

THE OPTIMAL METHOD FOR BUILDING DAMAGE FRAGILITY CURVE DEVELOPMENT

Apichat Janpila¹, *Piyawat Foytong², Supakorn Tirapat³, Nuttawut Thanasisathit⁴
and Anat Ruangrassamee⁵

^{1,2,3}Sustainable Infrastructure Research and Development Center, Department of Civil Engineering,
Faculty of Engineering, Khon Kaen University, Khon Kaen, Thailand.

⁴Department of Civil Engineering, King Mongkut's University of Technology North Bangkok,
Bangkok, Thailand.

⁵Department of Civil Engineering, Chulalongkorn University, Bangkok, Thailand.

*Corresponding Author, Received: 30 June 2019, Revised: 16 Nov. 2019, Accepted: 02 Jan. 2020

ABSTRACT: A fragility curve is a primary component in the risk assessment, which is useful for evacuation planning, estimation of potential losses, and estimation of the damage to residential buildings caused by natural hazards. In general, a fragility curve represents the relationship between the probability of exceeding a specific damage state of a structure and natural hazard intensity. For determining such a curve, two parameters: the median and standard deviation are estimated. A fragility curve can be constructed using empirical data and analytical data. Numerical fitting data is used to develop the fragility curve. Various methods have been proposed using numerical fitting data to approximate the fragility curves. However, the most widely used methods for developing fragility curves are the least-squares method and the maximum likelihood method. In this present study, these two different numerical fitting data methods for fragility curve development are analyzed and compared. Basic assumptions and limitations of each method are also discussed. The building damage data used in all methods to derive the fragility curve is generated from hypothetical damage data assuming a lognormal distribution. Finally, the maximum likelihood method is proven to be optimal for developing fragility curve based on structural damage data.

Keywords: *Optimal method, Fragility curve, Building damage, Risk assessment*

INTRODUCTION

Each year, several countries are affected by a variety of natural hazards, including hurricanes, tropical storms, strong winds, floods, earthquakes, and tsunamis. Therefore, risk assessment is the key to provide a way to prevent and mitigate damages or losses resulting from natural disasters in the future. The risk assessment process includes natural hazard identification, building inventory, and fragility curve. The Natural hazard is identified at any given site to determine the likely impacts in the form of a hazard map. The building inventory is a classification procedure in which buildings are grouped based on similar damage/ability/loss characteristics into a set of predefined building classes. Model building types are further constructed by developing the damage and loss prediction models to represent the average building population characteristics within each class. A fragility curve is utilized to estimate a vulnerability, which expresses a relationship between the probability of being in or exceeding building damage state and the intensity of natural hazard. Therefore, a fragility curve is the key parameters in risk assessment for evaluating the damage probability at a specific damage level

under the hazard. Statistical methodology is applied to establish a fragility curve. The building damage data, which is assumed to follow the lognormal distribution [1], is used in exploring the relationship of damage probability and hazard intensity. In general, methods most commonly used to develop the fragility curve are the least squares estimation (LSE) [2]-[3] and maximum likelihood estimation (MLE) [4]-[10]. Thus, in this study, the two most commonly used methods of estimation are compared, and the optimal method of fragility curve development is demonstrated based on statistical principles.

FRAGILITY CURVE

In this section, the concept and the key parameters of a fragility curve are described here to provide a basic understanding of analysis in fragility curve development for obtaining the appropriate method. A fragility curve describes a relationship between the level of damage probability and hazard. The probabilities defined in fragility curve are conditional probabilities as shown in Eq. (1):

$$F_d = P[D \geq d | X = x] \quad d \in \{1, 2, \dots, N_d\} \quad (1)$$

where F_d represents the fragility function for damaged state d evaluated at the hazard level of x . $P[A|B]$ is the probability that event A occurs given that event B has taken place. D is a damaged state of a particular component, which takes on a value of $\{0, 1, \dots, n_D\}$, and d is a damaged state. X is an uncertain excitation which is called demand parameter (DP) and x is a particular value of X .

The structural failure data follows a lognormal distribution as shown in Fig.1. The probability density function (PDF) as defined in Eq. (2) represents the probability of structural damage. The dispersion or width of distribution indicates the uncertainty of each damage state. Then the probability of the structural failure can be obtained by cumulative distribution function of the lognormal distribution with intensity measure on the horizontal axis. The probability of structural failure is obtained through the cumulative standard log-normal distribution function in which the horizontal axis represents the intensity measure. The log-normal cumulative distribution function expresses the fragility curve as defined in Eq. (3).

