Bayesian Analysis of Earthquake Fatality Rates

H. Noh
Carnegie Mellon University, Pittsburg, Pennsylvania, USA
A. Kiremidjian, D. Jaiswal
Stanford University, Stanford, CA USA
E. So
Cambridge University, Cambridge, UK

ABSTRACT: When developing vulnerability or fragility functions, information is often limited and comes in different forms such as expert opinion, data from post-earthquake investigations, or analytical simulations. Combining the information from various sources can be particularly challenging. In this paper a Bayesian statistical framework is presented to enable a systematic approach for combining disparate information when developing vulnerability functions. The framework is extensively illustrated through application to earthquake fatality data for four structural types obtained from the 2005 Pakistan earthquake combined with information from PAGER. Two different models are tested that include an exponential probability density fit over the entire range of fatality rates, a combination of Bernoulli and exponential fit and a combination of Bernoulli and gamma fit again to model respectively zero and non-zero fatality rates. It is shown that the uncertainties in the final fatality marginal distributions are lower than that in the expert opinion based PAGER information.

1 INTRODUCTION

Earthquake vulnerability functions that relate a damage or loss parameters to different values of a ground motionan intensity measure can be developed using expert opinion, or data from past observations, or through response simulations of analytical model structures. Expert opinion is most frequently used when data from observations are not available or when simulations of models structures are not feasible. In cases when several sources of information are available or repeated observations produce multiple sets of data, it is important to combine the different sources of information in a systematic way. Conventional statistical analysis methods enable the treatment of field data or analytically simulated data. Aggregating the various sources of information, however, can be particularly difficult when combining expert opinion information with other data sources. Also, treating the uncertainty from different data sources can be problematic using standard statistical techniques. In recent years, as heuristic information is increasingly being used in numerous fields, there has been a proliferation of methods that use Bayesian statistical analysis. While the statistical community used to be divided into frequentists and Bayesians, modern statisticians combine both to solve the complex issues in science and engineering that require consideration of information from disparate sources (Efron, 2005).

Bayesian statistical analysis methods are particularly well suited for combining the information from different sources. Singhal and Kiremidjian (1998) were first to use Bayesian statistics to combine damage functions developed using expert opinion (ATC-13, 1985) with analytically simulated responses for reinforced concrete frames of three different heights. Enright and Frangopol (1999) and Papadimitriu et al (2001) have treated various aspects of reliability and vulnerability analysis of structures using Bayesian methods. In addition to providing a formal framework for aggregating different sources of information, aleatory uncertainties, arising from the random nature of a process, and epistemic uncertainties, that are due to lack of sufficient data, can be treated separately through the implementation of Bayesian statistical analysis. Igusa et al. (2002) use Bayesian statistics to study the effect of these sources of variability on design parameters. Straub and Der Kiureghian (2008) introduce a Bayesian framework for fragility function development where the uncertainty in the intensity measure in observed data is treated systematically and illustrate their method through an application to failures of electrical substation components. Koutsourelakis (2010) introduces Bayesian methods for developing vulnerability functions and uncertainty bands on those functions using limited data. Moreover, they present the formulation of vulnerability functions that are dependent on several intensity measures.
The objective of this paper is to develop earthquake fatality functions utilizing Bayesian statistical analysis methods. A general formulation for Bayesian analysis of fragility functions is first presented and the method is applied to earthquake fatality rate. Information on earthquake fatality rates are particularly lacking for most regions in the world; however, subjective estimates have been developed in the Prompt Assessment of Global Earthquakes for Response (PAGER) (see http://earthquake.usgs.gov/earthquakes/pager/) tool provided by the US Geological Survey (USGS) and are presently widely used. Most recently, So et al. (2009) collected data on fatality rates after the Pakistan 2005 Earthquake. The data from Pakistan are rather limited and to increase the reliability of fatality rates for specific types of structure, these data will be combined with the information provided in (PAGER). Different functional forms for fatality rates are explored and tested. These include the exponential and gamma probability models. It is demonstrated that in order to have a realistic representation of both zero and non-zero fatality rates, it is necessary to use hierarchical model of combined distributions. Some of the difficulties in treating the epistemic uncertainties in the analysis are discussed providing guidance for potential future applications. The marginal distributions on fatality rates that integrate the epistemic uncertainties are also developed demonstrating one of the advantages of the Bayesian approach for combining different sources of information.

