Bayesian Updating of Earthquake Vulnerability Functions with Application to Mortality Rates

H.Y. Noh, Anne Kiremidjian, Luis Ceferino and Emily So

Vulnerability functions often rely on data from expert opinion, post-earthquake investigations, or analytical simulations. Combining the information can be particularly challenging. In this paper a Bayesian statistical framework is presented to combining disparate information. The framework is illustrated through application to earthquake mortality data obtained from the 2005 Pakistan earthquake and from PAGER. Three different models are tested including an exponential, a combination of Bernoulli and exponential and Bernoulli and gamma fit to model respectively zero and non-zero mortality rates. A novel Bayesian model for the Bernoulli-exponential and Bernoulli-gamma probability densities is introduced. It is found that the exponential distribution represents the zero casualties very poorly. The Bernoulli-exponential and Bernoulli-gamma models capture the data for both the zero and non-zero mortality rates. It is also shown that the Bernoulli-gamma model fits the 2005 Pakistan data the best and has uncertainties that are smaller than either the ones from the 2005 Pakistan data or the PAGER data.

INTRODUCTION

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. Singhal and Kiremidjian (1998) first proposed the use of Bayesian statistics to combine heuristic information with numerically simulated data for developing fragility functions for reinforced concrete structures. Several other studies have also deployed Bayesian statistical methods for fragility analysis of structures (e.g. Choe et al. 2007; Der Kiureghian 1999a, b, 2002; Gardoni et al. 2002; Gardoni et al. 2003; Huang et al. 2010; Koutsourelakis 2010; Zhong et al. 2008). In this paper a Bayesian statistical framework is presented to enable a systematic approach for combining disparate information for developing casualty rates from collapsed buildings. The focus of the paper is on the development of vulnerability rates that require different probability distributions to model zero and nonzero mortality rates. The Bayesian treatment of the exponential distribution fit to mortality rates is developed as an example of conventional Bayesian updating of probability distributions. The paper presents a novel approach to Bayesian updating of combination of probability distributions for mortality rate analysis, in this case the combination of Bernoulli with exponential and Bernoulli with gamma probability densities. The approach presented in the paper is general and can be used for fragility analysis whenever there is a need to have multiple probability distributions to represent the uncertainty of the fragility variables.

The framework is illustrated through application to earthquake mortality data for four structural types obtained from the 2005 Pakistan earthquake combined with information from PAGER (Earle et al. 2009; Jaiswal et al. 2009; Wald et al. 2010). This data will be used in the development of the simple exponential distribution first, and then for Bernoulli-exponential and Bernoulli-gamma models to better represent the zero and nonzero mortality rates. In these analyses the focus is to evaluate the uncertainties in the posterior mortality marginal distributions that combine the PAGER information and the observed data and to identify the reduction of epistemic uncertainty.

MORTALITY RATE MODELLING

Past observations following major earthquakes have shown that deaths and injuries occur primarily in structures that have either partially or completely collapsed (e.g. Ferreira, Olivera and Mota de Sá 2011; Shoaf et al. 2005; Petal 2011). 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 mortality rate, X, is defined in the equation (1) as the fraction of deaths of the total occupants in a building or a structure that has collapsed.

$$X = \frac{\text{Total number of deaths}}{\text{Total number of occupants}} \qquad (1)$$

Models and data on earthquake casualties are very limited and as a result mortality rates are highly uncertain. To capture this variability, the mortality rate, X, is modeled in this paper using parametric probability distribution. The form of the probability density function of X, $f_{X|\Theta}(x|\theta)$, is not known and there are several candidate functions examined in this paper. The parameters of the mortality rate probability distribution are typically obtained from limited data and the uncertainty on the parameters reflects the epistemic uncertainty in the model. Consequently, they are modeled as random variables.

In order to combine various sources of data and reduce the uncertainty in the mortality rate model, the Bayesian framework is utilized to estimate the parameters of the mortality rate probability distribution by combining information from different data sources. The general Bayesian formulation for information updating and its application are described in the following sections.

BAYESIAN UPDATING MODEL

The Bayesian formulation for vulnerability functions presented in this paper focuses on mortality rates, however, it is general and can be applied to any other vulnerability function analysis. Bayes' rule provides the basis for combining various sources of information and data and enables treating the uncertainties in the parameters of the parent distribution of the random variable X. Equation (2) states Bayes' rule as follows:

$$f_{\Theta|X}(\theta|x) = \frac{f_{X|\Theta}(x|\theta) f_{\Theta}(\theta)}{\int_{\theta} f_{X|\Theta}(x|\theta) f_{\Theta}(\theta)\, d\theta} \qquad (2)$$

where $x = \{x_1, x_2, \dots, x_n\}$ is the set of n 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 Θ that represents prior knowledge about the parameter before observing any data; $f_{\Theta|X}(\theta|x)$ is the 'posterior distribution' of the parameter that represents the knowledge about the parameter after observing data.

PRIOR/POSTERIOR DISTRIBUTION DISTRIBUTUIONS OF Θ

One of the main challenges of the Bayesian approach is to select appropriate distribution for the parameter(s) Θ and to estimate the distribution based on the information that is available. 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 the prior using the new information. Hyper-parameters are parameters that describe the prior distribution of Θ. For mathematical convenience, we use conjugate distributions to model the parameter distributions. In the next section, the exponential and gamma distributions will be used to model mortality rates and we will apply the conjugate distributions for their parameters. Subsequently, a novel approach for combining Bernoulli-exponential and Bernoulli-gamma distribution prior-posterior parameters will be presented. Further discussion on conjugate prior-posterior distributions can be found in Raiffa and Shlaiffer (2000), Rice (1999), and Ang and Tang (2007).