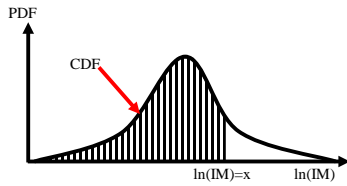


Fig. 1 Cumulative distribution function with intensity measure equal to x

$$f_d = \frac{1}{x\beta\sqrt{2\pi}} e^{(-1/2)[(\ln x - \alpha)/\beta]^2} \quad (2)$$

$$F_d = \int_{-\infty}^{\ln x} \frac{1}{x\beta\sqrt{2\pi}} e^{(-1/2)[(\ln x - \alpha)/\beta]^2} dx = \Phi\left(\frac{\ln(x/\alpha)}{\beta}\right) \quad (3)$$

where $\Phi[\cdot]$ represents the standard normal cumulative distribution function, α and β are parameters which controls the shape of the fragility curve. Thus, α is the median which controls the location of the fragility curve and β is the log-normal standard deviation which represents the dispersion of the fragility curve.

ESTIMATION METHOD FOR FRAGILITY CURVE

In this paper, we compare the results of two different estimation methods, LSE and MLE, for establishing the fragility curve. The optimal method is statistically proven based on the assumptions and properties inherent to each method.

Least Squares Estimation

Least squares estimation is the method which gives the minimum sum of squared errors in finding the optimal parameter values. The error term expresses the difference between the sample values, Y_i , and the estimated values, \hat{Y}_i . The estimated values \hat{Y}_i can be written as $E[Y_i|X_i]$. The least squares estimation is used for a simple linear regression. Given X is an independent variable, Y is a dependent variable, and β is a constant. Then, a linear relationship between Y and X can be set up as shown in Eq. (4):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad (4)$$

Considering the population regression model with p parameters and n observations written as follows:

$$\begin{aligned} \hat{Y}_1 &= E(Y_1|X_1) = \beta_0 + \beta_1 x_{1,1} + \beta_2 x_{2,1} + \dots + \beta_{p-1} x_{p-1,1} \\ \hat{Y}_2 &= E(Y_2|X_2) = \beta_0 + \beta_1 x_{1,2} + \beta_2 x_{2,2} + \dots + \beta_{p-1} x_{p-1,2} \\ &\vdots \\ \hat{Y}_n &= E(Y_n|X_n) = \beta_0 + \beta_1 x_{1,n} + \beta_2 x_{2,n} + \dots + \beta_{p-1} x_{p-1,n} \end{aligned} \quad (5)$$

Then, the matrix form is cast in the following form:

$$\hat{Y} = E(Y|X) = X\beta \quad (6)$$

$$\hat{Y} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix}; X = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} & \dots & x_{p-1,1} \\ 1 & x_{1,2} & x_{2,2} & \dots & x_{p-1,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & x_{2,n} & \dots & x_{p-1,n} \end{bmatrix}; \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$$

Estimates of the error terms refers to the differences between the observed value of the dependent variable and its predicted value in the model:

$$Y_i - \hat{Y}_i = \varepsilon_i \quad (7)$$

$$\varepsilon_i = Y_i - \hat{Y}_i = Y_i - E[Y_i|X_i] \quad (8)$$

$$S = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - E[Y_i|X_i])^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \quad (9)$$

Eq. (9) can be expressed in matrix form as follows:

$$S = \varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta) = Y'Y - \beta'X'Y - Y'X\beta + \beta'X'X\beta = Y'Y - 2\beta'X'Y + \beta'X'X\beta \quad (10)$$

It should be noted that $\beta'X'Y$ and $Y'X\beta$ are the same scalar. Then, the least squares estimation of β is shown in Eq. (11) and Eq. (12)

$$\frac{\partial S}{\partial \beta} = \frac{\partial \varepsilon'\varepsilon}{\partial \beta} = -2X'Y + 2X'X\hat{\beta} = 0 \quad (11)$$

$$\hat{\beta}_{OLS} = (X'X)^{-1} X'Y \quad (12)$$

where $\hat{\beta}_{OLS}$ is the least squares estimator of β .

