ESTIMATION OF INITIATING EVENT DISTRIBUTION AT NUCLEAR POWER PLANTS
BY BAYESIAN PROCEDURE

Guangming Chen
Department of Industrial Engineering
Morgan State University
Baltimore, MD 21239

T.-L. Chu
Department of Advanced Technology
Brookhaven National Laboratory
Upton, NY 11973

Key Words: Bayesian modelling, failure rate, nuclear power plant, risk assessment.

ABSTRACT

Initiating events at nuclear power plants such as human errors or component's failures may lead to a nuclear accident. The study of the frequency of these events or the distribution of the failure rate is necessary in probabilistic risk assessment for nuclear power plants. This paper presents Bayesian modelling methods for the analysis of the distribution of the failure rate. The method can also be utilized in other related fields especially where the data is sparse. An application of the Bayesian modelling in the analysis of distribution of the time to recover LOSP is discussed in the paper.

1. Introduction

Loss of off-site power (LOSP) accidents are relatively infrequent at nuclear power plants and data are sparse. The analysis of the initiating events is important because they may develop severe accidents - core meltdown. LOSP accidents involve the interruption of the preferred-off-site power supply to the essential and nonessential switch-gear buses. As a result, the emergency backup power supply must be used. However, if the off-site power is not recovered in a certain time period, the backup power supply may subsequently fail. That leads to a blackout at the plant and a core meltdown may occur. Probabilistic Risk Assessment (PRA) is used to conduct the reliability and safety analysis for nuclear power plants under the policies of Nuclear Regulatory Commission (NRC).

The goal of this paper is to investigate the application of Bayesian analysis in risk assessment and safety analysis for nuclear power plants, to develop and improve the method for estimating the distribution of the failure rate or the initiating events that may ultimately lead to LOSP accidents. Generally speaking, the methodology may also be applied to other fields after proper modifications such as estimating the distribution of the time to recover LOSP. The paper is organized as follows: Prior to the discussion of the approach, the generic assumptions used throughout the paper are given in next section. The justification for these assumptions will be presented or the assumptions will be appropriately addressed. Based on these underlying assumptions, the Bayesian modeling and the procedure for conducting Bayesian analysis will be developed. To demonstrate the utilization of the method, an application in
estimating the distribution of the time to recover LOSP is discussed using the data collected from the nuclear power plants in the United States.

2. Underlying Assumptions

Under stable conditions and given a failure rate \( \lambda \), the number of LOSP incidents during a time period \( t \), denoted by \( X \), can be reasonably assumed to have a Poisson distribution:

\[
P(\text{\# incidents in time } t \mid \text{given } \lambda) = e^{-\lambda t} \frac{(\lambda t)^x}{x!}
\]  

(1)

One of the objectives in performing PRA is to find the distribution of \( \lambda \) based on the available knowledge and data. The total number of LOSP incidents is the sum of the plant-centered LOSP incidents \( Y \) and the grid/severe weather LOSP incidents \( Z \) or \( X = Y + Z \). \( Y \) and \( Z \) are Poisson random variables with \( \lambda_y \) and \( \lambda_z \) respectively. Apparently, the overall failure rate or the intensity parameter for \( X \) is \( \lambda = \lambda_y + \lambda_z \). Hora and Iman use a convolution integral to develop the distribution of \( \lambda \). With an assumption of gamma distributions for \( X \) and \( X \), this convolution integral leads to a confluent hypergeometric function which is very complicated and must be evaluated numerically. Actually, \( X \), as the sum of \( Y \) and \( Z \), is also a Poisson random variable. This can be shown by finding the moment generating function for \( X \). The moment generating functions for \( Y \) and \( Z \) are \( E(e^{tY}) = \exp(K_y(e^t - 1)) \) and \( E(e^{tZ}) = \exp(K_z(e^t - 1)) \) respectively where \( K_y \) and \( K_z \) are the expected values of \( Y \) and \( Z \) respectively. Thus,

\[
E(e^{tX}) = E(e^{t(Y+Z)}) = E(e^{tY}e^{tZ})
= \exp(K_y(e^t - 1))\exp(K_z(e^t - 1))
= \exp(K(e^t - 1))
\]