2 FATALITY RATE MODEL

Past observations following major earthquakes have shown (Coburn & Spence 2002) that deaths and injuries occur primarily in structures that have either partially or completely collapsed. Structures that have minor to moderate damage but no collapse rarely cause deaths, the exception being deaths caused by falling objects or debris. Thus, the fatality rate, X, is defined in Equation 1 as the fraction of deaths of the total occupants in a building or a structure that has collapsed.

\[
X = \frac{\text{Total number of deaths}}{\text{Total Number of occupants}}
\]  

where \( x = \{x_1, x_2, \ldots, x_n\} \) is the set of independent observations of the random variable \( X \) that has the underlying probability distribution \( f_{X|\Theta}(x|\theta) \), \( f_\Theta(\theta) \) is the ‘prior distribution’ of the parameter \( \Theta \) that represents prior knowledge about the parameter before observing any data; \( f_{X|\Theta}(x|\theta) \) is the ‘posterior distribution’ of the parameter \( \Theta \) that represents the knowledge about the parameter after observing data; \( c \) is the normalizing constant, and \( L(X|\Theta) \) is the ‘likelihood function’ of the parameter \( \Theta \). In this study, \( X \) is the fatality rate, and \( \Theta \) is the vector of the parameters that describe the fatality rate distribution.

In this section, we first define the components of Equation 2. The probability distribution, \( f_{X|\Theta}(x|\theta) \), referred to as the parent distribution, is completely described by a set of a finite number of parameters \( \Theta = \{\theta_1, \theta_2, \ldots, \theta_k\} \). For the fatality rate density given that the structure has collapsed, this density can be written as follows:

\[
f_{X|\Theta}(x|\theta) = p(X|\text{collapse}, \Theta) = f_{X|C}
\]  

where \( \Theta \) is the vector an uncertain parameter. The form of the parent distribution \( f_{X|C} \) is not known. Review of past data, however, suggests that a distribution that has a fast decaying tail can be suitable.
Candidate distributions that will be examined in the application section of this paper include the exponential and gamma probability densities.

In general, the likelihood function of \( X \) is obtained using the parent distribution as follows:

\[
L(X | \Theta) = f_{X|\Theta}(x_1, x_2, ..., x_n | \theta) = \prod_{k=1}^{n} f_{X|\Theta}(x_k | \theta)
\]

where \( x = \{x_1, x_2, ..., x_n\} \) is the set of \( n \) independent observations of the fatality rate. The likelihood function is thus obtained by evaluating the parent distribution at each of the observations, and then, taking the product as defined in Equation 4.

3.1 Prior Distribution

One of the main challenges of the Bayesian approach is to select an appropriate distribution form for the parameter(s) \( \Theta \) and to estimate the hyper-parameter(s) of the distribution based on the information that is available. Hyper-parameters are parameters that describe the probability distribution of \( \Theta \). The variability captured by the distribution \( f_\Theta(\theta) \) reflects the epistemic uncertainty that with sufficient data can be reduced.

Computing the posterior distribution using Equation 2 can be algebraically complex and may not have a closed form solution. In special cases, the posterior distribution belongs to the same family of probability distributions as the prior distribution for a particular parent distribution (or likelihood function). This family of prior and posterior distributions is said to be ‘conjugate’ to the corresponding likelihood function. In such a case, the posterior distribution can be compactly represented by the hyper-parameters updated from those of the prior using the new information. For mathematical convenience, we use conjugate distributions to model the parameter distributions. In Section 4, the exponential and gamma distributions will be used to model fatality rates and we will apply the conjugate distributions for their parameters. Further discussion on conjugate prior-posterior distributions can be found in Raiffa and Shlaiffer (2000), Rice (1999), and Ang and Tang (2007).

3.2 Posterior Distribution

The posterior distribution of the parameter \( \Theta \), \( f_{\Theta|X}(\theta | x) \), will have the same form as the prior when conjugate distributions are used as stated in the previous section. This distribution, however, will contain the new information that comes either from observed data or from numerical simulations and will be combined with any prior information used to develop the hyper-parameters of the prior distribution on \( \Theta \). The use of the conjugate prior-posterior distributions will be illustrated in section 4.

