Nuclear Engineering and Technology 54 (2022) 2084–2093

Contents lists available at ScienceDirect

Nuclear Engineering and Technology

journal homepage: www.elsevier.com/locate/net

Original Article

Theoretical approach for uncertainty quantification in probabilistic safety assessment using sum of lognormal random variables

Gyun Seob Song, Man Cheol Kim*

Department of Energy Systems Engineering, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul, 06974, South Korea

ARTICLE INFO

Article history:
Received 29 June 2021
Received in revised form
12 November 2021
Accepted 25 December 2021
Available online 30 December 2021

Keywords:
Probabilistic safety assessment
Uncertainty analysis
Sum of lognormal random variables
Theoretical approach
Numerical integration
Change of variables

ABSTRACT

Probabilistic safety assessment is widely used to quantify the risks of nuclear power plants and their uncertainties. When the lognormal distribution describes the uncertainties of basic events, the uncertainty of the top event in a fault tree is approximated with the sum of lognormal random variables after minimal cutsets are obtained, and rare-event approximation is applied. As handling complicated analytic expressions for the sum of lognormal random variables is challenging, several approximation methods, especially Monte Carlo simulation, are widely used in practice for uncertainty analysis. In this study, a theoretical approach for analyzing the sum of lognormal random variables using an efficient numerical integration method is proposed for uncertainty analysis in probability safety assessments. The change of variables from correlated random variables with a complicated region of integration to independent random variables with a unit hypercube region of integration is applied to obtain an efficient numerical integration. The theoretical advantages of the proposed method over other approximation methods are shown through a benchmark problem. The proposed method provides an accurate and efficient approach to calculate the uncertainty of the top event in probabilistic safety assessment when the uncertainties of basic events are described with lognormal random variables.

1. Introduction

Uncertainty analysis aims to investigate the uncertainties in events of interest. In Level 1 probabilistic safety assessment (PSA), the event of interest corresponds to the top event of a single top fault tree comprising logic gates and basic events. Uncertainty analysis primarily focuses on the uncertainty associated with the occurrence of an event and its probability or frequency, which is modeled using a random variable with a probability distribution [1]. In the PSA of nuclear power plants, the uncertainties in the occurrence probabilities of events are often represented using lognormal distributions over the past decades [2,3]. The random variable for the uncertainty of the top event is given as a function of random variables for the uncertainties of basic events [4]. When the basic events are modeled with lognormal random variables, the failure frequency of the top event in Level 1 PSA, which is typically the core damage frequency, can be approximated with the sum of the lognormal random variables.

However, the closed-form probability density function of the sum of lognormal random variables is not known. Thus, several approximation methods have been developed to estimate the probability density function of the sum of lognormal random variables. Fenton-Wilkinson's method [5] and Schwartz-Yeh's method [6] approximated the sum of lognormal random variables to a single lognormal random variable through the method of moments. Moreover, in Ref. [7], an optimal lognormal approximation based on the minimax method was implemented using a linear function on the log probability. The probability density function of the sum of lognormal random variables was also approximated using orthonormal polynomials [8] and the moments of the distribution [9]. By contrast, Monte Carlo simulation can be used to approximate the probability density function of a function of random variables using samples and bins [10]. Among the approximation methods, El-Shanawany et al. [11] applied Fenton-Wilkinson's method to perform the uncertainty analysis of a nuclear power plant and compared the analysis with Monte Carlo simulations. For the uncertainty analysis in PSA, Monte Carlo simulations have been widely used.

Even though approximation methods are effective in numerous cases, these methods have inherent limitations, and hence, the theoretical approach with analytic expressions has unique and

* Corresponding author.
E-mail address: charleskim@cau.ac.kr (M.C. Kim).

https://doi.org/10.1016/j.net.2021.12.033

incomparable values. A method for finding the analytic expression of a function of random variables is the Jacobian transformation of the joint density [12]. In this method, the probability density function of the sum of lognormal random variables is represented as the multiple integration of a multivariate lognormal distribution. However, the integrand function is intricate, requiring numerical integration. As the number of random variables increases, the numerical integration may present the so-called "curse of dimensionality" problem. Genz [13] proposed an efficient numerical integration method to compute the cumulative distribution function of multivariate normal distribution even when the integration region is not bounded.

This study develops a new method to find the uncertainty of the top event in Level 1 PSA when the uncertainties of basic events are modeled with lognormal random variables. Even though Monte Carlo approach is widely used for the uncertainty analysis of Level 1 PSA for nuclear power plants because it is simple and easy to apply, it only provides a numerical approximation to the probability density function of the top event in the form of a histogram. The proposed method relies on the analytic expression for the probability density function of the top event and the transformation of variables to facilitate numerical integration. One of most important advantages of the proposed method over the traditional Monte Carlo approach is that it provides the theoretical foundation on which further theoretical development is possible.

The remainder of this paper is organized as follows. Section 2 describes the moments of the joint probability density function and how the multiple integral with correlated lognormal variables in a complicated integration region is transformed to that with independent uniform variables in a unit hypercube integration region. Section 3 demonstrates the application of the proposed method using a benchmark example concerning the uncertainty analysis of a nuclear power plant. A comparison with other approximation methods, such as the Monte Carlo simulation and Fenton-Wilkinson's method, is also included. Lastly, Section 4 provides the conclusions of the study.

2. Uncertainty quantification in PSA

In the uncertainty analysis of Level 1 PSA, the single top fault tree is transformed into a logically equivalent fault tree where the top event comprises the logical OR of minimal cutsets. Minimal cutsets are the smallest combinations of basic events leading to the occurrence of the top event. Because the basic events in a fault tree are assumed to be independent of each other, the occurrence probability of a minimal cutset is the product of all the occurrence probabilities of the basic events included in the minimal cutset, as follows:

$$X_i = \prod_{j=1}^{n_i} B_{ij} \tag{1}$$

where X_i is the random variable for the occurrence probability of the i-th minimal cutset, B_{ij} is the random variable for the occurrence probability of j-th basic event included in the i-th minimal cutset, and n_i is the number of basic events in the i-th minimal cutset. When the occurrence probabilities of basic events are modeled with lognormal random variables, the occurrence probability of a minimal cutset is also a lognormal random variable because the product of lognormal random variables is given as another lognormal random variable. In general, the occurrence probability of a minimal cutset is sufficiently low to apply a rare-event approximation. Hence, the occurrence probability of the top event is well approximated with the sum of the occurrence probabilities of minimal cutsets in various cases, as follows:

$$T \approx S = \sum_{i=1}^{n} X_i \tag{2}$$

where T is the top event probability, S is the rare-event approximation for T, and n is the number of minimal cutsets considered in S to approximate T.

