3단계 베이지안 처리절차 및 신뢰도 자료 처리 코드 개발†

임태진*

Development of the 'Three-stage' Bayesian Procedure And A Reliability Data Processing Code†

Tae-Jin Lim*

ABSTRACT

A reliability data processing program MPRDP (Multi-Purpose Reliability Data Processor) has been developed in FORTRAN language since Jan. 1992 at KAERI (Korean Atomic Energy Research Institute). The purpose of the research is to construct a reliability database (plant-specific as well as generic) by processing various kinds of reliability data in most objective and systematic fashion. To account for generic estimates in various compendia as well as generic plants' operating experience, we developed a 'three-stage' Bayesian procedure[1] by logically combining the 'two-stage' procedure[2] and the idea for processing generic estimates[3].

The first stage manipulates generic plant data to determine a set of estimates for generic parameters, e. g. the mean and the error factor, which accordingly defines a generic failure rate distribution. Then the second stage combines these estimates with the other ones proposed by various generic compendia (we call these generic book type data). This stage adopts another Bayesian procedure to determine the final generic failure rate distribution which is to be used as a priori distribution in the third stage. Then the third stage updates the generic distribution by plant-specific data resulting in a posterior failure rate distribution. Both running failure and demand failure data can be handled in this code.

In accordance with the growing needs for a consistent and well-structured reliability database, we constructed a generic reliability database by the MPRDP code[4]. About 30

* Dept. of Industrial Eng., Soong Sil Univ. 1-1 Sangdo-5 Dong, Dong-Jak Gu, Seoul Korea 156-743

generic data sources were reviewed and available data were collected and screened from them. We processed reliability data for about 100 safety related components frequently modeled in PSA. The underlying distribution for the failure rate was assumed to be lognormal or gamma, according to the PSA convention. The dependencies among the generic sources were not considered at this time. This problem will be approached in further study.

1. INTRODUCTION

The development of a generic data base as well as a plant specific data base is essential in performing a PSA (Probabilistic Safety Assessment). Moreover, the quality of a PSA depends heavily on the accuracy and consistency of the reliability data base employed. Since the first domestic PSA related research in 1986, there has been growing need to develop a reliability database in an objective way.

Two experts may show discrepancy in estimating a generic failure rate, even though they are mainly referring the same generic data source and the same plant. It seems that the difference stems from the analysts' interpretation of the data. This motivates the development of a systematic and consistent tool for processing reliability data. As Kaplan[2] mentioned, it is desirable to milk as much information as possible out of the available data, so that the impact of subjective judgment is minimized.

The purpose of this report is to document the development of the FORTRAN program MPRDP for the purpose of data specialization for the plant specific risk and reliability analyses. The computer program MPRDP is based on the "three-stage" Bayesian procedure[1] utilizing DPD (discretized probability distribution).

The first systematic procedure for analyzing failure data was proposed by Kaplan[2]. The DPD method was initially suggested by him in the 'two-stage' procedure and frequently used in many PSA studies[5, 6]. However, the 'two-stage' procedure can not handle generic book type data which are recorded in the form of the estimates for the parameters defining the failure rate distribution (Kaplan[2] named it the 'population variability curve'). This procedure was mainly for processing data from generic plants' experience and plant specific experience.

Later, Mosleh and Apostolakis[3] modified the 'two-stage' Bayesian procedure to handle generic book type data as well. However, there seems to be inconsistency in their way of getting the joint likelihood function for generic plant data and book data. We propose a sequential procedure to handle the plant data in the first stage, and the book data in the second stage. While our procedure is based on the above two procedures, significant improvements were made in the proposed procedure combining generic plant experience and generic book data.

In section 2, a brief introduction to the Bayes' theorem and its application to the problem of combining various types of data are explained. Section 3 describes the 'tree-stage' procedure. Here, we focus on the need for the development of 'three-stage' procedure instead of 'two-stage' one. Other characteristics of our procedure are also explained in Ref. [1]. Section 4 presents the framework and brief description for the subroutines of the MPRDP program. In section 5, the input requirements and some suggestions for using the MPRDP program are given. Various appendices are for the MPRDP source program, the sample input and the corresponding output. The contents of output files are also described. In section 6, the results of two specific applications are explained. In Section 7, brief conclusions and directions for further study are given.