(2)

where \( K = K_y + K_z \) is the expected value of \( X \). In Eq. (2), we utilize the fact that \( Y, Z \) are independent random variables. The moment generating function for \( X \) in Eq.(2) is identical to a one for a Poisson random variable so that we can conclude \( X \) is distributed by Poisson. As a result, we may not necessarily address the frequencies or the distributions of plant-centered LOSP and grid/severe weather LOSP incidents separately. Instead, the distribution of the failure rate for the total number of incidents can be discussed as a whole.

In developing the Bayesian procedure, Kaplan\(^{[8]}\) assumes that the failure rate has a lognormal distribution, while Hora and Iman\(^{[9]}\) uses a diffuse prior that ultimately leads to a gamma density for the failure rate. Chu\(^{[4]}\) has shown that Kaplan’s method and Hora & Iman’s method are similar except that they used different distributions and Kaplan used discrete space of the distribution of the parameters. Under the assumption that the failure rate has a general distribution, the Bayesian procedure will be discussed in next section.

3. Bayesian Modeling And Procedure

The data from an individual nuclear power plant is sparse. A much better estimation than the prior knowledge can be obtained for the frequency or the distribution of the failure rate by using Bayesian analysis. The Bayesian process can be considered as a learning process\(^{[9]}\). For each time using Bayesian analysis, we can improve the knowledge about the distribution of the failure rate. To construct the model and develop the procedure, we assume that there are \( m \) nuclear power plants. Each plant has a different operating time that is the time since starting operation, and a certain number of LOSP incidents during that operating period (maybe zero incidents). For instance, Plant 1 has been in service for time \( t_1 \) and it has \( X_1 \) LOSP incidents, so that data pair \( E_{i1} = (X_1, t_1) \) represent the event \( i \) that happened to Plant 1. We let \( \lambda_i \) denote the failure rate for Plant 1 or \( E = (E_{11}, E_{21}, ..., E_{m1}) \) (Bold letters are vector quantities unless otherwise specified). We also define \( E_i = (x_i, t_i) \) as the event that \( x_i \) incidents occur in time \( t_i \) to the specific plant of our interest. For the convenience of deriving the probability, \( E_i \) is excluded from \( E \). In other words, \( E \) and \( E_i \) are independent. According to Bayesian theorem, for any event \( E \) and respective \( \lambda \), we have

\[
P(\lambda \mid E) = \frac{P(E \mid \lambda)P(\lambda)}{\int_{\lambda} P(E \mid \lambda)P(\lambda)d\lambda}
\]

(3)

where \( P(\lambda | E) \) is the posterior probabilities of \( \lambda \), \( P(\lambda) \) is the prior probabilities and \( P(E | \lambda) \) is the probabilities that \( E \) happen given the inherent failure rate \( \lambda \). And the integration is over all possible values of \( \lambda \). Quantity \( P(E | \lambda) \int_{\lambda} P(\lambda)d\lambda \) here can improve our prior knowledge about \( \lambda \) or \( P(\lambda) \). If the prior is given as a function, the posterior will also be a function.
Thus, the problem boils down to determine the prior distribution. Kaplan [3] developed a two-stage procedure to perform Bayesian analysis. The first stage is to find the prior using available information under certain assumptions. The second stage is to find the posterior distribution of the failure rate.

Following [2], we consider two situations: (1) All plants share a common failure rate λ; (2) Each plant has an individual failure rate. In Case (1), we need to find the distribution of this common failure rate. In Case (2), we further divide the situation into two cases: all individual failure rates are from one population or each individual failure rate has a different population.

(i) All Plants Share A Common Failure Rate λ.