3.3 Marginal Distribution

As stated previously, the uncertainty in the parameters, \( \Theta \), of the distribution of fatality rates, \( X \), is captured through either the prior or the posterior distribution of the parameters. A distribution of \( X \) that includes these uncertainties can be obtained by integrating over the range of values of each of the parameters. This distribution is known as the marginal distribution and in general is computed by applying the total probability theorem using equation (3.4) given below

\[
f(x) = \int_{\Theta} f_{X|\Theta}(x | \theta) f_\Theta(\theta) d\theta
\]

Either the prior or the posterior distribution of \( \theta \) can be used as \( f_\Theta(\theta) \) in Equation 5 to obtain the marginal distribution of \( X \). The marginal distribution defined by Equation 5 cannot, in general, be obtained in a closed form and is typically estimated using numerical integration.

4 APPLICATION OF BAYESIAN UPDATING TO FATALITY RATE DATA FROM THE 2005 PAKISTAN EARTHQUAKE

In this section the Bayesian updating procedure is applied to the fatality rates. For this purpose information on structures and fatality rates collected after the 2005 Pakistan earthquake for Pakistan are investigated. The study considered three types of structures based on construction classes found in Pakistan. These include 244 concrete block walls construction (concrete), 155 cut stone masonry construction (stone), and 41 rubble stone and mud construction (mud and stone). In addition, a structural class that considers all structural types (444 structures) together is used and is termed “all” hereafter. The analysis was carried out for all four categories, but only the results for the all combined case are presented in this paper due to the space limitation.

Information for the prior distribution is obtained from PAGER. For all the construction classes considered in this paper, the information is given in terms of ranges. For all structural classes the mean is assumed to be the midpoint of the range and the coefficient of variation is assumed to be a constant of 0.3. In general, the fatality rate is bounded between 0 and 1. From this information the parameters of the prior distributions were obtained. Then fatality data from the 2005 Pakistan earthquake was collected and used in the development of the posterior and marginal distributions of the parameters.

4.1 Selection of Fatality Rate Distribution

In order to select appropriate distributions for fatality rates, existing fatality data were analyzed to find
the general trends were observed. Data were obtained for the four categories of construction classes defined in the previous section – Concrete, Stone, Mud and Stone, and All combined. The data from the 2005 Pakistan earthquake consists of the number of occupants and the number of fatalities for collapsed buildings in the affected area. Fatality rates, defined in Equation 1, were obtained for each structural type. After testing several distributions, it was found that the exponential and the gamma distributions fit the data well. Empirical cumulative distribution functions (eCDF) of the fatality rate data were developed from the given fatality rate data and were compared with the cumulative distribution functions (CDFs) of the fitted exponential and gamma distributions.

The equations of the PDF for exponential distribution is given by
\[ f(x|\lambda) = \lambda e^{-\lambda x} \] (6)
with mean and variance given as
\[ \mu = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = \frac{1}{\lambda^2} \] (7)
The equations of the PDF for gamma distribution is given by
\[ f(x|\kappa,\theta) = \frac{1}{\Gamma(\kappa)\theta^\kappa} x^{\kappa-1} e^{-x/\theta} \] (8)
with mean and variance as
\[ \mu = k\theta \quad \text{and} \quad \sigma^2 = k\theta^2 \] (9)

Figure 1. CDF from fatality rate data and exponential and gamma distributions fit

Equations 7 and 9 are used to estimate the parameters of the two distributions. From Figure 1 and 4-2 it can be seen that both the exponential and the gamma distributions follow the general pattern of the data, but they poorly estimate the zero and low fatality rates. In section 4.2 the Bayesian updating for only the exponential distribution will be shown to demonstrate the case when one distribution is used for representing the entire range of fatality rate values. The same procedure can be applied to use the gamma distribution for modeling the fatality rate. In section 4.3, the results for the combined Bernoulli and exponential and the Bernoulli and gamma distributions will be presented as these models can describe the variations between zero and non-zero fatality rates.

4.2 Exponential Bayesian Analysis

In the previous section, only the information from the observed data was used to model the fatality rate. In order to include the information obtained from PAGER, the Bayesian analysis is applied to the fatality data where the PAGER data is used to develop the prior distribution of the parameter \( \lambda \). The analysis is performed as follows.

The fatality rate \( X \) (\( 0 \leq x \leq 1 \)) is characterized by the exponential distribution given by Equation 6 with parameter \( \lambda \). The conjugate distribution on the parameter \( \lambda \) for the exponential distribution is the gamma given by Equation 8 redefined here with new parameters as follows
\[ f'(\lambda|\kappa',\theta') \sim \text{gamma}(\kappa',\theta') \] (10)
where \( \kappa \) and \( \theta \) are the hyper-parameters of probability distribution of \( \lambda \).