Assumptions of LSE

1. The dependent variable, Y , is continuous.
2. The relationship between the dependent variable, Y , and the independent variable, X , is linear.
3. Observations are independently and randomly drawn.
4. The error term has the expected value equal to zero and the variation equal to σ^2 .

$$E[\varepsilon_i] = 0; \quad Var[\varepsilon_i] = \sigma^2 \mathbf{I}$$

5. The error term is not auto correlated.

$$COV[\varepsilon_i, \varepsilon_j] = 0; \quad \text{if } i \neq j$$

6. The error term, ε , is independent of X .

$$COV[X_i, \varepsilon_j] = 0; \quad \text{for all } i \text{ and } j$$

7. The error term is approximately normally distributed.

$$\varepsilon_i \approx N(0, \sigma^2); \quad \text{for } i = 1, 2, \dots, n$$

We can write this in matrix form as $\varepsilon \approx N(0, \sigma^2 \mathbf{I})$ where \mathbf{I} is identity matrix. From this assumption, the distribution of the dependent variable, Y , is written as: $Y \approx N(X\beta, \sigma^2 \mathbf{I})$

The properties of LSE

1. The estimator of least squares method is unbiased estimator. The estimator is unbiased if and only if the expected value of an estimator, $E[\hat{\beta}]$ equals parameter β , that is,

$$E[\hat{\beta}] = \beta \quad (13)$$

From Eq. (12), we derive the following equation:

$$\begin{aligned} \hat{\beta}_{OLS} &= (X'X)^{-1} X'Y = (X'X)^{-1} X'(X\beta + \varepsilon) \\ &= (X'X)^{-1} (X'X)\beta + (X'X)^{-1} X'\varepsilon \\ &= \beta + (X'X)^{-1} X'\varepsilon \end{aligned} \quad (14)$$

$$E[\hat{\beta}_{OLS}] = E[\beta] + E[(X'X)^{-1} X'\varepsilon] = \beta + (X'X)^{-1} X'E[\varepsilon]$$

According to Assumption 4, if $E[\varepsilon_i] = 0$

then $E[\hat{\beta}_{OLS}] = \beta$

Therefore, $\hat{\beta}_{OLS}$ is an unbiased estimator of β .

2. The least squares estimator is an efficient estimator. The estimator is efficient if and only if the variance of estimator, $V[\hat{\beta}]$ is minimum

$$Var[\hat{\beta}_{OLS}] = E[(\hat{\beta}_{OLS} - \beta)(\hat{\beta}_{OLS} - \beta)'] \quad (15)$$

From Eq. (14), we now obtain:

$$\begin{aligned} \hat{\beta}_{OLS} - \beta &= (X'X)^{-1} X'\varepsilon \\ Var[\hat{\beta}_{OLS}] &= E[(X'X)^{-1} X'\varepsilon)((X'X)^{-1} X'\varepsilon)'] \\ &= (X'X)^{-1} X'E[\varepsilon\varepsilon']X(X'X)^{-1} \end{aligned}$$

According to Assumption 4, given

$$Var[\varepsilon_i] = \sigma^2 \mathbf{I} = E[(\varepsilon - E[\varepsilon])^2] = E[\varepsilon^2] \quad \text{then}$$

$$\begin{aligned} Var[\hat{\beta}_{OLS}] &= (X'X)^{-1} X'\sigma^2 \mathbf{I} X(X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1} \end{aligned} \quad (16)$$

3. The estimator of least square method is the best unbiased estimator.

Given $\hat{\beta}^* = [(X'X)^{-1} X' + A]Y$, where A is a matrix which treats $\hat{\beta}^*$ as the unbiased parameter.

$$E[\hat{\beta}^*] = \beta$$

$$\begin{aligned} \beta &= [(X'X)^{-1} X' + A]E[Y] \\ &= \beta + AX\beta \end{aligned}$$

Thus, if $AX = 0$ then $\hat{\beta}^* = (X'X)^{-1} X'Y = \hat{\beta}_{OLS}$

$$\begin{aligned} Var[\hat{\beta}^*] &= E[(\hat{\beta}^* - \beta)(\hat{\beta}^* - \beta)'] \\ &= E[(X'X)^{-1} X'Y + AY - \beta][(X'X)^{-1} X'Y + AY - \beta)'] \\ &= E[(X'X)^{-1} X'Y + AY - \beta][(X'X)^{-1} X'Y + AY - \beta)']; Y = X\beta + \varepsilon \\ &= E[(X'X)^{-1} X'\varepsilon + AX\beta + A\varepsilon][(X'X)^{-1} X'\varepsilon + AX\beta + A\varepsilon)']; AX = 0 \\ &= E[(X'X)^{-1} X'\varepsilon + A\varepsilon][(X'X)^{-1} X'\varepsilon + A\varepsilon)'] \\ &= [(X'X)^{-1} X'X(X'X)^{-1} + AA' + (X'X)^{-1} (AX)' + (AX)(X'X)^{-1}] \sigma^2; AX = 0 \\ &= [(X'X)^{-1} + AA'] \sigma^2 \end{aligned}$$

Given $AA' = B = (b_{ij})$, where $b_{ij} > 0$. Then, $Var[\hat{\beta}^*]$ is minimum if and only if $b_{ij} = 0$ for all $i=1, 2, \dots, p$.