The main idea for the uncertainty analysis when basic events are given as lognormal random variables is described in Eq. (2). In this case, the top event frequency of a single top fault tree, typically the core damage frequency in Level 1 PSA, can be approximated with the sum of a finite number of high-ranking lognormal random variables. In other words, the uncertainty in the top event frequency can be described with the probability density function of the sum of lognormal random variables. This is owing to the unique property of lognormal random variables, i.e., the product of the lognormal random variables for basic events becomes another lognormal random variable for a minimal cutset.

However, the lognormal random variables for minimal cutsets (X_i's) may be dependent to each other. Thus, some minimal cutsets may include the same initiating event or basic events. For example, several high-ranking minimal cutsets may commonly include the loss of offsite power as the initiating event and the basic events for the failure of emergency diesel generators. Because high-ranking minimal cutsets may be dependent to each other, the top event frequency is given as the sum of correlated lognormal random variables. To analyze the uncertainty in the top event frequency when basic events are given with lognormal random variables, we first consider Eq. (1). Let $X_1, \cdots, X_n$ be correlated lognormal random variables. The joint probability density function for $X_1, \cdots, X_n$ is given as follows:

$$f_X(x_1, \cdots, x_n) = \frac{1}{\sqrt{(2\pi)^n}\,|\boldsymbol{\Sigma}|^{\frac{1}{2}} \prod_{i=1}^{n} x_i} e^{-\frac{1}{2}\left(\begin{bmatrix} \ln(x_1)-\mu_1 \\ \vdots \\ \ln(x_n)-\mu_n \end{bmatrix}^T \boldsymbol{\Sigma}^{-1} \begin{bmatrix} \ln(x_1)-\mu_1 \\ \vdots \\ \ln(x_n)-\mu_n \end{bmatrix} \right)} \tag{3}$$

where μ_i's ($i = 1, \cdots, n$) and $\boldsymbol{\Sigma}$ are the means and covariance matrix of the logarithm of random variables, respectively.

The means of the logarithm of random variables are obtained from the first and second moments of random variables. Because the basic events in a fault tree are assumed to be independent, the first and second moments of the minimal cutset are the product of the first and second moments of all the included basic events. The mean of the logarithm of a minimal cutset is obtained as follows:

$$\mu_i = \ln\left(\frac{\prod_{j=1}^{n_i} E[B_{ij}]^2}{\sqrt{\prod_{j=1}^{n_i} E\left[B_{ij}^2\right]}} \right) \tag{4}$$

where μ_i is the mean of the logarithm of the random variable for the i-th minimal cutset (X_i).

The covariance between the logarithm of two lognormal random variables, which is an element of $\boldsymbol{\Sigma}$, can be derived from the covariance between two lognormal random variables, $\mathrm{Cov}(X_i, X_k)$ [14].

$$\boldsymbol{\Sigma}_{ik} = \mathrm{Cov}(\ln(X_i), \ln(X_k)) = \ln\left(\frac{\mathrm{Cov}(X_i, X_k)}{E[X_i]E[X_k]} + 1 \right) \tag{5}$$

The random variables for the basic events included in the minimal cutsets are classified into three categories. Let P_{Int} be the product of all the random variables for the basic events included in both minimal cutsets, that is the intersection of two minimal cutsets. Let $P_{X_i \backslash X_k}$ and $P_{X_k \backslash X_i}$ be the products of all random variables for the basic events included only in X_i and X_k, that is the set difference of two minimal cutsets, respectively. Then, X_i and X_k are given as follows:

$$X_i = P_{\text{Int}} P_{X_i \backslash X_k} \tag{6}$$

$$X_k = P_{\text{Int}} P_{X_k \backslash X_i} \tag{7}$$

The covariance between X_i and X_k is expressed as follows:

$$\begin{aligned}\text{Cov}(X_i, X_k) &= E[X_i X_k] - E[X_i]E[X_k] = E\left[{P_{\text{Int}}}^2 P_{X_i \backslash X_k} P_{X_k \backslash X_i}\right] \\ &\quad - E\left[P_{\text{Int}} P_{X_i \backslash X_k}\right] E\left[P_{\text{Int}} P_{X_k \backslash X_i}\right]\end{aligned} \tag{8}$$

The basic events are independent of each other and no basic event can be included in both P_{Int}, $P_{X_i \backslash X_k}$, and $P_{X_k \backslash X_i}$ simultaneously. Hence, P_{Int}, $P_{X_i \backslash X_k}$, and $P_{X_k \backslash X_i}$ are independent of each other.

$$E\left[{P_{\text{Int}}}^2 P_{X_i \backslash X_k} P_{X_k \backslash X_i}\right] = E\left[{P_{\text{Int}}}^2\right] E\left[P_{X_i \backslash X_k}\right] E\left[P_{X_k \backslash X_i}\right] \tag{9}$$

$$\begin{aligned}&E\left[P_{\text{Int}} P_{X_i \backslash X_k}\right] E\left[P_{\text{Int}} P_{X_k \backslash X_i}\right] = E[P_{\text{Int}}] E\left[P_{X_i \backslash X_k}\right] \cdot E[P_{\text{Int}}] E\left[P_{X_k \backslash X_i}\right] \\ &= (E[P_{\text{Int}}])^2 \cdot E\left[P_{X_i \backslash X_k}\right] E\left[P_{X_k \backslash X_i}\right]\end{aligned} \tag{10}$$

Eq. (8) becomes

$$\begin{aligned}\text{Cov}(X_i, X_k) &= \left(E\left[{P_{\text{Int}}}^2\right] - (E[P_{\text{Int}}])^2\right) E\left[P_{X_i \backslash X_k}\right] E\left[P_{X_k \backslash X_i}\right] \\ &= \text{Var}(P_{\text{Int}}) E\left[P_{X_i \backslash X_k}\right] E\left[P_{X_k \backslash X_i}\right]\end{aligned} \tag{11}$$

Considering Eqs. (4), (5) and (11), the means (μ_i's) and covariance matrix ($\boldsymbol{\Sigma}$) of the logarithm of the random variables for the specified high-ranking minimal cutsets for Eq. (3) can be calculated.

As defined in Eq. (2), let S be the sum of the correlated lognormal random variables as

$$S = X_1 + \cdots + X_n \tag{12}$$

The analytical mathematical formula for the probability density function of a rational function of random variables was derived by Gauss [15]. Assuming that the rational function is a simple sum of random variables, we can derive the probability density function for the sum of random variables. The cumulative distribution function for S is given as

$$F_S(s) = \int_0^s \cdots \int_0^{s - x_1 - \cdots - x_{n-1}} f_X(x_1, \cdots, x_n) dx_1 \cdots dx_n \tag{13}$$

The probability density function for S, which is the derivative of Eq. (13), is given as

$$f_S(s) = \int_0^s \cdots \int_0^{s - x_1 - \cdots - x_{n-2}} f_X(x_1, \cdots, x_{n-1}, s - x_1 - \cdots - x_{n-1}) dx_1 \cdots dx_{n-1} \tag{14}$$

When the number of random variables is large, the numerical integration of Eq. (14) requires substantial time; hence, it is unpractical. The region of integration in Eq. (14) is complicated because the upper bounds of the variables depend on their previous variables. Furthermore, the random variables may be correlated in practical applications.