2. BAYES' THEOREM AND ITS APPLICATION

2. 1 Bayes' Theorem and DPD

The Bayes' theorem is one of the most convenient ways to incorporate currently available pieces of information for updating our prior knowledge about the quantity of interest. If the quantity of interest is a failure rate Λ (failures per hour or per demand), the theorem states that:

$$f(\lambda|E) = \frac{f(\lambda)P(E|\lambda)}{\int_0^\infty f(\lambda)P(E|\lambda)d\lambda} \qquad (1)$$

where

$f(\lambda|E)$ = the probability density function of Λ given information E(posterior distribution),

$f(\lambda)$ = the probability density function prior to having information E(prior distribution), and

$P(E|\lambda)$ = the joint pdf of the information E given λ.

In most cases, it is impossible to get Eq. 1 in a closed form, since the denominator frequently requires numerical integration. By applying the DPD technique, we can get an approximate solution as in Eq. 2.

$$P(\Lambda \in C_\ell|E) \simeq \frac{P(\Lambda \in C_\ell)P(E|\lambda_\ell)}{\sum_{\ell=1}^{L} P(\Lambda \in C_\ell)P(E|\lambda_\ell)}, \ (\ell = 1,\cdots,L) \qquad (2)$$

where $\{C_1, C_2 \cdots, C_N\}$ is a partition for the space of λ, and each C_ℓ is represented by λ_ℓ.

2. 2 Information Types

The information E can be further divided into four types of information depending on its relevancy :

E_0 = general engineering knowledge or underlying assumptions,

E_1 = past information data from operating experience at similar plants,

E_2 = generic failure rate estimates or distributions contained in various industry compendia,

E_3 = plant specific performance data.

The first three of these information types, E_0, E_1 and E_2, may constitute the generic information. Without rigorous justification, the lognormal distribution has been used for the failure rate distribution in the PSA practice. In some cases like the frequencies for some transient events, the gamma distribution has been also used. This general engineering judgment may be an example of E_0.

Here, the role of E_0 is somewhat different from E_1 or E_2. The likelihood of λ can not be measured by E_0, while E_0 determines the a priori distribution of λ as well as other basic distributions. Therefore, in this situation, the Bayes formula in Eq. 1 should be rewritten by Eq. 3.

$$f(\lambda|E_0, E_1, E_2, E_3) = \frac{f(\lambda|E_0) \cdot P(E_1, E_2, E_3|\lambda, E_0)}{\int_0^\infty f(\lambda|E_0) \cdot P(E_1, E_2, E_3|\lambda, E_0)d\lambda} \tag{3}$$

The DPD version of this formula would be :

$$P(\lambda \in C_\ell|E_0, E_1, E_2, E_3) \simeq \frac{P(\lambda \in C_\ell|E_0) \cdot P(E_1, E_2, E_3|\lambda_\ell, E_0)}{\sum_{\ell=1}^{L} P(\lambda \in C_\ell|E_0) \cdot P(E_1, E_2, E_3|\lambda_\ell, E_0)} \tag{4}$$

The information E_1(generic plant data) represents the experience at similar plants and can be summarized by the doublets :

(X_m, T_m), m=1, . . . , M, for running type failure, or

(X_m, D_m), m=1, . . . , M, for demand type failur,

i. e., Xm failures in Tm operation hours or in Dm demands at the m-the plant.

As mentioned in Ref. [3], the third type of generic information E_2 (generic book data) is a set of estimates for parametes that define a generic failure rate distribution. It can be generalized by :

(ν^*_m, τ^*_m), (μ^*_m, τ^*_m), or (μ^*_m, σ^*_m) for m = 1, . . . , B ;

where ν^*_m, τ^*_m, μ^*_m, and σ^*_m are the estimates for the mean failure rate ν_m, the error factor τ_m, the median failure rate μ_m, and the logarithmic standard deviation σ_m, respectively.