Given $A = (a_{ij})$, $b_{ii} = \sum_{j=1}^n a_{ij}^2 = 0$ Therefore, $a_{ij} = 0$ for $i=1, 2, \dots, p$ and $j = 1, 2, \dots, p$

That is, when $A = 0$, $\hat{\beta}^* = (X'X)^{-1} X'Y = \hat{\beta}_{OLS}$ is the minimum variance parameter of β .

The least squares estimation has many assumptions. This estimation method is considered to be improper to develop the fragility curve using structural damage data due to some invalid assumptions. For instance, Assumption 4, $Var[\varepsilon_i] = \sigma^2 \mathbf{I}$, which assumes the error term has a constant variance and Assumption 6, $COV[X_i, \varepsilon_j] = 0$, which implies that the error is uncorrelated with all independent variables. The variance of the errors of structure damage is small when the IM (intensity measure) is small (most structures are not damaged). That is, the variance increases when the structure starts to damage and then the variance decreases again when IM is large because most structures are damaged. As a result, the variance of the errors is not constant and depends on IM. Thus, Assumption 7, $\varepsilon_i \approx N(0, \sigma^2)$ becomes invalid. In addition, review of the literature indicates that the data of structural

component failure appear to fit the log-normal distribution well. Although the log-normal distribution can be transformed to linear logarithm, properties of the error term are changed.

The Method of Maximum Likelihood

Maximum likelihood estimation is a method for parameter estimation by maximizing the likelihood function. Given X is a random variable with probability distribution function $f(X|\theta)$, where θ is a single unknown parameter. Let x_1, x_2, \dots, x_n be the observed values in a random sample of size n . Then, the likelihood function of a sample is defined as:

$$L(\theta|X) = L(\theta|x_1, x_2, \dots, x_n) = \prod_i f(x_i|\theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdots f(x_n, \theta) \quad (17)$$

The likelihood function depends on the unknown parameter θ only. Essentially, the maximum likelihood estimator is the value of θ that maximizes the probability of occurrence of the sample by differentiating $L(\theta|X)$ with respect to θ and setting that derivative equal to zero:

$$\frac{\partial L(\theta|X)}{\partial \theta} = 0 \quad (18)$$

Then, with the second derivative value being less than zero we have a maximum:

$$\frac{\partial^2 L(\theta|X)}{\partial \theta^2} < 0 \quad (19)$$

The assumptions of MLE

- 1) The observations are independently and randomly drawn.
- 2) The parameter $\hat{\theta}$ maximizes the likelihood of the sample, i.e., it represents the maximum likelihood estimator of θ in which $L(\hat{\theta}|X) \geq L(\theta|X)$.

Properties of MLE

1. The maximum likelihood estimator does not need to be unbiased since it can be corrected. The estimator is unbiased if and only if the expected value of an estimator, $E[\hat{\theta}]$ equals parameter θ .

We consider the random sample x_1, x_2, \dots, x_n , where n is the random sample size. The random sampling distribution is assumed to be normal which comprises two unknown parameters, μ and σ^2 . The probability density function of normal distribution can be expressed as:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}[(x-\mu)/\sigma]^2} \quad (20)$$

$$L(\mu, \sigma^2|X) = L(\mu, \sigma^2|x_1, x_2, \dots, x_n) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}[(x_i-\mu)/\sigma]^2} \right) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-\mu)^2} \quad (21)$$

Simplifying Eq. (21) by taking the log of both sides yields:

$$\ln L(\mu, \sigma^2|X) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \quad (22)$$

Then, the estimator of μ and σ^2 is

$$\frac{\partial \ln L(\mu, \sigma^2|X)}{\partial \hat{\mu}} = \frac{\partial \left(-\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \hat{\mu})^2 \right)}{\partial \hat{\mu}} = -\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \hat{\mu}) = 0$$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \quad (23)$$

