Estimating prior distribution for common cause failure parameters using empirical Bayes

Gyun Seob Song [a], Man Cheol Kim [a*]
[a]*School of Energy Systems Engineering, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul, 06974, Korea*
[*]*Corresponding author: charleskim@cau.ac.kr*

***Keywords :** Common cause failure, Uncertainty analysis, Empirical Bayes, Alpha factor model

1. Introduction

Redundancy is one of principles in nuclear power plant design. However, the redundancy can be threatened by common cause failure (CCF), which all or partial group of components becomes unavailable due to common cause. Therefore, it is recognized that the common cause events are significant risk contributors for the nuclear power plant. To quantify the frequency of CCF events, NUREG/CR-5485 [1] propose several models, alpha factor model, multiple Greek letter model, and so on. Among the analysis models, alpha factor model is widely used because of its statistical efficiency. As a part of probabilistic safety assessment (PSA), the model parameters for CCF events have uncertainties, and it should be quantified same as parameters of other basic events in the PSA model. INL/ETX-21-62940 [2] provide estimated CCF model parameters and their uncertainty distributions based on operational data and prior distributions provide by INL/EXT-21-43723 [3]. Although the prior distributions are derived based on the homogeneous assumption in the industry, it is possible that the performances of components have plant-to-plant variability. Empirical Bayes method is widely used model to analyze plant-to-plant variability in nuclear industry for component reliability data [4].

In this paper, the Empirical Bayes method is applied to statistical model for alpha factor model. And then, the result is compared to that of conventional constrained non-informative distribution (CNID) and Jeffrey's non-informative distribution (JNID) using an example data set.

2. Empirical Bayes for alpha factor model

Alpha factor model is one of CCF models that the probability of CCF event is represented as proportional to the total component failure probability. For a staggered testing scheme, the probability of i components failures in the common cause component group (CCCG) of size k is

$$p_i^{(k)} = \frac{1}{\binom{k-1}{i-1}} \alpha_i p_t \tag{1}$$

where p_t is total component failure probability.

Empirical Bayes model is a combination of hierarchical Bayes model and maximum likelihood model. In general, two-step approach is applied to the Empirical Bayes model for representing plant-to-plant variability and plant specific uncertainty. Therefore, the uncertainties are considered until parameters of the statistical model and the hyperparameters are determined by maximum likelihood estimator for the observation. For the alpha factor model, the observation process follows multinomial distribution and the uncertainty distribution for the multinomial distribution is modeled as Dirichlet distribution because the Dirichlet distribution is conjugate prior for the multinomial distribution. Then, the likelihood function for single plant observation is

$$L(\boldsymbol{A}; \boldsymbol{x}) = \int \left(\frac{n!}{x_1! \cdots x_k!} \prod_{i=1}^{k} \alpha_i^{x_i} \right) \frac{1}{\mathbf{B}(\boldsymbol{A})} \prod_{i=1}^{k} \alpha_i^{A_i - 1} \, d\boldsymbol{\alpha} \tag{2}$$

where n is total number of observations, k is the CCCG size, x_i's are the number of observations for i components failures in the CCCG, $\mathbf{B}$ is a beta function, $\alpha_i's$ are alpha factors, and $A_i's$ are the hyperparameters.

If there are N number of plants, the likelihood function can be represented as

$$L(\boldsymbol{A}; \boldsymbol{x_1}, \cdots \boldsymbol{x_N}) = \prod_{i=1}^{N} \frac{n_i!}{x_{1i}! \cdots x_{ki}!} \frac{\mathbf{B}(\boldsymbol{A} + \boldsymbol{x_i})}{\mathbf{B}(\boldsymbol{A})} \tag{3}$$

where x_{ij} is the number of observations for i components failures in *j*-th plant.

Then, the derivatives of the log-likelihood, which is used to derive the maximum likelihood estimator for the hyperparameters, are as follows:

$$\frac{\partial L}{\partial A_i} = N\psi(A_0) - N\psi(A_i) + \sum_{j=1}^{N} \psi(A_i + x_{ij}) - \sum_{j=1}^{N} \psi\left(A_0 + \sum_{l} x_{lj}\right) \tag{4}$$

where ψ is a digamma function, and A_0 is sum of all the hyperparameters.

The maximum likelihood estimators can be derived to solve the equation that all the derivatives in Eq. (4) equal to zero. The equation has a single solution unless there are no difference between the plants, which is homogeneous condition. To estimate plant specific CCF model parameters, the prior distribution which is derived by Empirical Bayes is updated by plant specific data. As the Dirichlet distribution is conjugate prior for the multinomial distribution, the updated posterior distribution is also Dirichlet distribution, and the parameters are updated by the observed data.

$$\boldsymbol{\alpha_i} | \boldsymbol{x_i} \sim Dirichlet(\widehat{\boldsymbol{A}} + \boldsymbol{x_i}) \tag{4}$$

where $\hat{\boldsymbol{A}}$ is the maximum likelihood estimator.