We transform the correlated lognormal random variables to independent identically distributed uniform random variables. The transformation of Genz [13] was intended for integrals with constant upper and lower bounds; by contrast, we develop a more advanced and complicated transformation for the integrals, where the upper bounds of the variables depend on their previous variables, as in Eq. (14).

Let $Y_1, \cdots, Y_{n-1}$ be the logarithm of $X_1, \cdots, X_{n-1}$ as

$$Y_i = \ln X_i \tag{15}$$

for $i = 1, \cdots, n-1$. Y_i's are the random variables of a multivariate normal distribution with μ_i's and a leading principal submatrix of order $n-1$ of $\boldsymbol{\Sigma}$ as the means and covariance matrix. Eq. (13) becomes

$$f_S(s) = \int_{-\infty}^{\ln(s)} \cdots \int_{-\infty}^{\ln(s - e^{y_1} - \cdots - e^{y_{n-2}})} f_Y(y, \cdots, y_{n-1}, s) dy_1 \cdots dy_{n-1} \tag{16}$$

where

$$f_Y(y_1, \cdots, y_{n-1}, s) = \frac{e^{-\frac{1}{2}\left(\begin{bmatrix} y_1 - \mu_1 \\ \vdots \\ y_{n-1} - \mu_{n-1} \\ \ln(s - e^{y_1} - \cdots - e^{y_{n-1}}) - \mu_n \end{bmatrix}^T \boldsymbol{\Sigma}^{-1} \begin{bmatrix} y_1 - \mu_1 \\ \vdots \\ y_{n-1} - \mu_{n-1} \\ \ln(s - e^{y_1} - \cdots - e^{y_{n-1}}) - \mu_n \end{bmatrix}\right)}}{\sqrt{(2\pi)^n} |\boldsymbol{\Sigma}|^{\frac{1}{2}} (s - e^{y_1} - \cdots - e^{y_{n-1}})} \tag{17}$$

$$dy_i = \frac{d \ln x_i}{dx_i} dx_i = \frac{1}{x_i} dx_i \tag{18}$$

for $i = 1, \cdots, n-1$.

The correlated normal random variables can be modeled with uncorrelated standard normal random variables via an affine transformation, as follows:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_{n-1} \\ \ln\left(S - e^{Y_1} - \cdots - e^{Y_{n-1}}\right) \end{bmatrix} = C \begin{bmatrix} Z_1 \\ \vdots \\ Z_{n-1} \\ \alpha \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_{n-1} \\ \mu_n \end{bmatrix} \tag{19}$$

where

$$\alpha = \frac{\ln\left(s - \sum_{j=1}^{n-1} e^{\mu_j + \sum_{k=1}^{j} C_{j,k} Z_k}\right) - \mu_n - \sum_{j=1}^{n-1} C_{n,j} Z_j}{C_{n,n}} \tag{20}$$

In Eq. (19), C is the lower triangular matrix from the Cholesky decomposition of the covariance matrix $\boldsymbol{\Sigma}$, and Z_i's are the standard normal random variables. By solving Eq. (19) for $Z_1, \cdots, Z_{n-1}$,

$$Z_i = \frac{Y_i - \mu_i - \sum_{k=1}^{i-1} C_{i,k} Z_k}{C_{i,\,i}} \tag{21}$$

for $i = 1, \cdots, n-1$, where $C_{i,\,k}$ is an element of the matrix C. The elements of the Jacobian matrix of transformation are

$$\mathrm{J}_{i,j} = \frac{\partial z_i}{\partial y_j} = \frac{\partial}{\partial y_j}\left(\frac{y_i - \mu_i - \sum_{m=1}^{j-1} C_{i,m} z_m}{C_{i,i}}\right) \tag{22}$$

for $i, j = 1, \cdots, n-1$. The Jacobian matrix is a lower triangular matrix. The absolute value of the determinant of the Jacobian matrix is

$$|\mathrm{J}| = \left|\prod_{i=1}^{n-1} \frac{1}{C_{i,i}}\right| = \prod_{i=1}^{n-1} \frac{1}{C_{i,i}} \tag{23}$$

where $C_{i,\,i}$ for $i = 1, \cdots, n-1$ is always positive because the covariance matrix C is real and positive-definite [16]. Eq. (16) becomes

$$f_S(t) = \int_{-\infty}^{\frac{\ln(t)-\mu_1}{C_{1,1}}} \cdots \int_{-\infty}^{\frac{\ln\left(t - \sum_{j=1}^{n-2} e^{\mu_j + \sum_{k=1}^{j} C_{j,\,k} z_k}\right) - \mu_{n-1} - \sum_{j=1}^{n-2} C_{n-1,\,j} z_j}{C_{n-1,\,n-1}}} f_Z(z_1, \cdots, z_{n-1}, s) dz_1 \cdots dz_{n-1} \tag{24}$$

where

$$f_Z(z_1, \cdots, z_{n-1}, s) = \frac{1}{\sqrt{(2\pi)^n}\, C_{n,n}\left(s - \sum_{j=1}^{n-1} e^{\mu_j + \sum_{k=1}^{j} C_{j,\,k} z_k}\right)} e^{-\frac{1}{2}\left(\begin{bmatrix} z_1 \\ \vdots \\ z_{n-1} \\ \alpha \end{bmatrix}^T I \begin{bmatrix} z_1 \\ \vdots \\ z_{n-1} \\ \alpha \end{bmatrix}\right)} = \frac{1}{\sqrt{(2\pi)^n}\, C_{n,n}\left(s - \sum_{j=1}^{n-1} e^{\mu_j + \sum_{k=1}^{j} C_{j,\,k} z_k}\right)} e^{-\frac{1}{2}\left(z_1^2 + \cdots + z_{n-1}^2 + \alpha^2\right)} \tag{25}$$

In Eq. (25), I is the n-dimensional identity matrix.