4.2.1 Prior distribution for \( \lambda \)

Prior mean values of the fatality rate are taken from the Proposed Fatality Rates for implementation into PAGER (see Table 1). These are the mid-values of the range provided in PAGER. The fatality rates are presented as a percentage of the total number of occupants in completely collapsed buildings. The mean values for all structural types is given as 15% or 0.15 (Evaluated by taking the weighted mean of concrete, stone, and mud and stone fatality rates where the weights are proportional to the number of structures in each group).

The coefficient of variation is assumed to be 0.3. Using these values for the mean and coefficient of variation, the hyper-parameters for the gamma distribution of the parameter \( \lambda \) are estimated and are listed in Table 2.

<table>
<thead>
<tr>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>0.106</td>
</tr>
<tr>
<td>Prior</td>
<td>0.15</td>
</tr>
<tr>
<td>Posterior</td>
<td>0.107</td>
</tr>
</tbody>
</table>

Table 1. Sample, prior, and posterior mean and standard deviations of fatality rates considering the exponential-gamma conjugate pair

<table>
<thead>
<tr>
<th>( \kappa )</th>
<th>( \theta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>50</td>
</tr>
<tr>
<td>Posterior</td>
<td>236.74</td>
</tr>
</tbody>
</table>

Table 2. Prior and posterior hyper-parameters for the gamma distribution on the parameter \( \lambda \) considering the exponential-gamma conjugate pair
4.2.2 Likelihood function

In general, the likelihood function will depend on the observations from actual data or from numerical simulations as explained in section 3. The data for this application was obtained for the Pakistan 2005 earthquake. The number of observations, the sample mean and sample standard deviation for each structural class are listed in Table 1. Since the gamma distribution is the conjugate distribution of the exponential, the likelihood function does not need to be developed explicitly.

4.2.3 Posterior distribution for \( \lambda \)

The hyper-parameters for the posterior distribution on the parameter \( \lambda \), which is also a gamma as the prior on \( \lambda \), are \( \theta'' \) and \( \kappa'' \). They can be obtained using the following equation (see Ang and Tang, 2007):

\[
\kappa'' = \kappa' + n \tag{11}
\]

\[
\frac{1}{\theta''} = \frac{1}{\theta'} + \sum_{i=1}^{n} x_i \tag{12}
\]

where \( x_i \)'s are the observed fatality rates from the data and \( n \) is the number of observations. Note that the hyper-parameter \( \kappa \) is closely related to the number of data. Thus, the \( \kappa \) value for the prior distribution represents the significance of the prior information quantified in terms of corresponding number of data. Credible prior information will assign a large value for \( \kappa \). Similarly, \( 1/\theta \) is related to the un-normalized mean of the data. Therefore, the prior distribution with the hyper-parameters \( \kappa' \) and \( \theta' \) implies that the prior information is equivalent to \( \kappa' \) observations whose sum is \( 1/\theta' \). We can select these prior hyper-parameters based on our confidence about the prior information. Table 1 lists number of observations from the 2005 Pakistan earthquake as well as the posterior mean and standard deviation computed by first updating the hyper-parameters using Equations 11 and 12 and then applying Equations 9. The posterior hyper-parameters are listed in Table 2.

From these estimates it is observed that the mean from posterior marginal distribution is closer to the

Table 3. Marginal prior and posterior mean and standard deviation of fatality rates

<table>
<thead>
<tr>
<th></th>
<th>Marginal mean, ( \mu_x )</th>
<th>Marginal standard deviation, ( \sigma_x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>0.15</td>
<td>0.045</td>
</tr>
<tr>
<td>Posterior</td>
<td>0.11</td>
<td>0.005</td>
</tr>
</tbody>
</table>

4.2.4 Marginal Distribution on Fatality Rates, \( X \)

Marginal probability density function \( f(x) \) of fatality rate was defined by Equation 5. As stated previously, Equation 5 cannot be obtained in closed form and is evaluated using numerical integration. Increments of \( 10^{-4} \) for \( \lambda \) and \( 10^{-2} \) for \( x \) were used for the purpose of numerical integration. Two versions of the marginal PDF for fatality rate were computed using the prior and posterior distributions of the parameters, and the corresponding prior and posterior means and standard deviations were obtained. The marginal prior and posterior means and standard deviations are shown in Table 3.
sample mean given in Table 1. The standard deviation from posterior marginal distribution is smaller than that the standard deviation from both the prior marginal distribution and the sample, which indicates that combining the heuristic information with the data enable us to make a more confident estimation about fatality rate.