Next, we check that the estimator $\hat{\mu}$ gives maximum L :

$$\frac{\partial^2 \ln L(\mu, \sigma^2|X)}{\partial \hat{\mu}^2} = \frac{\partial \left(-\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \hat{\mu}) \right)}{\partial \hat{\mu}} = -\frac{1}{\sigma^2} \frac{\partial \left(\sum_{i=1}^n x_i - \sum_{i=1}^n \hat{\mu} \right)}{\partial \hat{\mu}} = -\frac{n}{\sigma^2} < 0$$

$$\frac{\partial \ln L(\mu, \sigma^2|X)}{\partial \hat{\sigma}^2} = \frac{\partial \left(-\frac{n}{2} \ln(2\pi\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (x_i - \hat{\mu})^2 \right)}{\partial \hat{\sigma}^2} = -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^n (x_i - \hat{\mu})^2 = 0$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (24)$$

And if the estimator $\hat{\sigma}^2$ treats maximum L then

$$\frac{\partial^2 \ln L(\mu, \sigma^2|X)}{\partial (\hat{\sigma}^2)^2} = \frac{\partial \left(-\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^n (x_i - \hat{\mu})^2 \right)}{\partial \hat{\sigma}^2} = \frac{n}{2\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6} \sum_{i=1}^n (x_i - \hat{\mu})^2 = \frac{n}{2\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6} (n\hat{\sigma}^2) = \frac{n}{2\hat{\sigma}^4} - \frac{n}{\hat{\sigma}^4} = -\frac{n}{2\hat{\sigma}^4} < 0 \quad \text{OK.}$$

According to Eq. (19), we obtain:

$$E[\hat{\mu}] = E[\bar{x}] = \frac{1}{n} E[x_1 + x_2 + x_3 + \dots + x_n] = \frac{1}{n} \sum_{i=1}^n E[x_i] = \frac{1}{n} (n\mu) = \mu \quad (25)$$

Thus, $\hat{\mu}$ denotes an unbiased estimator of μ . From Eq. (24), we now derive the following form:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu + \mu - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n ((x_i - \mu) - (\bar{x} - \mu))^2$$

$$= \frac{1}{n} \left(\sum_{i=1}^n (x_i - \mu)^2 - 2n(\bar{x} - \mu) + \sum_{i=1}^n (\bar{x} - \mu)^2 \right)$$

$$E[\hat{\sigma}^2] = \frac{1}{n} E \left[\sum_{i=1}^n (x_i - \mu)^2 - 2n(\bar{x} - \mu) + \sum_{i=1}^n (\bar{x} - \mu)^2 \right]$$

$$= \frac{1}{n} \left(\sum_{i=1}^n V[x_i] - nV[\bar{x}] \right)$$

Based on the properties of normal distribution in which $V[x_i] = \sigma^2$ and $V[\bar{x}] = \frac{\sigma^2}{n}$, we obtain

$$E[\hat{\sigma}^2] = \frac{1}{n} \left(\sum_{i=1}^n \sigma^2 - n \frac{\sigma^2}{n} \right) = \frac{1}{n} \left(n\sigma^2 - n \frac{\sigma^2}{n} \right)$$

$$= \frac{(n-1)\sigma^2}{n} \neq \sigma^2 \quad (26)$$

$\hat{\sigma}^2$ is the biased parameter of σ^2 . Then, $\hat{\sigma}^2$ is corrected to be unbiased as follows

$$\hat{\sigma}^2 = \frac{1}{n} \left(\frac{n}{n-1} \right) \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (27)$$

2. The maximum likelihood estimator is efficient. The estimator is efficient if and only if the variance of an estimator, $V[\hat{\theta}]$ is minimal and unbiased:

$$\begin{aligned} V[\hat{\mu}] &= \text{Var} \left[\frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \right] \\ &= \frac{1}{n^2} (E[x_1] + E[x_2] + E[x_3] + \dots + E[x_n]) \\ &= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned} \quad (28)$$

