EFFICIENT STATISTICAL APPROXIMATION OF ENGINEERING DEMAND PARAMETERS IN BUILDING STRUCTURES

Huda MUNJY¹, Farzin ZAREIAN²

ABSTRACT

This paper aims to test the true statistical distribution of engineering demand parameters (EDPs), specifically interstory drift ratio and floor acceleration of buildings. A comparison among current and alternative methods that may be used to simulate EDPs for engineering applications is provided. Current loss estimation procedures assume a joint lognormal distribution between EDPs (FEMA, 2015); this assumption is tested using three statistical distribution verification tests: Mardia’s, Henze-Zirkler and Royston’s. Two statistical methods for generating realizations (i.e. factor analysis and copula probability theory) are evaluated for simulating interstory drift ratio (IDR) and peak floor accelerations (PFA). Realizations of EDPs from the two statistical methods suggested herein are compared with those generated by the commonly used FEMA P-58 approach, and the true population of EDPs. Two building models, 4-story and 8-story, at three ground motion hazard levels of 2%, 10%, and 50% probability of exceedance in 50 years are considered. A set of 100 selected and scaled ground motions for each hazard level was applied to each model to generate the true population of EDPs. Results show that interstory drift ratio and floor acceleration do not follow a joint lognormal distribution. The FEMA P-58 and copula probability theory approaches for generating realizations provide acceptable estimates of EDPs. These findings are useful for engineering practice, specifically for loss estimation. The validation approach suggested here can be used to determine if alternative methods for generating EDP realizations are credible.

Keywords: Performance-based Assessment; Statistical Analysis; FEMA P-58; Engineering Demand Parameters;

1. INTRODUCTION

Making assumptions and bracing against their consequences is one of the major challenges in engineering. In Performance Based Earthquake Engineering (PBEE), making assumptions is necessary for calculating probabilities and conducting statistical analyses. With a focus specifically aimed at seismic performance and loss estimation procedures, variables are commonly assumed to follow a lognormal distribution, represented by a median value (denoted by \( \theta \)) and dispersion (denoted by \( \beta \)). FEMA P-58 (2015) methodology for loss estimation assumes a joint lognormal distribution to represent many uncertain factors, including the distribution of IDR and PFA. IDR and PFA are often used to assess damage to structural and non-structural building components (Mitrami-Reiser, 2007; Aslani, 2005). Except for few anecdotal cases, no research has been conducted in this field to assess, through statistical tests, whether or not PFA and IDR in buildings follow a joint lognormal distribution. Another assumption in seismic performance assessment methodologies proposed by FEMA P-58 is the use of 7 to 11 analyses to obtain valid estimates of median response. More analyses are recommended when the geomean spectral shape of scaled motions does not accurately match the shape of the target spectrum. It is suggested that 11 analyses be used for ground motions that are selected with no consideration to the spectral shape (Huang et al, 2011). There is little reasoning provided as to why 11 ground motions are sufficient in producing unbiased estimates of median structural response. There are two types of uncertainties in variables: aleatory and epistemic. In PBEE, aleatory uncertainty can come from the inherent variability in ground motion records while epistemic uncertainty is due to structural modeling.

¹Ph.D. Candidate, University of California, Irvine, Irvine, Unites States, hmunjy@uci.edu
²Associate Professor, University of California, Irvine, Irvine, Unites States, zareian@uci.edu
assumptions (Yazdani et al., 2017). Research suggests that seismic building response is more sensitive to epistemic uncertainty than aleatory (Kazantzì et al, 2008; Kwon and Elnashai, 2006; Yazdani and Eftekharì, 2012). Cornell (2000) brings attention to the issue that as structures are more likely to collapse at high intensity values and therefore, there is no displacement data for these cases. Omitting this data can cause errors in reliability assessments. Stoica et al. (2007) propose a three-parameter distribution of variables (EDPs) as opposed to a two-parameter distribution used in current practice; their findings suggest that a lognormal distribution assumption is unable to accurately represent damage measures, specifically for components whose damage is sensitive to displacement. Yazdani et al. (2017) considered these factors and proposed a modified distribution that considers scenarios of collapse for high intensity measures. These studies propose modifications to the lognormal distribution but do not address the issue of what the actual distribution of EDPs are and if this actual distribution is able to provide more accurate estimates of seismic demand. In order to better understand and verify these assumptions, the statistics of EDPs simulated with 11 analyses as per FEMA’s suggestions are compared with 100 non-simulated or ‘population’ EDPs obtained through Nonlinear Response History Analysis (NRHA). This comparison is made by obtaining EDPs from a 4-story and 8-story steel moment frame building and considers three hazard levels: 2%, 10% and 50% in 50 years.

