1. Introduction

This appendix describes a statistical procedure for fitting fragility functions to structural analysis data, when the structural analysis is performed using different ground motions at each intensity level (e.g., when using Conditional Spectrum ground motion selection rather than Incremental Dynamic Analysis). In such a case, the most popular current method for fragility function fitting will not work, and so a maximum likelihood estimation (MLE) based approach is described here. The approach is easy to implement in any software setting, and Excel and Matlab tools for doing this fitting are provided. The approach can be used for fitting fragility functions for a variety of situations, but the focus here is on collapse fragility functions obtained from structural analysis data, as that is the application where this approach is used in Chapter 5.

A useful structural response quantity to estimate from dynamic structural analyses is the probability of collapse as a function of ground motion intensity (often quantified using spectral acceleration, $S_a$, at some specified period). This result can then be combined with a ground motion hazard to compute the mean annual rate of structural collapse (e.g., Deierlein and Haselton 2005; Ibarra and Krawinkler 2005; Shome 1999).

In most work of this type, the collapse fragility function is found by repeatedly scaling a ground motion (i.e., using incremental dynamic analysis) until the ground motion causes collapse of the structure. Using this method, each ground motion has a single spectral acceleration value at the specified period associated with its collapse. By repeating this process for a suite of ground motions, one can obtain a set of spectral acceleration values at a specified period $T$ (here denoted $S_a$, where the period is omitted from the notation for brevity) associated with the onset of collapse, as illustrated in Figure 1. The probability of collapse at a given $S_a$ level, $x$, can then be estimated as the fraction of records for which collapse occurs at a level lower than $x$. A plot of this estimate is shown in Figure 2. A lognormal cumulative distribution function is often fit to this data, to provide a continuous estimate of the probability of collapse as a function of $S_a$. The equation for this function is

$$P(C \mid S_a = x) = \Phi\left(\frac{\ln x - \mu}{\beta}\right)$$

(1)
where \( P(C \mid Sa = x) \) is the probability of collapse, given a ground motion with \( Sa = x \), \( \Phi() \) is the normal cumulative distribution function and \( \mu \) and \( \beta \) are the mean and standard deviation of \( \ln Sa \) (note that \( \beta \) is also sometimes called the “dispersion” of \( Sa \), and that \( e^\mu \) is an estimate of the median of the \( Sa \)’s at collapse).

Calibrating equation 1 for a given structure requires estimating \( \mu \) and \( \beta \) from the IDA results (i.e., the data of Figure 1). We will denote estimates of those parameters as \( \hat{\mu} \) and \( \hat{\beta} \), and in general use the ^ notation to denote an estimate of a parameter.

In the context of Incremental Dynamic Analysis, the parameters \( \hat{\mu} \) and \( \hat{\beta} \) can be estimated by taking logarithms of each \( Sa \) value associated with collapse of a record. The mean and standard deviation of the \( \ln Sa \) values can then be calculated as used as the estimated parameters (e.g., Ibarra and Krawinkler 2005).

\[
\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \ln Sa_i \tag{2}
\]

\[
\hat{\beta} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (\ln Sa_i - \hat{\mu})^2} \tag{3}
\]

where \( n \) is the number of ground motions considered, and \( Sa_i \) is the \( Sa \) value associated with onset of collapse for the \( i \)th ground motion. This approach is denoted “Method A” by Porter et al. (2007). A similar alternative is to use counted fractiles of the \( Sa_i \) values to estimate the mean and standard deviation using the same \( Sa \) data (Vamvatsikos and Cornell 2004). This approach has also been used to calibrate fragility functions for data other than structural collapse (e.g., Aslani and Miranda 2005).