To determine the variance of $\hat{\sigma}^2$, Chi-square distribution is utilized. Chi-square distribution is the distribution of the sum of squared standard normal deviates. Given the random variable for a Chi-square distribution with k degrees of freedom; that is,

$$Q = Z_1^2 + Z_2^2 + Z_3^2 + \dots + Z_k^2 \sim \chi_{(k)}^2 \quad (29)$$

where Z_i is a standard normal distribution which is independently and identically distributed, that is, $Z_i \sim N(0,1)$. The expected value and variance of Q are written as:

$$\begin{aligned} E(Q) &= k \\ \text{Var}(Q) &= 2k \end{aligned} \quad (30)$$

According to Eq. (20), given $z = \frac{x-\mu}{\sigma}$ which can be written as $\frac{x-\mu}{\sigma} \sim N(0,1)$, it can be shown that

$$\begin{aligned} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 &: \chi_{(n)}^2 \\ \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 &= \sum_{i=1}^n \left(\frac{x_i - \mu - \bar{x} + \bar{x}}{\sigma} \right)^2 = \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} + \frac{\bar{x} - \mu}{\sigma} \right)^2 \\ &= \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 + \left(\frac{\bar{x} - \mu}{\sigma} \right)^2 \\ &= \chi_{(n-1)}^2 + \chi_{(1)}^2 \end{aligned} \quad (31)$$

Considering Eq. (28), the second term on the right-hand side of the equation is the normal distribution of \bar{X} which gives a Chi-square distribution with 1 degree of freedom. Then, the formula becomes

$$\sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 = \chi_{(n-1)}^2 \quad (32)$$

From Eq. (27), the unbiased estimator of parameter σ^2 , denoted as $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ or alternatively written as $\sum_{i=1}^n (x_i - \bar{x})^2 = (n-1)\hat{\sigma}^2$, is substituted into Eq. (32):

$$\begin{aligned} \frac{(n-1)\hat{\sigma}^2}{\sigma^2} &= \chi_{(n-1)}^2 \\ \text{Now, from Eq. (30) we have} \\ V \left[\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \right] &= 2(n-1) \\ V[\hat{\sigma}^2] &= \frac{2\sigma^4}{(n-1)} \end{aligned} \quad (33)$$

3. The maximum likelihood estimation method appears to be the best possible or optimal estimator because its variance of estimator equals the Cramér -Rao Lower Bound which yields a minimum variance estimator. Let $f(x|\theta)$ be the probability density function of the population and $\hat{\theta}$ denotes an unbiased estimator of parameter θ . The Cramér -Rao Lower Bound can be expressed as:

$$V(\hat{\theta}) \geq \frac{1}{nE \left[\frac{\partial \ln f(x|\theta)}{\partial \theta} \right]^2} \quad (34)$$

Assuming normal distribution, it follows that

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} \left(\frac{x-\mu}{\sigma} \right)^2}, -\infty < x < \infty$$

$$\ln f(x|\mu, \sigma^2) = -\frac{1}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} (x-\mu)^2$$

$$\frac{\partial \ln f(x|\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} (x-\mu)$$

$$E \left[\left(\frac{\partial \ln f(x|\mu, \sigma^2)}{\partial \mu} \right)^2 \right] = \frac{1}{\sigma^4} E(x-\mu)^2 = \frac{V[x]}{\sigma^4} = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}$$

Taking into account Eq. (28), substituting the above equation into Eq. (34) yields

$$\begin{aligned} \frac{1}{nE \left[\left(\frac{\partial \ln f(x|\mu, \sigma^2)}{\partial \mu} \right)^2 \right]} &= \frac{1}{n \left(\frac{1}{\sigma^2} \right)} = \frac{\sigma^2}{n} = V[\hat{\mu}] \\ \frac{\partial \ln f(x|\mu, \sigma^2)}{\partial \sigma^2} &= -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} (x-\mu)^2 \end{aligned} \quad (35)$$

$$E \left[\left(\frac{\partial \ln f(x|\mu, \sigma^2)}{\partial \sigma^2} \right)^2 \right] = \frac{1}{4\sigma^4} + \frac{1}{4\sigma^4} \left[V \left[\left(\frac{x-\mu}{\sigma} \right)^2 \right] + \left(E \left[\left(\frac{x-\mu}{\sigma} \right)^2 \right] \right)^2 \right] - \frac{1}{2\sigma^6} V[x]$$

$$\text{From Eq. (30), since } V \left[\left(\frac{x-\mu}{\sigma} \right)^2 \right] = 2, \quad E \left[\left(\frac{x-\mu}{\sigma} \right)^2 \right] = 1$$

then,

$$E \left[\left(\frac{\partial \ln f(x|\mu, \sigma^2)}{\partial \sigma^2} \right)^2 \right] = \frac{1}{4\sigma^4} + \frac{1}{4\sigma^4} (2+1) - \frac{1}{2\sigma^6} = \frac{1}{2\sigma^4}$$

