation, and further analysis must be done to determine the proper curve for any given earthquake, and more importantly to investigate the correlation structure of the process noise (as well as the measurement noise). Using the $Mw$ regression analysis and training earthquake nonlinear regression analysis the following values have been assigned to the reported loss model parameters based on the 20 earthquakes' training data:

$$\beta_0 = 2.752, \quad \beta_1 = -0.375,$$
$$\phi_0^2 = 0.170, \quad \tau_0^2(t) = \tau_{Base}^2(2t)^{-1},$$
$$\tau_{Base} = .571$$

For demonstrative purposes, the measurement noise will be set equal to the process noise, again not a realistic assumption, but one that should equally weigh the two noises. Because the observed loss cannot be directly observed, having the measurement and process noise leads to an underdetermined system for which it would be difficult to distinguish one from the other. Determining a more realistic measurement noise is also beyond the scope of this
thesis and left for future work.

5.2 Kalman Filtering

This section seeks to describe how the Kalman filter based loss updating model works. This process is illustrated with a more basic version of Equation 5.3 which assumes that the decay rate is a known value $\alpha$. Information about Kalman filters used in the following formulations can be found in [26] and [27].

5.2.1 The Kalman filter loss updating model

Assume that some constant value $\alpha$ is used to estimate the random variable $A$, so that it is treated as a deterministic quantity. The observed loss at time $t_2$ based off of the observed loss at time $t_1$, with $t_1 < t_2$ can be found by solving Equation 5.3 for $\Theta_\infty$ at time $t_1$ and plugging it into the same equation solved for $t_2$ giving

$$\Theta(t_2) = \Theta(t_1) + \log \left( \frac{1 - e^{-\alpha t_2}}{1 - e^{-\alpha t_1}} \right) + \mathcal{E}_\Theta(t_2) - \mathcal{E}_\Theta(t_1).$$

Furthermore, a distinction can be made between the losses that are observed versus those that are reported by the media. This means that the reported loss at time $t$, call it $X(t)$, is the observed loss $\Theta(t)$ plus some residual error, call it $\mathcal{E}_X(t)$, or $X(t) = \Theta(t) + \mathcal{E}_X(t)$ for any $t$. That is, $\mathcal{E}_X(t)$ represents the fact that the reported values have error compared to the actual observed values over time. While before this distinction between observed and reported losses was not made, henceforth it will in that the observed losses can never be directly observed. In order to help make this distinction, this work will often refer to $\mathcal{E}_X_i$ as a measurement noise, which also more closely aligns to the traditional Kalman filter system.

Suppose that an earthquake has $n$ different reported losses $x_1, ..., x_n$ taken at times $t_1, ..., t_n$. In what follows, let $\Theta_i := \Theta(t_i)$ and $X_i := X(t)$, $i = 1, ..., n$. Given the first observed loss, a recursive relationship for each observed loss can be found.
\[
\Theta_i = \Theta_{i-1} + \log \left( \frac{1 - e^{-\alpha t_i}}{1 - e^{-\alpha t_{i-1}}} \right) + \mathcal{E}_\Theta(t_2) - \mathcal{E}_\Theta(t_1), \\
X_i = \Theta_i + \mathcal{E}_{X_i},
\]
(5.6)
for \(i = 2, ..., n\). The Kalman filter works by assuming each observation is only correlated to its preceding observation, and provides a way of injecting a basic correlation structure in the observed loss temporal process. The term inside the logarithm of 5.6 can be seen as a forcing term which shifts the temporal observed loss process. The Kalman filter uses what is known as a Markov assumption in order to say that conditioned on \(\Theta_i\), each observation \(X_i\) is independent of all other \(\Theta_j\) and \(X_j\) for \(j \neq i\). Similarly, another Markov assumption is made between the different log reported losses over time, in that conditioned on the previous observation \(\Theta_{i-1}\), \(\Theta_i\) is independent of all observed \(X_i\), \(i = 1, ..., N\) and all \(\Theta_j\), \(j < i - 1\).

The basic Kalman filter process will now be developed slowly step by step.

Using Equation 5.3 the observed loss at time \(t_1\) is

\[
\Theta_1 = \Theta_\infty + \log \left( 1 - e^{-\alpha t_1} \right) + \mathcal{E}_\Theta(t_1),
\]
where \(\Theta_\infty\) is assumed to be some estimated random variable (such as the PAGER initial estimate). According to Bayes’ equation, for an a realization \(x_1\) of \(X_1\) at \(t_1\),

\[
f_{\Theta_1 | X_1}(\theta_1 | x_1) = \frac{f_{X_1 | \Theta_1}(x_1 | \theta_1) f_{\Theta_1}(\theta_1)}{C_1},
\]
where \(C_1\) is a normalization constant. Using this distribution of \(\Theta_1\) conditioned on the observation \(X_1 = x_1\), Equation 5.6 can be used to predict the total loss \(\Theta_2\) at time \(t_2\), conditioned on \(x_1\). Using the law of total probability,

\[
f_{\Theta_2 | X_1}(\theta_2 | x_1) = \int_{\Theta_1} f_{\Theta_2 | \Theta_1, X_1}(\theta_2 | \theta_1, x_1) f_{\Theta_1 | X_1}(\theta_1 | x_1) d\theta_1
\]
\[
= \int_{\Theta_1} f_{\Theta_2 | \Theta_1}(\theta_2 | \theta_1) f_{\Theta_1 | X_1}(\theta_1 | x_1) d\theta_1
\]
due to the Markov assumption regarding the independence of \(\Theta_2\) and \(X_1\) conditioned on \(\Theta_1\). Now
\[ \Theta_2 = \Theta_1 + \log \left( \frac{1 - e^{-\alpha t_2}}{1 - e^{-\alpha t_1}} \right) + \mathcal{E}_\Theta(t_2) - \mathcal{E}_\Theta(t_1) \]

means that given a realization \( \theta_1 \) of \( \Theta_1 \), The distribution of \( \Theta_2 | \Theta_1 \) is simply the distribution of \( \mathcal{E}_\Theta(t_2) - \mathcal{E}_\Theta(t_1) \) shifted by an amount \( \theta_1 + \log \left( \frac{1 - e^{-\alpha t_2}}{1 - e^{-\alpha t_1}} \right) \) in addition, by letting \( t_2 \to \infty \) one can obtain the total loss based on observed loss \( \Theta_1 \).

Using the Bayes equation one can find the density of observed loss \( \Theta_2 \) at time \( t_2 \) conditioned on the reported losses \( x_1 \) and \( x_2 \):

\[
 f_{\Theta_2|X_1X_2}(\theta_2|x_1, x_2) = \frac{f_{X_2|\Theta_2X_1}(x_2|\theta_2, x_1) f_{\Theta_2|X_1}(\theta_2|x_1)}{C_2} = \frac{f_{X_2|\Theta_2}(x_2|\theta_2) f_{\Theta_2|X_1}(\theta_2|x_1)}{C_2},
\]

where the second equality is because of the Markov assumption

\[
 f_{X_2|\Theta_2X_1}(x_2|\theta_2, x_1) = f_{X_2|\Theta_2}(x_2|\theta_2).
\]

This can be extended for any \( i > 2 \) by using Equation 5.6. The Kalman filter total loss updating model acts in three steps, first the logarithm of loss at time \( t_i \) is updated given the observation of loss \( x_i \), called the updating step, and then the reported loss model is used to predict the reported loss \( \Theta_{i+1} \) at time \( t_{i+1} \), called the predicting step. The final step, called the total loss prediction step, is when the reported loss model is used to determine a new logarithm of total loss \( \Theta_\infty \) using the updated reported loss \( \Theta_i | X_1, ..., X_i \). For \( i > 2 \), given the predicted distribution \( f_{\Theta_i|X_1,...,x_{i-1}}(\theta_i|x_1, ..., x_{i-1}) \) and data distribution \( f_{X_i|\Theta_i}(x_i|\theta_i) \), the updated distribution becomes

\[
 f_{\Theta_i|X_1,...,x_i}(\theta_i|x_1, ..., x_i) = \frac{f_{X_i|\Theta_i}(x_i|\theta_i) f_{\Theta_i|X_1,...,x_{i-1}}(\theta_i|x_1, ..., x_{i-1})}{C_i},
\]

and the predicted density of observed loss at time \( t_{i+1} \) is given using the law of total probability:
\[ f_{\Theta_{i+1}|X_1,\ldots,X_i}(\theta_{i+1}|x_1,\ldots,x_i) = \int_{\Theta_i} f_{\Theta_{i+1}|\Theta_i}(\theta_{i+1}|\theta_i) f_{\Theta_i|X_1,\ldots,X_i}(\theta_i|x_1,\ldots,x_i) d\theta_i. \]

Again, letting \( t_{i+1} \to \infty \) one can obtain the distribution of \( \Theta_\infty|X_1,\ldots,X_i \) by taking the updated distribution of loss \( \Theta_i \) and using Equation 5.6.

The complete basic Kalman filter model can now be summarized. For notational simplicity, let \( \Theta_{i|j} := \Theta_i|[X_1,\ldots,X_j], \quad j = 1,\ldots,i \), and \( \Theta_\infty|j := \Theta_\infty|[X_1,\ldots,X_j], \quad j = 1,\ldots,n \).

Furthermore, let \( f_{i|j} := f_{\Theta_i|X_1,\ldots,X_j}, \quad j = 1,\ldots,i \) and \( f_{\infty|j} := f_{\Theta_\infty|X_1,\ldots,X_j}, \quad j = 1,\ldots,n \). Then

\begin{align*}
\text{Step 1 :} & \quad \Theta_{1|0} = \Theta_\infty|0 + \log \left(1 - e^{-\alpha t_1}\right) + \mathcal{E}_\Theta(t_1), \\
\text{Step 2 :} & \quad \Theta_{i|i-1} = \Theta_{i-1|i-1} + \log \left(\frac{1 - e^{-\alpha t_i}}{1 - e^{-\alpha t_{i-1}}}\right) + \mathcal{E}_\Theta(t_i) - \mathcal{E}_\Theta(t_{i-1}), \quad i > 1, \\
& \quad X_i = \Theta_{i|i-1} + \mathcal{E}_X, \quad (5.7) \\
\text{Step 3 :} & \quad f_{i|i}(\theta_i|x_1,\ldots,x_i) = \frac{f_{X_i|\Theta_{i|i-1}}(x_i|\theta_i)f_{i|i-1}(\theta_i|x_1,\ldots,x_{i-1})}{C_i}, \\
\text{Step 4 :} & \quad \Theta_{\infty|i} = \Theta_{i|i} - \log \left(1 - e^{-\alpha t_i}\right), \quad i = 1,\ldots,n.
\end{align*}

These different updating steps and prediction steps are summarized in Figure 5.7, which depicts a flow chart of how each piece of information is used to predict or update the loss at each time step.