A standard normal random variable can be modeled with a uniform random variable by using its cumulative distribution function, as follows:

$$U_i = \Phi(Z_i) \tag{26}$$

for $i = 1, \cdots, n-1$, where $\Phi(Z_i)$ is the cumulative distribution function of a standard normal random variable Z_i, and U_i is a uniform random variable. Eq. (24) becomes

$$f_S(t) = \int_0^{e_1} \cdots \int_0^{e_{n-1}} f_U(u_1, \cdots, u_{n-1}, s) du_1 \cdots du_{n-1} \tag{27}$$

where

$$f_U(u_1, \cdots, u_{n-1}, s) = \frac{1}{\sqrt{2\pi}\, C_{n,n}\left(s - \sum_{j=1}^{n-1} e^{\mu_j + \sum_{k=1}^{j} C_{j,\,k} \Phi^{-1}(u_k)}\right)} e^{-\frac{1}{2}\alpha^2} \tag{28}$$

$$du_i = \frac{d\Phi(Z_i)}{dz_i} dz_i = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z_i^2} dz_i \tag{29}$$

and

$$e_i = \Phi\left(\frac{\ln\left(s - \sum_{j=1}^{i-1} e^{\mu_j + \sum_{k=1}^{j} C_{j,k} \Phi^{-1}(u_k)}\right) - \mu_i - \sum_{j=1}^{i-1} C_{i,j} \Phi^{-1}(u_j)}{C_{i,\,i}}\right) \tag{30}$$

for $i = 1, \cdots, n-1$.

The interval of a uniform random variable U_i is determined by the upper bounds of its former variables. The sampling region can be simplified by representing the uniform random variables U_i's as the standard uniform random variables W_i's via an affine transformation, which is related to the convex combination of two points, as given by

$$U_i = e_i W_i \tag{31}$$

where e_i is the upper bound of U_i for $i = 1, \cdots, n-1$. The lower bounds in Eq. (27) are 0. Eq. (27) becomes

$$f_S(s) = \int_0^1 \cdots \int_0^1 f_W(w_1, \cdots, w_{n-1}, s) dw_1 \cdots dw_{n-1} \tag{32}$$

where

$$f_W(w_1, \cdots, w_{n-1}, s) = e_1 e_2 \cdots e_{n-1} \frac{1}{\sqrt{2\pi}\, C_{n,n} \left(s - \sum_{j=1}^{n-1} e^{\mu_j + \sum_{k=1}^{j} C_{j,k} \Phi^{-1}(w_k e_k)} \right)} e^{-\frac{1}{2}\alpha^2} \tag{33}$$

$$dw_i = \frac{1}{e_i} du_i \tag{34}$$

for $i = 1, \cdots, n-1$.

Fig. 1 summarizes the change of variables described above from correlated random variables with a complicated region of integration to independent random variables with a unit hypercube region of integration.

For example, the probability density function for the sum of two lognormally-distributed random variables ($n = 2$) is given as follows:

$$f_S(s) = \int_0^s f_X(x_1, s - x_1) dx_1 = \int_0^s \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2} x_1 (s - x_1)} e^{-\frac{1}{2\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \ln x_1 - \mu_1 \\ \ln(s - x_1) - \mu_2 \end{bmatrix}^T \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} \begin{bmatrix} \ln x_1 - \mu_1 \\ \ln(s - x_1) - \mu_2 \end{bmatrix}}$$

$$dx_1 = \int_0^1 \Phi\left(\frac{\ln s - \mu_1}{\sigma_1}\right) \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2} \left(s - e^{\sigma_1 \Phi^{-1}\left(\Phi\left(\frac{\ln s - \mu_1}{\sigma_1}\right) w_1\right) + \mu_1} \right)} - \frac{1}{2} \left(\frac{\ln\left(s - e^{\sigma_1 \Phi^{-1}\left(\Phi\left(\frac{\ln s - \mu_1}{\sigma_1}\right) w_1\right) + \mu_1} \right) - \mu_2 - \rho\sigma_2 \Phi^{-1}\left(\Phi\left(\frac{\ln s - \mu_1}{\sigma_1}\right) w_1\right)}{\sigma_2\sqrt{1-\rho^2}} \right)^2 dw_1 \tag{35}$$

The Monte Carlo integration is an efficient method for performing multi-dimensional integration numerically. Thus, we could apply the Monte Carlo integration method to Eq. (32) to calculate the probability density function $f_S(s)$. Eq. (32) can be expressed by the expected value of the integrand function with respect to the multivariate random variable $\boldsymbol{W} = (W_1, \cdots, W_{n-1})^T$.

$$f_S(s) = V \cdot E_{\boldsymbol{W}}[f_W(W_1, \cdots, W_{n-1}, s)] = E_{\boldsymbol{W}}[f_W(W_1, \cdots, W_{n-1}, s)] \tag{36}$$

where V is the volume of the sample space, that is the hypercube with uniform random variables, which equals 1. Eq. (36) is approximated considering M independent and identical uniform samples for $W_1, \cdots, W_{n-1}$, which are denoted as $w_{1,m}, \cdots, w_{n-1,m}$ for $m = 1, \cdots, M$:

$$f_S(s) \approx \bar{f}_W(\boldsymbol{W}, s) = \frac{1}{M} \sum_{m=1}^{M} f_W(w_{1,m}, \cdots, w_{n-1,m}, s) \tag{37}$$

where $\bar{f}_W(\boldsymbol{W}, s)$ indicates the sample mean of $f_W(w_1, \cdots, w_{n-1}, s)$. According to the central limit theorem, the sample mean approaches a normal distribution when the sample size is sufficiently large. The standard error of the mean is approximated with the standard deviation of the samples, as follows:

$$\sigma_{\bar{f}_W(\boldsymbol{W},s)} = \frac{\sigma_{f_W(\boldsymbol{W},s)}}{\sqrt{M}} = \frac{\sqrt{\mathrm{Var}(f_W(\boldsymbol{W}, s))}}{\sqrt{M}} \approx \underset{\bar{f}_W}{\widehat{\sigma}}(\boldsymbol{W}, s) = \frac{1}{\sqrt{M}} \cdot \sqrt{\frac{1}{M-1} \sum_{m=1}^{M} \left(f_W(w_{1,m}, \cdots, w_{n-1,m}, s) - \bar{f}_W(\boldsymbol{W}, s) \right)^2} \tag{38}$$

where $\underset{\bar{f}_W}{\widehat{\sigma}}(\boldsymbol{W}, s)$ indicates the estimator for the standard error of the mean.

When the confidence level is given as $(1 - c)$, the confidence interval for $f_S(s)$ is given with a t-distribution because the true standard error of the mean is not known and hence it is estimated by the standard deviation of the samples, as follows:

$$\Pr\left(\bar{f}_W(\boldsymbol{W}, s) - t_{c/2} \cdot \underset{\bar{f}_W}{\widehat{\sigma}}(\boldsymbol{W}, s) < f_S(s) < \bar{f}_W(\boldsymbol{W}, s) + t_{c/2} \cdot \underset{\bar{f}_W}{\widehat{\sigma}}(\boldsymbol{W}, s) \right) = 1 - c \tag{39}$$

For a 95% confidence interval, $t_{c/2}$ is 1.96. Moreover, the error of the Monte Carlo integration for the confidence level is defined as follows:

$$\varepsilon = \frac{t_{c/2} \cdot \underset{\bar{f}_W}{\widehat{\sigma}}(\boldsymbol{W}, s)}{\bar{f}_W(\boldsymbol{W}, s)} \cdot 100\% \tag{40}$$

The sample mean calculation is repeated until the error is lower than the user-specified tolerance error.