We use information E_1 and E_2 for the specification of parameters in the lognormal or gamma distribution. Note that both distributions require two parameters. For the lognormal distribution, the parameters are μ and σ, where μ is the median and σ the logarithmic standard deviation as given by the following pdf :

$$\phi(\lambda|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma\lambda} \cdot \exp\left[-\frac{(\log(\lambda/\mu))^2}{2\sigma^2}\right], \text{ for}(\lambda>0 ; \mu, \sigma>0) \tag{5}$$

For the gamma distribution, the pdf is determined by two parameters, α, and β as in Eq. 6.

$$\psi(\lambda|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\lambda^{\alpha-1} \cdot \exp\{-\lambda/\beta\}, \text{ for } (\lambda>0 ; \alpha, \beta>0) \tag{6}$$

3. THE 'THREE-STAGE' BAYESIAN PROCEDURE

Intially E_0 determines the type of the underlying distribution ofr λ. Since we consider only two types of distributions, we may have :

$$f(\lambda|E_0) = \begin{cases} \phi(\lambda|\mu, \sigma), \text{ where } \mu \text{ and } \sigma \text{ are unspecified ;} \\ \psi(\lambda|\alpha, \beta), \text{ where } \alpha \text{ and } \beta \text{ are unspecified.} \end{cases} \quad (7)$$

3. 1 The First Stage

In the 'three−stage' Bayesian data update, the first stage concentrates on the estimation for the parameters of the posterior distribution of λ, given information E_0 and E_1. the basic formula required in this stage would be :

$$P(\theta|E_0, E_1) = \frac{P(\theta|E_0)P(E_1|\theta, E_0)}{\int P(\theta|E_0)P(E_1|\theta, E_0)d\theta} \quad (8)$$

where $\theta = (\mu, \sigma)$, or (α, β).

The calculation of the denominator would be even more complicated. First we discretize the parameter space of θ into a finite set of values, $\theta_{ij}(i = 1, \ldots I ; j = 1, \ldots, J)$. Then we calculate the likelihood of θ_{ij} given information E_0 and E_1 by the following formula.

$$P(\theta_{ij}|E_0, E_1) \simeq \frac{P(\theta_{ij}|E_0)P(E_1|\theta_{ij}, E_0)}{\sum_{i=1}^{I}\sum_{j=1}^{J}P(\theta_{ij}|E_0)P(E_1|\theta_{ij}, E_0)} \quad (9)$$

As usual, we employ the uniform prior for $P(\theta_{ij}|E_0)$. Then we have :

$$P(\theta_{ij}|E_0, E_1) \simeq \frac{P(E_1|\theta_{ij}, E_0)}{\sum_{i=1}^{I}\sum_{j=1}^{I}P(E_1|\theta_{ij}, E_0)}, \quad (10)$$

where

$$P(E_1|\theta_{ij}, E_0) = \begin{cases} \prod_{m=1}^{M}P(x_m|t_m, \theta_{ij}, E_0), \text{ for running type failure ;} \\ \prod_{m=1}^{M}P(x_m|d_m, \theta_{ij}, E_0), \text{ for demand type failure,} \end{cases} \quad (a)$$

$$P(x_m|t_m, \theta_{ij}, E_0) = \int_0^\infty P(x_m|t_m, \lambda)P(\lambda|\theta_{ij}, E_0)d\lambda,$$

$$P(x_m|d_m, \theta_{ij}, E_0) = \int_0^\infty P(x_m|d_m, \lambda)P(\lambda|\theta_{ij}, E_0)d\lambda, \quad (b)$$

$$P(\lambda|\theta_{ij}, E_0) = \begin{cases} \phi(\lambda|\mu_i, \sigma_j) \\ \psi(\lambda|\alpha_j, \beta_i), \end{cases} \quad (c)$$

$$P(x_m|t_m, \lambda) = e^{-\lambda tm}(\lambda t_m)^{xm}/x_m! \text{ for } x_m = 0, 1, 2, \cdots \quad (d)$$

$$P(x_m|d_m, \lambda) = \binom{dm}{xm}\lambda^x(1-\lambda)^{dm-xm} \text{ for } x_m = 0, 1, 2, \cdots \quad (e)$$

From the likelihood function in Eq. 10, either the maximum likelihood estimate (MLE) or the best estimate for θ_{ij} can be determined approximately.