MARGINAL DISTRIBUTION

For the purposes of practical applications, it is desirable to develop one distribution that combines the uncertainty in the random variable X described by the conditional distribution $f_{X|\Theta}(x|\theta)$ and the uncertainty in the parameters Θ described by $f_{\Theta}(\theta)$. Integrating over the range of values of each of the parameters of the parent distribution of X we the marginal distribution of X is given by equation (3).

$$f_X(x) = \int_{\theta} f_{X|\Theta}(x|\theta) f_{\Theta}(\theta)\, d\theta \tag{3}$$

Either the prior or the posterior distribution of Θ can be used in equation (3) to obtain the marginal distribution of X. Equation (3) cannot be obtained in closed form and is typically estimated using numerical integration.

DEVELOPMENT OF BAYESIAN MODELS FOR MORTALITY RATES

SELECTION OF MORTALITY RATE DISTRIBUTION

In order to select appropriate distributions for mortality rates, analysis was performed first on existing mortality rates data and the general trends were observed. Mortality rates are defined as number of deaths divided by the number of occupants. Data were obtained for the four categories of construction classes defined as Concrete, Stone, Mud and Stone, and All combined. The data from the 2005 Pakistan earthquake consists of number occupants and number of deaths for collapsed buildings in the affected area. The Pakistan earthquake data were collected by household questionnaire surveys carried out in June 2016 in collaboration with the University of Peshawar in the local language. In all, 530 surveys were done. The interviews assembled two different kinds of information, descriptive and factual. In many ways the descriptive accounts provide the best evidence of what occurred and what was observed by the survivor. There are some limitations to this data set as a representative sample of those affected. For example, we could only interview families who survived. For obvious reasons families with no survivors or those that had migrated could not be interviewed. This could mean there are more reports from people living in better building types, and from those on flatter lands in our data set. Returned forms for the uninjured were few for Pakistan. This could be due to the bias introduced with the selection of heavily damaged villages in Pakistan or to do with the responders and interviewers not understanding that information on the uninjured and their survival were just as important as those seriously injured. Mortality rates defined as number of deaths divided by the number of occupants were obtained for each structural type. The number of mortality rates number of data used for each construction type is listed in Table 1. The sample mean, standard deviation and coefficient of variation (CoV) for Mortality rates defined as number of deaths divided by the number of occupants were obtained for each structural typemortality rates in each building category are also listed in that table.,

Table 1. Sample Statistics for Casualties Mortality Rates by Building Construction Class

Construction Class	Number of Samples	Sample Mean	Sample Std. Dev.	Sample CoV
Concrete	248	0.125	0.195	1.556
Stone	155	0.092	0.140	1.530
Mud and Stone	41	0.072	0.114	1.590
All	444	0.107	0.181	1.684

Formatted: Left, Line spacing: single, No widow/orphan control, Don't adjust space between Latin and Asian text, Don't adjust space between Asian text and numbers

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Asian) Japanese

Several distributions were tested and it was found that the exponential and the gamma parametric distributions fit the data well. Empirical cumulative distribution functions (eCDF) of the mortality rate data were developed from the given mortality rate data and were compared with the cumulative distributions (CDFs) of exponential and gamma distribution.

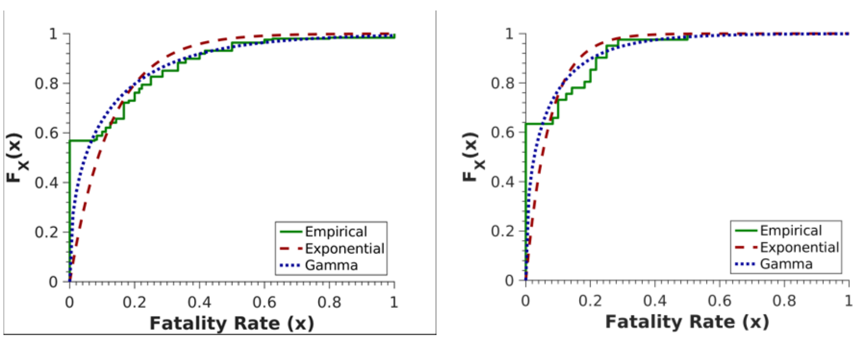
The equations of the probability density function for the exponential and gamma probability densities are given by equations (4) and (6), respectively. The mean values and standard deviations for each distribution are given in equation (5) and (7).

$$f(x|\lambda) = \lambda e^{-\lambda x} \tag{4}$$

$$\mu = 1/\lambda$$
$$\sigma = 1/\lambda \tag{5}$$

$$f(x|\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \tag{6}$$

$$\mu = \alpha/\beta$$
$$\sigma = \alpha/\sqrt{\beta} \tag{7}$$



a) Concrete b) Mud and stone structures

Figure 1. Empirical CDF from fatality rate data and exponential and gamma distributions ~~fits~~ fitted CDFs for (a) concrete structures (b) mud and stone structures.

Formatted: Figure Caption, Space After: 0 pt, Line spacing: single

Figures 1a and 1b show examples of eCDF's and fitted exponential and gamma distributions to the Pakistan 2005 data for concrete and stone structures. From that figure it can be seen that both the exponential and the gamma distributions follow the general pattern of the data, however, they underestimate the zero mortality rates. The same results were observed with the remaining building categories in the data. In the next section, the Bayesian updating only for the exponential distribution will be shown when using one distribution for the entire range of mortality rate values. In the following sections, the results for the combined Bernoulli and exponential and the Bernoulli and gamma distributions will be presented as these capute the variations between zero and non-zero mortality rates.