2. EVALUATION OF ENGINEERING DEMAND PARAMETER DISTRIBUTION

Three tests are used to assess Multi-Variate Normality (MVN) between EDP variables. To use these tests, the log of the data was tested for MVN. If data have a multivariate normal distribution then each of the variables has a univariate normal distribution, but the opposite does not have to be true (Korkmaz et al, 2014). For this reason, it is not enough to perform a simple distribution check on each individual variable because we are interested in the joint distribution. A test of skewness was done for the individual distribution check which proved the distributions of individual variables to possess a lognormal distribution. After this simple test of skewness was complete, an in-depth analysis was conducted to assess multivariate normality using a series of three statistical tests: Mardia’s MVN test, Henze-Zirkler’s test, and Royston’s test. All equations in this section are obtained from Korkmaz et al (2014) to provide a detailed statistical background of each MVN test.

2.1 Mardia’s MVN Test

Mardia’s MVN test is a popular test for determining multivariate skewness and kurtosis and is based on the third and fourth moments. The skewness ($\hat{\gamma}_{1,p}$) and kurtosis ($\hat{\gamma}_{2,p}$) are measured with Equations 1 and 2, where $m_{ij} = (x_i - \bar{x})'(S^{-1})(x_j - \bar{x})$, the squared Mahalanobis distance, and $p$ is the number of variables.

$$\hat{\gamma}_{1,p} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij}^3$$

$$\hat{\gamma}_{2,p} = \frac{1}{n} \sum_{i=1}^{n} m_{ii}^2$$

2.2 Henze-Zirkler’s MVN Test

The Henze-Zirkler MVN test is based on the measured distance between two distribution functions. The test statistic obtained is lognormally distributed if the data is normal. For the estimation of the $p$-value, the mean and variance are log-normalized (Korkmaz et al, 2014). The Henze-Zirkler test statistic is computed using Equation 3, where $p$ represents the number of variables, $\beta$ is a function of the number of variables, and $D_i$ and $D_j$ are the squared Mahalanobis distance of the $i$th observation to the centroid and the $i$th and $j$th observations, respectively.

$$HZ = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} e^{\frac{-\beta^2}{2} D_{ij}^2} - 2(1 + \beta^2)^{-\frac{p}{2}} \sum_{i=1}^{n} e^{-\frac{\beta^2}{2(1+\beta^2)} D_{i}^2} + n(1 + 2\beta^2)^{-\frac{p}{2}}$$
\[ \beta = \frac{1}{\sqrt{2}} \left( \frac{n(2p+1)}{4} \right)^{\frac{1}{p+4}} \] (4)

\[ D_{ij} = (x_i - x_j)^{S^{-1}}(x_i - x_j) \] (5)

\[ D_i = (x_i - \bar{x})^{S^{-1}}(x_i - \bar{x}) = m_{ii} \] (6)

For normally distributed multivariate data, the HZ test statistic is log normally distributed and therefore represented by two parameters: mean and variance. Thus, this test statistic for multivariate normality is calculated from Equation 7.

\[ z = \frac{\log(HZ) - \log(\mu)}{\log(\sigma)} \] (7)

2.3 Royston’s MVN Test

Royston’s test uses the Shapiro-Wilk/Shapiro-Francia statistic. The Royston test statistic is calculated using Equation 8, where \( e \) is the equivalent degrees of freedom (edf) and \( \Phi(.) \) is the cumulative distribution function for standard normal distribution with \( e = \frac{p}{1 + (p - 1)\bar{c}} \).

