16th World Conference on Earthquake Engineering, 16WCEE 2017

Santiago Chile, January 9th to 13th 2017

Paper N° 2402

Registration Code: S-O1465852734

ACCOUNTING FOR UNCERTAINTY IN EARTHQUAKE FRAGILITY CURVES

D. Lallemant[1], A. Kiremidjian[2]

[1] *PhD Student, Stanford University, davidcbl@stanford.edu*

[2] *Professor, Stanford University, ask@stanford.edu*

Abstract

This paper describes statistical procedures for characterizing and accounting for uncertainty in earthquake fragility models. Both fully analytical and non-parametric bootstrap methods are used to describe the conditional probability distribution of damage exceedance given an intensity measure. This enables the development of confidence intervals for fragility curves for any confidence level of interest. When analyzing annual collapse rate, the uncertainty in fragility curves gets propagated when integrated with the seismic hazard curve. This study therefore proposes methods to estimate the moments as well as the full distribution of the resulting annual damage exceedance rate. This is a significant improvement from current practice, which only use the "expected fragility" to integrate with the hazard curve, thus producing a single value for annual collapse rate. Using an example for a building analyzed through incremental dynamic analysis for a site in Oakland CA, this study demonstrates the significant uncertainty surrounding the annual collapse rate and demonstrates simplified methods to characterize this uncertainty through a closed-form beta-distribution model.

Keywords: Fragility curves, uncertainty modeling, beta distribution, bootstrap method

1. Introduction

In performing seismic loss estimation it is necessary to characterize the vulnerability of the exposed portfolio. This vulnerability is usually described by relating a ground motion intensity measure (IM) to the probability of exceeding various damage states (DS), a process which has become fundamental for analyzing the risk of buildings and structures **[1]–[5]**. These damage to ground-motion relationships can be developed through expert judgment **[6]**, based on structural analysis **[7]–[9]** , or empirical information **[10]–[13]**. Regardless of the source of data, the creation of a fragility "model" involves some form of regression-based analysis of the data, resulting in a one-to-one relationship between a ground-motion intensity measure and a probability of experiencing a certain level of damage, taking the form of a continuous fragility curve or a discreet damage probability matrix.

When discussing fragility, the response of a structural system can be described in binary terms; a building either exceeds or does not exceed a specific damage state. A fragility curve therefore describes the mean conditional response of the system as a function of a ground-motion intensity measure. This conditional mean response is the probability of exceeding the damage state of interest. Due to the limited data used to fit any regression model, the fragility curve (mean response) has uncertainty, which can be characterized in terms of confidence intervals. The uncertainty in future response is the combination of the uncertainty in the fragility curve, as well as the variability in response around this curve. It is described in terms of prediction intervals. In a strict sense, the prediction interval for any single building is always {0,1}, since the data itself is binary (e.g. collapse or non-collapse).

This paper describes both analytical (closed form) and simulation-based methods for calculating the uncertainty in fragility models. While applicable for any damage state, this paper focuses on collapse as the main example. In addition, the study explores methods to propagate the uncertainty in the fragility model so as to characterize the probability distribution of collapse rate. Results from this paper make evident the importance of accounting for uncertainty in fragility models. In particular, it demonstrates that collapse rate is very sensitive to uncertainty in fragility models. A simplified method is proposed to model the probability distribution of collapse rate without the need for any simulation.

Two data-sets are used to demonstrate the methods described. A hypothetical empirical earthquake damage data set is shown in Table 1. Since it is an empirical data set (presumably gathered as part of a post-earthquake damage survey), it contains different number of buildings at different and non-uniformly distributed intensity measures. The buildings are of the same structural building type and only the collapse damage state is used as demonstration.

The data from Table 2 are obtained from the collapse performance assessment of an 8-story infill frame building using the Incremental Dynamic Analysis technique **[14]**, based on the methodology developed by Burton and Deierlein **[15]** for simulating seismic collapse in non-ductile reinforced concrete frame buildings with infill. Nonlinear dynamic analysis was conducted on a two-dimensional model developed in OpenSees **[16]** using 44 far-field ground motions. The first mode spectral acceleration (Sa) is used as the ground motion intensity measure. The results from this analytical model take the form of a count of the number of ground motions leading to collapse for every incremental intensity measure.

Table 1. Empirical earthquake damage data (hypothetical data set)

IM (PGA)	0.1	0.11	0.12	0.14	0.19	0.2	0.22	0.3	0.4	0.42	0.46	0.6
Collapse	1	0	0	1	1	1	4	3	6	9	1	29
NonCollapse	31	9	20	15	10	13	38	12	44	24	4	11
IM (PGA)	0.6	0.72	0.8	1	1.04	1.2	1.2	1.22				
Collapse	7	32	17	18	3	8	3	5				
NonCollapse	4	12	3	8	3	5	4	1				

Table 2. Analytical data from Incremental Dynamic Analysis of an 8-storey concrete frame building with infill.