EXPONENTIAL BAYESIAN ANALYSIS

In the previous section, only the information from the observed data was used. Prior to any data collection, information for Pakistan was available primarily from the Prompt Assessment of Global Earthquakes for Response (PAGER), (Jaiswal et al. 2009). The United States Geological Serviey (USGS) developed the PAGER framework to estimate fatalities caused by earthquakes. In PAGER, fatality rates were estimated using a combination of fatality statistics from previous earthquakes and expert judgment. These rates were estimated for multiple construction materials and they represent global rates since they were obtained from worldwide data (Jaiswal et al., 2009). For the Bayesian analysis, data from PAGER were used to estimate the prior distribution of the parameter λ. The analysis is performed as follows.

The mortality rate X, $(0 \leq x \leq 1)$ is characterized by the exponential distribution given by equation (4) with parameter λ. The conjugate distribution on the parameter λ for the exponential distribution is the gamma distribution given by equation (8).

$$f_{\Lambda}(\lambda|\omega,\varphi) = \frac{\varphi^{\omega}}{\Gamma(\omega)}\lambda^{\omega-1}e^{-\varphi\lambda} \quad (8)$$

where α and β are called the hyper-parameters of the probability distribution of λ. It should be noted that these hyper-parameters are different from the ones that were used in equation (6).

PRIOR DISTRIBUTION FOR λ

Prior mean values of the mortality rate are taken from the Proposed Fatality Rates for implementation into PAGER. These are the fatality rate values given structural collapse

Formatted: Font: Times New Roman, Not Italic

Formatted: Font: Times New Roman, Not Italic

Formatted: Font: Times New Roman, Not Italic

Formatted: Font: Times New Roman, Not Italic provided in PAGER (Jaiswal et al. 2009). The coefficient of variation, CoV, is assumed to be have a value of 0.30 and is kept constant for all construction types., with a value of 0.30. The fatality rates are presented as a ratio of the total number of occupants in completely collapsed buildings. The mean values for each structural type are given as follows:

- Concrete: 0.15
- Stone: 0.06
- Mud and stone: 0.06
- All: 0.11 (Evaluated by taking the weighted mean of concrete, stone, and mud and stone fatality rates where the weights are the number of structures in each group corresponding to which are 248, 155 and 41, respectively).

These values are also listed in Table 2 for completeness. The coefficient of variation (CoV) is assumed to be 0.3 for all the four construction types. Using these values for the mean and coefficient of variation, the hyper-parameters for the gamma distribution of the parameter λ are estimated and are listed in Table 2.

Table 2. Prior and posterior hyper-parameters for the gamma distribution on the parameter λ considering exponential-gamma conjugate pairs for different structural types.

Construction Class	ω_{prior}	φ_{prior}	Prior Mean λ	Prior CoV λ	$\omega_{posterior}$	$\varphi_{posterior}$	Posterior Mean λ	Posterior CoV λ
Concrete	11.11	1.667	6.67	0.30	259.11	32.78	7.90	0.06
Stone	11.11	0.667	16.67	0.30	166.11	14.87	11.17	0.08
Mud and stone	11.11	0.667	16.67	0.30	52.11	3.62	14.39	0.14
All	11.11	1.222	9.07	0.30	285.11	30.65	9.30	0.59

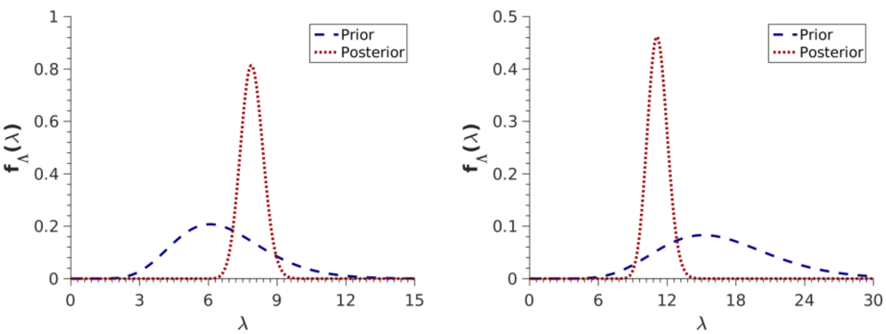
POSTERIOR DISTRIBUTION FOR λ

Since the gamma distribution is the conjugate distribution of the exponential probability density, the likelihood function does not need to be developed explicitly. The posterior distribution follows the same gamma distribution form as the prior, but with the posterior values of the hyper-parameters substituted for the prior ones. The hyper-parameters ω and φ for the posterior distribution on the parameter λ can be obtained using the following equation (see Ang and Tang, 2007):

$$\omega_{posterior} = \omega_{prior} + n$$

$$\varphi_{posterior} = \varphi_{prior} + \sum_{i=1}^{n} x_i \qquad (9)$$

where x_i's are the observed mortality rates from the data and n is the number of observations. Table 1 listed the number of observations from the 2005 Pakistan earthquake as well as the sample mean and standard deviation for each building type. The posterior hyper-parameters are listed in Table 2.



a) Concrete **b) Stone**

Figure 2. Prior and posterior gamma distributions of the exponential distribution parameter λ for (a) concrete structures and (b) stone structures.

Figure shows the prior and posterior distribution on the parameter λ for concrete and stone building types. The mean in each case has moved closer to the sample mean, and in all cases the standard deviation for the posterior is smaller than the prior showing that the epistemic error in the estimates has been reduced using this approach.