\[ H = \frac{e^{\sum_{j=1}^{p} \psi_{ij}}}{p} \sim X_e^2 \] (8)

\[ \psi_{ij} = \left\{ \Phi^{-1} \left[ \frac{\Phi(-z_j)}{z} \right] \right\}^2, j = 1, 2, ..., p \] (9)

Equations 11 and 12 are used to calculate \( c \), which must be calculated from the correlation between two parameters.

\[ c_{ij} = \begin{cases} g(r_{ij}, n) & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} \] (11)

\[ g(r, n) = r^\lambda \left[ 1 - \frac{\mu}{\nu} (1 - r)^\mu \right] \] (12)

Ross (1980) determined \( \mu, \nu, \lambda \) as \( \mu=0.715, \lambda=5 \) Equation 13 for calculating \( \nu \) through simulations for a sample size between 10 and 2000.

\[ \nu(n) = 0.21364 + 0.015124x^2 - 0.0018034x^3 \] (13)

\[ x = \log(n) \] (14)

2.4 MVN Summary of Tests

Since there are three tests used here to confirm or deny multivariate normality, a Q-Q (quantile-quantile) plot is used to assist with any inconsistent results between the three tests. A Q-Q plot is a commonly used graphical representation of the agreement between probability distributions (Korkmaz et al, 2014). The Mahalanobis distance is a representation of the distance between a point and a distribution. It is a multi-dimensional measure of how many standard deviations the point is from the mean of the distribution. One axis, the distance, represents theoretical (hypothesized) quantiles while the other axis represents the observed quantiles. The more of a fit there is between actual data and the assumed distribution, the better the points and the line will match. For all 6 cases (2 buildings, 3 hazard levels) the 3 multivariate normality tests indicated that the data does not follow a normal distribution. Since this data was in the log domain, this means that there was no indication that the EDP distribution is actually...
lognormal. Six Q-Q plots are shown in Figure 1, each representing one building and one hazard level considered. When examining these plots, a large amount of deviation is seen, is seen for all three hazard levels in the 8-story building. The 4-story building shows the most deviation for the 2% in 50 years hazard level. For all three tests, the p values were significantly below 0.05, indicating no sign of MVN. One case, the 4-story building with a 50% in 50 years hazard level, had a p value of 0.05 for Royston’s test. This p value is on the border and barely passed at passed at 5% confidence. This result can be disregarded as Mardia’s test and the Henze-Zirkler’s test showed the 4-story 50% in 50 years data as not following a normal distribution.

![Figure 1: Quantile-Quantile plots for 8-story and 4-story building representing deviations of EDPs from the exhibiting a lognormal distribution](image)

3. EDP SIMULATION METHODOLOGIES

3.1 Lognormal EDP Simulation

FEMAP-58 2015 methodology assumes that the EDPs are jointly lognormally distributed. Monte Carlo is used to generate a large number of demands from a small number of analyses. The general procedure is to obtain the demand matrix, get median values for each parameter and a covariance matrix and then simulate a large number of demand vectors mathematically using a random number selection process and the median and covariance matrices. The required number of analyses depends on many factors but 11 analyses are mentioned in FEMAP-58 as a possibly appropriate number. Yang (Yang et al. 2006,Yang et al. 2009) developed the algorithm used to generate the simulated demands. The peak absolute value of each EDP is assembled into a one by n vector. A jointly lognormal distribution can be fully explained using the mean and covariance of the parameters, which are used in this analysis method. Equation 15 represents the simulated demands where $Z$ is the vector of natural log demand parameters, $X$ is the matrix of demand parameters, $Y$ is the natural log of $X$, $M_e$ is the mean vector of the demands,
and $U$ is a vector of uncorrelated standard normal variables with a mean of 0 and a covariance matrix which is the identity matrix.

$$Z = AU + B = L_{np}D_{pp}U + M_y$$

The rank of the covariance matrix, $k$, is assessed by counting the number of non-zero eigenvalues of the covariance matrix. If the matrix is full rank then $p$ equals the number of analyses considered. Covariance matrices are not observed to be full rank when the number of analyses is less than the number of variables and/or when one or more vectors of variables are linear combinations of other vectors.