IM (Sa)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2
Collapse	0	0	0	0	1	6	9	18	22	27	32	33
NonCollapse	44	44	44	44	43	38	35	26	22	17	12	11
IM (Sa)	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2	2.1	2.2	2.3	2.4
Collapse	37	38	38	40	41	41	42	43	43	43	44	44
NonCollapse	7	6	6	4	3	3	2	1	1	1	0	0

2. A simple fragility curve model

The most common functional form used to describe earthquake fragility is the lognormal cumulative density distribution (CDF), as it has been found to provide good representation of earthquake damage fragility [9], [17]. It further has the characteristic of multiplicative reproducibility (the product of lognormally distributed random variables are lognormally distributed), which is convenient for reliability-based analysis [18], [19]. Several methods have been used for regression of lognormal CDF models, including method of moments [3], least squares estimation (LSE), and maximum likelihood estimation (MLE) . Alternative models have also been used, including other parametric models such as logistic regression [18], [20]–[23], semi-parametric models such as generalized additive models [24], [25] and fully non-parameteric models such as kernel-based regression [26]. This paper will use the parametric lognormal CDF model to illustrate characterization and propagation of uncertainty in damage prediction. The maximum likelihood estimation method is used to estimate the model parameters, as it is statistically robust and generalizable to any data type.

The dataset shown in Table 1 describes the exceedance and non-exceedance of a particular damage state for buildings at a specific ground-motion intensity (assumed recorded or estimated). While this represents a hypothetical empirical damage data set, the exact same analysis can be done on data from incremental dynamic analysis (IDA), replacing “number of buildings” with “number of ground motions.” Also while this example is conducted for the collapse damage state, the same process would be conducted for all other damage states of interest.

A lognormal-CDF fragility curve fit by MLE can also be obtained through a generalized linear model (GLM) based regression. This alternative formulation is more convenient, as it is readily available in standard statistics software and GLM theory is very well established in statistics literature. Specifically, the lognormal CDF model fit with MLE is exactly equivalent to a GLM model with probit link function fit with the log of intensity measure [20], [24], [26], [23].

$$E[y] = \mu_y = g^{-1}(\eta) = g^{-1}(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n) \tag{1}$$

where $E[y]$ is the expected response given predictor variables $X_1, X_2, \ldots, X_n$, and can therefore be thought of as the mean response μ_y. The term η is the linear predictor, which is related to the expected response through the link function $g()$. For developing a fragility function, Eq. 1 reduces to:

$$\mu_y(IM) = P(DS > ds|IM) = g^{-1}(\hat{\beta}_0 + \hat{\beta}_1 log(IM)) = \Phi(\hat{\beta}_0 + \hat{\beta}_1 log(IM)) \tag{2}$$

where $g^{-1}()$ is the inverse probit link function equal to the cumulative distribution function of the standard normal distribution $\Phi()$, and $\widehat{\beta_0}$ and $\widehat{\beta_1}$ are the MLE parameter estimates for the regression model. This GLM model will be used in the next section for the development of closed-form confidence intervals.

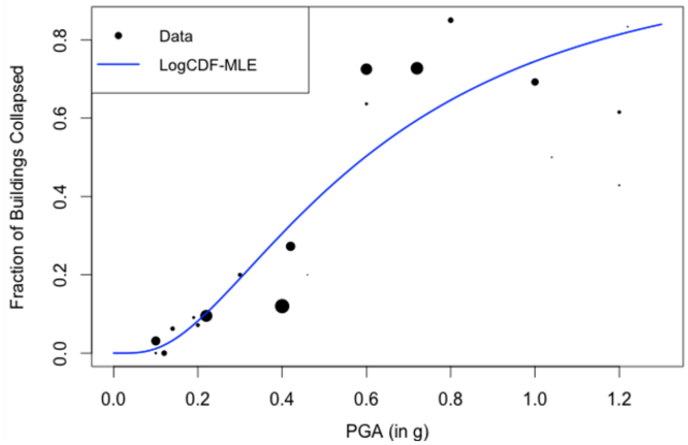

Fig. 1 Lognormal CDF fragility curve fitted using Maximum Likelihood Estimation (MLE) method

3. Uncertainty in Fragility Curves

One of the most common methods to measure and communicate uncertainty is by developing confidence intervals. Two methods are proposed for developing confidence intervals on fragility curves. The first consists of deriving the full analytical conditional distribution of expected probability of collapse. This derivation can be applied for any generalized linear model, including the lognormal CDF model described previously, as well as logistic, probit or cloglog models. The second approach utilizes the non-parametric bootstrap method and is applicable to all models including semi-parametric generalized additive models and the non-parametric kernel smoothing [23].