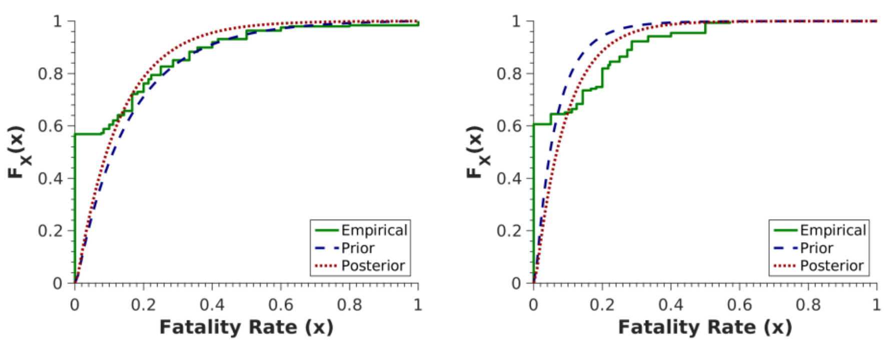
MARGINAL DISTRIBUTION ON MORTALITY RATES, *X*

Marginal probability density function of mortality rate was defined by equation (3). Equation (3) was evaluated using numerical integration with increments of 10^{-4} for λ and 10^{-2} for x. After computing the prior and posterior marginal distributions, the prior and posterior mean and standard deviation on mortality rates X are computed also using numerical integration. The results from these analyses for each of the structural types are given in Table 3. From these estimates, it is observed that the mean from the posterior marginal distribution is closer to the sample mean given in Table 1. The standard deviation from the 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 reduced the overall uncertainty of mortality rate.

Table 3. Marginal prior, sample, and marginal posterior mean and coefficient of variation of mortality rates considering exponential-gamma conjugate pairs for different structural types.

Construction Class	Prior Marginal Mean, μ_x	Prior Marginal CoV, σ_x/μ_x	Posterior Marginal Mean, μ_x	Posterior Marginal CoV, σ_x/μ_x	Sample Mean	Sample CoV
Concrete	0.159	1.038	0.127	1.016	0.125	1.556
Stone	0.066	1.144	0.090	1.036	0.092	1.530
Mud and Stone	0.066	1.144	0.071	1.059	0.072	1.590
All	0.119	1.089	0.108	1.025	0.107	1.684

Figure 3 shows the prior and posterior cumulative distribution functions (CDF) for concrete and stone building types. It can be observed that the posteriors functions fit the data well for values higher than zero. The zero mortality rate is not predicted well by either the prior or the posterior marginal distributions. A different approach will be considered in subsequent sections to capture this difference.



a) Concrete **b) Stone**

Figure 3. Prior and posterior marginal cumulative densities for exponentially distributed mortality rate for (a) concrete structures and (b) stone structures

BERNOULLI-EXPONENTIAL BAYESIAN MODEL OF MORTALITY RATES

In this section, zero mortality rates have been considered separately and Bernoulli Bayesian analysis performed on zero mortality rates is presented. Non-zero mortality rates are modeled again using exponential Bayesian analysis.

The mortality rate X, defined over the range ($0 \leq x \leq 1$) is now defined over two ranges. For that purpose, we use a Bernoulli distribution to describe the occurrence of a mortality rate equal to 0, p_0, and we use the exponential probability to model the PDF of the mortality rate given that it is greater than 0. Therefore, the probability distribution of X is defined in equation (10).

$$f(x|\lambda, p_0) = \begin{cases} p_0 & \text{for } x = 0 \\ (1-p_0)\lambda e^{-\lambda x} & \text{for } 0 < x \leq 1 \end{cases} \tag{10}$$

Prior and Posterior Distributions for p_0 and λ

A beta distribution was used to model the prior distribution of p_0 since it is the conjugate of a Bernoulli distribution and is given by equation (11).

$$f(p_0|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} {p_0}^{a-1}(1-p_0)^{b-1} \text{ for } 0 < p_0 < 1 \tag{11}$$

where 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 density function, the posterior distribution on the parameter p_0 will also be a beta with posterior hyper-parameters parameters computed using the following equations (see Ang and Tang, 2007):

$$\begin{aligned} a_{\text{posterior}} &= a_{\text{prior}} + \sum_{i=1}^{n} 1\{x_i = 0\} \\ b_{\text{posterior}} &= b_{\text{prior}} + n - \sum_{i=1}^{n} 1\{x_i = 0\} \end{aligned} \tag{12}$$

where n is the total number of data points and $1\{x_i = 0\}$ takes the value of 1 when $x_i = 0$ and 0 otherwise. The prior and posterior distributions of λ are again gamma as described previously. The prior Bernoulli-exponential distribution was set by using PAGER information. The values of the hyper-parameters for the gamma conjugate probability density of the exponential parameter λ were chosen randomly, and the prior mean value of p_0, μ_{p_0}, was chosen so that the mean value of the prior distribution of X for all values greater than or equal to zero, given in Equation (3), approaches to the mortality rates reported by PAGER. The coefficient of variation *CoV* of p_0 is assumed to be 0.3 for all the four construction types. It was found that the results do not vary significantly with random selection of the hyper-parameters. For the particular case of the Pakistan earthquake data, the posterior is governed by the data. Table 4 lists the sample statistics for the zero fatality observations for each of the structural types considered in the study.