$E_1$ is the event that there incidents occur in time $t_1$ to the specific plant of our interest. Using Poisson model (1) and Bayesian theorem, we have the posterior distribution of the failure rate:

$$P(\lambda | E_s) = \frac{P(E_s | \lambda) P(\lambda)}{\int P(E_s | \lambda) P(\lambda) d\lambda}$$

$$= \frac{e^{-\lambda X_s} \lambda^{x_s} P(\lambda)}{\int e^{-\lambda x_s} \lambda^{x_s} P(\lambda) d\lambda}$$

If a good estimated prior is available, the posterior distribution of λ can be determined by the above equation. However, if only a subjective prior based on expert knowledge judgement is available, we can utilize the data set $E_s$ to improve this prior. Thus, an intermediate posterior of λ is introduced by using Bayesian theorem and the Poisson model (1):

$$P(\lambda_1 | E_s) = \int P(\lambda | E_s) P(\lambda) d\lambda$$

$$= \frac{P(E_s | \lambda_s) P(\lambda_s)}{\int e^{-\lambda_s x_s} \lambda_s^{x_s} P(\lambda_s) d\lambda_s}$$

As long as $P(\lambda)$ is given, the posterior distribution of λ can be determined by Equations (6) and (7).

(ii) Each Plant Has An Individual Failure Rate From The Same Population.

In this case, the posterior distribution of $\lambda_i$ after using $E_s$ is given by:

$$P(\lambda_i | E_s) = \frac{P(E_s | \lambda_i) P(\lambda_i)}{\int e^{-\lambda_i x_i} \lambda_i^{x_i} P(\lambda_i) d\lambda_i}$$

Since all failure rates are from the same population, $P(\lambda_i | E_s)$
can be considered as the posterior for any failure rate or $P(\lambda | E)$, and $P(\lambda)$ can be considered as the prior for any failure rate or $P(\lambda)$, for which an intermediate posterior $P(\lambda | E)$ can substitute by employing the data set $E$ in the same way as in Case (i). As a result, the procedure developed in Case (i) can be applied to this case.

(iii) Each Plant Has An Individual Failure Rate From Different Populations.

The failure rate at the specific plant in this case is completely independent of those at other plants. Thus, the data set $E$ from other plants provide no information for improving the knowledge about the failure rate at the specific plant. The analysis should be only based on the data from this specific plant:

$$P(\lambda_s | E_s) = \frac{\int P(\lambda_s | E_s, \lambda_s) P(\lambda_s) d\lambda_s}{\int \int P(\lambda_s | E_s, \lambda_s) P(\lambda_s) d\lambda_s}$$

$$= \frac{e^{-\lambda_s^2 \lambda_s^2} P(\lambda_s)}{\int e^{-\lambda_s^2 \lambda_s^2} P(\lambda_s) d\lambda_s}$$

By using Eq. (9), we can find the distribution of failure rate for each specific plant.

4. Application in Estimating Distribution of Time to Recover LOSP

In last section, the Bayesian modelling for estimating the distribution of the failure rate is studied. However, the method can be applied to other related areas. For instance, Chu, etc. utilized the Bayesian method to find the distribution of the time to recover LOSP during the SURRY studies. They used the data of the initiating events at nuclear power plants collected up to 1988 in the United States that led to LOSP. The study presents the distribution curves for the median, 5th percentile and 95th percentile of the time to recover LOSP as given in Figure 1. For instance, the probability that the median of the time to recover LOSP exceeds four hours is about 0.13 by reading the figure. These curves can be utilized in PRA of the nuclear power plants.

5. Conclusion

Bayesian modelling is a useful tool in PRA for nuclear power plants when data is sparse. The available limited data can improve our subjective or judgmental knowledge about the distribution of the initiating events. This paper is aimed at improving and extending Kaplan's two-step procedure for Bayesian modelling and strengthening the explanation and understanding of the model. Basically, any subjective distribution can be used in this model instead of limiting the distribution to lognormal or gamma. The use of the method in estimating the distribution of the time to recover LOSP is also discussed to give a scenario of possible applications.

Acknowledgement

This research was supported by NRC/ORISE HBCU Faculty Research Participation Program.

References


Figure 1. Non-recovery curves (curves of probabilities that time to recover LOSP exceeds the given value. POS 2 to 14).