3.1 Closed form confidence intervals for fragility curves based on asymptotic normality:

Recall from Eq. (1) that GLMs relate the mean of a response variable linearly to independent variables through a link function transformation. For developing a fragility function, the GLM model has only a single independent variable, as described by Eq. (2). The model is solved by MLE method, which provides parameter estimates $\widehat{\beta_0}$ and $\widehat{\beta_1}$, based on assumptions of a conditional distribution of the exponential family.

Confidence intervals, which define the uncertainty in the expected value (uncertainty in the mean curve itself), can therefore be defined based on the uncertainty in parameter estimates. Standard MLE theory states that parameter estimates are asymptotically normal, with covariance matrix equal to the inverse of the Fisher-information matrix [28]. More information on this derivation, called the "Cramér-Rao bound" can be found in various statistics textbooks [29]. For general intuition the Fisher-information measures the amount of information that an observation carries about unknown parameters needed to define the probability distribution of that observation. For fragility modeling, GLM theory hence states that the estimates $\widehat{\beta_0}$ and $\widehat{\beta_1}$ are unbiased and have bivariate normal distribution of the form:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \sim N\left[\begin{matrix} \beta_0 \\ \beta_1 \end{matrix}, \begin{matrix} \sigma^2_{\beta_0} & \sigma_{\beta_0\beta_1} \\ \sigma_{\beta_0\beta_1} & \sigma^2_{\beta_1} \end{matrix} \right] = N\left[\begin{matrix} \beta_0 \\ \beta_1 \end{matrix}, I^{-1} \right] \quad (3)$$

where I^{-1} is the inverse of the Fisher-information matrix.

Since the parameter estimates have an asymptotically multivariate normal distribution, their linear combination is also normal. In other words, the linear predictor $\eta = \widehat{\beta_0} + \widehat{\beta_1}\log(IM)$ has normal distribution centered at $\widehat{\beta_0} + \widehat{\beta_1}\log(IM)$ such that:

$$\mu_\eta(IM)=\beta_0+\beta_1 log(IM) \tag{4}$$

$$\sigma_\eta^2(IM)=\sigma_{\beta_0}^2+\sigma_{\beta_1}^2 log(IM)^2+2log(IM)\sigma_{\beta_0\beta_1} \tag{5}$$

The variance and covariance terms are obtained from the corresponding indices of the inverse of the Fisher-information matrix I^{-1}:

$$I_{i,j}=-E\left[\frac{\partial^2}{\partial\beta_i\,\partial\beta_j}\ell(\beta)\right] \tag{6}$$

where $I_{i,j}$ is the $(i,j)^{th}$ term of the Fisher-information matrix, and $\ell(\beta)$ is the log-likelihood function used for the MLE estimation.

While the Fisher-information equations can sometimes be solved analytically by taking the partial derivatives of the log-likelihood function, in most cases they need to be solved numerically. Conveniently, this covariance matrix is actually a bi-product of the MLE estimation used to fit the original curve, as it is needed for the Iteratively Reweighted Least Square algorithm. It can therefore be readily extracted from standard software packages (R Statistical Computing software, Matlab).

Given the basic formulation of the GLM described by Eq. (1), it follows that the mean probability of collapse μ_y is the transformation of a normally distributed random variable η through the link function $g^{-1}() = \Phi$. Since this transformation is strictly monotonic, the distribution of μ_y can be related to the distribution of η through substitution:

$$F_{\mu_y}(\mu_y)=F_\eta(\eta(\mu_y)) \tag{7}$$

$$\mu_y=\Phi(\eta)\Rightarrow\eta(\mu_y)=\Phi^{-1}(\mu_y) \tag{8}$$

We know that η is normally distributed with parameters described in equations Eq. (4) and Eq. (5). Therefore:

$$F_{\mu_y}(\mu_y)=F_\eta(\eta(\mu_y))=\Phi\left(\frac{\eta(\mu_y)-\mu_\eta}{\sigma_\eta}\right)=\Phi\left(\frac{\Phi^{-1}(\mu_y)-\mu_\eta}{\sigma_\eta}\right) \tag{9}$$

Any percentile of interest p can then be obtained by solving the inverse distribution function. We find the value of Q for which $F_y(Q) = p$:

$$\begin{aligned} F_{\mu_y}(Q) &= p=\Phi\left(\frac{\Phi^{-1}(Q)-\mu_\eta}{\sigma_\eta}\right) \\ \Rightarrow Q(p) &= \Phi(\Phi^{-1}(p)\sigma_\eta+\mu_\eta) \end{aligned} \tag{10}$$

The moments $\mu_\eta$and σ_η terms from equations Eq. (4) and Eq. (5). can be substituted into Eq. (10) to define the confidence interval corresponding to any percentile limits of interest.