Figure 3 and Figure 4 show the marginal PDF and CDF, respectively. Both prior and posterior functions are shown in the figures. It is observed that the posterior fits the data closely for values higher than zero, but the zero fatality rate is not represented well by either the prior or the posterior marginal distributions. A different approach will be considered in subsequent sections to capture this difference.

4.3 Bernoulli-Exponential Bayesian Analysis of Fatality Rates

In section 4.2, Figure 3 and Figure 4 show that although the posterior distribution fits the non-zero fatality rates well, the zero fatality rates are not well represented by either the prior or the posterior marginal distributions.

In this section, zero fatality rates have been considered separately and a hierarchical model that utilizes Bernoulli Bayesian analysis performed on zero fatality rates is presented. This concept is illustrated in Figure 5. Non-zero fatality rates are modelled again using exponential Bayesian analysis.

The fatality rate $X$, defined over the range $0 \leq x \leq 1$ is now divided into two ranges. For that purpose we introduce an indicator function $K$, where $K = 1$ for $x = 0$ and $K = 0$ for $x > 0$. Then the probability distribution of $K$ is modelled with a Bernoulli distribution given as

$$f(p_0, k) = p_0^k(1-p_0)^{1-k} \text{ for } (x = 0, k = 1) \quad (13)$$

where $p_0$ is the probability that $K = 1$.

Conditioned on that the fatality rate is non-zero, it is modelled with the exponential distribution given by equation (4.1) and modified as follows:

$$f(x|\lambda) = \lambda e^{-\lambda x} (1-k) \quad (14)$$

for $0 < x \leq 1$ and $K = 0$. Note that the parameter in Equation 14 is again $\lambda$.

4.3.1 Prior and Posterior Distributions for $p_0$ and $\lambda$

The conjugate prior distribution of $p_0$ is beta given by

$$f(p_0|a,b) = \frac{r(a+b)}{r(a)r(b)} p_0^{a-1}(1-p_0)^{b-1} \quad (15)$$

where $0 \leq p_0 \leq 1$ and $a$ and $b$ are the hyper-parameters of the beta probability distribution of $p_0$. Since the beta distribution is the conjugate of the Bernoulli distribution, the posterior distribution will also be a beta distribution with hyper-parameters $a''$ and $b''$ computed using the following (Ang and Tang, 2007):

$$a'' = a' + \sum k_i \quad (16)$$

$$b'' = b' + n - \sum k_i \quad (17)$$

where $n$ is the total number of data points and $k_i$ is the indicator function as defined above. Note that the prior distribution with the hyper-parameters $a'$ and $b'$ implies that the prior information is equivalent to $a'$ and $b'$ observations of $K = 1$ and 0, respectively. Prior mean value of $p_0$, $\mu_{p_0}'$, was chosen randomly from values between 0 and 1 because no information was available. Several tests were performed with different prior values of $\mu_{p_0}'$ and it was found that the results do not vary significantly with random selection of $\mu_{p_0}'$. For the particular case of the Pakistan earthquake data, the posterior is governed by the data. Table 4 lists the prior, sample and posterior mean values for the Bernoulli distribution of zero fatality rates.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean ($p_0$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>0.62</td>
</tr>
<tr>
<td>Prior</td>
<td>0.3</td>
</tr>
<tr>
<td>Posterior</td>
<td>0.58</td>
</tr>
</tbody>
</table>

The prior and posterior distributions of $\lambda$ are again gamma as described in section 4.4. Figure 6 shows the prior and posterior distributions for the parameter $\lambda$. The prior and posterior distributions of the parameter $p_0$ are shown in Figure 7. The prior and posterior hyper-parameter values for the gamma conjugate probability density of the exponential parameter $\lambda$ are listed in Table 5. Table 6 provides the means and standard deviations for the posterior and prior distributions using the exponential-gamma conjugate pairs and non-zero fatality rates. From Table 4 and 6, it is observed that the posterior mean values are between the prior and the sample means but considerably closer to the sample means. This is due to the large amount of data and low confidence
on the prior information, represented by the large variance in prior distribution. The standard deviations for the posterior non-zero fatality rates are again smaller than the prior values. More importantly, these values are also smaller than the standard deviations when the exponential was used to model the fatality rates over the entire range of values. Thus, modelling the zero and non-zero values separately has resulted in a large reduction in the uncertainties in the model.