Even though the proposed approach may not be applicable to all minimal cutsets of the top event in the PSA of a nuclear power

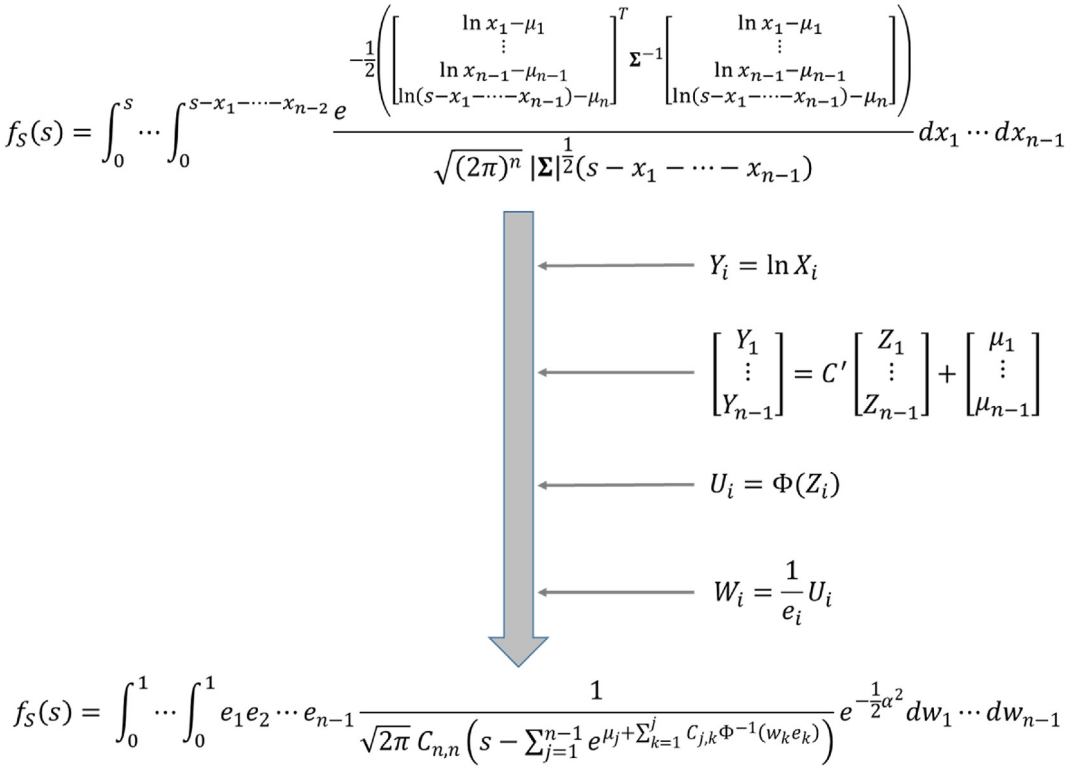

Fig. 1. Transformation of variables to independent random variables with a unit hypercube region of integration.

plant, it can be applicable to a pre-defined number of high-ranking minimal cutsets that may dominate the frequency of the top event. Also, the proposed approach can be applicable to the uncertainty analysis of a simple system in which a small number of minimal cutsets are obtained.

The proposed method can also be applicable to other problems with sum of lognormal random variables. For example, the proposed method is able to provide theoretical probability distribution of the sum of operator action times while each operator action time is assumed to follow a lognormal distribution such as the time uncertainty analysis approach in Level 2 human reliability analysis provided by Suh et al. [17].

3. Application to a benchmark problem

In this section, the accuracy of the proposed method is highlighted through a benchmark problem. We consider the example presented in Ref. [11], which is associated with the uncertainty analysis of a Level 1 PSA of a nuclear power plant in the United Kingdom. The top event of the benchmark is a core damage event caused by the short-term loss of offsite power. In a practical PSA, the fault tree for the top event involves hundreds or thousands of basic events and an almost infinite number of minimal cutsets. Therefore, when the fault tree is large, a cut-off value should be applied to make the fault tree manageable [17]. Also, there are many Level 1 PSA studies that the top event probability/frequency (e.g. core damage frequency) is dominated by a finite number of high ranking minimal cutsets.

As a simple example, only nine minimal cutsets with 12 basic events are considered for the benchmark. Tables 1 and 2 present the minimal cutsets and the parameters for the relevant basic events, respectively. The contribution of the minimal cutsets (Table 1) to the core damage frequency due to the short-term loss of offsite power exceeds 83%. All the uncertainties of basic events are

Table 1
Minimal cutsets for core damage frequency due to short-term loss of offsite power [11].

1	ABC
2	ABD
3	ABE
4	AFG
5	ABH
6	ABIJ
7	ABF
8	ABK
9	ALMNH

Table 2
Relevant basic events and parameters [11].

	Event	Parameters	
		E[*X*]	EF
A	Short-term loss of offsite power	6.00E-2	5
B	CCF of batteries	6.60E-6	5
C	Operator fails to start backup DG by local action	1.00E-2	5
D	Operator fails to start backup DG or to close breakers	2.13E-3	5
E	CCF to run backup DG	8.33E-4	5
F	CCF of batteries via 2 h discharge	5.20E-5	5
G	CCF to start EDG	6.10E-5	5
H	CCF to run EDG	4.20E-5	5
I	Fail to run backup DG-A	1.58E-3	5
J	Fail to run backup DG-B	1.58E-3	5
K	CCF to start backup DG	1.00E-4	5
L	Damage to O-ring seals	9.00E-2	5
M	Severe seal damage on all RCP	1.00E-1	5
N	CCF to run HPSI pump	1.20E-4	5

Abbreviations: CCF: common cause failure; DG: diesel generator; EDG: emergency diesel generator.

modeled with lognormal random variables.