The prior and posterior distributions of the parameter p_0 are shown in Figure 4. The prior and posterior parameters and the respective means and standard deviations are listed in Table 5. The prior and posterior hyper-parameter values for the gamma conjugate probability density of the exponential parameter λ are listed in Table 6. The corresponding probability densities for concrete and stone building types are shown in Figure 5. Table 6 provides posterior mean and standard deviations for the exponential distribution for each of the four structural classes and non-zero mortality rates. From Table 4, 5 and 6, it can be observed that the posterior mean values in all cases are considerably closer to the sample means. The standard deviations for the posterior distribution of p_0 and λ are again smaller than the prior values. These show that the incorporation of new data reduces the initial epistemic uncertainties.

Table 4. Prior and posterior parameters for the zero mortality rate parameter p0 for different structural types

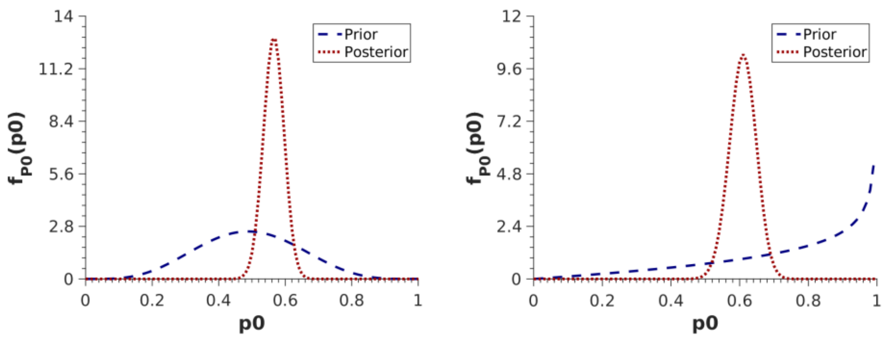
Construction Class	Number of samples	Houses with zero fatalities	Fraction with zero fatalities
Concrete	248	141	0.569
Stone	155	94	0.606
Mud and Stone	41	26	0.634
All	444	269	0.606

Table 5. Prior and posterior parameters for the zero mortality rate parameter p_0 for different structural types

Construction Class	a_{prior}	b_{prior}	Prior Mean p_0	Prior CoV p_0	$a_{posterior}$	$b_{posterior}$	Posterior Mean p_0	Posterior CoV p_0
Concrete	5.170	5.370	0.491	0.30	146.17	112.37	0.565	0.054
Stone	2.004	0.661	0.752	0.30	99.00	61.66	0.576	0.064
Mud and stone	1.948	0.627	0.757	0.30	27.95	15.63	0.577	0.112
All	3.793	2.484	0.604	0.30	169.79	110.48	0.606	0.048

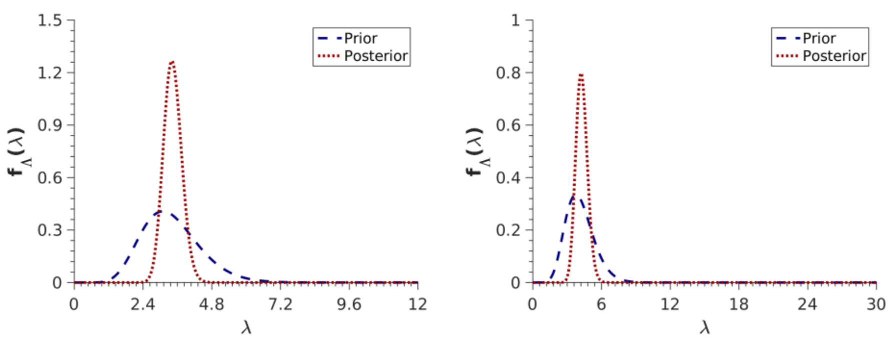
Table 6. Prior and posterior parameters for the gamma distributed parameter λ considering exponential-gamma conjugate pairs for different structural types

Construction Class	ω_{prior}	φ_{prior}	Prior Mean λ	Prior CoV λ	$\omega_{posterior}$	$\varphi_{posterior}$	Posterior Mean λ	Posterior CoV λ
Concrete	11.11	3.27	3.40	0.30	118.11	34.38	3.44	0.09
Stone	11.11	2.69	4.13	0.30	72.11	16.89	4.27	0.12
Mud and stone	11.11	2.74	4.06	0.30	26.11	5.69	4.59	0.20
All	11.11	3.09	3.60	0.30	119.11	32.51	3.66	0.09



a) Concrete **b) Stone**

Figure 4. Prior and posterior beta distributions of p_0 for the Bernoulli distribution for zero fatalities for (a) concrete structures and (b) stone structures.



a) Concrete **b) Stone**

Figure 5. Prior and posterior gamma distributions of the exponential distribution parameter λ for (a) concrete structures and (b) stone structures.

Marginal Distributions of Bernoulli-Exponential Mortality Rates

Marginal probability density function of mortality rate is computed using the total probability theorem and is given in equation (13)

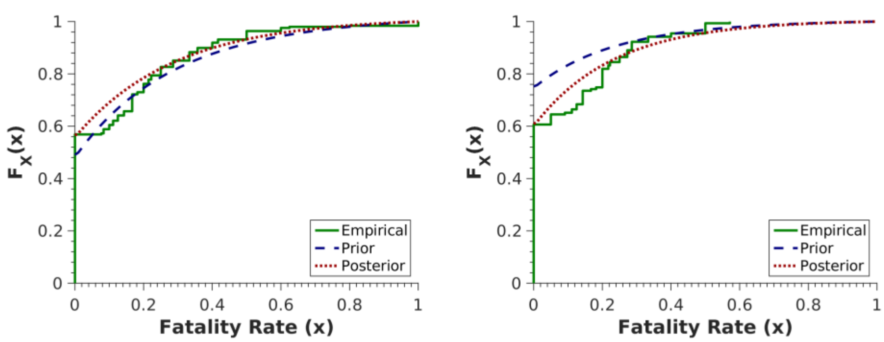
$$\begin{cases} \mu_{p0} & for\ x\ =\ 0 \\ (1-\mu_{p0})\int_{\lambda} f(x|\lambda)f(\lambda)d\lambda & for\ 0 < x \leq 1 \end{cases} \tag{13}$$

where $f(x|\lambda)$ is the probability distribution of the mortality rate X and $f(\lambda)$ is either the prior or the posterior of the probability density function of the parameter λ. μ_{p0} is the mean value of p_0 obtained from the beta distribution of p_0. The marginal distribution defined by Equation (13) was again evaluated using numerical integration.