### 3.2 Factor Analysis EDP Simulation

Factor Analysis is a statistical procedure that can be used to detect interrelationships between variables. This procedure has a number of applications that revolve around data reduction and identifying links among different factors used to assess data. The data is effectively reduced by determining a smaller number of unobserved factors from a larger number of variables. Factor Analysis therefore aims to explain correlations among a large number of observed variables by describing these variables with a smaller number of unobserved ‘factors’. Principle component analysis is the form of factor analysis used in this study. MATLAB was used to generate simulations using the factor analysis model for the 4- and 8-story buildings with three considered hazard levels. From the 100 data points, 11 were used to generate the 100 simulated data points for the 4-story building. For the 8-story building, 18 data points from the population of 100 considered data points were used to generate the 100 simulated EDPs. For the 8-story building, this number was necessary due to the fact that factor analysis requires a full rank covariance matrix of the data. For an 8-story building, considering 9 values for acceleration and 8 values for drift makes a total of 17 variables and therefore the minimum amount of data points required for factor analysis is 18. These data points were standardized and then used to calculate the factor loadings. When the original unstandardized data was used to generate factor scores, there was a gap between the factor scores for accelerations and for drifts due to the large difference in magnitudes for these two measurements. Therefore, in order to more easily interpret the meaning of the results, the 11 cases were standardized and then Factor Analysis was done. After the simulations were gathered, the data was converted back to express the values with the original units.

After this data is computed the EDP simulations can then be generated. Equation 16 is used to compute simulated EDPs where, EDP is an array of 10000 simulated engineering demand parameters, $\alpha_1$ and $\alpha_2$ are an arrays of 10000 factor scores for factor 1 and 2; $\psi_1$ and $\psi_2$ are an arrays of 5 loadings for factor 1 and 2, $R=10000\times1$ matrix of ones, $\epsilon=1\times5$ matrix of specific variance.

$$EDP = \alpha_1\psi_1 + \alpha_2\psi_2 + R\epsilon$$

### 3.2 Copula Distribution EDP Simulation

Copulas are used to illustrate how marginal distributions can be linked together to describe a joint distribution. Therefore, given a copula, multivariate distributions can be determined by selecting different marginal distributions, which refer to the probability distributions of each variable in the sample, as the inputs. According to Sklar’s theorem (1959) the joint distribution function, $F_{XY}(x,y)$, can be expressed as a function of $F_X(x)$ and $F_Y(y)$, which represent the marginal distributions of $X$ and $Y$, respectively (Nelson, 1999). Consider, the joint distribution function as represented in Equation 17 where $C(u,v)$ is the copula.

$$F_{XY}(x,y) = C(F_{X}(x),F_{Y}(y))$$

$$C(u,v) = F_{XY}(F_{X}^{-1}(u),F_{Y}^{-1}(v))$$

$$C(F_{X}(x),F_{Y}(y)) = F(F_{X}^{-1}(F_{X}(x)),F_{Y}^{-1}(F_{Y}(y))) = F_{XY}(x,y)$$
Using marginal distributions is another form of data reduction when there are many variables being considered in the study. Different distributions for the copula can be assumed when simulating data. Initially for this study the Gaussian copulas are assumed and implemented to simulate EDPs. This assumption is checked and modified using the Vine Copula method. In order to conduct a simulation using a copula distribution, three items need to be identified: the copula distribution, the rank correlation of the data, and the marginal distributions of each random variable in the dataset. Each of these will be discussed in further detail, below.