As mentioned above, the mean values of the logarithm of random variables are calculated from the first and second moments of the random variables. The relationship between the second moment of random variables and the parameters listed in Table 2 is as follows:

$$\boldsymbol{\Sigma}_{ii} = \left(\frac{\ln(EF)}{1.64}\right)^2 \tag{41}$$

$$E\left[X^2\right] = E[X]^2(e^{\boldsymbol{\Sigma}_{ii}} - 1) \tag{42}$$

Hence, the means and covariance matrix of the logarithm of random variables for minimal cutsets can be calculated using Eq. (4) and Eq. (5), as follows:

$$\boldsymbol{\mu} = [-20.79 - 22.34 - 23.28 - 23.83 - 26.26 - 29.57 - 26.05 - 25.40 - 29.04]^T \tag{43}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 2.89 & 1.93 & 1.93 & 0.96 & 1.93 & 1.93 & 1.93 & 1.93 & 0.96 \\ 1.93 & 2.89 & 1.93 & 0.96 & 1.93 & 1.93 & 1.93 & 1.93 & 0.96 \\ 1.93 & 1.93 & 2.89 & 0.96 & 1.93 & 1.93 & 1.93 & 1.93 & 0.96 \\ 0.96 & 0.96 & 0.96 & 2.89 & 0.96 & 0.96 & 1.93 & 0.96 & 0.96 \\ 1.93 & 1.93 & 1.93 & 0.96 & 2.89 & 1.93 & 1.93 & 1.93 & 1.93 \\ 1.93 & 1.93 & 1.93 & 0.96 & 1.93 & 3.85 & 1.93 & 1.93 & 0.96 \\ 1.93 & 1.93 & 1.93 & 1.93 & 1.93 & 1.93 & 2.89 & 1.93 & 0.96 \\ 1.93 & 1.93 & 1.93 & 0.96 & 1.93 & 1.93 & 1.93 & 2.89 & 0.96 \\ 0.96 & 0.96 & 0.96 & 0.96 & 1.93 & 0.96 & 0.96 & 0.96 & 4.82 \end{bmatrix} \tag{44}$$

In general, the uncertainty in the top event probability is expressed with the probability density function given in the analytic form in Eq. (32). Numerical results for the probability density function can be calculated using Eq. (37). In other words, after calculating the means and covariance matrix for the minimal cutsets, the numerical value of the theoretical probability density function at a specific point in the domain can be obtained with Eq. (37). El-Shanawany et al. [11] compared the resultant cumulative distribution function and percentile calculated from Monte Carlo simulations, Fenton-Wilkinson's method, and Wilks's method.

To verify the accuracy of the proposed method, we compared the proposed method with Monte Carlo simulations, Fenton-Wilkinson's method, and Wilks's method considering the resultant probability density function, cumulative distribution function, and percentiles. For calculating the probability density function, the tolerance error of the numerical integration with a 95% confidence interval was set to 0.0001% of the sample mean, and the number of iterations was limited to 1,000,000. The cumulative distribution function was approximated from the probability density function using the trapezoidal rule. The percentiles were approximated using linear interpolation from the calculated cumulative distribution function.

In the Monte Carlo simulation, samples were generated for basic events, while the top event frequency for the samples was calculated using a rare-event approximation. The number of simulations was set to 1,000,000. The percentiles were estimated with Monte Carlo simulations via linear interpolation.

Fenton-Wilkinson's method approximates the sum of lognormal random variables to a single lognormal random variable with the same first and second moments [18]. The first and second moments of the sum of lognormal random variables are as follows:

$$E[S] = \sum_{i=1}^{n} E[X_i] \tag{45}$$

$$E\left[S^2\right] = \sum_{i=1}^{n} E\left[X_i^2\right] + 2\sum_{i=1}^{n-1}\sum_{k=i+1}^{n} E[X_i]E[X_k]e^{\mathrm{Cov}(\ln(X_i),\ln(X_k))} \tag{46}$$

The first moment of the top event was 5.40E-9, and the second moment of the top event was 3.76E-16. Hence, the random variable approximated from Fenton-Wilkinson's method is as follows:

$$X_{FW} \sim LN(-20.31, 2.55) \tag{47}$$

Wilks's method is applicable to estimate the percentile based on order statistics with a specified confidence level. The cumulative distribution function at the k-th smallest value ($X_{(k)}$) of a sample set with sample size of n is a random variable which follows a beta distribution.

$$F_X\left(X_{(k)}\right) \sim \mathrm{Beta}(k, n-k+1) \tag{48}$$

Then, the probability that $X_{(k)}$ is larger than the p-th percentile can be calculated as follows:

$$\Pr\left(X_{(k)} \le X_p\right) = \Pr\left(F_X\left(X_{(k)}\right) \le p\right) = I_p(k, n-k+1) \le \alpha \tag{49}$$

where X_p is the p-th percentile and $I_{1-p}(k, n-k+1)$ is an incomplete beta function. Wilks' method defines the order k and the sample size n to satisfy the probability in Eq. (49) with confidence level α. The confidence level was set to 95%, and 1,000 sample sets with a sample size of 1,000 were used to estimate the percentile according to Wilks's method.

Fig. 2 shows a comparison of the cumulative distribution functions of core damage frequency for the benchmark using Monte Carlo simulations, Fenton-Wilkinson's methods, Wilks's method, and the proposed method in the log and linear scales. The percentiles calculated with different methods and their relative errors compared to the theoretical values calculated with the proposed method are presented in Table 3 and Table 4, respectively. The relative percent error, $\delta_{\%}$, is defined as follows:

Table 3
Percentiles in cumulative distribution function of core damage frequency with Monte Carlo simulations, Fenton-Wilkinson's method, Wilks's method, and the proposed method.

Percentile	Monte Carlo simulation	Fenton-Wilkinson's method	Wilks's method	Proposed method
5th	1.4621E-10	1.0869E-10	1.7032E-10	1.4568E-10
25th	5.9542E-10	5.1261E-10	6.5825E-10	5.9720E-10
50th	1.6262E-09	1.5066E-09	1.7986E-09	1.6275E-09
75th	4.5166E-09	4.4277E-09	5.0246E-09	4.5220E-09
95th	2.0447E-08	2.0882E-08	2.4433E-08	2.0259E-08

Table 4
Relative error of percentiles for Monte Carlo simulations, Fenton-Wilkinson's method, and Wilks's method with respect to the proposed method.

Percentile	Monte Carlo simulation	Fenton-Wilkinson's method	Wilks's method
5th	0.3606%	25.3900%	16.9113%
25th	0.2974%	14.1642%	10.2241%
50th	0.0841%	7.4335%	10.5129%
75th	0.1179%	2.0837%	11.1156%
95th	0.9267%	3.0737%	20.6021%

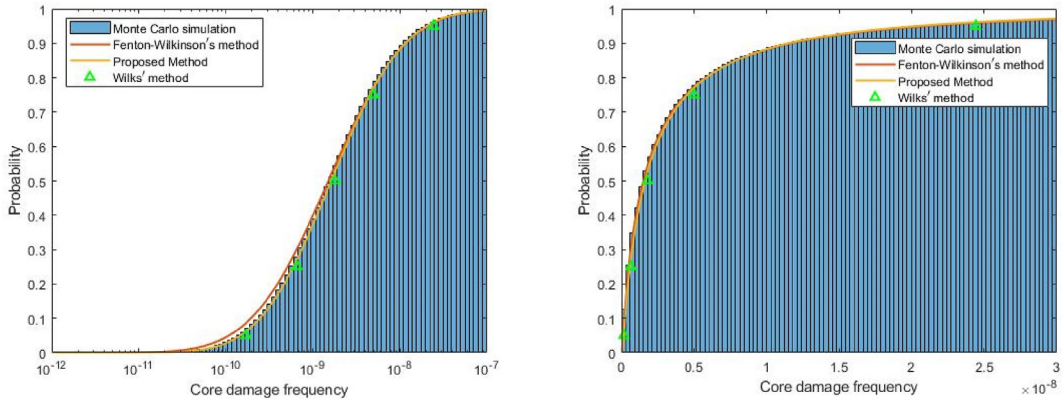

Fig. 2. Cumulative distribution functions for benchmark using Monte Carlo simulations, Fenton-Wilkinson's method, Wilks's method, and the proposed method in the (a) log scale and (b) linear scale.