3.2 The Second Stage

This set of parameter estimates takes the role of a generic compendium, E_2. If other estmates (generic book data) are not available as is the case in Ref. [2], the estimates from the first stage can be directly used as a prior in the third stage. Otherwise, we perform the second stage with all the estimates available including those from the first stage. Likewise, even when information type E_1(plant experience) is not available, the second stage can be performed with other estimates to determine a prior generic failure rate distribution.

The second stage characterize our procedure. We can think of two formulae for calculation the conditional likelihood of θ_{ij} given information E_0, E_1, and E_2.

$$P(\theta_{ij}|E_0, E_1, E_2) \simeq \frac{P(\theta_{ij}|E_0)P(E_1, E_2|\theta_{ij}, E_0)}{\sum_{i=1}^{I}\sum_{j=1}^{J}P(\theta_{ij}|E_0)P(E_1, E_2|\theta_{ij}, E_0)} \tag{11}$$

$$P(\theta_{ij}|E_0, E_1, E_2) \simeq \frac{P(\theta_{ij}|E_0, E_1)P(E_2|\theta_{ij}, E_0, E_1)}{\sum_{i=1}^{I}\sum_{j=1}^{J}P(\theta_{ij}|E_0, E_1)P(E_2|\theta_{ij}, E_0, E_1)} \tag{12}$$

Eq. 11 is employed in Ref. [3]. It might be possible to handle E_1 and E_2 at the same time by this equation. However, we notic some problem in determining the hoint distribution $P(E_1, E_2|\theta_{ij}, E_0)$. Even assuming the independence of E_1 and E_2, we can not guarantee that each item in E_1 and E_2 has the equal weight, which was also assumed in Ref. [3]. Eq. 12 would be appropriate in the 'three-stage' cotext. However, it is even more difficult to get the from of $P(E_2|\theta_{ij}, E_0, E_1)$. since each item in E_2 can be regarded as a colection of E_1, we may replace E_1 by θ_F, the resulting estimate from the first stage, in the can devise a simple formula by modifying Eq. 11 as :

$$\begin{aligned} P(\theta_{ij}|E_0, E_1, E_2) &\simeq P(\theta_{ij}|E_0, \theta_F, E_2) \\ &\simeq \frac{P(\theta_{ij}|E_0)P(\theta_F, E_2|\theta_{ij}, E_0)}{\sum_{i=1}^{I}\sum_{j=1}^{J}P(\theta_{ij}|E_0)P(\theta_F, E_2|\theta_{ij}, E_0)} \end{aligned} \tag{13}$$

In Eq. 13, we can easily have the formula for $P(\theta_F, E_2|\theta_{ij}, E_0)$ as in Eq. 14, since θ_F has the equal weight as each item in E_2.

$$\begin{aligned} P(\theta_F, E_2|\theta_{ij}, E_0) &= P(\theta_F, \theta_1^{*}, \theta_2^{*}, \cdots, \theta_B^{*}|\theta_{ij}, E_0) \\ &= P(\theta_F|\theta_{ij}, E_0) \cdot \prod_{b=1}^{B}(\theta_b^{*}|\theta_{ij}, E_0). \end{aligned} \tag{14}$$

3.3 The Third Stage

In the third stage, the prior thus obtained is updated using Eq. 15.

$$P(\Lambda \in C_l|E) \simeq \frac{P(\Lambda \in C_l|E_0, E_1, E_2)P(E_3|\lambda_l, E_0, E_1, E_2)}{\sum_{l=1}^{L}P(\Lambda \in C_l|E_0, E_1, E_2)P(E_3|\lambda_l, E_0, E_1, E_2)} \tag{15}$$

The conditional pdf of E_3, $P(E_3|\lambda)$ can be either a bionomial or a Poisson distribution depending on the information type E_3. If E_3 is given by 'x failures in t hours,' then the pdf will be the form of Eq. 10d. On the other hand, if E_3 is given by 'k failures in d trials,' then the pdf will be the form of Eq. 10e.

This stage is a direct application of the Bayes theorem. Examples of this third stage update are given in Refs. [5] and [6].