In the statistical inference literature, this approach is termed the method of moments (Rice 1995), as the normal distribution parameters in equation 1 were determined by equating the moments (i.e., mean and variance) of the target distribution to the sample moments from a set of data. A resulting fitted distribution using this approach is shown in Figure 2. The following section describes a fitting procedure that is very different from this approach, and applicable to a more general set of problems.
2. Collapse fragility data from Conditional Spectrum ground motion selection

The above method is not possible if, instead of using Incremental Dynamic Analysis, we select different ground motions at each $Sa$ level. This is needed, for instance, when using the Conditional Spectrum as a target spectrum, because the target spectrum changes at each $Sa$ level. In this case, instead of an $Sa$ value...
associated with the onset of collapse for each ground motion, we have instead the fraction of ground motions at each $Sa$ level that caused collapse. Also, in general, the analysis may not be carried out up to $Sa$ amplitudes where all ground motions cause collapse. Data of this type is illustrated in Figure 3. For this figure, probabilities of collapse were obtained from records selected to match target Conditional Spectra at each $Sa$ level (i.e., different ground motions are used at each $Sa$ level). With data of this type, we cannot use the method of moments approach described earlier, because we do not have the $Sa_i$ values associated with the onset of collapse that are needed in equations 2 and 3.

![Figure 3](image)

**Figure 3** Observed fractions of collapse of a 20 story concrete frame building discussed in Chapter 6, obtained from dynamic analysis using sets of 40 ground motions that match Conditional Spectrum targets at each of ten $Sa$ amplitudes.

When fitting a lognormal distribution to the observations of Figure 3, the goal is to identify the lognormal distribution parameters ($\mu$ and $\beta$ in equation 1) so that the fitted distribution predicts probabilities that are “most consistent” with the observed fractions of ground motions causing collapse at each $Sa$ level.

The observed probability of collapse at level $Sa = x_j$ is

$$P(C | Sa = x_j)_{observed} = \frac{\text{number of collapses when } Sa = x_j}{\text{number of ground motions}}$$

and the desired collapse fragility function is defined by equation 1.

In order to account for the non-constant variance of the observed fractions of collapse, a more appropriate fitting technique is one relying on the method of maximum likelihood (e.g., Rice 1995; Shinozuka et al. 2000; Baker and Cornell 2005; Straub and Der Kiureghian 2008). This approach is described in detail in the following section.
2.1. **Maximum Likelihood Estimate (MLE) Formulation**

First, we note that the probability of observing $z_j$ collapses out of $n_j$ ground motions with $Sa = x_j$ is given by the binomial distribution

$$P(z_j \text{ collapses in } n_j \text{ ground motions}) = \binom{n_j}{z_j} p_j^{z_j} (1 - p_j)^{n_j - z_j} \quad (5)$$

where $p_j$ is the probability that a ground motion with $Sa = x_j$ will cause collapse of the structure, as specified by a fragility function (which was previously denoted $P(C \mid Sa = x)$ above and is shortened here for brevity). Our goal is to identify the fragility function that will provide $p_j$, and the maximum likelihood estimation (MLE) approach tells us that the way to find this function is to choose the function that gives us the highest probability of observing the collapse data that we have observed. When we have analysis data at multiple $Sa$ levels, then we take the product of the binomial probabilities at each $Sa$ level to get the so-called likelihood function

$$\text{Likelihood} = \prod_{j=1}^{m} \binom{n_j}{z_j} p_j^{z_j} (1 - p_j)^{n_j - z_j} \quad (6)$$

where $m$ is again the number of $Sa$ levels, and $\prod$ denotes a product over all $m \ Sa = x_j$ levels. We will choose a fragility function that maximizes this likelihood.

To perform this maximization, we first substitute in the fragility function defined in equation 1 in place of $p_j$, so that the fragility function parameters are explicit in the likelihood function.

$$\text{Likelihood} = \prod_{j=1}^{m} \binom{n_j}{z_j} \Phi \left( \frac{\ln x_j - \mu}{\beta} \right)^{z_j} \left( 1 - \Phi \left( \frac{\ln x_j - \mu}{\beta} \right) \right)^{n_j - z_j} \quad (7)$$

The estimates of our fragility function parameters are then obtained by maximizing this likelihood function

$$\{\hat{\mu}, \hat{\beta}\} = \max_{\mu, \sigma} \prod_{j=1}^{m} \binom{n_j}{z_j} \Phi \left( \frac{\ln x_j - \mu}{\beta} \right)^{z_j} \left( 1 - \Phi \left( \frac{\ln x_j - \mu}{\beta} \right) \right)^{n_j - z_j} \quad (8)$$