Table 7 lists the marginal prior and posterior means and standard deviations for the mortality rates with Bernoulli-exponential distribution model. From these estimates it is observed that the mean from posterior marginal distribution is closer to the sample mean given in Table 1.

Table 7. Marginal prior and posterior means and coefficients of variation for the Bernoulli-exponential mortality model for different structural types.

Construction Class	Prior Marginal Mean, μ'_x	Prior Marginal CoV, σ'_x/μ'_x	Posterior Marginal Mean, μ''_x	Posterior Marginal CoV, σ''_x/μ''_x	Sample Mean	Sample CoV
Concrete	0.138	1.569	0.113	1.744	0.125	1.556
Stone	0.058	2.539	0.087	1.924	0.092	1.530
Mud and Stone	0.058	2.561	0.076	2.056	0.072	1.590
All	0.010	1.876	0.098	1.875	0.107	1.684



a) Concrete b) Stone

Figure 6. Prior and posterior marginal cumulative densities for the Bernoulli-exponentially distributed mortality rate for (a) concrete structures and (b) stone structures.

Figure 6 shows the marginal cumulative distribution functions (CDF) for the concrete and stone structural classes. It can be observed that in all the cases the posterior distributions fit the data closely for all values of mortality rate. More importantly, the distributions appear to reflect the zero mortality rate better than when only the exponential distribution was used to model X. Thus, modeling the zero and nonzero mortality rates separately can have a significant impact on the fitting of the Pakistan data.

BERNOULLI-GAMMA BAYESIAN ANALYSIS

From previous analysis, in almost all cases it was found that the gamma distribution fits the data better than the exponential distribution particularly for high values of mortality rates. Thus, a combined Bernoulli-gamma model is developed for the zero and non-zero mortality rates.

The occurrence of a mortality rate X equaling 0 is again defined as a Bernoulli distribution with probability p_0, and the distribution of X given that it is greater than 0 is defined as a gamma distribution. Therefore, in this case the probability distribution of X is defined according to equation (14).

$$f(x|\alpha,\beta,p_0) = \begin{cases} p_0 & \text{for } x = 0 \\ (1-p_0)\dfrac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\kappa-1} e^{-\beta x} & \text{for } 0 < x \leq 1 \end{cases} \quad (14)$$

The Bayesian analysis of the Bernoulli distribution remains the same as described in the previous section. The conjugate distribution of the gamma distribution for the non-zero mortality rates is given by equation (15). This conjugate distribution has $p, q, r,$ and s as hyper-parameters of the joint probability distribution of α and β. Then, the posterior distribution on the parameter will have the same form as equation (14) with the values of the hyper-parameters updated with the information obtained from the field observations. The relationship between the prior and the posterior hyper-parameters for the probability density are given by equation (16).

$$f(\alpha,\beta) = \frac{p^{\alpha-1} e^{-\beta q}}{K\Gamma(\alpha)^{r} \beta^{-\alpha s}} \quad \text{for} \quad \alpha > 0, \beta > 0 \quad (15)$$

$$
\begin{gathered}
K = \int_0^\infty \frac{p^{\alpha-1}\Gamma(s\alpha+1)}{\Gamma(\alpha)^r q^{s\alpha+1}} \\
p_{\text{posterior}} = p_{\text{prior}} P \\
q_{\text{posterior}} = q_{\text{prior}} + S \\
r_{\text{posterior}} = r_{\text{prior}} + n_{\text{non-zero}} \\
s_{\text{posterior}} = s_{\text{prior}} + n_{\text{non-zero}} \\
P = \prod\nolimits_{\{i=1,2,\ldots,n | x_i \neq 0\}} x_i \\
S = \sum\nolimits_{\{i=1,2,\ldots,n | x_i \neq 0\}} x_i \\
n_{\text{non-zero}} = \sum\nolimits_{i=1}^{n} 1\{x_i \neq 0\}
\end{gathered}
\tag{16}
$$

Prior and Posterior Distributions for p_0 and the Gamma Parameters α and β

As before the prior mean values of mortality rate are taken from PAGER. The prior values of parameters $p, q, r,$ and s are chosen randomly and different values were tested. It was observed that the posterior distribution is not affected by this choice of prior values as the data dominate the posterior parameters. Again, the prior mean value of p_0, μ_{p_0}, was chosen so that the mean value of the prior distribution of X for all values greater than or equal to zero, given in Equation (3), approaches to the mortality rates reported by PAGER, and the coefficient of variation of p_0 is assumed to be 0.3 for all the construction types.

In this analysis, the prior, sample and posterior mean of the zero mortality rates were taken identical to the values reported in Table 5 and are repeated here for completeness. The prior and posterior hyper-parameters for the beta conjugate distribution of the Bernoulli model are given in Table 8. Table shows the values of the prior, sample and posterior means for the Bernoulli-gamma combined distributions for the four structural types considered in this analysis.