It should be noted that traditional Kalman filtering involves linear equations of normal random variables, whereas Equation 5.7 has not yet made any distributional assumptions and serves a generalization of the traditional Kalman filter in which any distributions can be assumed for the different variables. However, the loss of normality means that the entire updated distributions must be calculated, rather than just the mean and variances as with traditional Kalman filtering. In order to test this model, the next section makes several distributional assumptions to the parameters involved in Equation 5.7. In addition, the reported loss model parameters \( \alpha \) and \( \Theta_\infty \) must also be estimated in order to apply the model to actual earthquake reported loss data.
Figure 5.7: Diagram showing the proposed Kalman filter model for predicting total loss given reported losses. The $x_1, x_2, \ldots$ refer to observed loss, $\theta_{i|i-1}$ refers to the predicted accumulated loss at time $i$ given observations at time $t_1, \ldots, t_{i-1}$, and $\theta_{i|i}$ refers to the updated accumulated loss at time $i$ after incorporating measurement error of $x_i$ conditioned on $\theta_{i|i-1}$ through the Bayes equation.
5.3 The Kalman Filter Model for Updating Total Loss

In order to apply the total loss updating model just described, it is necessary to make some assumptions about the distributions of the variables involved, namely the logarithm of total loss \( \Theta_\infty \), the observed loss process noise \( \mathcal{E}_\Theta(t) \), and the reported loss process noise \( \mathcal{E}_X(t) \). It is also necessary to estimate the decay rate \( \alpha \) through some means. Again, due to the assumption that the PAGER estimated loss \( \Theta_{PAGER} \sim N(\mu_{PAGER}, \sigma_{PAGER}^2 = \zeta^2) \), \( \Theta_\infty \) will be assumed lognormal. In addition, for mathematical convenience both the process noise and measurement noise at each time will be treated as normal random variables and this work will not assume any correlation between these noises and between the same noise type at different times. In reality, these distributional and correlation assumes are far from the truth and is only done as for demonstrative purposes and as a starting, basic assumption.

The random variables involved in Equation 5.7 are as such:

\[
\begin{align*}
\Theta_\infty &\sim N(\mu_\Theta, \sigma_\Theta^2), \\
\mathcal{E}_\Theta(t) &\sim N(0, \tau_\Theta^2(t)), \\
\mathcal{E}_X(t) &\sim N(0, \tau_X^2(t)).
\end{align*}
\]

Given these assumptions, it is necessary to find estimates for four parameters in Equation 5.8 as well as the decay rate parameter \( \alpha \). In order to demonstrate how the Kalman filter works, this thesis chose 20 different earthquakes with reported loss data to train the model by estimating all of the parameters mentioned.

5.3.1 Formulation of Model for Normal Assumptions

Suppose that an earthquake has the following reported loss model:

\[
\Theta(t) = \Theta_{PAGER} + \mathcal{R}_{\Theta_\infty} + \log \left( 1 - e^{-\alpha t} \right) + \mathcal{E}_\Theta(t)
\]  

(5.9)

where \( \log(\alpha) = \beta_0 + \beta_1 M_w \). That is, this model assumes a single exponential decay rate \( \alpha \) found from the weighted least-squares regression of \( \log(\alpha) \) based on \( M_w \) for this earthquake.

Recall the complete Kalman filter loss updating model:
Step 1: \( \Theta_{1|0} = \Theta_{\infty|0} + \log \left( 1 - e^{-\alpha t_1} \right) + \mathcal{E}_\Theta(t_1) \),

Step 2: \( \Theta_{i|i-1} = \Theta_{i-1|i-1} + \log \left( \frac{1 - e^{-\alpha t_i}}{1 - e^{-\alpha t_{i-1}}} \right) + \mathcal{E}_\Theta(t_i) - \mathcal{E}_\Theta(t_{i-1}), \quad i > 1, \)

\[ X_i = \Theta_{i|i-1} + \mathcal{E}_{X_i} \]

Step 3: \( f_{ij}(\theta_i|x_1, ..., x_i) = \frac{f_{X_i|\Theta_{ij-1}}(x_i|\theta_i)f_{ij-1}(\theta_i|x_1, ..., x_{i-1})}{C_i} \),

Step 4: \( \Theta_{\infty|i} = \Theta_{ij} - \log \left( 1 - e^{-\alpha t_i} \right), \quad i = 1, ..., n. \)

Looking at the model and Equation 5.9 one can see that \( \Theta_{\infty|0} = \Theta_{PAGER} + \mathcal{R}_\Theta \). Because the initial total loss estimate and process noise are assumed to be normally distributed, \( \Theta_{1|0} \) is also normally distributed with \( \mu_{\Theta_{1|0}} = \mu_{PAGER} + \log \left( 1 - e^{-\alpha t_1} \right) \) and \( \sigma^2_{\Theta_{1|0}} = \zeta^2 + \phi^2_{\Theta} + \tau^2_{\Theta}(t_1) \).

Because the measurement noise is assumed to be normally distributed, \( X_i|\Theta_{i|i-1} \) is also normally distributed, meaning that the posterior \( \Theta_{1|1} \) will also be normally distributed due to the normal conjugate pair relationship discussed in the previous chapter. Using the recursive relationship of the Kalman filter model, every distribution involved is normal. Furthermore, using the relationships between predictive marginal distributions and Bayesian posterior distributions from Chapter 4, one can obtain the mean and standard deviation for \( \Theta(t) \) or \( \Theta_\infty \) at any time using the following steps.

Update Step:

\( \Theta_{ij} \sim N(\mu_{\Theta_{ij}}, \sigma^2_{\Theta_{ij}}), \) where

\[
\mu_{\Theta_{ij}} = \frac{1}{\sigma^2_{\Theta_{ij-1}} + \frac{1}{\tau_{X_i}}} \left( \frac{\mu_{\Theta_{ij-1}}}{\sigma^2_{\Theta_{ij-1}}} + \frac{x_i}{\tau_{X_i}} \right),
\]

\[
\sigma^2_{\Theta_{ij}} = \frac{1}{\sigma^2_{\Theta_{ij-1}} + \frac{\tau^2_{X_i}}{\tau_{X_i}}}. 
\]

Prediction Step:
\[ \Theta_{i+1|i} \sim N\left( \mu_{\Theta_{i+1|i}}, \sigma^2_{\Theta_{i+1|i}} \right), \text{ where} \]

\[ \mu_{\Theta_{i+1|i}} = \mu_{\Theta_{i|i}} + \log \left( \frac{1 - e^{-\alpha t_i}}{1 - e^{-\alpha t_{i+1}}} \right), \]
\[ \sigma^2_{\Theta_{i+1|i}} = \sigma^2_{\Theta_{i|i}} + \tau^2_{\Theta}(t_{i+1}) + \tau^2_{\Theta}(t_i). \]

and

**Total Loss Prediction Step:**

\[ \Theta_{\infty|i} \sim N\left( \mu_{\Theta_{\infty|i}}, \sigma^2_{\Theta_{\infty|i}} \right), \text{ where} \]

\[ \mu_{\Theta_{\infty|i}} = \mu_{\Theta_{i|i}} - \log \left( 1 - e^{-\alpha t_i} \right), \]
\[ \sigma^2_{\Theta_{\infty|i}} = \sigma^2_{\Theta_{i|i}} + \tau^2_{\Theta}(t_i). \]

The information from the nonlinear regression fitting of the training earthquakes can be used to estimate the reported loss model for a future earthquake. Recall that the basic Kalman filter model is:

\[ \Theta(t) = \Theta_{PAGER} + R_{\Theta} + \log(1 - e^{-\alpha t}) + E_{\Theta}(t), \]
\[ \log(\alpha) = \beta_0 + \beta_1 Mw, \]

where

\[ \Theta_{PAGER} \sim N(\mu_{PAGER}, \sigma^2_{PAGER}), \]
\[ R_{\Theta} \sim N(0, \phi^2_{\Theta}), \]
\[ E_{\Theta}(t) \sim N(0, \tau^2_{\Theta}(t)), \]

with the parameter values given in Equation 5.5 and measurement noise set equal to the process noise. With these crude parameter estimates, the basic Kalman filter is tested on actual reported loss data from the 2015 Gorkha magnitude 7.8 earthquake in Nepal.
5.3.2 Illustration of Kalman Filter Loss Updating Model

The Gorkha earthquake had a recorded moment magnitude of 8.2, which corresponds to an estimated mean $\alpha$ value of 0.72046 using the linear regression relationship between $\log(\alpha)$ and $M_w$. Given the full Gorkha reported loss data set, the exponential cdf model can be fitted using nonlinear regression to give an estimated value for $\alpha$ of , the full reported loss data set and estimated loss model curves are shown below in Figure 5.8.

![Figure 5.8: Plot of the nonlinear regression model fit for the Gorkha earthquake along with the reported loss data from which it was fitted.](image)

The best fit $\alpha$ value using the full data set is almost 1/3 that of the $M_w$ regression estimate. This leads one to the conclusion that assuming no uncertainty in $\alpha$, the Kalman filter model using the $M_w$ estimate should be under estimating the total loss, because it thinks the decay rate is larger. However, the model should still be estimating total losses greater than the updated loss values at each point in time.

Below are some figures corresponding to the Kalman filter using the best fit $\alpha$ value, which would never be available without the full reported loss data set. The first (Figure 5.9 shows the estimated total loss at each timestep along with its standard deviation curves, the actual final reported loss (green line), and the reported loss data itself. For clarity in how the updating works compared to the initial PAGERER estimate (the very first points in each figure), there are two plots showing different scales for the independent and dependent
variables. The left shows a plot of the loss on a logarithmic scale, and the right plot shows everything from a log-log perspective. This second plot is useful at examining how the total loss updates given the earlier data sets, keeping in mind that 1 day is the crucial time to still be able to change aid and recovery. This best fit $\alpha$ scenario uses the model uncertainty in $\Theta_\infty$ and the sample standard deviation of its log residuals from the nonlinear regression fit.

Figure 5.9: Plots showing the updated total loss estimates for Gorkha using the basic KF model and best fit decay rate $\alpha$ along with their uncertainties, the loss data, and the total estimated loss (green line).

This plot shows that given the best fit $\alpha$ the Kalman filter model does well at maintaining the correct total loss that PAGER was assumed to have estimated. The earlier times have the largest changes in estimated total loss. In order to see how this model performs given different PAGER estimates, Figure 5.10 shows the estimated median total loss at each timestep (in the log-log view) for PAGER estimates equal to $E_{final} \times [1/100, 1/10, 1, 10, 100]$, where $E_{final}$ is the final reported loss.

This plot shows that all of the median estimated total losses become very close after only one or two updates. The reason for this is because compared to the uncertainty in the first data points, the uncertainty in the reported loss value at those times due is very large, mostly
Figure 5.10: Plot of the range of the best fit $\alpha$, KF basic total losses given different PAGER estimates for the Gorkha earthquake. Estimates are displayed in the log-log scale, and each initial estimate differs by an order of magnitude.

because of the relatively large uncertainty in the PAGER estimate. This to the model means that the updating process is heavily weighted around the observation, and because $\alpha$ is not being updated, this forces a large change in the estimated total loss regardless of the initial PAGER estimation. This is not ideal, as the information contributed by PAGER is drowned out after only a few observations to the point where they are almost indistinguishable after about a day, and is the motivation for including the uncertainty in $\alpha$ and updating it along with the reported losses at each timestep. Because all variables are normally distributed, the estimation of the PAGER mean has no affect on the standard deviation estimates of the updated total losses.

Now it is time to compare this with the results obtained using the $M_w$ regression $\alpha$ value. In Figure 5.11 and Figure 5.12 the same plots can be seen with this different $\alpha$ value.

These plots comparatively show the median total loss estimates underestimating with each data set, regardless of initial PAGER estimate. This is exactly what was predicted, and the affects of having the wrong $\alpha$ value can be readily seen. In addition, this curve has a
Figure 5.11: Plots showing the updated total loss estimates for Gorkha using the basic KF model and $M_w$ regression decay rate $\alpha$ along with their uncertainties, the loss data, and the total estimated loss (green line).