3.2.1 Copula Distribution

The copula distribution represents the link between the marginal distributions of the variables and the dependencies of the variables that make up the joint distribution. There are several different copula families that can be considered (Nelson, 1999); R provides built in functions for 40 of these copulas (Kojadinovic and Yan, 2010). The copula should be selected based on the type of data and any observed trends that can be identified. Due to the variability of many data sets when considering real world simulations, there is an element of trial and error involved when picking a copula distribution, especially since the copula does not necessarily follow the same distribution of the data. For example, a set of multivariate normally distributed random variables may fit best with a Clayton copula. This means that although the joint distribution of the data is Gaussian, the dependencies between the variables are best described by a Clayton distribution. When this is the case, simulating data with just the assumption of a joint multivariate normal distribution may cause the loss of key elements regarding the relationship between the variables and therefore lead to less accuracy in results. A Gaussian copula was selected (to begin the trial and error process). The data was converted to the log domain for all copula simulations; this was done in order to standardize the magnitude of all considered variables. Also, the histograms of the variables all exhibited right-hand skew. In order to make right skew data follow more of a Gaussian distribution the log of the data should be taken. Although this did not give the data a normal distribution it did severely impact the results of the Gaussian copula simulation, shifting the mean and standard deviations to match the sample (11 observations) and population (100 observations) very well. Two different approaches were considered for the determination of the copula distribution: Gaussian copula assumption, for ease of calculations, and the vine copula method, for higher accuracy in results. The Gaussian copula assumption was not based on any observed trends in the data but rather done to begin the trial and error process used to determine the copula that best fits the data.

3.2.2 Rank Correlation

Rank correlation must be specified when simulating data using a copula distribution. Therefore the default correlation, which is usually Pearson’s correlation, cannot be used here. A rank correlation is required when dealing with copulas because it measures the association based on the copula and not the marginal distributions of the variables. The common Pearson’s correlation, measures the linear relationship between variables based on the marginal distribution and is not preserved by the copula. This means that two correlated variables that have the same copula can have different correlations. In contrast, the rank correlations are preserved. Kendall’s tau, a rank correlation measure, is used. Tau is calculated for the data using R.

3.2.3 Marginal Distribution

Marginal distributions must be specified in order to simulate data using a copula distribution. Similar to the distribution of the copula, the marginal distributions of the random variables in a dataset do not necessarily reflect the joint distribution. Therefore, the marginal distributions need to be calculated. According to Aslani and Miranda (2005), PFA and IDR follow marginal lognormal distributions. Therefore, lognormal marginals were used for the copula simulation method. The Inverse Transform Sampling method is used to generate random numbers following the same distribution as the input data using the inverse cumulative distribution function (CDF). The ‘ksdensity’ kernel smoothing function is used via MATLAB for this estimate which makes a kernel estimate of a distribution and evaluates the inverse CDF at the copula points in one step. A typical random number generator does not represent real world data because the probability of getting any value from 0 to 1 is equal where our generated points
need to represent the same distribution as the input data. Inverse transform sampling simulates random numbers with the same distribution as the input data by obtaining the probability distribution functions (PDFs) of the random variables and creating a CDF from this PDF (a Riemann sum can be used to accomplish this) and inverting this CDF. An arbitrary value, \( e \), can be chosen with a uniform distribution and the inverted value of this is equal to \( y \), \( \text{invCDF}(e) = y \). This value of \( y \) is our desired random number from our original random variable with the desired PDF.

4. RESULTS
The three simulation methodologies previously outlined are: (1) joint lognormal distribution assumption suggested by FEMAP-58, (2) factor analysis method, (3) copula distribution method. These methods will be referred to as FEMA, FA, and Copula, respectively, throughout the rest of this report. The results provide a comparison of the statistics of simulated EDPs to recorded EDPs. From the population of 100 selected and scaled recorded ground motions used to generate EDPs via NRHA, 11 cases were randomly selected as a sample to generate the simulated EDPs. This was done for the 4-story building for FEMA, Copula and FA. For the 8-story building, 11 samples were used for FEMA and Copula but 18 were used for FA due to the requirement of a full rank covariance matrix, as mentioned previously. The ratios of the mean and standard deviations of recorded to simulated EDPs are compared to show how closely these statistical measures match. Furthermore, the empirical distribution of each simulation method is compared to the population distribution using a two sample Kolmogorov-Smirnov goodness of fit test (KS test).