We can further obtain the full probability density function for the expected collapse probability. Once again, since the transformation function $\Phi()$is strictly monotonic, the distribution of μ_y can related to the distribution of η:

$$f_{\mu_y}(\mu_y)=f_\eta(\eta(\mu_y))\left|\frac{\partial\eta}{\partial\mu_y}\right| \tag{11}$$

We conduct the appropriate substitution:

$$\eta(\mu_y) = \Phi^{-1}(\mu_y) \tag{12}$$

$$f_\eta(\eta(\mu_y)) = \phi\left(\frac{\Phi^{-1}(\mu_y) - \mu_\eta}{\sigma_\eta}\right) \tag{13}$$

$$\frac{\partial \eta}{\partial \mu_y} = \frac{\partial}{\partial \mu_y}(\Phi^{-1}(\mu_y)) = \frac{1}{\phi(\Phi^{-1}(\mu_y))} \tag{14}$$

Combining results from Eqs. 12, 13, 14 into 11, we obtain:

$$f_{\mu_y}(\mu_y) = f_\eta(\eta(y)) \left| \frac{\partial \eta}{\partial \mu_y} \right| = \frac{\phi\left(\frac{\Phi^{-1}(\mu_y) - \mu_\eta}{\sigma_\eta}\right)}{\phi(\Phi^{-1}(\mu_y))} \tag{15}$$

Which can also be rewritten as:

$$f_{\mu_y}(\mu_y) = \frac{1}{\sigma_\eta} e^{-\frac{1}{2\sigma_\eta^2}[(1-\sigma_\eta^2)(\Phi^{-1}(\mu_y))^2 - 2\mu_\eta \Phi^{-1}(\mu_y) + \mu_\eta^2]} \tag{16}$$

Fig. 2 shows an example of the curve for expected probability of collapse, its 90% confidence interval and its complete conditional distribution at two intensity measures of interest.

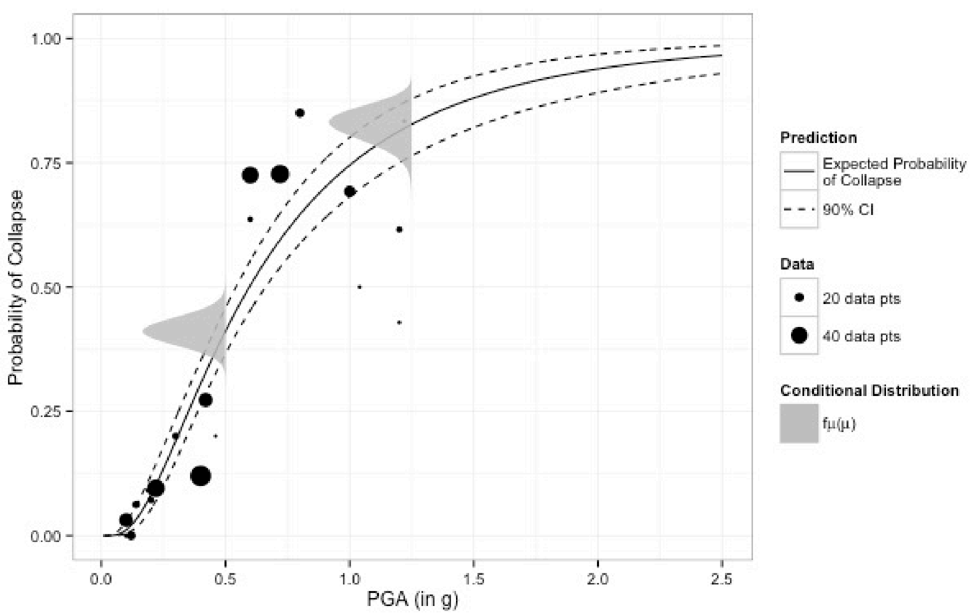

Fig. 2 Confidence interval (dashed lines) and conditional distribution of expected probability of collapse at PGA=0.5g and 1.25g.

Eq. (10) and Eq. (15) can be generalized for other link functions $g()$, such that:

$$Q(p) = \Phi(g(p)\sigma_\eta + \mu_\eta) \tag{17}$$

$$f_{\mu_y}(\mu_y)=\frac{\phi\left(\frac{g(\mu_y)-\mu_\eta}{\sigma_\eta}\right)}{g^{'-1}(g(\mu_y))} \tag{18}$$

where $g'^{-1}()$ is the first derivative of the inverse link function.

3.2. Confidence intervals through non-parametric bootstrap method

Another common method to obtain the confidence intervals on regression curves is through the non-parametric bootstrap method. In this approach, regression analysis is conducted iteratively for numerous random sub-samples of the data. The process involves randomly sampling datasets with replacement from the training dataset, each sample of equal size to the original dataset [28]. The sampling is done on the binary data, not the fraction of buildings exceeding the damage state. Regression is conducted for each bootstrap dataset, from which a distribution of regression curves (or parameter estimates) is obtained. The pointwise confidence band can then be obtained by taking the $p^{th}$and $(1-p)^{th}$ percentile of all bootstrap curves at each IM level (the bootstrap percentile interval), or by plugging in the mean, variance and correlation of the bootstrap parameter estimates into Eq. (4) and Eq. (5) and plugging those in the quantile function of Eq. 10 (the normal interval).