Figure 5.12: Plot of the range of the $M_w$ regression $\alpha$, KF basic total losses given different PAGER estimates for the Gorkha earthquake. Estimates are displayed in the log-log scale, and each initial estimate differs by an order of magnitude.
larger uncertainty in the estimates, due to the fact that the best fit analysis had the actual model uncertainty and process error estimated from that earthquake’s loss data. Regardless, the initial estimates from this data were sufficient to have changed the PAGER alert level to its correct value, regardless of its initial estimate. However, the uncertainty around those estimates is too tight, and this difference between the two $\alpha$ values is not as easy to see until their respective uncertainties are added in the enhanced Kalman filter model. Gorkha represented a very good fit to the exponential model so it is not surprising that the updating stayed so smooth for the best fit $\alpha$ case. However, for earthquakes in which ones single $\alpha$ seems to vary rapidly this will not be the case.

This Kalman filter loss updating model illustrates how the estimated decay rate can greatly affect the total loss updating procedure, and highlights that in some cases the best fit $\alpha$ value can perform worse than the $Mw$ regression $\alpha$ for portions of any given data set. However, by not including the uncertainty in this $\alpha$ value, the spread of these total loss estimates is much smaller than it should be realistically. This is because this model is a very simplified way of explaining a complex process, and without including the necessary uncertainties in its parameters, the Kalman filter model gives too much confidence in this model, so much so that the PAGER estimated information is lost after only one or two observations. In order to better capture this uncertainty in the model and its parameters, an generalization of the Kalman filter model is proposed in the following chapter which accounts for both the uncertainty in $\alpha$ using the exponential cdf model gathered from the nonlinear regression fit, as well as the uncertainty in $\alpha$ from using the $Mw$ regression when no reported loss data is available.
The results from Chapter 5’s Kalman filter example indicated that incorporating the uncertainty in the decay rate $\alpha$ is necessary when predicting and updating total loss for a future earthquake in order to get more accurate estimates of the spread around these estimates. This section seeks to enhance this Kalman filter model by dealing with these uncertainties in the rate. However, when including this uncertainty, the normality assumptions that define a traditional Kalman filter are lost. As such, the Kalman filter model from the previous chapter must be generalized to account for this change. Due to the loss of normality, the entire distribution of loss and projection rate must be modeled as the conjugate relationships used in the previous chapter are no longer valid. While this enhanced model is not a Kalman filter, but rather a generalization of it, it will be referred henceforth as either the enhanced Kalman filter model, generalized Kalman filter model, or more appropriately the recursive Bayes loss updating model. For an earthquake in which the reported loss data is known, the nonlinear regression applied to the exponential cdf model produced an estimation of the decay rate with some uncertainty. However, there is an additional uncertainty in the rate for earthquakes in which no reported loss observations are available. In these cases, it is necessary to use estimate the distribution of the decay rate. In the previous chapter the linear weighted least-squares regression on the moment magnitude for that earthquake was used to determine $\log(\alpha)$ with some uncertainty. However, the loss model uncertainty in the decay rate is also not known for an earthquake without any reported loss data, and so it too must be estimated from the training earthquakes.

The last chapter mentioned that because there were uncertainties in the $\alpha$ values used to determine the $Mw$ linear regression relationship, that an additional uncertainty existed when estimating the decay rate for a future relationship. This uncertainty, denoted $R_A$,
arises from using the nonlinear regression for the 20 training earthquakes, similar to $R_x$.

This chapter models this nonlinear regression uncertainty as $R_A \sim N(0, \phi_A^2)$, where $\phi_A^2$ is estimated as the sample mean of the variances in $\log(\alpha)$ from the 20 training earthquakes similarly to how the model uncertainty $R_\Theta$ of $\Theta_\infty$ was estimated in the basic Kalman filter example. Using a sample mean for an estimate has its downsides, and a more enhance model would treat these model and process variances themselves as random variables with distributions trained to match the respective sample mean and standard deviations of the different model and process noise variances.

This section will incorporate these additional uncertainties in the decay rate $\alpha$ into the basic Kalman filter used in the previous section. While the results will indicate that including this uncertainty provides more realistic results, it does not treat the variability in this decay rate $\alpha$ over time. That is, all models used thus far assume that there is some fixed (but unknown) decay rate $\alpha$ indicating how the reported loss should accumulate. In reality however, the idea of a fixed rate is unrealistic and in fact the idea that this exponential cdf model could capture the reported loss process perfectly for any given earthquake is also unrealistic. While the limitations of the exponential cdf model will not be addressed.

### 6.1 Incorporating $\alpha$ Uncertainty in the Kalman Filter Model

As before, let $\Theta_i = \log(N(t_i))$, be random variables describing the actual observed loss at each time step $t_i$, $i = 1, ..., n$. Also, let $X_1, ..., X_n$ represent the noisy reported logarithm of loss made at corresponding times. Because the decay rate is estimated with some uncertainty, it will be treated as a random variable, denoted $A$, for which $\alpha$ will represent a specific realization. Because the rate is now a random variable, which is dependent on $\Theta(t)$ through the reported loss model, it can also be updated given the reported loss $X(t)$. As such it is convenient to calculate the joint distribution of $\Theta(t)$ and $A$ given each observation. This formulation uses similar Markov assumptions as the basic model, except that it incorporates the assumes that the data $X_i$ conditioned on $\Theta_i$ is independent of $A$ as well as incorporating
the conditional dependence of each $\Theta_i$ on $A$:

$$f_{X_i|\Theta_1,...,\Theta_i,A}(x_i|\theta_1, ..., \theta_i, \alpha) = f_{X_i|\theta_i}(x_i|\theta_i),$$

$$f_{\Theta_i|\Theta_{i-1},X_1,...,X_{i-1},A}(\theta_i|\theta_{i-1}, x_1, ..., x_{i-1}, \alpha) = f_{\Theta_i|\Theta_1,...,\Theta_{i-1},A}(\theta_i|\theta_1, ..., \theta_{i-1}, \alpha),$$

for $i = 2, ..., n$. As mentioned, this model will still incorporate all uncertainties in the logarithm of total loss $\Theta_\infty$ and process noise $E_\Theta(t)$. The PAGER predicted total loss distribution will again be used as an estimation for $\Theta_\infty$ along with the model uncertainty $R_\Theta$ of $\Theta_\infty$, with variance estimated to be the sample mean of the corresponding model variances from the 20 training earthquakes.

This formulation begins by laying out the specific reported loss model with all included uncertainties, assuming the same distributional assumptions and parameter values as in the end of the previous chapter. This formulation makes the additional assumption that $R_A \sim N(0, \phi_A^2)$ and uses the linear regression of $\log(\alpha)$ on $Mw$ to estimate the mean value of $\log(A)$. Again, these assumptions are not very realistic and were crudely determined. However, the following Kalman filter model can be readily changed for different distributions and means of estimating the decay rate, particularly because this chapter’s formulation does not take advantage of any conjugate pair relationships as in the previous chapter. This is because the incorporation of the uncertainty in the decay rate breaks the linear sum of normal random variables which was used in the basic Kalman filter model.

Similar to the previous chapter, the following notation is introduced to simplify the following formulation. Let $\Theta_{ij} := \Theta_i|X_1,...,X_j$, $A_j = A|X_1,...,X_j$, and let the joint updated distribution of $\Theta(t_i)$ and $A$ given $x_1, ..., x_i$ be $f_{i|i} := f_{\Theta_i,A|X_1,...,X_i}$ and the predicted distribution for the loss at time $t_{i+1}$ as $f_{i+1|i} := f_{\Theta_{i+1},A|X_1,...,X_i}$. For any time $t$ the logarithm reported loss $\Theta(t)$ and logarithm of decay rate $A$ can be estimated as:
$\Theta(t) = \Theta_{\text{PAGER}} + R_{\Theta} + \log(1 - e^{-A_0 t}) + E_{\Theta}(t)$, 
$\log(A_0) = \log(A_{\text{reg}}) + R_A$,

where

\begin{align*}
\Theta_{\text{PAGER}} &\sim N(\mu_{\text{PAGER}}, \sigma^2_{\text{PAGER}}), \\
R_{\Theta} &\sim N(0, \phi^2_{\Theta}), \\
E_{\Theta}(t) &\sim N(0, \tau^2_{\Theta}(t)), \\
E_{R_A} &\sim N(0, \phi^2_A), \\
\log(A_{\text{reg}}) &\sim N(\beta_0 + \beta_1 Mw, \sigma^2_{A_{\text{reg}}}),
\end{align*}

(6.2)

where $\sigma^2_{A_{\text{reg}}}$ is the uncertainty in $\log(A_0)$ from the regression with $Mw$ estimated as the residual standard deviation, and $\phi^2_{\Theta}, \phi^2_A$, and $\tau^2_{\Theta}(t)$ are determined from the corresponding sample means of these parameters for all 20 training earthquakes. As mentioned, because $A$ is now a random variable, the assumption of normality of all involved variables is no longer valid, because there is no reason to assume that if $\log(A) \sim N(\beta_0 + \beta_1 Mw, \phi^2_{R_A} + \sigma^2_{A_{\text{reg}}})$, then $\log(1 - e^{-A t})$ is normally distributed. In fact, observing this distribution using Monte Carlo simulation reveals a distribution closer to a negative lognormal random variable.

Just as with the basic Kalman filter model, the first step is to describe the initial state at $t = 0$, and how to predict the first state at $t_1$. Because the reported losses are known to be zero at $t = 0$, this means that $f_{\Theta_0, A_0}(\theta_0, \alpha) = f_{A_0}(\alpha)$, and the above equation gives

$\Theta_{1|0} = \Theta_{\infty|0} + \log(1 - e^{-A_0 t}) + E_{\Theta}(t_1),$

which implies that

$f_{\Theta_1, A}(\theta_1, \alpha) = \int \theta_1 \mid A, \Theta_{\infty|0}(\theta_1 | \alpha, \Theta_{\infty}) f_{A_0}(\alpha) f_{\Theta_{\infty}}(\Theta_{\infty|0}) \, d\Theta_{\infty},$

assuming that $A_0$ and $\Theta_{\infty|0}$ are independent (which again realistically they are not). Note that conditioned on $A_0 = \alpha$ and $\Theta_{\text{PAGER}}$, the distribution of $\Theta_1$ is simply the distribution of $R_{\Theta} + E_{\Theta}(t_1)$ shifted by an amount equal to $\theta_{\text{PAGER}} + \log(1 - e^{-\alpha t_1})$. Thus,
\[
\Theta_1|[\Theta_\infty|0 = \theta_\infty, A = \alpha] \sim N\left(\mu_{\Theta_1|\Theta_\infty|0, A}, \sigma_{\Theta_1|\Theta_\infty|0, A}^2\right),
\]

where

\[
\mu_{\Theta_1|\Theta_\infty|0, A} = \theta_\infty + \log \left(1 - e^{-\alpha t_1}\right),
\]

\[
\sigma_{\Theta_1|\Theta_\infty|0, A}^2 = \phi_\Theta^2 + \tau_1^2 - 2\text{Cov}(E_{\Theta}(t_1), R_{\Theta}),
\]

although this thesis does not assume any correlation between $R_{\Theta}$ and $E_{\Theta}(t)$ for simplicity. Although it is expected that they are correlated with one another, that kind of reformulation is left for future work and is not accounted for in this thesis. The motivation of computing the joint distributions of each $\Theta_i$ and $A$ is so that $A$ can also be updated given the data. This is especially useful when the estimate of $A$ is highly uncertain due to the additional regression term as was found with the basic Kalman filter. Given the first observed logarithm of loss $x_1 \in X_1$ at $t_1$, with $X_1|[\Theta_1 = \theta_1] \sim N(\theta_1, \tau_1^2 X_1)$, the updated joint distribution of $\Theta_1$ and $A$ given $X_1 = x_1$ is given by Bayes’ equation:

\[
f_{1|1}(\theta_1, \alpha|x_1) = \frac{f_{X_1|\Theta_1, A}(x_1|\theta_1, \alpha)f_{1|0}(\theta_1, \alpha)}{C_1}
\]

\[
= \frac{f_{X_1|\Theta_1}(x_1|\theta_1)f_{1|0}(\theta_1, \alpha)}{C_1}
\]