4.1 Mean and Standard Deviation Comparisons
Figures 2 and 3 show the mean and standard deviation ratios for the 4-story building considering the 10% and 50% in 50 years hazard levels. The ratio of the population mean or standard deviation to the sample or simulation mean or standard deviation (\( \mu_{EDP}^{\text{pop}} / \mu_{EDP}^{\text{sim}} \)) is used.

![Figure 2](image)

Figure 2: Mean ratio of drift for recorded motions to simulated motions for each sampling method with 10 trials per method

Each point represents one building EDP; there are 5 values of PFA and 4 values of drift (9 total EDPs)
for the 4-story building. These points express the range of the ratio compared with the population to see how accurately the simulations are able to express the statistics of the original set they were taken from.

Figure 3: Standard deviation ratio of drift for recorded motions to simulated motions for each sampling method with 10 trials per method

For both hazard levels in the 4-story building, the EDP mean ratios are generally centered around 1.0, indicating some over- and under-estimation in terms of matching the population EDPs. For the 10% in 50 years hazard level, the sample-11, FEMA and Copula EDPs all match with a spread of 0.95-1.1. FA EDPs overestimate the population EDPs with spread starting at 0.8 and reaching 1.05. For the 50% in 50 years hazard, the sample-11, FEMA and Copula EDPs lie within the range of 0.95-1.15, which, similar to the 10% in 50 years hazard, are centered around 1.0. FA EDPs, again, underestimate the EDPs more often with a spread starting at 0.8 and reaching 1.1. For both hazard levels, the sample-11, FEMA and Copula EDPs for the 8-story building fall within the range of 0.95-1.4. FA EDPs fall within the same range for the 50% in 50 years hazard level but have more instances of overestimation for the 10% in 50 years hazard level with a range from 0.8-1.1.

Sample-11, FEMA and Copula EDPs all exhibit closely matching standard deviation ratios and fall within the range of 0.7-1.3 for the 10% in 50 years hazard level and 0.4-2 for the 50% in 50 years hazard level. FA underestimates EDPs for both hazard levels but exhibits ratios that are generally closer to 1.0 for the 10% in 50 years hazard level indicating a close match to the population standard deviation. Sample-11 and FEMA EDP standard deviations match well with a similar range of 0.8-2.3 for the 10% in 50 years hazard level but Copula EDPs match more closely to the population with a range of 0.7-1.8. For the 50% in 50 years hazard level, sample-11, FEMA and Copula EDP standard deviation ratios all match and exhibit a range of 0.7-3.5. FA EDPs fall within a smaller range of 0.8-2 which matches the
population EDPs more closely.

All of the mean and standard deviation ratios presented provide one significant conclusion: the statistics of the simulated EDPs are highly dependent on the selected ground motions and actual EDPs that they are being simulated from. They hold very little significance in terms of other motions that were scaled and selected with the same parameters that the sample comes from. Regardless of the method, FEMA orCopula, the EDP statistics match that of the sample almost perfectly in all cases. Along with this, bothFEMA and Copula are able to capture the statistics of the sample EDP very accurately even though they both follow very different approached in terms of simulation. As mentioned earlier, for the 8-story building, FA EDPs are simulated from 18 GMs not 11 because the covariance matrix needs to be full rank (number of simulations must be greater than number of variables). Therefore, because 18 GMs are used for the FA in the 8-story building, the mean and standard deviation ratios are generally centered around 1.0 more often than the other cases which shows that the simulated EDPs match the population of 100 EDPs closely.