Now, the above equation is substituted into Eq. (34) and Eq. (33) is taken into consideration:

$$\frac{1}{nE\left[\left(\frac{\partial \ln f(x|\mu, \sigma^2)}{\partial \sigma^2}\right)^2\right]} = \frac{1}{n\left(\frac{1}{2\sigma^4}\right)} = \frac{2\sigma^4}{n} < V[\hat{\sigma}^2] = \frac{2\sigma^4}{(n-1)} \quad (36)$$

4. The maximum likelihood estimator is consistent because it converges to the true value of parameter being estimated.

From Eq. (26), despite having $E[\hat{\sigma}^2] = \frac{(n-1)\sigma^2}{n} \neq \sigma^2$,

but $E[\hat{\sigma}^2]$ converges to σ^2 for a large sample size ($n-1 \approx n$). In the same way, $2\sigma^4/n$ converges to $V[\hat{\sigma}^2] = \frac{2\sigma^4}{(n-1)}$ as shown in Eq. (36).

The development of fragility curve using the maximum likelihood method have been adopted in many previous studies [1], [4]-[10]. The review of literature shows that the structural component failure data fits with the lognormal distribution well. A log-normal distribution appears to set robust precedent in risk analysis and assessment. Furthermore, theoretically, applying a lognormal distribution yields zero probability density at and below zero intensity measure (IM).

TESTING ESTIMATION METHOD FOR FRAGILITY CURVE

As a first step of the method for the fragility curve development, the parameter values of μ and σ^2 assuming a lognormal distribution is chosen. Next, 100 groups of 50 data is randomly generated from these parameters. Finally, estimated values of the parameter mean and variance ($\hat{\mu}$ and $\hat{\sigma}^2$) are obtained from using the MLE and LSE methods for each group of data. Precisely, $\hat{\mu}$ is not equal to μ for all groups of data, as well as the estimated value of $\hat{\sigma}^2$. A summary of findings is presented in Table 1. The distribution of estimated parameter value is shown in Fig.2. The estimated probability density function, which is inherent in Eq. (2), appears to have a bell-shaped curve. The maximum likelihood estimation gives $\mu = -0.80$ and $\sigma^2 = 0.40$ which has a smaller variance than the variance of the least square estimation. The results show that the MLE method based on the lognormal approach yields better estimates than the LSE method.