(6.3)

where the second equality is due to the Markov assumption $f_{X_1|\Theta_1, A}(x_1|\theta_1, \alpha) = f_{X_1|\Theta_1}(x_1|\theta_1)$ and $C_1$ is a normalization constant equal to the integral of the numerator with respect to $\Theta_1$ and $A$. The probability density of the predicted joint distribution of $\Theta_2$ and $A$ given $X_1 = x_1$ is then determined as in the basic Kalman filter method using the method of total probability:
\[ f_{2|1}(\theta_2, \alpha|x_1) = \int_{\Theta_1} f_{\Theta_2|\Theta_1,A,x_1}(\theta_2|\theta_1, \alpha, x_1) f_{1|1}(\theta_1, \alpha|x_1) d\theta_1 \]

\[ = \int_{\Theta_1} f_{\Theta_2|\Theta_1,A}(\theta_2|\theta_1, \alpha) f_{1|1}(\theta_1, \alpha|x_1) d\theta_1, \]

where the second equality is made using the other Markov assumption that
\[ f_{\Theta_2|\Theta_1,A,x_1}(\theta_2|\theta_1, \alpha, x_1) = f_{\Theta_2|\Theta_1,A}(\theta_2|\theta_1, \alpha). \]

Note the presence of the previously updated posterior distribution from Equation 6.3 appearing inside the integral in Equation 6.4. Now \( \Theta_2 \) can be determined as a function of \( \Theta_1 \) and \( A \) by solving Equation 6.2 for \( \Theta_1 \) and \( \Theta_2 \), and then combining the two equations to eliminate \( \Theta_{PAGER} + \mathcal{R}_\Theta \), giving

\[ \Theta_2 = \Theta_1 + \log \left( \frac{1 - e^{-\alpha t_2}}{1 - e^{-\alpha t_1}} \right) + \mathcal{E}_\Theta(t_2) - \mathcal{E}_\Theta(t_1). \]

This means that conditioned on \( \Theta_1 \) and \( A \), then

\[ \Theta_2|[\Theta_1 = \theta, A = \alpha] \sim N \left( \mu_{\Theta_2|\Theta_1,A}, \sigma_{\Theta_2|\Theta_1,A}^2 \right), \]

where

\[ \mu_{\Theta_2|\Theta_1,A} = \theta + \log \left( \frac{1 - e^{-\alpha t_2}}{1 - e^{-\alpha t_1}} \right), \]

\[ \sigma_{\Theta_2|\Theta_1,A}^2 = \tau_{t_2}^2 + \tau_{t_1}^2 - 2\text{Cov}(\mathcal{E}_\Theta(t_1), \mathcal{E}_\Theta(t_2)). \]

The temporal correlation between the process noise \( \mathcal{E}_\Theta(t) \) for any two times is not accounted for in this work, and is assumed to be zero. This is mainly for mathematical simplicity, although it is expected that these process noises in log loss are much more correlated with the corresponding process noise in \( A \) as a function of time, which will be investigated in the next section. These process noises in \( A \) are expected to be correlated with each other over time, and deciding upon this correlation structure as well as the correlation between the process noise of \( \Theta(t) \) and \( A(t) \) will naturally force the log loss process noises to be correlated.
over time. However, precisely determining these relationships is left for future work as it is beyond the scope of this thesis and any correlation structures will be estimated or chosen for simplicity.

In addition to predicting the log loss $\Theta_2$ at time $t_2$, the model can also be used to estimate the total log loss at $t = \infty$. Following a similar approach used to estimate $\Theta_2$, $\Theta_\infty$ can be written in terms of $\Theta_1$ and $A$ as

$$
\Theta_\infty = \Theta_1 - \log \left(1 - e^{-At_1}\right) + \mathcal{E}_\Theta(\infty) - \mathcal{E}_\Theta(t_1).
$$

At $t = \infty$ no process noise is expected, as the process noise is expected to decay over time. This is because regardless of the uncertainty in the total loss reported, it will not continue to vary over time because people are no longer investigating these losses several months or years after the event. Thus, while the total loss number can be uncertain, its estimation would be expected to remain the same after a certain amount of time, implying a process noise of zero. Therefore $\mathcal{E}_\Theta(\infty)$ can be assumed to be zero, as all uncertainty in the total loss number is already contained in the random variable $\Theta_1$. Recall that the model’s uncertainty in predicting total loss using the exponential cdf was described by the regression error $\mathcal{R}_\Theta$, which is already included in $\Theta_1$. The updated total loss distribution given $x_1 \in X_1$ is found again using the law of total probability:

$$
f_{\Theta|X_1}(\theta|x_1) = \int_{\Theta_1} \int_A f_{\Theta,A}(\theta|x_1) d\alpha d\theta_1,
$$

which again uses the posterior joint distribution of $\Theta_1$ and $A$ given $X_1 = x_1$. This completes the first step of the Kalman filter, and due to the recursive relationship involved, the updating, prediction, and total loss prediction steps can be written for any $i = 1, \ldots, N$. Suppose that the Kalman filter model gives the predictions and updates for all data up to say $i$, where $1 \leq i \leq N - 1$. Using the recursive relationship of $\Theta_i$ in terms of $\Theta_{i-1}$, the following Kalman filter steps can be used to obtain the different distributions. In order to further simplify the notation let $x_j := x_1, \ldots, x_j$, $j = 1, \ldots, n$.

\textbf{Prediction Step}
\[ f_{i|i-1}(\theta_i, \alpha|x_{i-1}) = \int_{\Theta_i-1} f_{\Theta_i|\Theta_{i-1}, A}(\theta_i|\theta_{i-1}, \alpha)f_{i-1|i-1}(\theta_{i-1}, \alpha|x_{i-1})d\theta_{i-1}, \quad (6.6) \]

based off of the second Markov assumption. The updated loss estimate at time \( t_i \) given the observed \( X_i = x_i \) is given by Bayes’ equation:

**Update Step**

\[ f_{ij|i}(\theta_i, \alpha|x_i) = \frac{f_{X_i|\Theta_i}(x_i|\theta_i)f_{i|i-1}(\theta_i, \alpha|x_{i-1})}{C_i}, \quad (6.7) \]

where \( C_i \) is a normalization constant. Given this updated loss value, the updated total loss can now be predicted:

**Total Loss Predict Step**

\[ f_{\Theta_{\infty}|X_i}(\theta_{\infty}|x_i) = \int_{\Theta_i} \int_{A} f_{\Theta_{\infty}|\Theta_i, A}(\theta_{\infty}|\theta_i, \alpha)f_{ij|i}(\theta_i, \alpha|x_i)d\alpha d\theta_i. \quad (6.8) \]

for all \( i = 1, \ldots, n \). This completes the formulation for this more enhance Kalman filter total loss updating model. As mentioned, the involvement of the exponential cdf reported loss model eliminates any quick and easy updating and prediction methods, as the many involved distributions are not normal and conjugate prior relationships do not exist. As such this thesis uses numerical integration via the trapezoid quadrature rule. However, this can be time consuming and expensive for fine meshed numerical integration. Thus, the Kalman filter updating model with uncertain decay rate \( A \) is:

\[
\begin{align*}
\Theta_{1|0} &= \Theta_{\infty|0} + \log \left( 1 - e^{-A_0 t_1} \right) + \mathcal{E}_\Theta(t_1), \\
\Theta_{i|i-1} &= \Theta_{i-1|i-1} + \log \left( \frac{1 - e^{-A_{i-1} t_i}}{1 - e^{-A_{i-1} t_{i-1}}} \right) + \mathcal{E}_\Theta(t_i) - \mathcal{E}_\Theta(t_{i-1}), \quad i > 1, \\
X_i &= \Theta_{i|i-1} + \mathcal{E}_X, \\
f_{i|i}(\theta_i, \alpha|x_i) &= \frac{f_{X_i|\Theta_i}(x_i|\theta_i)f_{i|i-1}(\theta_i, \alpha|x_{i-1})}{C_i}, \\
\Theta_{\infty|i} &= \Theta_{i|i} - \log \left( 1 - e^{-A_{t_i}} \right), \quad i = 1, \ldots, n.
\end{align*}
\quad (6.9)\]

Now that this enhanced Kalman filter model has been formulated, it is time to get an idea of how it works and how the initial PAGER estimates affect the updating process. This will be done using the three test earthquakes: Gorkha, L’Aquila, and Mexico City.
6.2 Using the Generalized Kalman Filter Model on the Test Earthquakes

Similar to the basic Kalman filter model, this enhanced model will be looking at how the total loss updating works using both the best fit nonlinear regression information for each earthquake, as well as estimated parameters obtained by the $M_w$ regression and nonlinear regression information from the 20 training earthquakes. It is important to note that the additional uncertainties in $A$ and its relationship to loss via the exponential cdf will complicate the results. This is because one can no longer assume that the total losses are lognormally distributed due to the contribution from the uncertain decay rate $A$. This is similar to Chapter 3 in which the addition of the uncertainties in shaking led to results that could not be accurately fitted using a method of moments fitting to a lognormal approximation. In Chapter 3, this motivated us to find an unbiased estimate of the enhanced total loss distribution, which was found by determining the median of the logarithm of total loss. Because the median is an unbiased estimator, and the exponential transformation of a variable preserves the median value, this led to results that were more accurate than those found by simple calculating the mean and variance of total loss and using those to fit a lognormal distribution.

As mentioned, Chapter’s 3 and 4 while useful, were mainly to help lay the foundation for some of the ideas and problems that would be present in this chapter. Because of this loss of normality, one is no longer justified in simply calculating the mean and variance of the updated total loss distributions and fitting them to a lognormal distribution. As in Chapter 3, one must instead construct the entire distribution of these estimates so that median estimates can be obtained, which remain unbiased whether one is looking at the logarithm of total loss or the total loss itself. The work and results done in Chapter 4 will be used to help understand how the updating given this uncertain $\alpha$ affects the different updated distributions. Because the normal distribution is the only distribution in which its mean and variance are independent, one should expect any updating of a non normal
distribution to behave similarly to the final case in Chapter 4, in which both the mean and variance were assumed to be unknown. As the results indicated in the previous chapter, this tends the updated distribution to both shift in mean towards the data as well as increasing its tail on the side of the data in order to better capture the measurement. This is the exact kind of trend expected to be seen with updating in this section, because the distributions involved are no longer all normal. As such, this section will also be looking at the higher statistical moments of the different updated distributions in order to see how close or far they are from a normal distribution. In addition, with this loss of normality, the conjugate relationships that were able to give closed form posterior and predictive distributions are no longer present. As such these distributions must be approximated numerically, which will be done via numerical integration in this section.