Therefore, the focal point in the three-stage Bayesian update is the sequential use of information E_1 and E_2 for the determination of a priori failure rate distribution by applying Bayes theorem. In Ref. [2], E_2 type information was not considered, while it is mentioned in Ref. [3]. In Ref. [3], however, each set of parameter estimates was treated with equal weight as each set of plant data. In our procedure, available plant data are processed first to propose a set of estimates, then it is treated with equal weight as each set of other estimates. This new procedure is based on our belief that a set of estimates from a generic compendium is based on a group of generic plant data, so it has more information than that from a plant's experience. The flow diagram of the 'three-stage' procedure is given in figure 1.

3. 4 Illustrating Formulae

Here we illustrate some updating formulae for the running type failure when the lognormal underlying distribution is assumed. The first stage can be performed by :

$$P(E_1|\mu_i, \sigma_j, E_0) = \prod_{m=1}^{M} \left\{ \int_0^\infty \phi\ (\lambda|\mu_i, \sigma_j) \cdot f(x_m|t_m, \lambda)d\lambda \right\}$$

$$\simeq \sum_{\ell=1}^{L} \int_{z\in s_\ell} \frac{1}{\sqrt{(2\pi)}} \exp\ (\ -z^2/2)dz \times \frac{(\mu_i e^{-\sigma_j z_\ell} t_m)^{x_m} \exp\ (\mu_i e^{-\sigma_j z_\ell} t_m)}{x_m!} \tag{16}$$

Here the parameters μ_i and σ_j define the distribution for $\wedge$ as in Eq. (2), when the lognormal distribution is assumed, and S_ℓ denotes the fixed sub-interval corresponding to C_ℓ for the DPD of the standard normal probability.

When the underlying distribution is assumed to be the lognormal, it would be appropriate to use the lognormal distribution for the conditional density of θ as illustrated in Eq. 17 and Eq. 18.

$$P(\theta|\theta_{ij}, E_0) = \int_0^\infty P(\theta|\lambda, \theta_{ij}, E_0)P(\lambda|\theta_{ij}, E_0)d\lambda$$

$$= (\sqrt{2\pi(\sigma^2+\sigma_j^2)} \cdot v)^{-1} \cdot \exp\left[-\frac{1}{2} \left(\frac{\log(v/\mu_i)}{\sigma^2+\sigma_j^2}\right)\right], \tag{17}$$

where $\theta = (v, \sigma)$, $\theta_{ij} = (\mu_i, \sigma_j)$, and

$$P(\theta|\lambda, \theta_{ij}, E_0) = P((v, \sigma)|\lambda, (\mu_i, \sigma_j), E_0)$$

$$= (\sqrt{2\pi}\sigma v)^{-1} \cdot \exp\ \left[-\frac{1}{2}\left(\frac{\log(v/\lambda)}{\sigma}\right)^2\right]. \tag{18}$$

The second stage can be performed by Eq. 13 and 14 with Eq. 17 and 18.