The bootstrap method has the advantage of characterizing the uncertainty of any regression model, including non-parametric models. The analytic and bootstrap-based confidence bands for parametric models (including all GLMs) will coincide as the number of bootstrap samples approaches infinity [28].

The confidence intervals describe the uncertainty in the expected probability of collapse for any ground shaking intensity. The next section describes a method to characterize the uncertainty in annual collapse rate, an important measure for understanding the risk of structural collapse over the lifetime of a building.

4. Uncertainty in Collapse Rate

The rate of collapse is computed by integrating the collapse fragility curves with the ground motion hazard curve, as shown in Eq.19.

$$\lambda_c=\int_{IM_{\min}}^{IM_{\max}} P(C|IM=im)\cdot|d\lambda_{IM}(im)| \tag{19}$$

where $P(C|IM=im)$ is the probability of collapse at a ground motion intensity im and $\lambda_{IM}(im)$ is the mean annual frequency of exceedance of a ground motion intensity im. In practice this is computed by summing over all intensity measures the product of the probability of collapse conditioned on IM and the change in hazard curve over the IM increment:

$$\lambda_c=\sum_{im_i=IM_{\min}}^{IM_{\max}} P(C|im_i)\cdot|\lambda(im_i)-\lambda(im_{i+1})| \tag{20}$$

Typically, the annual collapse rate is computed with the expected fragility curve and therefore does not account for the uncertainty embeded in the fragility model. The uncertainty in annual collapse rate can be computed exactly by conducting bootstrap simulations similar to the process for obtaining non-parametric bootstrap confidence interval on the fragility curve described earlier. In this way, each bootstrap simulated fragility curve is integrated over the hazard curve following Eq. (19). The distribution over numerous bootstrap samples defines the empirical distribution of annual collapse rate.

Alternatively, the distribution of annual collapse rate can be estimated. It follows from Eq. (20) that the annual collapse rate λ_C is the sum of the product of random variable terms $P(C|im_i)$ with constants terms $\lambda(im_i)-\lambda(im_{i+1})$. The distribution of collapse probability conditioned on IM was derived earlier and is described in Eq. (15). This distribution is centered on the mean collapse fragility curve and is bounded between 0 and 1. It can in fact be closely approximated by a beta distribution. Following this line of thinking, λ_C can be approximated as the sum of beta distributed random variables scaled by constant terms. While this has no simple solution, Johannesson (1995) [30] provided a method to estimate the sum of beta distributed random variables as another beta distribution with matching moments. For our purposes, the moments of distribution for annual collapse rate are estimated as:

$$\mu_{\lambda_c}=\sum_{i=1}^{IM_{\max}}\mu_{P(C|im_i)}\cdot|\lambda_i-\lambda_{i+1}|=\sum_{i=1}^{IM_{\max}}\mu_{\mu_y}(im_i)\cdot|\lambda_i-\lambda_{i+1}| \tag{21}$$

$$\sigma_{\lambda_c}^2=\sum_{i=1}^{IM_{\max}}\sum_{j=1}^{IM_{\max}}|\lambda_i-\lambda_{i+1}|\,|\lambda_j-\lambda_{j+1}|\,\rho_{ij}\sigma_{\mu_y,i}\sigma_{\mu_y,j} \tag{22}$$

Since the collapse probability distributions conditioned on any two intensity measures im_i and im_j are linked to the same parameters $\widehat{\beta_0}$ and $\widehat{\beta_1}$, these distributions are therefore perfectly correlated. Therefore the correlation coefficient $\rho_{ij}=1$. The terms $\mu_{\mu_y}(im_i)$ and $\sigma_{\mu_y,i}$ are the moments of the probability of collapse conditioned on $IM=im_i$ and can be quickly estimated using the first-order mean centered approximation:

$$\mu_{\mu_y}=g^{-1}(\mu_\eta)=\Phi(\mu_\eta) \tag{23}$$

$$\sigma_{\mu_y}\approx\left.\frac{\partial g}{\partial\eta}\right|_{\eta=\mu_\eta}\sigma_\eta=\phi(\mu_\eta)\sigma_\eta \tag{24}$$

We can now obtain the approximate moments of the collapse rate distribution by plugging in the conditional moments from equations Eq. (23) and Eq. (24) into Eq. (21) and Eq. (22), and obtain its estimated beta distribution parameters:

$$\lambda_c\sim Beta(\alpha,\beta)$$

$$\alpha=\mu_{\lambda_c}\left(\frac{\mu_{\lambda_c}(1-\mu_{\lambda_c})}{\sigma_{\lambda_c}}-1\right) \tag{25}$$

$$\beta=(1-\mu_{\lambda_c})\left(\frac{\mu_{\lambda_c}(1-\mu_{\lambda_c})}{\sigma_{\lambda_c}}-1\right) \tag{26}$$

These results will be used in the following sections to quantify the uncertainty in collapse risk of a building.