3. Application

In the uncertainty analysis for CCF model parameters, homogeneous assumption is generally used. JNID used when there is little information about the model parameters, and CNID is used to constrained expected value for the prior distribution [4]. When there is variability in the plants, the uncertainties for the parameters and the following CCF probabilities can be underestimated. To compare the effects of population variability, an example data set which is based on the total CCF data collected from 1997 to 2015 is used [3]. The number of plants is assumed as 5 and the data is randomly distributed to the plants. As the data is collected from the plants which have different number of components, the data is mapping to the plant which have 4 redundant components. Table I shows the adjusted CCF data and Table II shows the calculated parameters with respect to the model parameters based on the data.

Table I: Adjusted CCF events observation data

Plant	n_1	n_2	n_3	n_4
1	2201.1	3.1746	4.6912	1.6519
2	645.54	20.076	4.7146	5.6078
3	1836.6	23.810	4.9429	1.8677
4	452.64	19.044	4.6282	0.9440
5	689.77	14.833	6.8804	4.8016

Table II: Model parameters for the prior distributions

	A_1	A_2	A_3	A_4
EB	157.62	2.7172	1.2365	0.8018
CNID	90.676	1.2597	0.4024	0.2315
JNID	0.5	0.5	0.5	0.5

As the model parameters have high dimensions, marginal distribution is widely used to represent uncertainties of the parameters. The marginal distribution for the Dirichlet distribution is beta distribution.

$$\alpha_i \sim Beta(A_i, A_0 - A_i) \quad (5)$$

Fig. 1 to Fig. 4 show the marginalized posterior Dirichlet distribution for the Plant 1 specific parameters with respect to the prior distributions. It is shown that the model parameters expect for α_1 are underestimated when the homogeneous assumption is used. It means that not only the uncertainties of the parameters but also the expected values for the CCF probabilities are underestimated.

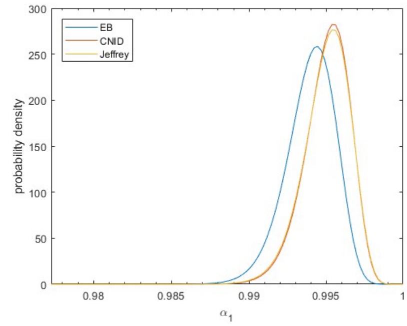

Fig. 1. Uncertainty distribution for the CCF parameter α_1

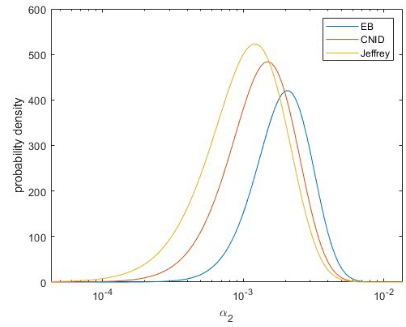

Fig. 2. Uncertainty distribution for the CCF parameter α_2

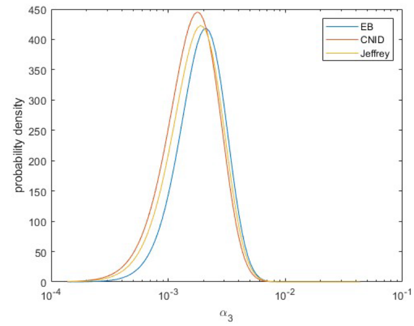

Fig. 3. Uncertainty distribution for the CCF parameter α_3

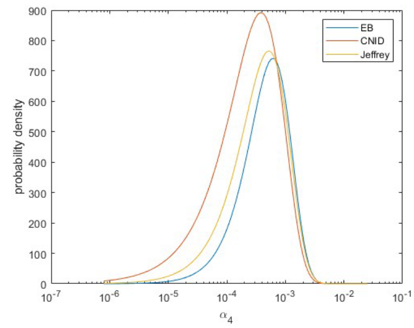

Fig. 4. Uncertainty distribution for the CCF parameter α_4

3. Conclusions

In the probabilistic safety assessment, there are two types of uncertainties, aleatory uncertainty and epistemic uncertainty. Although CCF event is related to aleatory uncertainty, the epistemic uncertainty for the CCF event should be considered. In the conventional analysis, the homogeneous assumption is used because of mathematical convenience. However, there are variability in the population, especially environmental cause CCF events. In this paper, the prior distribution which consider the population variability is derived based on empirical bayes method and the result is compared to other prior distributions with an example data set. It is shown that the uncertainties and the estimated CCF probability can be underestimated when the homogeneous assumption is used. Therefore, the homogeneous assumption should be carefully used, and the population variability should be considered.

Acknowledgement

This work was supported by the Nuclear Safety Research Program of the Korea Foundation of Nuclear Safety, funded by the Korean Government's Nuclear Safety and Security Commission [grant number 2203027-0122-CG100, 2104028-0222-SB120]

REFERENCES

[1] A. Mosleh, D. M. Rasmuson, F. M. Marshall, Guidelines on modeling common cause failures in probabilistic risk assessment, NUREG/CR-5485, United States Nuclear Regulatory Commission, 1998.
[2] Z. Ma, K. J. Kvarfordt, CCF parameter estimations 2020 update, INL/EXT-21-62940, Idaho National Laboratory, 2022.
[3] Z. Ma, C. L. Atwood, J. A. Schroeder, Developing generic prior distributions for common cause failure alpha factors and causal alpha factors, INL/EXT-21-43723, Idaho National Laboratory, 2021.
[4] C. L. Atwood, Constrained Noninformative priors, INEL-94/0074, Idaho National Laboratory, 1994.