Table 8. Prior and posterior hyper-parameters of the conjugate distribution of the gamma-distributed non-zero mortality rates

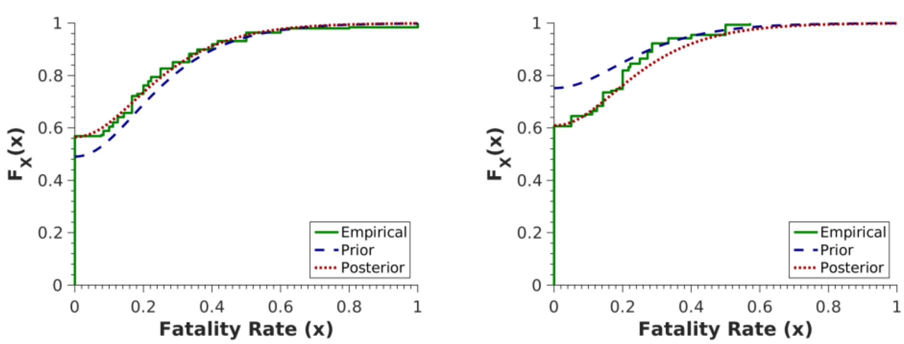
Construction Class	p_{prior}	q_{prior}	r_{prior}	s_{prior}	$p_{posterior}$	$q_{posterior}$	$r_{posterior}$	$s_{posterior}$
Concrete	2	1.23	1.50	1	1.43e-66	32.34	108.5	108
Stone	2	1.23	1.50	1	1.50e-43	15.43	62.5	62
Mud and stone	2	1.23	1.50	1	9.13e-12	4.18	16.5	16
All	2	1.23	1.50	1	3.18e-70	30.65	109.5	109

Marginal Distributions of Bernoulli-Gamma Mortality Rates

Marginal probability density function of non-zero mortality rate can be computed using the total probability theorem using equation (15).Th

$$f(x) = \begin{cases} \mu_{p0} & \text{for } x = 0 \\ (1-\mu_{p0}) \iint f(x|\alpha,\beta) f(\alpha,\beta)\, d\alpha\, d\beta & \text{for } 0 < x \leq 1 \end{cases} \quad (17)$$

where $f(x|\alpha,\beta)$ is the parent distribution of the mortality rate X and $f(\alpha,\beta)$ is the joint probability density function of the parameters α,β. μ_{p0} is the mean value of p_0 obtained from the beta distribution of p_0. The marginal distribution defined by equation (17) is again evaluated using numerical integration. Figure 7 shows the cumulative posterior marginal distribution function for all four structural classes as compared to the data. In all the cases the posterior distributions fit the data very closely for all values of mortality rate. This fit is significantly better than any of the previous combinations of analyses.



a) Concrete **b) Stone**

Figure 7. Posterior marginal cumulative densities for the Bernoulli-Gamma distributed mortality rate for (a) concrete structures (b) stone structures (c) mud and stone structures and (d) all structures

Table 9. Marginal prior and posterior means and standard deviations for the Bernoulli-Gamma distributed mortality rate model for different structural types

Construction Class	Prior Marginal Mean, μ'_x	Prior Marginal CoV, σ'_x/μ'_x	Posterior Marginal Mean, μ_x	Posterior Marginal CoV, σ_x/μ_x	Sample Mean	Sample CoV
Concrete	0.150	1.292	0.1278	1.460	0.125	1.556
Stone	0.060	2.163	0.1026	1.436	0.092	1.530
Mud and Stone	0.060	2.281	0.1043	1.424	0.072	1.590
All	0.110	1.566	0.1145	1.475	0.107	1.684

Table 11 shows the marginal posterior mean and standard deviation for the various structural types considered in the study. The mean values of the marginal distributions are very close to the mean values obtained from the posterior Bernoulli-gamma distribution. The standard deviation of the marginal mortality rate is slightly lower than the standard deviations obtained from the posterior mortality rate analysis before integrating over the uncertainties on the parameters.

CONCLUSIONS

In this paper a Bayesian formulation is presented for developing vulnerability functions and the methodology is used for treating the uncertainties in earthquake mortality rate data. Three different models are presented and tested with information from PAGER and from the 2005 Pakistan earthquake. The models included exponentially distributed mortality rates for all values greater than or equal to zero, a combined Bernoulli-exponential model and a combined Bernoulli-gamma model for zero and non-zero mortality rates. 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 mortality rates 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 Bayesian 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 mortality rates are not well represented and introduce a bias on the mortality rates greater than zero prompting a separate treatment of the zero and non-zero mortality rates. The Bernoulli distribution is used for the zero mortality rate and either an exponential or gamma distribution is used for the non-zero mortality rate referred to respectively as the Bernoulli-exponential and the Bernoulli-gamma models. Novel Bayesian formulations are introduced for the Bernoulli-exponential and the Bernoulli-gamma models. The conjugate distribution for the Bernoulli distribution parameter is the beta. The exponential model remains the same as in the first analysis. A conjugate distribution for the gamma non-zero rates is then defined and used for the Bayesian analysis of the Bernoulli-gamma model. The formulations for the Bayesian marginal distributions that integrate over all the parameter pdf's are developed for all models. These distributions are evaluated numerically as closed form solutions cannot be obtained. Prior and posterior mean values, standard deviations and coefficients of variation are also estimated for all models.