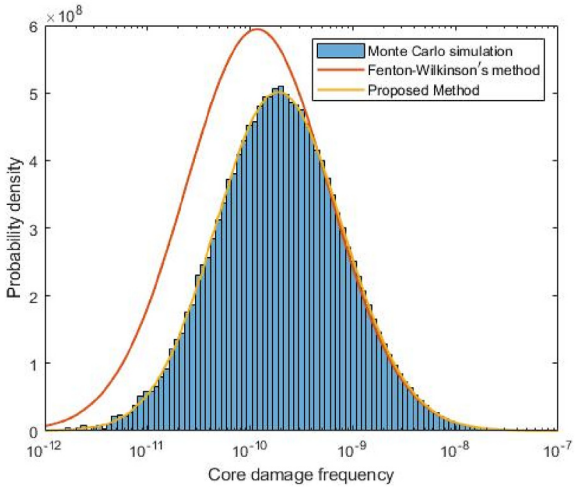

Fig. 3. Probability density functions for the example with Monte Carlo simulations, Fenton-Wilkinson's method, and the proposed method.

$$\delta_{\%} = \left| 1 - \frac{P_{\text{approximation}}}{P_{\text{proposed}}} \right| \cdot 100\% \tag{50}$$

where P_{proposed} and $P_{\text{approximation}}$ are the percentile values from the proposed method and other approximation methods, respectively. The results listed in Tables 3 and 4 indicate that the percentiles and relative errors are consistent with those reported by El-Shanawany et al. [11]. The proposed method provides theoretical results that can serve as a baseline to compare other methods. Particularly, the relative errors when using the Monte Carlo simulation are less than 1%; thus, the Monte Carlo estimation approaches the exact results. Fenton-Wilkinson's method shows a relatively high accuracy in the right tail of the distribution, whereas the method shows poor accuracy in the left tail of the distribution. This is because Fenton-Wilkinson's method relies on matching the first and second moments; thus, it is appropriate for estimating the middle and high regions of the distribution [5]. Wilks's method provides results that are more conservative because a 95% confidence level was adopted. However, significant differences between the methods are not recognized in the cumulative distribution functions or percentiles.

Fig. 3 shows the benchmark results from the probability density functions of the core damage frequency obtained using Monte Carlo simulations, Fenton-Wilkinson's methods, and the proposed method. The proposed method provides a theoretical probability density function that has not been revealed in most previous PSA studies. Furthermore, the Monte Carlo simulation is in good agreement with the proposed method.

Fenton-Wilkinson's method provides less accurate results for the proposed benchmark. This is because the method is applied when the standard deviations of the logarithm of the summand random variables are less than 0.9210 (4 dB) for an independent case [6] and 2.7631 (12 dB) for an identical and correlated case [18]. Furthermore, the accuracy of Fenton-Wilkinson's method is deteriorated when the mean values of the summands are similar, the standard deviations of the summands have large variance, the correlation between the summands is low, and the number of summands is large [19]. Therefore, Fenton-Wilkinson's method could provide inaccurate results if the top event has more minimal cutsets and the standard deviations of minimal cutsets have large variance, especially when the lower cut-off value is applied. In general, the resultant probability density function derived from Fenton-Wilkinson's method is expected to have a larger variance than the exact one [20], as seen in Fig. 3.

Fig. 4 shows the relative error of the Monte Carlo simulation and Fenton-Wilkinson's method, as compared to the proposed method, for the probability density function (Fig. 3) and cumulative

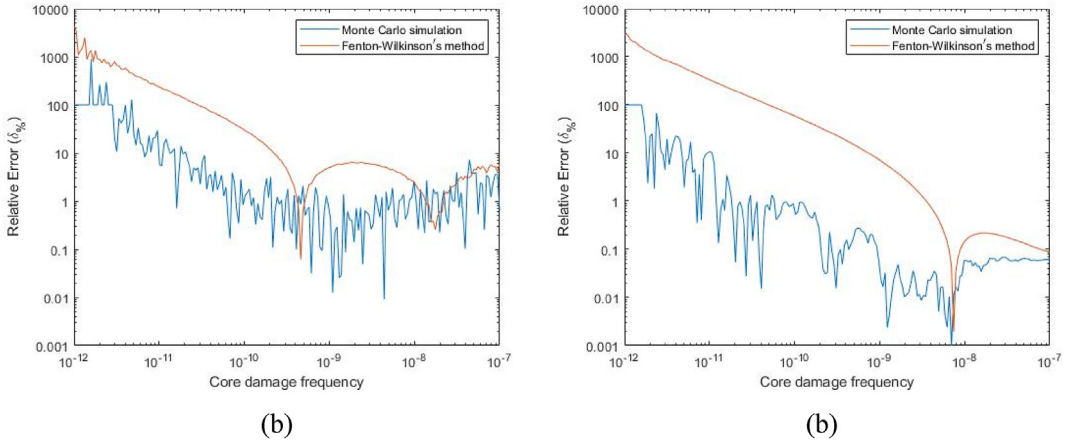

Fig. 4. Relative errors of the Monte Carlo simulation and Fenton-Wilkinson's method compared to the proposed method for the (a) probability density function and (b) cumulative distribution function.

distribution function (Fig. 2). Because the proposed method provides theoretically more accurate results, Fig. 4 quantitatively shows the amount of error Monte Carlo simulation or Fenton-Wilkinson's method may have produced in estimating the probability density function and cumulative distribution function when applied to the benchmark problem. The relative error of the Monte Carlo simulations in the left tail of the distribution reaches up to 100% because there are bins with no samples. It should be noted that the relative error of the Monte Carlo simulation fluctuates considerably over the entire abscissa range, whereas the relative error of Fenton-Wilkinson's method remains relatively smooth. The minima of the relative error of Fenton-Wilkinson's method indicate the points where the probability density functions and the cumulative distribution functions from Fenton-Wilkinson's method and the proposed method intersect. The relative errors of both methods increase in the left tail of the distributions.