The parameters which maximize this likelihood function will also maximize the log of the likelihood, and numerically it is typically easier to optimize this log likelihood function

$$\{\hat{\mu}, \hat{\beta}\} = \max_{\mu, \sigma} \sum_{i=1}^{m} \left\{ \ln \binom{n_i}{z_i} + z_i \ln \Phi \left( \frac{\ln x_i - \mu}{\beta} \right) + (n_i - z_i) \ln \left( 1 - \Phi \left( \frac{\ln x_i - \mu}{\beta} \right) \right) \right\} \quad (9)$$

The information on the right hand side of this formula is all readily available from dynamic analysis results, and the optimization to determine the maximum value is easily performed using many computational software programs.
A few comments can now be made regarding this formulation. First, equations 8 and 9 are defined using a lognormal cumulative distribution function for the fragility function, but that is not required to use this approach. Any other function could be substituted in equations 8 and 9 without changing the fitting approach.

Second, this formulation does not require multiple observations at each $Sa$ level of interest. A single value of collapse or non-collapse at each given $Sa$ level is still sufficient for fitting (i.e., $n_i$ can equal 1 in the above formulas). This may be useful when fitting collapse results from unscaled ground motions (each of which has a unique $Sa$ amplitude) or when using field observations (where only one observed collapse or non-collapse is available at each location).

Third, we note that this formulation assumes independence of observations, in order to take the overall likelihood as the product of likelihoods at each $Sa$ level. This may not be strictly true if the same ground motion is used for structural analysis at multiple $x_i$ levels, although quantifying this dependence may be somewhat challenging. Straub and Der Kiureghian (2008) present a generalization of the above approach that allows consideration of dependent samples, though the formulation is more complex to implement.

The values of $\mu$ and $\beta$ that maximize equation 9 can be found using optimization algorithms available widely in software tools such as Excel or Matlab. The numerical optimization may occasionally converge to a local minima if unreasonable starting estimates for $\mu$ and $\beta$ are provided; this will be apparent to the user, as the resulting fragility function will provide an extremely poor fit to the data. A robust alternative approach to finding the optimum $\mu$ and $\beta$ in equation 9 is to use generalized linear regression with a Probit link function to predict $P(C)$ as a function of $\ln Sa$ (Agresti 2002). This provides an equivalent solution because generalized linear regression uses maximum likelihood as its optimization scheme and the Probit link function is equivalent to the Lognormal Cumulative Distribution Function. This regression function is available in many statistical software packages (e.g., R Team 2010; The Mathworks 2010). The Probit model fits a normal CDF, and so the observed $P(C)$ values should be paired with $\ln Sa$ values in the regression function in order to obtain a lognormal CDF for (non-log) $Sa$.

Estimates obtained using this scheme have been observed to always find the global minimum, unlike the manually implemented maximum likelihood method that may rely on a less robust numerical optimization scheme. The global maxima found by the two approaches are identical, so the robustness of the regression result comes at no cost. An illustration of the fitted distribution is displayed in Figure 4.

Note that the generalized linear regression approach is a general approach to calibrate probabilistic models from categorical data (i.e., “collapse” or “no collapse”), and it also does not require the use of a lognormal CDF for fitting. The logit distribution, associated with logistic regression, is another general function that is popular for use in categorical data and has been used occasionally for fragility function calibration (Basöz and Kiremidjian 1998). Logistic regression is implemented using the same procedure as the Probit Regression described above, with only a different data transformation (“link function”) utilized in the regression algorithm.
Although this proposed method is slightly more complicated than the method of moments approach used commonly today, it is much more flexible. In addition to handling the more general form of data considered here, other generalizations can be used. For example, if one is primarily interested in the lower tail of the distribution (as is often the case, because lower $Sa$ values occur much more frequently), then only this portion of the data need be obtained for fitting. For example, dynamic analysis results could be obtained for increasing $Sa$ levels until a certain fraction (e.g. 40% or 50%) of the records cause collapse. Then the analysis can be stopped, and the collapse fragility function fitted without knowledge of the effect of larger $Sa$ values. This may be advantageous when structural analyses are expensive.