In order to understand how the number of quadrature points for each quantity affect the updated distributions, a convergence study will implemented to observe how these estimates change with increasing quadrature point size. Now that it is clear how the ideas put forth in Chapters 3 and 4 will be used, it is now time to investigate this enhanced Kalman filter model in regards to the three test earthquakes. For all three test earthquakes this work makes the following distributional assumptions regarding Equation 6.9

\[
\Theta_{\infty|0} = \Theta_{PAGER} + \mathcal{R}_\Theta,
\]

\[
\log(A) = \beta_0 + \beta_1 Mw + \mathcal{R}_A + \mathcal{E}_{A_{\text{reg}}},
\]

where

\[
\Theta_{PAGER} \sim N(\mu_{PAGER}, \sigma^2_{PAGER}),
\]

\[
\mathcal{R}_\Theta \sim N(0, \phi^2_{\Theta}), \quad \mathcal{E}_\Theta(t) \sim N(0, \tau^2_{\Theta}(t)),
\]

\[
\mathcal{R}_A \sim N(0, \phi^2_{A}), \quad \mathcal{E}_{A_{\text{reg}}} \sim N(0, \sigma^2_{A_{\text{reg}}}),
\]

where \(\beta_0 + \beta_1 Mw\) and \(\sigma^2_{A_{\text{reg}}}\) are the mean and variance of \(\log(\alpha)\) from the regression using the training earthquakes, \(\phi^2_{\Theta}\) and \(\phi^2_{A}\) are the sample mean of the variances of \(\Theta_{\infty}\) and \(\log(\alpha)\) from the training earthquakes’ nonlinear regression model fits respectively, and \(\tau^2_{\Theta(t)}\) depends

142
on the log residuals from those training earthquakes. Using the training earthquakes, the
following values have been estimated for the model parameters:

\[
\begin{align*}
\beta_0 &= 2.752, \\
\beta_1 &= -0.376, \\
\phi_0 &= 0.412, \\
\phi_A &= 0.403, \\
\tau_{Base}^2 &= \tau_{Base}^2(2 \ast t)^{-1}, \\
\tau_2 &\mid \Theta = \tau_2^{\ast t} - 1, \\
\sigma_{A reg} &= 1.022,
\end{align*}
\]

As with the basic Kalman filter model, the measurement noise will be set equal to the process
noise as a stand-in.

For all three earthquakes, evaluation and numerical integration involving \(\theta(t), \theta_\infty\), and \(\alpha\)
will be performed with 320 uniformly spaced quadrature points. The ranges for \(\theta_\infty\) change
depend on the initial PAGER estimate and are set to be the PAGER mean log loss \(\mu_{PAGER}\)
plus or minus 5\(\sigma_{PAGER}\). The other discretizations are currently set to have fixed ranges
for each timestep, although a more adaptive approach would significantly reduce the com-
putation time to obtain more accurate estimates with smaller quadrature point numbers.
Presently, they are both fixed with ranges \(\theta(t) \in [-10, 12]\), which covers an exponentiated
range of \(N(t) \in [4.5 \times 10^{-5}, 1.6 \times 10^5]\), and the ranges for \(\alpha\) are set to be the median estimated
\(\alpha\) from the regression plus or minus 6\(\sigma_A\) where \(\sigma_{\log(A)} = \sqrt{\phi_A^2 + \sigma_{A reg}^2}\). The 320 quadrature
points gives reasonable results as can be seen in the numerical integration analysis performed
at the end of this chapter, although further analysis needs to be done in terms of analyzing
different quadrature sizes for different quantities. Now that the information regarding the
specific settings of integration for the problem, it is time to use this enhanced model. The
first test earthquake to look at is the Gorkha earthquake.

### 6.2.1 April 2015 magnitude 7.8 earthquake in Gorkha, Nepal

Recall that the best fit \(\alpha \approx .24\) and that the \(Mw\) regression \(\alpha \approx .72\) for Gorkha. That
is, the estimated decay rate is larger than the best fit by a factor of three. This means
that the model will assume a much faster rate than the overall average, which will tend to underestimate the total loss. The incorporation of the uncertainty in $A$, particularly due to the $M_w$ regression, should give results which better use the initial PAGER information so that even if the median loss shifts significantly, the distribution will be such to still give weight to the PAGER estimated loss. However, because only observations of loss are used as data, and there are not any measurements of this decay rate to supply, while the updated distribution of $A$ will change significantly, its overall variance will only marginally decrease with each observation. In Figure 6.1 the updated total loss at each timestep is shown for the case supposing PAGER estimated the reported final loss exactly, including its uncertainty. For clarity in how the process behaves, views are given in respect to the logarithm of loss and are shown both linearly and logarithmically in time. In order to understand how the data affects the estimates, the reported losses over time are also shown along with the green line indicating the reported final loss.

Figure 6.1: Plot of the $M_w$ regression predicted total loss for the Gorkha earthquake assuming exact PAGER estimate and its uncertainty over time using a Kalman Filter with parameters set using the information gathered from the 20 training earthquakes.

Similarly to the basic Kalman filter, the medians of the estimations are significantly affected by the first observations, which is expected because the measurement error is small in relation to the model and parameter uncertainties. However, because the rate is also being allowed to change, this does not force the median to shift as much as it did for the basic filter, because the distribution is also changing its shape (skew, kurtosis) to explain the data.
The next figure (Figure 6.2) shows a 3D plot illustrating this updating process for different PAGER estimations up to 2 orders of magnitude above and below the reported total loss. Included is another plot showing the updated variance for these updates over time.

![3D plot of predicted total loss over time and standard deviation](image1.png)

Figure 6.2: Plot of the $M_w$ regression predicted total loss of the Gorkha earthquake for different PAGER estimates (left) and its uncertainty over time (right) using a Kalman Filter with parameters estimated by the regression on the 20 training earthquakes.

The plot illustrates that there is a general trend of the median estimated loss towards the total value over time, and the estimate variance also is decreasing over time, varying only slightly for different PAGER estimations. There does appear to be discrepancy in the median value plot for different PAGER estimates, and a look at the range plot of these (Figure 6.3) illustrates the predicted median value over time for them.

![Range plot of predicted loss values](image2.png)

Figure 6.3: Plot of the range of $M_w$ regression predicted total loss for the Gorkha earthquake for different PAGER estimates using a Kalman filter with parameters set using the 20 training earthquakes.
The range plot shows that the different estimation medians differ more than they did when the decay rate was assumed to be constant. This is good because it means that the PAGER information is not all lost in the noise as it was with the basic filter. Unfortunately the initial points have a large affect on all of the estimates, mainly because the rate at these times is very different from the overall average decay rate. Additionally the first point will have adverse affects on the median estimates due to the time lag issue. While some lag was assumed, which tends decrease the model overestimation at the first point, it was simply set to be a 1 hour shift in this work. Ideally this factor should depend on the time of day of the event as well as numerous other factors to determine at what point the exponential process truly begins. Note that the median estimates are still mostly in agreement by about the end of the first day, which is also a desired trait for this model. As mentioned, this total loss is no longer expected to be normally distributed, and to illustrate this, Figure 6.4 shows the total loss skew (left plot) and kurtosis (right plot). Recall that a normal distribution has zero skew and kurtosis of 3.

![Figure 6.4: Plots of the skew and kurtosis of the Mw regression predicted total loss for the Gorkha earthquake for different PAGER estimates using a Kalman filter with parameters set using the 20 training earthquakes.](image)

These plots show that the total loss distributions are far from being normally distributed, due to the large positive skews and kurtosis’ seen in the previous figure. Notice that the peak of the skew and kurtosis falls around the point where the saturation takes in, and as subsequent reports differ less of larger amounts of time, both the skew and kurtosis approach
the values for a normal distribution. This is expected because the most uncertainty in the model from the nonlinear regressions tended to occur at this saturation level. These plots are the motivation behind seeking the median logarithm of total loss, rather than its mean, to estimate the total loss, ensuring that this too is a median estimate. As such, the ability to update $A$ allows for the distribution to skew significantly in the direction of the data (or prior information) while changing the median value relatively less than for the basic Kalman filter model.

Next the updated distributions of the decay rate $A$ are investigated. Similarly to the total loss results, the first plot (Figure 6.5) shows the update median of $A$ at each timestep, with its uncertainty, compared to the best fit value obtained using the nonlinear regression, assuming the PAGER estimated the reported total loss exactly.

![Updated a rate Over Time Assumming Correct PAGER Estimated Loss Value](image1)

Figure 6.5: Plot of the $M_w$ regression predicted decay rate $A$ for the Gorkha earthquake assuming exact PAGER estimate and its uncertainty over time using a Kalman Filter with parameters set using the information gathered from the 20 training earthquakes.

The results indicate that the median value is updating in the direction of the best fit parameter, although it is doing so quite slowly. As mentioned, because there is not rate data, the mean shouldn’t be changing dramatically because the data has the largest affect on the log loss. Looking at different PAGER estimates, Figure 6.6 shows how alpha updates given different estimated PAGER losses, both in terms of its median (left) and log standard deviation (right).
The median plot shows that the direction in which the median updates is inversely related to the estimated PAGER loss. This is realistic, because if PAGER overestimates the loss, and the first observation is much lower than estimated, then it will decrease the total loss estimate while simultaneously decreasing its rate, which will estimate a slower accumulation of loss over time. Conversely, for underestimations of total loss, the reported value increases the median decay rate so that the reported loss is expected to grow more rapidly than originally assumed. Regardless of the estimate, there is a decreasing trend over time for all estimates in the direction of the best fit $\alpha$ value. The uncertainty in $A$ decreases slowly over time, with only a 10% change in its the standard deviation of $\log(A)$ over the entire data set. This uncertainty does depend on the PAGER estimate, and for any given time, the standard deviation at the larger PAGER estimates is greater than that for the smaller estimates.

Next, in order to understand how these updated total loss distributions and different initial loss estimates affect the PAGER total loss histograms, a series of PAGER histograms is shown below in Figure 6.7 for different estimated PAGER losses and different times. The time-steps shown correspond to the initial PAGER estimate, the first reported loss, the reported loss closest in time to 1 day, and the final estimated loss after all reported losses.
Figure 6.7: PAGER total loss histograms of the $M_w$ regression updated total loss for the Gorkha earthquake assuming exact different initial loss estimates computed using a Kalman Filter with parameters set using the information gathered from the 20 training earthquakes.

The results indicate that the PAGER histogram and the PAGER alert level are converging to their true values over time, and at about 1 day (3rd row), they are all estimating the correct ‘red’ alert level. These results indicate that the loss estimates are achieving good results rather quickly for this earthquake, while still carrying the initial PAGER information for longer amounts of time. This indicates that the model is combining these two pieces of information (PAGER estimate and reported loss function) as desired.

In order to see how the Kalman filter model performs when the assumed uncertainties in $A$ and the regression error $R_\Theta$ are found from the nonlinear regression of the entire reported loss data set for the Gorkha earthquake. In Figure 6.8 below, the updated median loss is shown for the case where PAGER was assumed to have estimated the final reported loss exactly.
Figure 6.8: Plot of the NLR best fit predicted total loss for the Gorkha earthquake assuming exact PAGER estimate and its uncertainty over time using a Kalman Filter with parameters set using a nonlinear regression on the full reported loss data set.

Figure 6.8 shows a very stable median loss estimate, which stays close to the final reported loss except for two places where the rate is clearly different from the best fit value. In this light, this can be seen as an actual change in rate found after filtering out the various measurement and process noises, and could be useful in determining the stochastic process of the decay rate over time.