The calculation time of the proposed method depends on many factors such as the number of high ranking minimal cutsets considered in the calculation, the number of data points in the domain, and the number of samples for each data point. For the benchmark problem, it took 1,803 s for the proposed method to calculate probability density of 200 data points with 1,000,000 sample sets for each data point, that is 200,000,000 sample sets in total. On the other hand, it took 3 s to perform Monte Carlo simulations with 1,000,000 sample sets in total. The long computation time of the proposed method is also closely related to the increased number of sample sets which also contributes to the enhanced accuracy of the proposed method. When the proposed method uses 100,000 sample sets for each data point, it took 214 s to calculate probability density at 200 data points.

In the current initial idea development stage, the authors believe that the proposed approach is able to provide the theoretical probability density function with up to 25 high ranking minimal cutsets in an acceptable calculation time. As the proposed approach is further developed, the number of high ranking minimal cutsets considered in the calculation is expected to increase significantly.

4. Conclusions

Uncertainty quantification is a fundamental element in PSA. When the uncertainties of basic events are modeled with lognormal random variables, the top event frequency or probability is well approximated as the sum of the correlated lognormal random variables. We proposed a new method for analyzing the uncertainty in the top event frequency or probability in PSA when basic events are given with lognormal random variables by calculating the probability density function of the sum of correlated lognormal random variables. The proposed method is characterized by the complex transformation of the analytic expression for the sum of correlated lognormal random variables to an equivalent multiple integral form with a unit hypercube region of integration. A benchmark problem considering uncertainty analysis showed that the proposed method provided an accurate probability density function and cumulative distribution function, as compared to the other approximation methods. As a theoretical approach for the uncertainty analysis in PSA, the proposed method can be used to validate the uncertainty analysis results with other approximate methods when the basic events are given with lognormal random variables.

The limitations of the proposed method mainly originate from relatively long computation time owing to increased number of sample sets proportional to the number of data points to calculate the probability density. However, the increased number of sample sets also contributes to enhanced accuracy in probability density calculation because more sample sets are used to calculate the probability density of a single data point. Future researches need to include new algorithms for reducing the number of sample sets per data point while the enhanced accuracy is maintained.

The proposed method provided reasonable and nearly continuous results over the entire abscissa, whereas the other approximation methods provided inaccurate or discrete results. The analytical expression with the traditional numerical integration may suffer from the curse of dimensionality as the number of random variables increases. Thus, the proposed method is applicable to a broader range of dimensions than the traditional numerical integration.

In summary, the proposed method is an effective approach providing the probability density function of the top event frequency or probability with high accuracy. As a theoretical approach, the proposed method could serve as a mathematical framework for developing advanced approximate methods for uncertainty analyses in PSA. Moreover, the proposed method can be used for a wide range of topics in PSA when the results are associated with the sum of lognormal random variables.

Declaration of competing interest

The authors declare that they have no known competing

financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the Nuclear Safety Research Program of the Korea Foundation of Nuclear Safety, granted financial resource from the Multi-Unit Risk Research Group (MURRG), with funding by the Korean government's Nuclear Safety and Security Commission [grant number 1705001] and the Nuclear Research & Development Program of the National Research Foundation of Korea, funded by the Korean government's Ministry of Science and ICT [grant number NRF-2017M2B2B1071973].

References

[1] G. Apostolakis, The concept of probability in safety assessments of technological systems, Science 250 (1990) 1359–1364.
[2] U.S. Nuclear Regulatory Commission, Reactor Safety Study: an Assessment of Accident Risks in US Commercial Nuclear Power Plants, WASH-1400, 1975 (NUREG 75-014), 1975.
[3] American Nuclear Society and Institute of Electrical and Electronics Engineers, PRA Procedures Guide: A Guide to the Performance of Probabilistic Risk Analyses for Nuclear Power Plants, NUREG/CR-2300, 1983.
[4] H.E. Lambert, Fault Trees for Decision Making in Systems Analysis, Lawrence Livermore National Lab. (LLNL), Livermore, CA (United States), 1975.
[5] L. Fenton, The sum of log-normal probability distributions in scatter transmission systems, IRE. Trans. Commun. Syst. 8 (1960) 57–67.
[6] S.C. Schwartz, Y.S. Yeh, On the distribution function and moments of power sums with log-normal components, Bell.Syst. Tech. J. 61 (1982) 1441–1462.
[7] N.C. Beaulieu, Q. Xie, An optimal lognormal approximation to lognormal sum distributions, IEEE Trans. Veh. Technol. 53 (2004) 479–489.
[8] S. Asmussen, P.O. Goffard, P.J. Laub, Orthornomal Polynomial Expansions and Lognormal Sum Densities, arXiv:1601.01763 [Math], 2016.
[9] A.S.H. Mahmoud, New quadrature-based approximations for the characteristic function and the distribution function of sums of lognormal random variables, IEEE Trans. Veh. Technol. 59 (2010) 3364–3372.
[10] M. Cox, P. Harris, B.R.L. Sibert, Evaluation of measurement uncertainty based on the propagation of distributions using Monte Carlo simulation, Meas. Tech. 46 (2003) 824–833.
[11] A.B. El-Shanawany, K.H. Ardron, S.P. Walker, Lognormal approximations of fault tree uncertainty distributions, Risk Anal. 38 (2018) 1576–1584.
[12] A. Papoulis, Probability, Random Variables and Stochastic Processes, second ed., McGraw-Hill, New York, 1985.
[13] A. Genz, Numerical computation of multivariate normal probabilities, J. Comput. Graph Stat. 1 (1992) 141–150.
[14] L.J. Halliwell, The lognormal random multivariate, in: Casualty Actuarial Society E-Forum, 2015. Spring, U.S, June 1-2.
[15] O.B. Sheynin, C. F. Gauss and the theory of errors, Archive for History of Exact Sciecnes 20 (1979) 21–72.
[16] G.H. Golub, C.F. Van Loan, Matrix Computations, Johns Hopkins University Press, Baltimore, Maryland, 1996.
[17] Young A. Suh, Jaewhan Kim, Soo Yong Park, Time uncertainty analysis method for level 2 human reliability analysis of severe accident management strategies, Nucl. Eng. Technol. 53 (2021) 484–497.
[18] M. Modarres, H. Dezfuli, A truncation methodology for evaluating large fault trees, IEEE Trans. Reliab R-33 (1984) 325–328.
[19] A.A. Abu-Dayya, N.C. Beauliue, Outage probabilities in the presence of correlated lognormal interferes, IEEE Trans. Veh. Technol. 43 (1994) 164–173.
[20] P. Cardieri, T.S. Rappaport, Statistics of the sum of lognormal variables in wireless communications, in: VCT2000-Spring. 2000 IEEE 51st Vehicular Technology Conference Proceedings, vol. 3, 2000, pp. 1823–1827.