4. Estimation of the full probability distribution of collapse rate

Much of the uncertainty in the collapse risk of a structure arises from the limited number of ground-motions used to fit the analytical collapse fragility curve. Table 2 contains the incremental dynamic analysis results for an 8-story infill frame building subjected to 44 ground-motions at IM increments of 0.1g. This data-set will be used to develop a collapse fragility curve for this building and evaluate its annual collapse rate.

The building of interest is located in Oakland, California. The hazard curve for the site is computed using the open source seismic hazard assessment tool OpenSHA [31] for $IM=Sa(1s)$ at a site with $V_{s30}=760\,{}^{m}/_{s}$.

In a first demonstration, the GLM model from Eq. (2) (lognormal CDF fit) is fit to the data from Table 2. This data was obtained by performing 1056 dynamic non-linear analyses (44 ground-motions at 24 increments). The non-parametric bootstrap simulation method is used to generate 10,000 fragility curves, each integrated with the hazard curve following Eq. (19) to obtain an empirical distribution of collapse rate. The algorithm below further describes the steps to obtain the estimated beta distribution of collapse rate.

Algorithm 1 Probability Distribution of Collapse Rate

1: Obtain β_0, β_1 (parameter estimates)

2: Obtain $\sigma^2_{\beta_0}, \sigma^2_{\beta_1}$ and $\sigma_{\beta_0\beta_1}$ (parameter covariance) ▷ from GLM fitting or bootstrap

3: $IM = 0, 0.01, 0.02, ..., 3$ ▷ IM is sequence from 0 to large IM value.

4: Obtain $\lambda(IM)$ ▷ site-specific hazard curve

5: **for** $i = 1 \rightarrow \{\text{all IM}\}$ **do**

6: $\mu_\eta(IM_i) = \beta_0 + \beta_1 log(IM_i)$ ▷ from equation 4

7: $\sigma^2_\eta(IM_i) = \sigma^2_{\beta_0} + \sigma^2_{\beta_1} log(IM_i)^2 + 2log(IM_i)\sigma_{\beta_0\beta_1}$ ▷ from equation 5

Using results above

8: $\mu_{\mu_y}(IM_i) = \Phi\left(\mu_\eta(IM_i)\right)$ ▷ from equation 23

9: $\sigma_{\mu_y}(IM_i) = \phi\left(\mu_\eta\left(IM_i\right)\right)\sigma_\eta$ ▷ from equation 24

10: **end for**

11: $\mu_{\lambda_c} = \sum_{i=1}^{IM_{\max}} \mu_{\mu_y}(IM_i) \cdot \left|\lambda_i - \lambda_{i+1}\right|$ ▷ from equation 21

12: $\sigma^2_{\lambda_c} = \sum_{i=1}^{IM_{\max}} \sum_{j=1}^{IM_{\max}} \left|\lambda_i - \lambda_{i+1}\right|\left|\lambda_j - \lambda_{j+1}\right|\rho_{ij}\sigma_{\mu_y,i}\sigma_{\mu_y,j}$ ▷ from equation 22

where μ_{λ_c} and $\sigma^2_{\lambda_c}$ are the mean and variance of collapse rate

13: $\alpha = \mu_{\lambda_c}\left(\frac{\mu_{\lambda_c}(1-\mu_{\lambda_c})}{\sigma_{\lambda_c}} - 1\right)$ ▷ from equation 25

14: $\beta = (1-\mu_{\lambda_c})\left(\frac{\mu_{\lambda_c}(1-\mu_{\lambda_c})}{\sigma_{\lambda_c}} - 1\right)$ ▷ from equation 26

15: $\lambda_c \sim Beta\left(\alpha, \beta\right)$

where α and β are the estimated beta distribution parameters for the approximate full distribution of collapse rate