Table 5. Prior and posterior parameters for the gamma distributed parameter $\lambda$ considering exponential-gamma conjugate pairs

<table>
<thead>
<tr>
<th></th>
<th>$\kappa$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>50</td>
<td>3.33</td>
</tr>
<tr>
<td>Posterior</td>
<td>158</td>
<td>22.21</td>
</tr>
</tbody>
</table>

Table 6. The means and standard deviations of fatality rate considering exponential-gamma conjugate prior

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>0.106</td>
<td>0.171</td>
</tr>
<tr>
<td>Prior</td>
<td>0.15</td>
<td>0.045</td>
</tr>
<tr>
<td>Posterior</td>
<td>0.113</td>
<td>0.0020</td>
</tr>
</tbody>
</table>

4.3.2 Marginal Distributions of Bernoulli-Exponential Fatality Rates

Marginal PDF, $f(x)$, of fatality rate is computed using the law of total probability as

$$f(x) = \begin{cases} \mu_p \quad & x = 0 \\ (1 - \mu_p) \int f(x|\lambda)f(\lambda)d\lambda \quad & x > 0 \end{cases}$$

(18)

where $f(x|\lambda)$ is the parent distribution of the fatality rate $X$, $f(\lambda)$ is either the prior or the posterior distribution of the parameter $\lambda$, and $\mu_p$ is the mean of $p_0$. The marginal distribution defined by Equation 18 cannot be obtained in closed form and is evaluated using numerical integration. Increments of $10^{-4}$ for $\lambda$ and $10^{-2}$ for $X$ have been used for the purpose of numerical integration as before. The marginal mean and standard deviation of fatality rate can also be computed using standard statistical methods and numerical integration. Table 7 lists the marginal prior and posterior means and standard deviations for the fatality rates with Bernoulli-exponential distribution model.

Table 7 Marginal prior and posterior means and standard deviations for the Bernoulli-exponential fatality model for different structural types

<table>
<thead>
<tr>
<th></th>
<th>$\mu_x$</th>
<th>$\sigma_x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>0.15</td>
<td>0.045</td>
</tr>
<tr>
<td>Posterior</td>
<td>0.11</td>
<td>0.0020</td>
</tr>
</tbody>
</table>

From these estimates it is observed that the mean from posterior marginal distribution is closer to the sample mean given in Table 1. The standard deviation from posterior marginal distribution is considerably smaller than that the standard deviation from prior marginal distribution, which again indicates that combining the heuristic information with data reduced the overall uncertainty of fatality rate estimation. Note that this marginal standard deviation is smaller than the one using the exponential distribution, shown in Table 3.
5 CONCLUSIONS

In this paper a Bayesian formulation is presented for developing vulnerability functions and the methodology is used for treating the uncertainties in fatality data. Two different models are presented and tested with information from PAGER and from the 2005 Pakistan earthquake. These models include exponentially distributed fatality rates using the exponential and gamma conjugate pair and a combined hierarchical Bernoulli-exponential model to separate zero and non-zero fatality rates. The exponential and gamma distributions are selected based on initial data fitting through the Pakistan data.

The Bayesian formulation for the exponential distribution with gamma-distributed hyper-parameters is first presented. Values for hyper-parameters of the prior distributions are developed using the midpoint of the fatality ranges provided in PAGER and by assuming that the coefficient of variation is fixed at 0.3. The posterior hyper-parameters are obtained by combining the prior and data from Pakistan. In addition, the marginal prior and posterior probability densities and cumulative distributions are computed through numerical integration. Based on the analysis of the exponential model, it is found that the zero fatality rates are not well represented and introduce a bias to the non-zero fatality rates. Thus, an additional hierarchical model is introduced to enable modeling the zero and non-zero fatality rates separately.

The Bernoulli-exponential model consists of a Bernoulli distribution that models the zero fatality rate and an exponential distribution for the non-zero fatality rates. The conjugate distribution for the Bernoulli distribution parameter is the beta distribution. The exponential model remains the same as in the first analysis. The formulations for the prior and posterior marginal distributions of the fatality rate are developed by integrating over all the parameter space. These distributions are evaluated numerically as closed form solutions cannot be obtained. Prior and posterior means and standard deviations are also estimated. Based on the analysis conducted for the Bernoulli-exponential model, it is first observed that separating the zero and non-zero rates significantly reduces the bias. This model has the posterior standard deviation that is considerably smaller than the prior standard deviation pointing to one of the main advantages of using Bayesian analysis.

REFERENCES