2.2. Potential alternative methods

One alternative to the proposed method for estimating $\mu$ and $\beta$ would be to minimize the sum of squared errors (SSE) between the observed fractions of collapse and probabilities of collapse predicted by equation 1. Mathematically, this would be stated:

$$\{\hat{\mu}, \hat{\beta}\} = \min_{\mu, \sigma} \sum_{j=1}^{m} \left( P(C \mid Sa = x_j)_{\text{observed}} - \Phi \left( \frac{\ln x_j - \mu}{\beta} \right) \right)^2$$

(10)

where all variables are defined earlier. This is again a nonlinear optimization that can be performed with a variety of widely available algorithms. Example results are shown in Figure 5 for the data considered above. An additional set of data is provided in Figure 6, to illustrate that the differences in results can be less trivial in some cases.

The differences arise because the least-squares method ignores a fundamental property of the data: the variance of the observations is non-constant. That is, if zero collapses are observed at a given $Sa$ level and the fitted probability of collapse is 0.1, then this error is much larger than fitting a probability of collapse
of 0.6 at an $S_a$ level where 50% collapses are observed. This result is apparent in Figure 6, where higher probabilities of collapse are predicted at low $S_a$ levels that the data suggests are very unlikely to cause collapse. This lower tail of the fragility function is known to have a key role in resulting performance analyses (i.e., computation of the mean annual frequency of collapse), because ground motions with low $S_a$ levels are much more frequent than high $S_a$ levels, so this difference may have a significant impact on performance assessments.

Another alternative would be to fit the function using the “Method B” approach of Porter et al. (2007). This method transforms the observed fractions of collapse into a form where linear regression can be used to estimate the fragility function parameters. As with the above two alternatives, this approach minimizes an error metric between observations and a fitted function—in this case the sum of squared errors in the transformed space. As with the error metric of equation 10, this metric is not optimally defined with respect to the nature of the data being considered in this particular application. Results from this fitting approach are shown in Figure 5 and Figure 6, and indicate that the error metric used by this method tends to result in poorer-fitting fragility functions at low $S_a$ levels.

These alternate approaches are thus not recommended for use in fitting this type of data. They are only mentioned briefly here to illustrate the potential problems with seemingly appropriate alternate approaches for fitting this collapse data.
Figure 6  An alternate set of observed fractions of collapse from dynamic analyses and fitted fragility function obtained using the proposed Maximum Likelihood Estimation (MLE) approach and two alternative methods. The data shown here is observed fractions of collapse of the 20 story concrete frame building discussed in Chapter 6, using sets of 40 ground motions that match Conditional Spectrum targets at each of ten $Sa(0.45s)$ amplitudes.

3. Software tools

To facilitate the use of this approach, a set of simple software tools have been created that do the above calculations. The user need only provide observed data (i.e., number of collapses and non-collapses at each $Sa$ level), and the maximum likelihood functions will be evaluated to identify the fragility function parameters. The procedure has been implemented in Microsoft Excel (using the Goal Seek function to perform the optimization) and Matlab (where one implementation uses an optimization of the likelihood function and another uses Probit regression). All three tools produce identical results for a given data set—the three versions are provided simply to suit a range of user software preferences. The tools are available online at http://www.stanford.edu/~bakerjw/fragility.html.

4. Conclusions

A technique has been described to estimate collapse fragility functions from structural analysis data using maximum likelihood estimation. Unlike the currently popular approach of finding the mean and standard deviation of $\ln Sa$ values associated with the onset of collapse from incremental dynamic analysis, this approach is general in that differing ground motions can be used at each $Sa$ level and the analysis does not have to be performed until one reaches an $Sa$ level at which all ground motions cause collapse. This is beneficial, because fragility functions can then be developed using Conditional Spectrum concepts, which use different ground motions at each $Sa$ level. This also means that fragility functions can be calibrated based only on $Sa$ levels at which there are lower probabilities of collapse, saving analysis time and focusing the fitting on the region of the curve most important for seismic risk assessments.
The above procedure is useful for situations other than structural collapse modeling. For example, Shinozuka et al. (2000) first proposed this approach to fit fragility functions to observed bridge damage data in the Northridge earthquake. But given that previous work had not discussed the application of this principle to structural analysis collapse results where one has repeated observations of collapse or non-collapse at user-specified $S_a$ levels, a description of the approach for that specific situation has been provided here.

The proposed procedure is very easy to implement, and several simple software tools have been developed and provided for users, to facilitate this implementation.

5. References