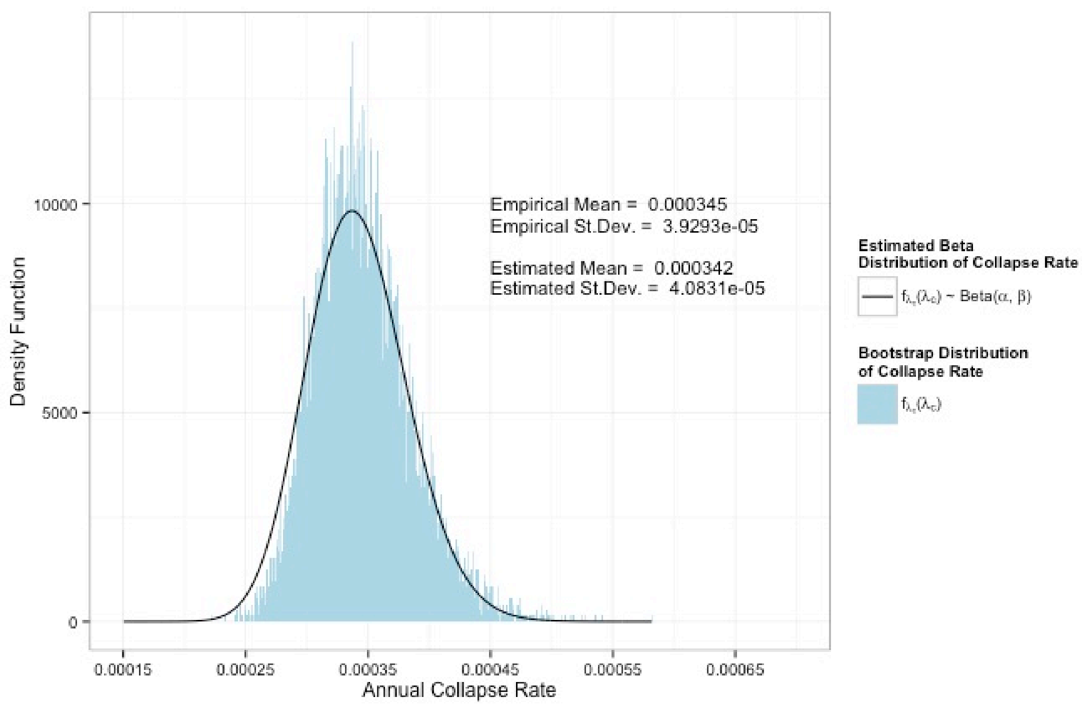

Fig. 3 - Bootstrap distribution of collapse rate (histogram) compared with the estimated distribution modeled as a beta distribution (black curve).

Fig. 3 demonstrates that the annual collapse rate has significant uncertainty, even while it resulted from over 1000 dynamic analyses. It further demonstrates that the estimated beta distributed collapse rate distribution is in very close agreement with the empirical distribution.

5. Conclusion

This paper provides exact analytical formulations for the conditional probability distribution of damage-state exceedance for any fragility curve that can be represented as a generalized linear model (including the lognormal CDF fragility curve formulation). This allows for the development of confidence intervals for any confidence level of interest. The uncertainty in fragility curves gets propagated when integrated with the seismic hazard curve. This study proposes methods to estimate the moments as well as the full distribution of the resulting annual damage exceedance rate. This is a significant improvement from current practice, which only use the "expected fragility" to integrate with the hazard curve, thus producing a single value for annual collapse rate. Using an example for a building analyzed through incremental dynamic analysis for a site in Oakland CA, this study demonstrates the significant uncertainty surrounding the annual collapse rate and demonstrates methods to measure this uncertainty.

6. References

[1] R. K. McGuire, "Seismic Hazard and Risk Analysis," *EERI Monograph*, 2004.

[2] C. A. Kircher, A. A. Nassar, O. Kustu, and W. T. Holmes, "Development of Building Damage Functions for Earthquake Loss Estimation," *Earthquake Spectra*, vol. 13, no. 4, pp. 663–682, Nov. 1997.

[3] K. Porter, R. Kennedy, and R. Bachman, "Creating Fragility Functions for Performance-Based Earthquake Engineering," *Earthquake Spectra*, vol. 23, no. 2, pp. 471–489, May 2007.

[4] R. P. Kennedy and M. K. Ravindra, "Seismic fragilities for nuclear power plant risk studies," *Nuclear Engineering and Design*, vol. 79, no. 1, pp. 47–68, May 1984.

[5] G. M. Calvi, R. Pinho, G. Magenes, and J. J. Bommer, "Development of seismic vulnerability assessment methodologies over the past 30 years," *ISET Journal of Earthquake Technology*, 2006.

[6] K. S. Jaiswal, W. Aspinall, and D. Perkins, "Use of Expert Judgment Elicitation to Estimate Seismic Vulnerability of Selected Building Types," *Proc 15th World Conference on Earthquake Engineering*, pp. 1–10, Jul. 2012.

[7] A. Singhal and A. S. Kiremidjian, "Method for Probabilistic Evaluation of Seismic Structural Damage," *Journal of Structural Engineering*, vol. 122, no. 12, pp. 1459–1467, Dec. 1996.

[8] T. Rossetto and A. Elnashai, "A new analytical procedure for the derivation of displacement-based vulnerability curves for populations of RC structures," *Engineering Structures*, vol. 27, no. 3, pp. 397–409, Feb. 2005.

[9] L. F. Ibarra and H. Krawinkler, "Global collapse of frame structures under seismic excitations," *Blume Center Technical Report*, 2005.