This concludes the results for the Gorkha earthquake. Again, the updating appears to be doing what it should be, in the time frame of 1 day desired. However, deviations from the best fit decay rate had pronounced affects on the first updates, particularly during the first day. These sizable variability in decay rates during the first hours after the event is present in all of the earthquake data and seems a natural hazard to the reported loss scheme. In fact, in [1] they mentioned that when using the NLR regression relationship to estimate total loss, they would ignore the first few data points because the rate variability during these times was so large. However, because these are the only data points that are truly of use to the PAGER updating framework, they must be used, and better describing the variability in decay rate during these times should stabilize the model further, although that is left for future work and involves including the time varying process noise in $A$ (or $\log(A)$ rather). This decay rate process noise would not generally be a zero-mean process, as some earthquakes show steady decreases in the decay rate over time. This and other improvements
will be described in the final conclusions chapter, but first the remaining earthquakes and quadrature integration convergence must be investigated. The next two earthquakes will not show all of the figures that were used for the Gorkha case, as some of them don’t exhibit any new or interesting features compared to the Gorkha results.

6.2.2 April 2009 magnitude 6.3 earthquake in L’Aquila, Italy

Recall that the best fit $\alpha \approx .988$ and that the $Mw$ regression $\alpha \approx 1.47$ for L’Aquila. This is similar to the Gorkha earthquake, except the $Mw$ decay rate is only about 50% larger as opposed to three times larger. In the basic model, this tended to underestimate the total losses just like in the Gorkha case.

In Figure 6.9 the updated total loss at each time-step is shown for the case supposing PAGER estimated the median reported final loss exactly.

![Figure 6.9: Plot of the $Mw$ regression predicted total loss for the L’Aquila earthquake assuming exact PAGER estimate and its uncertainty over time using a Kalman Filter with parameters set using the information gathered from the 20 training earthquakes.](image)

The updated losses in Figure 6.9 do not suffer from the underestimation that the basic model did for this earthquake, in contrast to how it behaved for the Gorkha earthquake. this is because the beginning losses do not have the wide variability in decay rates in the first few hours, which leads to smaller changes in rate and median. The range plot of the median updated total loss estimates given different magnitudes of initial estimates can be
seen in Figure 6.10.

Figure 6.10: Plot of the range of $M_w$ regression predicted total loss for the L’Aquila earthquake for different PAGER estimates using a Kalman filter with parameters set using the 20 training earthquakes.

The range plot shows that the different estimation medians differ more than they did when the decay rate was assumed to be constant. The wider variability in these median estimates over time compared to the Gorkha earthquake is a result of having much fewer reports to update from. This makes sense because the more data available, the stronger its influence compared to the prior PAGER information. Note that the median estimates are still relatively close by about the end of the first day as desired.

Similarly to the Gorkha results, the peak of the skew and kurtosis falls around the point where the saturation takes in and did not exhibit any new and interesting features compared to the Gorkha kurtosis and skew figures.

Next the updated distributions of the decay rate $A$ are investigated. Similarly to the Gorkha results, the first plot (Figure 6.11) shows the update median of $A$ at each time-step, with its uncertainty, compared to the best fit value obtained using the nonlinear regression, assuming the PAGER estimated the reported total loss exactly in both the semilog and log-log views.

The results indicate that the median value is updating in the direction of the best fit parameter, although it is doing so quite ever slower than with the Gorkha results. This is because there is a smaller difference between the $M_w$ regression rate and the best fit rate.
Figure 6.11: Plot of the $M_w$ regression predicted decay rate $A$ for the L’Aquila earthquake assuming exact PAGER estimate and its uncertainty over time using a Kalman Filter with parameters set using the information gathered from the 20 training earthquakes.

While the median isn’t shifting as much the standard deviation is decreasing much faster than with the Gorkha data. This suggests that when the rate is farther away, it will tend to shift the median estimate more as well as increase the kurtosis, while for closer rates the variance is more affected.

The features of these plots are the same as were found in the Gorkha example, except on different scales. Because the median decay rate is larger for this earthquake than it was for the Gorkha case, the L’Aquila rate plots have a much larger range of updated values, which is a feature of assuming a lognormally distributed rate. Again the standard deviation of log rate indicates that the uncertainty is decreasing about twice as fast as it was with the Gorkha example.

Next, in order to understand how these updated total loss distributions and different initial loss estimates affect the PAGER total loss histograms, a series of PAGER histograms is shown below in Figure 6.12 for different estimated PAGER losses and different times. The time-steps shown again correspond to the initial PAGER estimate, the first reported loss, the reported loss closest in time to 1 day, and the final estimated loss after all reported losses have been obtained.

Just as with the Gorkha example, the results indicate that the PAGER histogram and the PAGER alert level are converging to their true values over time, and at about 1 day (3rd
Figure 6.12: PAGER total loss histograms of the $M_w$ regression updated total loss for the L’Aquila earthquake assuming exact different initial loss estimates computed using a Kalman Filter with parameters set using the information gathered from the 20 training earthquakes.

(row), they are all estimating the correct ‘orange’ alert level with increased precision. These results indicate that the loss estimates are achieving good results rather quickly for this earthquake as well, while still carrying the initial PAGER information for longer amounts of time, again indicating that the model is combining these two pieces of information (PAGER estimate and reported loss function) as desired.

Similarly to the Gorkha example, the nonlinear regression model estimates for the L’Aquila earthquake will also be run through the Kalman filter in order to see how updating occurs when the rate is guessed correctly with larger precision. In Figure 6.13 below, the updated median loss is shown for the case where PAGER was assumed to have estimated the final reported loss exactly.

Figure 6.13 shows a very stable median loss estimate, even more than with the Gorkha example actually, which stays close to the final reported loss except for one or two places where the rate is clearly different from the best fit value and can easily be explained by
looking at the peaks or dips in the data. While not stated with the Gorkha earthquake, notice that both of these best fit plots do not have dramatically smaller variances than their $M_w$ regression counterparts. This is because the process noises and measurement noises dominate the updated uncertainties, and the updated uncertainties cannot be significantly smaller than them at any time. Finally, it is time to look at the last test earthquake, the 2017 Mexico City earthquake.

### 6.2.3 September 2017 magnitude 7.1 earthquake in Mexico City, Mexico

Recall that the best fit $\alpha \approx 2.76$ and that the $M_w$ regression $\alpha \approx 1.09$ for Mexico City earthquake. This example was one in which the $M_w$ decay rate was much smaller than the best fit. However, this earthquake’s data showed that while the best fit $\alpha$ best described the average behavior, the actual behavior appeared to have a decreasing rate over time. Looking at the data, this large rate mainly describes the initial growth, and not the final saturation, which is at a slower rate comparable to the $M_w$ regression estimation for this earthquake. Therefore, it is expected that the updated total losses will over predict in the beginning around where the decay rate starts noticeably decreasing. However, when the rate is closer to the regression estimate, the total losses should be much closer to the reported final loss.
In Figure 6.14 the updated total loss at each time-step is shown for the case supposing PAGER estimated the reported final loss exactly, along with its uncertainty.

Figure 6.14: Plot of the $M_w$ regression predicted total loss for the Mexico City earthquake assuming exact PAGER estimate and its uncertainty over time using a Kalman Filter with parameters set using the information gathered from the 20 training earthquakes.

These results are what were expected, as the model shows some difficult adapting to the steadily decreasing decay rate. This is because the data grows so much faster than initially estimated that the model updates the decay rate too much. Then when the rate saturates, which it does much more sharply than the other test earthquakes, the rate takes longer to update back towards its original value. Figure 6.15 shows the predicted median value over time for given different initial PAGER loss estimates.

Figure 6.15: Plot of the range of $M_w$ regression predicted total loss for the Mexico City earthquake for different PAGER estimates using a Kalman filter with parameters set using the 20 training earthquakes.
Just as with the previous earthquakes, the enhanced Kalman filter model does a much better job of keeping the PAGER information throughout the updating process, and the smaller number of data points makes these differences pronounced even up to the final estimation. Unlike the other two earthquakes, not all of the updated medians are within close range of the true level after one day, although only the overestimated PAGER losses. This is because the decay rate initially was so fast that the larger PAGER estimated losses were updated less because the data agreed with their estimates. However, as the decay rate began to decrease, these overestimates were not able to shift enough towards the actual value given the number of data points. If more data was available, this would most likely not be the case. However, this example illustrates that for earthquakes with decay rates which have a trend of change over time, the Kalman filter model will not perform as well if the PAGER estimate happens to agree with the initial decay rate but not the final decay rate.

Similarly to the total loss results, the first plot (Figure 6.16) shows the update median of \( A \) at each time-step, with its uncertainty, compared to the best fit value obtained using the nonlinear regression, assuming the PAGER estimated the reported total loss exactly.

![Updated \( A \) rate Over Time Assuming Correct PAGER Estimated Loss Value](image1)

![Log Log Plot of Updated \( A \) Over Time Assuming Correct PAGER Estimated Loss Value](image2)

Figure 6.16: Plot of the \( M_w \) regression predicted decay rate \( A \) for the Mexico City earthquake assuming exact PAGER estimate and its uncertainty over time using a Kalman Filter with parameters set using the information gathered from the 20 training earthquakes.

The results indicate that the median value is updating in the direction of the best fit parameter, except more slowly in the beginning that at the end. Again these plots have the same features as were found with the other test earthquakes, but with an even wider range.
in the decay rates the PAGER estimates are updating towards.

Next, in order to understand how these updated total loss distributions and different initial loss estimates affect the PAGER total loss histograms, a series of PAGER histograms is shown below in Figure 6.17 for different estimated PAGER losses and different times.

![Figure 6.17: PAGER total loss histograms of the $M_w$ regression updated total loss for the Mexico City earthquake assuming exact different initial loss estimates computed using a Kalman Filter with parameters set using the information gathered from the 20 training earthquakes.](image)

The results indicate that the PAGER histogram and the PAGER alert level are converging to their true values over time, and at about 1 day (3rd row) for only the underestimation and correct estimation. The overestimation, as noted earlier, has more difficult obtaining the same results because it happened to overestimate the losses for an earthquake which happened to have a smoothly decaying decay rate. This is the first major limitation of this model, and is a perfect example of why assuming a constant (but unknown) decay rate will not work well for earthquakes which do not have a constant average rate over time. Again, by including process noise in the decay rate $A$, this could help, however this steadily changing rate complicates the issue as it is not a constant mean process. The kinds of modifications
left to rectify problems such as this are left for future work, although some explanation as to achieving results will be discussed in the conclusion chapter.

Similarly to the previous earthquakes, the nonlinear regression model estimates for the Mexico City earthquake will also be run through the Kalman filter in order to see how updating occurs when the rate is guessed correctly with larger precision. In Figure 6.18 below, the updated median loss is shown for the case where PAGER was assumed to have estimated the final reported loss exactly.

Figure 6.18: Plot of the NLR best fit predicted total loss for the Mexico City earthquake assuming exact PAGER estimate and its uncertainty over time using a Kalman Filter with parameters set using a nonlinear regression on the full reported loss data set.

Figure 6.18 shows a very stable median loss estimate, even moreso than the previous two earthquakes albeit for the peak in decay rate initially.