4.2 Kolmogorov-Smirnov Goodness of Fit

Since the statistics of the simulated EDPs are highly dependent on the sample they are simulated from, it is important to recognize whether or not samples from the same selected and scaled population exhibit similar trends. The Kolmogorov-Smirnov (KS) goodness of fit test is use in order to determine this. The KS test is used to compare the empirical distribution of each simulation with the population of EDPs. A two sample KS test is used to compare each set of samples. The KS test defines the maximum distance between the empirical distributions and assigns a p value to this distance which either confirms or negates the null hypothesis. A two sample KS test is one of the most versatile statistical tools that can be used to compare samples as it considers differences in both location and shape of the sample distributions (Massey, 1951). A KS test was used to compare the distributions of 100 data points simulated from randomly selected samples of 5, 11, 15, 25, 50 and 100. For each sample size, 100 random selections were simulated 10 times and tested to check how the sampling affects the outcome of the distributions. For Figure 4, PFAi corresponds to PFA and Di refers to IDR. There are six columns of measurements for each EDP, representing 5-100 samples used to generate the EDPs. From left to right the columns correspond to 5, 11, 15, 25, 50 and 100 data points. The null hypothesis for this test states that no significant difference exists between the population and sample distributions. The p-value for the KS test is plotted on the y axis. Considering a 5% significance level, any p value below 5% will lead to a rejection of the null hypothesis. The dashed red line represents the 5% significance level where every point above that line means that for that simulation the null hypothesis was accepted and the distribution of that simulation and the population is the same. In terms of PFA, a trend can be seen for both FEMA and Copula where with increasing sample size used for simulations, there are less points that fall below 0.05. In terms of IDR, a similar trend can be seen for Copula but for FEMA increasing the sample size does not make a difference in terms of significance; many of the trials fall below 0.05 indicating that the sample does not follow the same distribution as the population. Increasing the sample size, especially past 15, for Copula leaves very little trials that reject the null hypothesis therefore confirming that the sample and population distributions are the same.
CONCLUSIONS

Of the three methods implemented for the simulation of EDPs (PFA and IDR), FA is the least versatile as it is unable to generate EDPs when the number of simulations are less than the number of EDPs, making it an inefficient means of simulation. Although in some cases, FA is able to match the population statistics better than the other methods, this is due to the fact that more analyses are used, not that this method outperforms the others. Copula and FEMA both perform well in terms of obtaining the correct mean and standard deviation to match the sample-11 and in some cases, the population. It can also be seen, however, that Copula and FEMA EDPs are highly dependent on how well the random sample-11 matches the population. In general, if the random sample-11 matches the population then FEMA and Copula are accurate in replicating the statistics of the population EDPs but if the sample differs from the population this is no longer the case. This is further shown by the KS test comparing various random samples which shows that, depending on the chosen sample, the null could either be accepted or rejected and increasing sample size does provide more accurate measures of EDPs. In summary, (1) EDPs do not exhibit a jointly lognormal distribution according to three statistical measures to determine multivariate normality, (2) factor analysis is not as versatile as FEMA and Copula in simulating EDPs and cannot be used for non-full rank covariance matrices, (3) Copula and FEMA both accurately replicate the statistics of the population EDPs when the sample matches the population.

Figure 4: Plot of 5, 11, 15, 25, 50 and 100 samples sizes with 10 trials each representing P values for peak floor acceleration and drift

5. CONCLUSIONS

Of the three methods implemented for the simulation of EDPs (PFA and IDR), FA is the least versatile as it is unable to generate EDPs when the number of simulations are less than the number of EDPs, making it an inefficient means of simulation. Although in some cases, FA is able to match the population statistics better than the other methods, this is due to the fact that more analyses are used, not that this method outperforms the others. Copula and FEMA both perform well in terms of obtaining the correct mean and standard deviation to match the sample-11 and in some cases, the population. It can also be seen, however, that Copula and FEMA EDPs are highly dependent on how well the random sample-11 matches the population. In general, if the random sample-11 matches the population then FEMA and Copula are accurate in replicating the statistics of the population EDPs but if the sample differs from the population this is no longer the case. This is further shown by the KS test comparing various random samples which shows that, depending on the chosen sample, the null could either be accepted or rejected and increasing sample size does provide more accurate measures of EDPs. In summary, (1) EDPs do not exhibit a jointly lognormal distribution according to three statistical measures to determine multivariate normality, (2) factor analysis is not as versatile as FEMA and Copula in simulating EDPs and cannot be used for non-full rank covariance matrices, (3) Copula and FEMA both accurately replicate the statistics of the population EDPs when the sample matches the population.
7. REFERENCES