Based on the analysis conducted for the Bernoulli-exponential and the Bernoulli-gamma combined models, it is found that the Bernoulli-gamma functions are closest to the observed data. In majority of cases, both the Bernoulli-exponential and the Bernoulli-gamma models have posterior CoV's that are considerably smaller than those for the prior or data pointing to one of the main advantages of using Bayesian analysis. For the Bernoulli-gamma model, the posterior marginal CoV's of the casualty rates are smaller than the prior CoV's, except for in one case. It is recommended that the Bernoulli-gamma model be used as this model appears to best capture the zero and the non-zero rates.

ACKNOLEDGEMENT

This research was partially supported by the Global Earthquake Model, by the National Science Foundation Grant CMMI 1233694 and the Shah Family Graduate Fellowship.

REFERENCES

Ang, A. H.-S., and Tang, W. H. (1975). *Probability concepts in engineering planning and design*, Vol. I, John Wiley & Sons, New York, NY.

Choe, D.-E., Gardoni, P., & Rosowsky, D. (2007). Closed-Form Fragility Estimates, Parameter Sensitivity, and Bayesian Updating for RC Columns. *Journal of Engineering Mechanics*, *133*(7), 833–843.

Der Kiureghian, A. (1999a). "Fragility estimates for electrical substation equipment," *Proc. 5thU.S. Conference on Lifeline Earthquake Engineering*, August 1999, Seattle, WA, W.Elliott and P. McDonough, Eds, 643-652.

Der Kiureghian, A. (1999b). "A Bayesian framework for fragility assessment," *Proc. 8th Int.Conf. On Applications of Statistics and Probability* (ICASP) *in Civil Engineering Reliability and Risk Analysis*, Sydney, Australia, December 1999, R.E. Melchers and M. G. Stewart, Eds., 1, 219-226.

Der Kiureghian, A. (2002). "Bayesian methods for seismic fragility assessment of lifeline components," *In Acceptable Risk Processes: Lifelines and Natural Hazards, ASCE Council on Disaster Reduction and Technical Council on Lifeline Earthquake Engineering* Monograph N0. 21, C. Taylor and E. VanMarcke, Editors, March 2002, 61-77.

Earle, P.S., Wald, D.J., Jaiswal, K.S., Allen, T.I.; Hearne, M.G., Marano, M.G., Hotovec, A.J., and Fee, J.M (2009). *Prompt Assessment of Global Earthquakes for Response (PAGER)—A system for rapidly determining the impact of earthquakes worldwide: U.S*. Geological Survey Open-File Report 2009-1131, 15 p.

Ferreira, M. A., Oliveira C.S., and Mota de Sá, F. 2011. *"Estimating Casualties for the Southern California Shakeout."* Pp. 255–66 in Human Casualties in Earthquakes: Progress in Modelling and Mitigation. Springer.

Gardoni, P., Mosalam, K. M. and Der Kiureghian, A. D. (2003). "Probabilistic Seismic Demand Models and Fragility Estimates for RC Bridges," *Journal of Earthquake Engineering*, 7:S1, 79-106.

Gardoni, P., Der Kiureghian, A., & Mosalam, K. M. (2002). Probabilistic Capacity Models and Fragility Estimates for Reinforced Concrete Columns based on Experimental Observations. *Journal of Engineering Mechanics*, *128*(10), 1024–1038.

Huang, Q., Gardoni, P. and Hurlebaus, S. (2010). "Probabilistic Seismic Demand Models and Fragility Estimates for Reinforced Concrete Highway Bridge with One Single-Column Bent," *Journal of Engineering Mechanics*, 136(11): 1340-1353.

Jaiswal, K., Wald, D.J., Earle, P.S., Porter, K., and Hearne, H. (2009). "Earthquake casualty models within the USGS Prompt Assessment of Global Earthquakes for Response (PAGER) system". Pp. 1–*10 in Second International Workshop on Disaster Casualties.*

Koutsourelakis, P.S. (2010). Assessing structural vulnerability against earthquakes using multi-dimensional fragility surfaces: A Bayesian framework, *Probabilistic Engineering Mechanics*, Vol. 25, pp.49–60

Petal, M. 2011. "Earthquake Casualties Research and Public Education." Pp. 25–50 in Human Casualties in Earthquakes: Progress in Modelling and Mitigation. Springer.

Raiffa, H. and Schlaifer, N. (2000). *Applied Statistical Decision Theory*, John Wiley & Sons, New York, NY.

Rice, J. A. (1999). *Mathematical Statistics and Data Analysis*, Second Edition, Duxbury Press, Second Edition, New York.

Singhal, A., and Kiremidjian, A. S. (1998). Bayesian Updating of Fragilities with Application to RC Frames. *Journal of Structural Engineering,* Vol. 124, No. 8,pp 922-929.

Shoaf, K., H. Seligson, M. Kano, and Ramirez M. (2005). "Fatality Model for Non-Ductile Concrete Frame Structures Developed From Golcuk Population Survey Data." in *Van Nuys Hotel Building Test Bed Report: Exercising Seismic Performance Assessment*. (H. Krawinkler, Ed.) Pacific Earthquake Engineering Research Center.

Wald, D.J., Jaiswal, K.S., Marano, K.D., Bausch, D.B., and Hearne, M.G (2010). *PAGER — Rapid assessment of an earthquake's impact*. U.S. Geological Survey Fact Sheet 2010–3036, 4 p.

Zhong, J., Gardoni, P., Rosowsky, D. and Haukaas, T. (2008). "Probabilistic Seismic DemandModels and Fragility Estimates for Reinforced Concrete Bridges with Two-ColumnBents," *Journal of Engineering Mechanics,* 134(6): 495-504.