The results from three test earthquakes illustrate the model’s ability to incorporate reported loss data (however uncertain) into an exponential cdf model describing how loss accumulates (however uncertain) with parameters found studying past earthquake loss data and using the PAGER estimated total loss, while incorporating the uncertainties associated with these estimates. One last bit of analysis needs to be done in regards to this loss updating problem, which has to do with the reliability of results produced using numerical integration.
6.3 Analysis of Quadrature Size in Numerical Integration

The number of points in which the different independent variables were discretized into has an affect on the accuracy of the estimated distributions and quantities. Because all distributions mentioned were for continuous random variables, explaining their behavior using any discretization will necessarily accrue errors in the truncation process. For sufficiently smooth functions, this error will decrease as the discretizations get more and more fine. In the enhanced Kalman filter model, numerical integration is used to calculate the predicted and updated distributions with respect to the variables $\theta(t)$, $\theta_{\infty}$, and $\alpha$, which represent realizations of their respective random variable quantities. The model discretizes these variables into finite vectors of what this paper has been calling quadrature points, the length of these vectors being called the quadrature number. The goal of this section is to analyze how the enhanced Kalman filter behaves given a set quadrature size for all three variables. That is, this assumes some $M$ quadrature points to describe each variable $\theta(t)$, $\theta_{\infty}$, and $\alpha$. These vectors of $M$ sized points can then be used for integration or evaluation of the posterior update and predictive distribution at a time $t$.

The idea is that if $M$ increases, the error in the quantities that use these variables will decrease in some manner depending on the variable and the norm with which the error is described, such as $L^1$, $L^2$, or $L^\infty$. After first verifying that the error is decreasing with increased quadrature number, this section then seeks to quantify this convergence giving an estimated rate for each time-step. In order to describe error for a quantity which is unknown, a high resolution solution with a large quadrature size is calculated, and subsequently lower quadrature numbered results are compared to this as though it was the actual answer. If the error is indeed converging, then comparing any solution to the fine mesh solution will be within some error range of the actual solution.

The fine mesh Kalman filter results were obtained using a quadrature size of 640 for $\theta(t)$, $\theta_{\infty}$, and $\alpha$, which represents the limits of calculation doing brute force matrix formulation and functional evaluation on all 640x640x640 total quadrature points, at least in a reasonable
amount of time without reducing everything to loops. This convergence analysis begins with \( M = 20 \) and runs the enhanced Kalman filter process for consecutively doubling quadrature sizes, so \( M = 20, 40, 80, 160, 320 \), the last of which is the quadrature size of the Kalman filter results described in the previous section. It should be noted that keeping the same quadrature size for all three variables is not the best choice, as some variables are used much more, or are relatively smooth or complex, requiring respectively fewer or more quadrature points for more accurate calculations. The relationship between these variables and their corresponding quadrature sizes is not covered in this work, and is left for future work to be done regarding this problem, the full amount of which will be discussed in the final chapter. Additionally there are relationships and dependencies depending on the range of values that these discretized vectors cover, and their values also affect both the error and the rate it converges, which also must be left for future work, specifically the adaptive mesh sizes touched upon earlier in regards to \( \theta(t) \).

This work looks at this error convergence process using the Mexico City earthquake and its reported loss data and \( Mw \) regression model. For each quadrature number \( M = 20, 40, ..., 320 \), the Kalman filter model is run and the updated log loss density \( f_{\Theta_{\infty}|Data} (\theta_f^{(M)} | x_1, ..., x_i) \), the log loss cdf \( F_{\Theta_{\infty}|Data} (\theta_f^{(M)} | x_1, ..., x_i) \), and the decay rate probability density \( f_A|Data (\alpha^{(M)} | x_1, ..., x_i) \), for each \( i = 1, ..., N \), where \( \theta_f^{(M)} \) and \( \alpha^{(M)} \) are the vectors corresponding to the continuous variables \( \theta_{\infty} \) and \( \alpha \) and \( M \) is the size of the vector. For each \( M \), the absolute difference between corresponding quadrature points of the \( M \) quadrature number and the fine 640 quadrature number are found for each of these functions, leading to vectors of size \( M \) describing the estimated error in the solutions. The functional norms \( L^1 \), \( L^2 \), and \( L^\infty \) of these functions are then calculated and compared both in feature and the rate at which these errors converge with increasing quadrature number \( M \).

In Figure 6.19, Figure 6.20, and Figure 6.21 the errors in the pdf and cdf of total loss and pdf of decay rate \( A \) for the Mexico City earthquake are compared for each data point \( x_i \) and \( M \), each plot describing a different functional norm.
Figure 6.19: Plots of the error convergence of the updated total loss pdf for different quadrature integration sizes and using different functional norms. Data used is from the Mexico City earthquake.

Figure 6.20: Plots of the error convergence of the updated total loss cdf for different quadrature integration sizes and using different functional norms. Data used is from the Mexico City earthquake.

Figure 6.21: Plots of the error convergence of the updated decay rate $A$ pdf for different quadrature integration sizes and using different functional norms. Data used is from the Mexico City earthquake.
In all three functions, the error can be seen decreasing with increased quadrature size $M$ for all three norms. The smoothest of these is $A$, which seems to be decreasing at an increasing rate over time. However, the error looks to be increasing as the number of data updates increases, and appears to be growing exponentially for the smallest quadrature sizes, which just cannot capture the small variances of the final updates. The other two equations have some these same features but on a more dramatics scale because they are functions of $A$. The more slight increased error in time-step for the pdf of $A$ leads to large errors in the final time-steps of both the pdf and cdf of $N_{\infty}$. This is good because it shows that given large enough quadrature numbers, the error should be able to be predicted.

These predictions would require an estimate of what rate these errors are decreasing with each doubling quadrature size, called the estimated order of convergence (EOC). This can be estimated (for increasing $M$ of factor 2) by taking

$$ (EOC)_{g,L}^{(M_k)} \approx \frac{1}{\log 2} \log \left( \frac{|err_g(M_{k+1})|_L}{|err_g(M_k)|_L} \right), \quad M_k = 20, 40, 80, 160, 320, \quad (6.11) $$

where $k = 1, 2, 3, 4, 5$, $g$ is some function such as $f_{N_{\infty}}$, $F_{N_{\infty}}$, or $f_A$, $| \cdot |_L$ is a functional norm with respect to the function space $L = L^1, L^2, \text{or } L^\infty$, and $err_g(\cdot)$ is the error of function $g$ with respect to the fine mesh solution. An EOC of 1 means that every time the quadrature size doubles, the error halves, while an EOC of 2 means that the error quarters each time the quadrature size doubles. Figure 6.22, Figure 6.23, Figure 6.24 show plots of the EOC for each the error of each function of interest for different functional norms. Each plot is a 3D plot describing the EOC at each quadrature size $M$ and time-step $t_i, \quad i = 1, ..., N$.

Note that ignoring the final data updates on the lowest quadrature numbers, the EOC for all of these plots is between one and two for all of the functions and norms, and appears to be linearly increasing with quadrature number. The final data updates on the lowest quadrature numbers are so large because these quadrature sizes are not large enough to capture any of the actual behavior of the different functions and because the variance is so small for these points. As mentioned early a simple adaptive mesh on the $\theta(t)$ discretization
Figure 6.22: Plots of the estimated order of convergence (EOC) of the updated total loss pdf for different quadrature integration sizes and using different functional norms. Data used is from the Mexico City earthquake.

Figure 6.23: Plots of the estimated order of convergence (EOC) of the updated decay rate $A$ cdf for different quadrature integration sizes and using different functional norms. Data used is from the Mexico City earthquake.

Figure 6.24: Plots of the estimated order of convergence (EOC) of the updated total loss pdf for different quadrature integration sizes and using different functional norms. Data used is from the Mexico City earthquake.
which depended on $t$ would eliminate this issue, as well as allow for answers with similar error convergence with less quadrature points, the same can be said of $\alpha$. Similarly, choosing different quadratures yield different convergence rates. Rather than a uniform quadrature, one could use a Gaussian quadrature, which is like a uniform quadrature that gets stretched out exponentially on both sides right down the middle of the domain, or perhaps Simpson’s quadrature or many other different ones. Nonetheless, the current trapezoidal quadrature integration used in this thesis provides results who’s rates of convergence increase as the combined quadrature size of $\theta(t)$, $\theta_\infty$, and $\alpha$ increase. Finally, Figure 6.25 shows the amount of time taken to perform each quadrature size $M$ in order to illustrate the computational expense with each quadrature size doubling.

Figure 6.25: Plot showing the time (in seconds) it took to perform the Kalman updating filter on the Mexico City loss data set consisting of 11 points for each doubling quadrature size $M$.

Here one can see that doubling the quadrature of all three variables is very expensive each iteration. This plot shows that the computation cost with each doubling of quadrature point size is exponentially increasing with increased $M$. This motivates the search for more computationally efficient methods for estimating high precision discretizations of the Kalman filter model, particularly if a large amount of reported loss data is present.

This concludes the results and research conducted for this chapter as well as this thesis. The final chapter is designed to wrap everything up, while also highlighting the limitations,
adjustments, improvements, and future work related to this problem, before touching on the impact that this Kalman filter model (specifically the enhanced version) could have in terms of providing more accurate and precise estimates of total loss over time, which also depend on the initial PAGER loss distribution for that earthquake.
CHAPTER 7
CONCLUSIONS, IMPROVEMENTS, AND FUTURE WORK

This thesis has developed and demonstrated different methods for incorporating additional information into the preexisting PAGER loss prediction model in order to obtain estimations of the total loss in a manner has not yet been done. The general framework of improved PAGER estimates was carefully constructed using probabilistic theory regarding conditional probability, Bayes theorem, and characteristics of both linear and nonlinear functions of independent and dependent random variables. It has also illustrated that care needs to be taken when trying to produce unbiased (median) estimates of total loss once the distribution loses its lognormality assumption and illustrates different techniques for achieving this throughout. In addition, it provided useful means of incorporating additional levels of uncertainty into preexisting methodologies, such as allowing for the fatality rate uncertainty to depend on an uncertain shaking intensity, or including uncertainty in the prior parameters in a Bayesian framework for updating prior distributions given noisy data measurements. There are numerous advantages and potential impacts from these original methods, as well as issues, challenges, and improvements that need to be further investigated before this framework can be of practical use in the field of earthquake loss estimation.

Chapter 3 sought to enhance the PAGER model by accounting for the uncertainties in the shaking intensity, population, and fatality rate data it uses to estimate loss. This chapter also illustrated the need to develop different methods for obtaining accurate estimations of the median loss value and PAGER histogram using more robust techniques than just calculating and matching moments to different distributions. It also highlighted the fact that the lognormal nature of shaking intensity and fatality rate mean that small changes in their values can lead to orders of magnitude difference in the estimation compared to the actual number of fatalities.
By incorporating these uncertainties into the PAGER model for estimating a country’s fatality rates, it might reduce the error in the residuals it uses to determine the $\theta$, and $\beta$ parameters, which in turn could reduce the loss distribution uncertainty parameter $\zeta$. However, this would require accurate estimates of population and shaking intensity for all past earthquakes used to train the model. In general, one would expect older earthquakes to have less reliable data, meaning they would have larger uncertainties in the population, and shaking, and would contribute less to the fatality rate parameter regression. This is actually beneficial because for many countries have dramatically improved their building codes, materials, and designs over time to be able to sustain larger amounts of shaking, leading to dramatically smaller fatality rates. In addition, this information can be used to test the reliability of the country model parameters for an earthquake when PAGER greatly under- or overestimates the total losses. Because the additional uncertainties included, the model has more power to be able to explain these outlying estimates by making them more probable (wider variance).