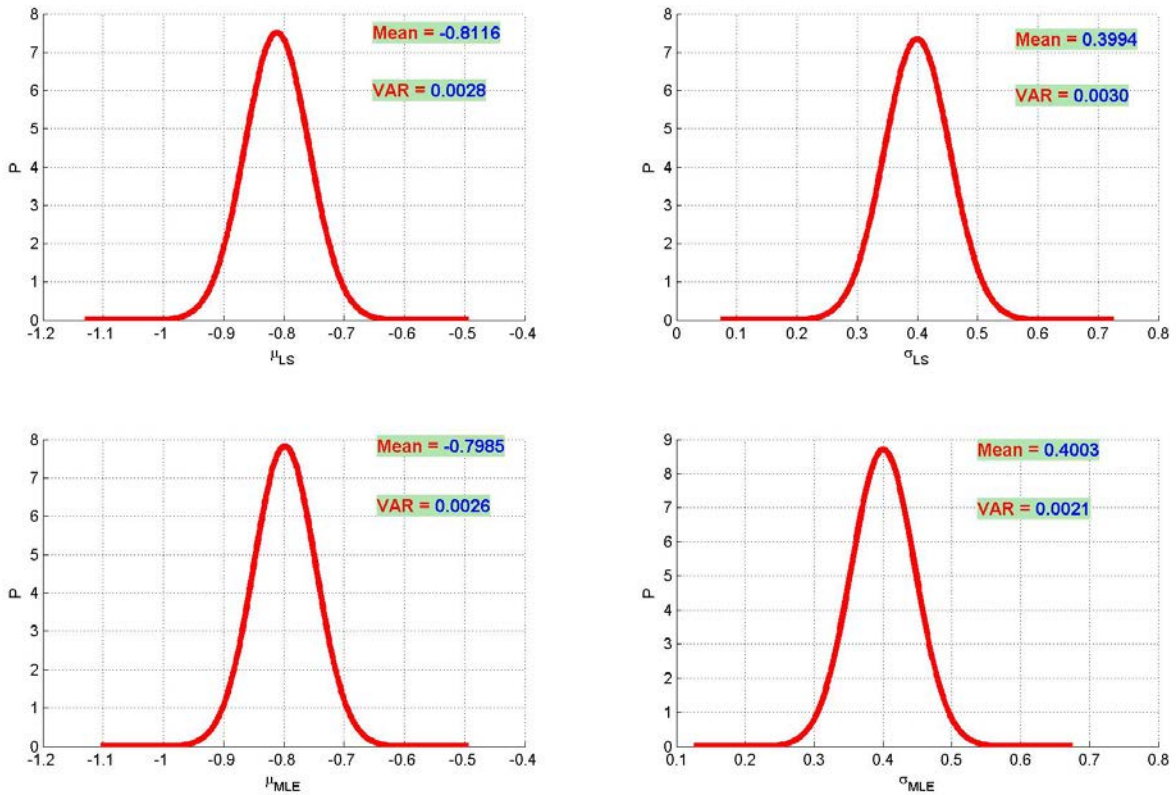


Fig. 2 The distribution of estimated value.

Table 1 The estimated parameters by LSE and MLE

Methods in numerical fitting	Estimated parameter		Error (%)	
	μ	σ^2	μ	σ^2
Least Squares	-0.8116	0.3994	0.0065	0.0016
Maximum Likelihood	-0.7985	0.4003	0.0063	0.0016

3.2 CONCLUSIONS

In this paper, the estimated parameter of mean (μ) and variance (σ^2) of a lognormal distribution are determined either by the maximum likelihood estimation (MLE) or by the least squares estimation (LSE). Then, the estimated results obtained from these two methods are compared with each other. To develop a fragility curve, the method starts by choosing the parameter values of μ and σ^2 , assuming a lognormal distribution. The results indicate that the size of variance obtained with MLE is smaller than the variance obtained with LSE. In summary, the MLE method based on the lognormal approach gives better outcomes of the estimated parameter of mean (μ) and variance (σ^2) of lognormal distribution. In other words, fragility curves developed by the MLE method appears to be more consistent and efficient than those developed by the LSE method.

REFERENCES

- [1] Porter K., Kennedy R. and Bachman R., Creating fragility functions for performance-based earthquake engineering. *Earthquake Spectra*, 2007, Issue 23, pp. 471-489.
- [2] Mackie K. and Stojadinovic B., Fragility curves for reinforced concrete highway overpass bridges. 13th World Conference on Earthquake Engineering, Vancouver, B.C., Canada August 1-6, 2004, Paper No. 1553.
- [3] Tien Y.M. and Pai C.H., Development of fragility curves for the 1999 Chi-Chi, Taiwan, earthquake. *Proceedings of the 8th U.S. National Conference on Earthquake Engineering*, San Francisco, California, USA April 18-22, 2006, Paper No. 401.
- [4] Foytong P. and Ruangrassamee A., Fragility curves of reinforced-concrete buildings damaged by the 2004 tsunami. *Engineering and Applied Science Research*, Vol. 43, Issue S3, 2016, pp. 419-423.
- [5] Karim K.R. and Yamazaki F., A simplified method of constructing fragility curves for highway bridges. *Earthquake Engineering and Structural Dynamics*, 2003, Vol. 32, pp. 1603-1626.
- [6] Mander B., Fragility curve development for assessing the seismic vulnerability of highway bridges. *University at Buffalo*, 1999, State University of New York., pp. 1-10.
- [7] Saxena V., Deodatis G., Shinozuka M. and Feng M.Q., Development of fragility curves for multi-span reinforced concrete bridges. *International conference on Monte Carlo simulation Technical Program*, Balkema Publishers, 2000, pp. 499-504.
- [8] Shinozuka M., Feng M.Q., Kim H.K. and Kim S.H., Nonlinear static procedure for fragility curve development. *Journal of Engineering Mechanics*, 2000, Vol. 126, pp. 1287-1295.
- [9] Singhal A. and Kiremidjian A.S., Bayesian updating of fragilities with application to RC frames. *Journal of Structural Engineering*, 1998, pp. 922-929.
- [10] Yamazaki, F. and Murao, O., Fragility curves for buildings in Japan based on experience from the 1995 Kobe Earthquake. *Institute of Industrial Science*, 2000, University of Tokyo, Tokyo, Japan.

Copyright © Int. J. of GEOMATE. All rights reserved, including the making of copies unless permission is obtained from the copyright proprietors.