[10] M. Colombi, B. Borzi, H. Crowley, M. Onida, F. Meroni, and R. Pinho, "Deriving vulnerability curves using Italian earthquake damage data," *Bulletin of Earthquake Engineering*, vol. 6, no. 3, pp. 485–504, Jul. 2008.

[11] F. Braga, M. Dolce, and D. Liberatore, "A Statistical Study on Damaged Buildings and Ensuing Review of the M.S.K.-76 Scale," *Proceedings of the 7th European Conference on Earthquake Engineering*, pp. 1–11, 1982.

[12] N. Lantada, J. Irizarry, A. H. Barbat, X. Goula, A. Roca, T. Susagna, and L. G. Pujades, "Seismic hazard and risk scenarios for Barcelona, Spain, using the Risk-UE vulnerability index method," *Bulletin of Earthquake Engineering*, vol. 8, no. 2, pp. 201–229, Aug. 2009.

[13] F. Sabetta, A. Goretti, and A. Lucantoni, “Empirical Fragility Curves from Damage Surveys and Estimated Strong Ground Motion,” *Proceedings of the 11th European Conference on Earthquake Engineering*, 1998.

[14] D. Vamvatsikos and C. A. Cornell, “Incremental dynamic analysis,” *Earthquake Engineering & Structural Dynamics*, vol. 31, no. 3, pp. 491–514, 2002.

[15] H. Burton and G. Deierlein, “Simulation of Seismic Collapse in Nonductile Reinforced Concrete Frame Buildings with Masonry Infills,” *Journal of Structural Engineering*, vol. 140, no. 8, p. A4014016, Aug. 2014.

[16] S. Mazzoni, F. McKenna, M. H. Scott, and G. L. Fenves, “Open System for Earthquake Engineering Simulation (OpenSees),” *Pacific Earthquake Engineering Research Center*, 2006.

[17] B. A. Bradley and R. P. Dhakal, “Error estimation of closed-form solution for annual rate of structural collapse,” *Earthquake Engineering & Structural Dynamics*, vol. 37, no. 15, pp. 1721–1737, Dec. 2008.

[18] M. Shinozuka, M. Q. Feng, J. Lee, and T. Naganuma, “Statistical Analysis of Fragility Curves,” *Journal of Engineering Mechanics*, vol. 126, no. 12, pp. 1224–1231, Dec. 2000.

[19] Nuclear Regulatory Commission, *PRA Procedures Guide: A Guide to the Performance of Probabilistic Risk Assessments for Nuclear Power Plants*. The American Nuclear Society, 1983.

[20] J. W. Baker, “Efficient analytical fragility function fitting using dynamic structural analysis,” *Earthquake Spectra*, p. 141208072728004, Dec. 2014.

[21] D. Straub and A. Der Kiureghian, “Improved seismic fragility modeling from empirical data,” *Structural Safety*, vol. 30, no. 4, pp. 320–336, Jul. 2008.

[22] I. Ioannou, T. Rossetto, and G. N. Damian, “Use of regression analysis for the construction of empirical fragility curves,” in *Proceedings of 15th world conference on Earthquake Engineering*, 2012.

[23] D. Lallemant, A. Kiremidjian, and H. Burton, “Statistical procedures for developing earthquake damage fragility curves,” *Earthquake Engineering & Structural Dynamics*, vol. 44, no. 9, pp. 1373–1389, Jul. 2015.

[24] T. Rossetto, I. Ioannou, D. N. Grant, and T. Maqsood, “GEM Guidelines for Empirical Vulnerability Assessment,” *Global Earthquake Model foundation report*, pp. 1–108, Mar. 2014.

[25] D. Lallemant, A. Kiremidjian, and H. Burton, “Statistical Procedures for Developing Earthquake Damage Fragility Curves,” 2013.

[26] H. Y. Noh, D. Lallemant, and A. S. Kiremidjian, “Development of empirical and analytical fragility functions using kernel smoothing methods,” *Earthquake Engineering & Structural Dynamics*, vol. 44, no. 8, pp. 1163–1180, Oct. 2014.

[27] A. Agresti, *Categorical Data Analysis*. Wiley-Interscience, 2002.

[28] T. J. Hastie, R. J. Tibshirani, and J. J. H. Friedman, *The Elements of Statistical Learning*. Springer, 2009.

[29] B. R. Frieden, *Science from Fisher Information*. Cambridge University Press, 2004.

[30] B. Jóhannesson and N. Giri, “On approximations involving the beta distribution,” *Communications in Statistics - Simulation and Computation*, vol. 24, no. 2, pp. 489–503, Jan. 1995.

[31] E. H. Field, T. H. Jordan, and C. A. Cornell, “OpenSHA: A Developing Community-modeling Environment for Seismic Hazard Analysis,” *Seismological Research Letters*, vol. 74, no. 4, pp. 406–419, Jul. 2003.