However, this model also has many assumptions and issues that need to be addressed. The main assumption made in this section, and throughout the thesis, is the normality or lognormality assumption for many of the random variables involved. Some of these distributional assumptions were made based on prior information or expert opinion, others were made only for simplicity in demonstrating the various methods. For instance, there is no reason to believe that the population should be lognormally distributed for a given subregion, in fact it would more likely be much more centered around its mean. That is, one would expect the mass of its distribution to be on the order of the median estimated value, rather than the lognormal assumption that it is spread across multiple orders of magnitude. However, the results indicated that the population uncertainty contributed much less to the distribution than did the shaking intensity, due to the logarithmic behavior of the fatality rate equation. Thus, one would expect an even smaller contribution from the population uncertainty if it was considered to be normally distributed.
Another assumption that was made is that the shaking intensity dependent fatality rate error can be reasonably approximated without including the shaking intensity, without dramatically changing the total loss estimate. The motivation behind this was that the distribution off loss for a particular subregion showed marginal differences between the two fatality rate error distributions. The problem is that there are 100’s of thousands of these subregions for any given earthquake, and one should not assume that the sum off all these distribution errors is still insignificant. One way to determine this would be to use the actual fatality rate uncertainty in the Monte Carlo scheme. However, this had to be done by drawing realizations of fatality rate for each subregion using the Markov chain Monte Carlo method. The problem is that the MCMC requires optimizing its search parameters to ensure that the target distribution is being well covered without throwing away to many candidate realizations. This can be done by hand for one subregion, but running the MCMC method across all subregions would require automating this optimization process and is left for future work.

In addition to these assumptions, many of the variances and relationships in Chapter 3’s enhanced model were unknown and unavailable, such as the population uncertainty. In addition, this model did not account for the spatial correlation in population, shaking intensity, and fatality rate error, which is not realistic. Future work on these notes would need to investigate the role of of these subregional correlations. Another possible improvement is investigating the effect that the time of day has on these uncertainties, particularly in the population distribution, could lead to more accurate loss estimates. Because the method in Chapter 3 was for mainly for demonstrative purposes used to build an understanding of concepts that would be covered in Chapters 5 and 6, the issues and enhancements for this method, as well as testing it on a diverse sample of past earthquake data, is left for future work.

Chapter 4 was the shortest of the principle chapters and as stated was more to demonstrate how incorporating uncertainty in the prior parameters themselves affected the up-
dating process. This chapter did not produce any new ideas or concepts related to this problem, however it did illustrate how to use this process with the total loss distribution given some noisy total loss observation. The techniques and analysis in this chapter helped set the stage for how additional prior uncertainties in the decay rate and total loss value affect the Kalman filter updating model, which was shown in Chapter 6. Again this chapter made many distributional assumptions as well as assumptions about the parameter values for these distributions. Future work regarding Chapter 4’s model would require studying the updating process for a variety of possible measurement error distributions, prior distributions, and prior parameter distributions in order to understand how the results change when the conjugate pair assumption is no longer being made. In addition, in order for the methods presented in Chapter 4 to be used for an actual earthquake, the uncertainty in the observed total loss and possible uncertainties in the PAGER loss parameters needs to be better estimated and quantified before the process can be expected to give reliable results. However, earthquakes are not a repeatable random variable, and so this reliability of the model would be difficult to test without developing a suitable means of simulating the earthquake data, which also applies to the model created in Chapter 3. As stated earlier, being able to update the initial PAGER loss distribution (or the enhanced version from Chapter 3) would provide a statistically viable way of determining how reliable the prior PAGER information was. For instance, if reported loss was much larger than the PAGER median estimate, traditional maximum likelihood estimation would predict how probable the reported loss was given the PAGER information. If the reported loss was deemed statistically improbable, it would indicate that the PAGER distribution was off. However, when the reported loss is itself uncertain, this must be handled in the Bayesian setting, which ends up increasing the probability of such an outcome. Additionally, having incorporated the shaking intensity, etc., a major underestimation in the total loss could point to the shaking intensity, fatality rates, or even population being larger than initially estimated, which could lead to improving those models as well.
As stated throughout, the main goal of this thesis was to develop a means of updating the total loss estimates given reports of partial losses over time in a Bayesian setting. As such the bulk of the content herein was left for Chapters 5 and 6, which developed and explored this model. This enhancement to the PAGER model has not been done before, and although it uses many well know methods such as Kalman filters and the rate loss accumulates [1], its enhancements and adaptations to these models help mediate some of their limitations.

For instance, traditional Kalman filtering is performed on linear functions of normal random variables, in which everything can be reduced to solving for the mean vectors and covariance functions of multivariate normal random variables. However, in this paper the additional uncertainty the decay rate violated this normality assumption.

The results in Chapter 5 and 6 showed that the incorporation of the PAGER model, the reported loss model, and the uncertainties contained within them required a more robust estimation of the distribution than simply determining the statistical moments and fitting them to a candidate distribution. One key finding was that even using a similar approach to Chapter 3, wherein the mean of the logarithm of loss was calculated and exponentiated to estimate the median total loss value. This worked reasonably well for the loss distribution in Chapter 3 because it was somewhat close to being normally distributed, which was evident in the closeness of the median matching and mean matching approximations. However, when the logarithm of loss is far from being normally distributed, as was found in Chapter 6, then exponentiating its mean no longer provides an unbiased estimate of total loss. This is because for an asymmetric continuous distribution, the median and mean are not equivalent, and only the median value maps to the median value under the exponential transformation of variables. Thus exponentiating the mean of the log loss distribution would not produce an unbiased estimate of the actual loss, particularly if the log loss was highly skewed. However, exponentiating the median of the log loss will give the median of the total loss, which again, is an unbiased estimator.
The recursive Bayesian updating model made the most distributional assumptions, assuming normality for every random variable involved in calculation. While this is reasonable for many of these variables, such as the decay log decay rate and the process noise, the measurement noise, which is the uncertainty in the logarithm of reported losses given the actual losses that were observed at the time, would most likely not be lognormally distributed. There is no reason to suspect that these errors should be lognormally distributed and are most likely closer to a normal distribution in that their spread is concentrated on the same order of magnitude as the observed loss. In addition, the assumptions for the the evolution of the process and measurement noises over time was chosen based off of loose conjecture, noting throughout this piece that proper care and further evaluation and analysis is required to fully understand these two temporal processes, especially their correlation structure over time as well as with each other and any other variables such as $\Theta_\infty$ or $A$.

One of the main benefits in incorporating uncertainty in the decay rate when updating total loss was the ability to adapt to changes on the decay rate over time, as was seen with the Mexico City data. This ability to detect different rates and to project from them could also be of use in updating earthquakes which have major after shocks, as well as those with tsunami related deaths. While the initial estimate doesn’t account for these estimates, the reported loss model will be able to use them regardless. Another benefit of this uncertainty in decay rate was that the updated distributions were more dependent on the PAGER estimate, because they were able to change the estimated decay rate as well.

There is a great deal more future work to do regarding these problems, specifically it has been mentioned that incorporating some random temporal process to the decay rate as well as the reported losses would help better explain the high degree of variability in early loss reports. Perhaps the most important piece of future work would be determining a much better way of estimating the $\alpha$ decay rate for any given earthquake. The residual plot from the linear regression of $\log(\alpha)$ with $Mw$ showed only slight correlations, and more possible regressors and relationships need to be investigated in hopes of finding a more reliable way of
determining this rate. Similarly to the PAGER fatality rate model, it is reasonable to assume that this rate should be somewhat dependent on the region it occurred. That is, for places like southern California, this decay rate would be expected to be large, with the majority of loss being reported in a short amount of time, while for places with high density and less efficient recovery capabilities this rate should be expected to be substantially less. In addition, this rate should depend on the amount of help coming from outside, such as international aid. Many countries worldwide use the PAGER estimates to inform the necessary parties of when international aid is recommended. As such, there could be a loose correlation between this decay rate and the PAGER alert level. For instance, the 2017 Mexico City earthquake mentioned throughout the last two chapters had an actual PAGER estimation of over 1000 fatalities, while the actual amount was substantially less. The decay rate for Mexico City was noticed to be very large to begin with, and then decreased over time. It would be interesting to investigate if international aid was provided for this earthquake initially, and then tapered off after a few days when it was clear the losses would be well below 1000.

Another task would be to try and find a model which can describe the yellow earthquakes as well, which as mentioned, tend to grow at similar exponential rates initially, except they plateau in a very short amount of time, possibly instantly. This type of behavior could be modeled using some higher parameter reported loss model, although predicting each parameter for a future earthquake would become even more difficult than for the decay rate, assuming that improved regressors were found to predict it with smaller uncertainty.

This updating model has the potential to be of great use in trying to improve the PAGER loss estimates in a relatively short time following the event. With proper care and expert analysis on the particular distributions of the variables involved, it may be able to provide loss estimates with a much higher precision than is currently given by the PAGER model. That is, within a few days of the event, it may be able to provide more details about the value of the median estimated loss, which currently is not shared because it misrepresents the power of the PAGER estimate to be applicable for a wide variety of events, performing
quite well despite not taking the actual built environment characteristics of the affected region into account. Unfortunately this is a problem with scarce data and what is available has large uncertainties. Fortunately, the PAGER fatality rate model has laid the groundwork with its various country models which perform quite well. Perhaps finding a stronger correlation between the reported loss rate and information from the country model, a more well-grounded framework can be achieved.

The framework presented herein represents an analysis into a specific problem, however the techniques and models presented could be used for a much larger variety of problems. Once a problem is broken down and viewed as stochastic process with different random variables interacting with each other with some assumed relationships and behaviors, it is only a matter of carefully constructing the proper resultant distributions. Fundamentally, this work illustrates the typical results of incorporating additional noise into the different parameters and even hyperparameters that make up any given problem. The difficulty is that for problems with sparse amounts of data, nonparametric analysis is difficult and often unreliable, while parametric models work well with sparse data but require assumptions about what specific model is used, as well as its parameters, and possibly their own hyperparameters, as was seen in Chapter 3 regarding the $\epsilon_{\nu}(S_i)$ fatality rate error, which was a random variable with a random valued variance composed of a function of an uncertain shaking intensity.

By incorporating additions uncertainties and observed (possibly noisy) data, the framework presented demonstrates the ability to produce updated PAGER estimates which can better explain the high variability associated with earthquake loss, and filter through these variabilities to give accurate total loss estimates in a reasonable amount of time, even if the initial PAGER estimate is substantially different than the final reported value. While under and over estimations do not indicate the PAGER estimate is wrong, as it has a large probability of being off by one to two orders of magnitude, it is nonetheless important to be able to find a means of estimating a more accurate total loss as quickly as possible. By incorporating this reported loss model with the initial PAGER estimates, this work has shown
that these estimates can be done reasonably, however, the distributional assumptions and
different approximations presented throughout must be reconciled before this model could
be trusted as a real-time, working product.

However, this work has illustrated multiple techniques regarding the different uncertain-
ties and data trying to be used, and one need only plug in more accurate estimations of the
different variables to obtain even better loss estimates. Unfortunately, the assumptions were
made primarily for mathematical and computational ease, and more involved models with
more levels of uncertainty would need to be solved with more brute force methods than nu-
merical integration, which was slow even for three dimensions. The prime candidate is Monte
Carlo simulation, however this work showed that it can take a very long time to get reason-
ably precise estimations, particularly with difficult multifaceted problems. Other techniques
such as quasi-random Monte Carlo and MCMC simulation techniques, metaheuristic algo-
rithms for searching random spaces, to spectral distribution decomposition methods could
be better suited to tackling the more rigorous versions of this problem. This particular area
of incorporating different kinds of data and uncertainties to improve different estimates re-
lated to earthquakes is not only of fundamental importance for helping save lives and better
inform the public and authorities, but it is a rich and fascinating scientific frontier.
REFERENCES CITED