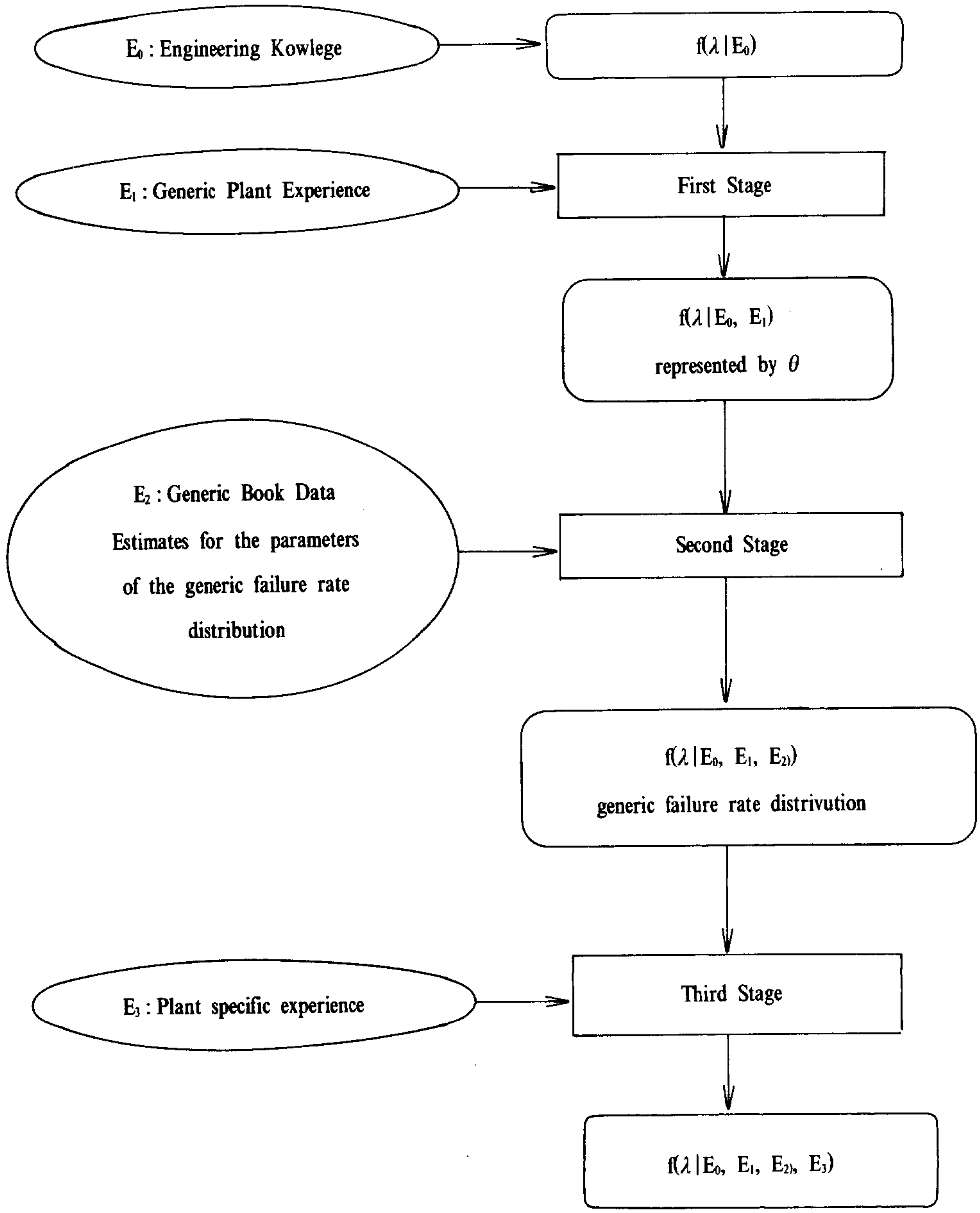

Figure 1. The Flow Diagram of the 'Three—Stage' Baesian Procedurere

At the last stage the DPD for $\wedge$ can be obtained by Eq. 15 with the numerator given by :

$$\int_{z \in s_\ell} \frac{1}{\sqrt{2\pi}} \exp\ (-z^2/2)dz \times \frac{(\mu_s e^{-\sigma_s z_\ell} t_s)^{x_s} \exp\ (-\mu_s e^{-\sigma_s z_\ell} t_s)}{x_s!}, \tag{19}$$

where(μ_s, σ_s) denotes the parameter estimates from the second stage, and (x_s, t_s) denotes the plant specific data.

4. THE PROCEDURE OF MPRDP

4. 1 The Structure of MPRDP

MPRDP code was developed to provide wide range of options, and so it is composed of various subroutines. According to the input options, only a part of these subroutines are to be executed. The flow diagram of the MPRDP is in figure 2. Even though the relation among subroutines looks cmplicated due to its multi-purposed tasks, the subroutines can be classified into 4 stages ; initialization for DPD, the first stage update, the second stage update, and the third stage update. The brief description of the framework is as follows.

4. 1. 1 Initialization Stage

Step 1.

Read necessary input data and options, to check their consistency, and to write out inpur data for users to check them easily.

Step 2.

Calculate the confidence interval of the failure rate from generic plant data and to provide a sufficient range of parameters under the given prior distribution.

Step 3−1.

Establish standard normal intervals according to a given rule and the number of intervals provided by users, and to assign probability for each interval.(* for the lognormal prior)

Step 3−2.

Establish standard gamma intervals for each discretized α parameter and to assign probability for each interval.(* for the lognormal prior)

4. 1. 2 First Stage

Step 1.

If generic plant data are provided, update them using Eq. 10 to produce a set of parameter estimates for the generic failure rate distribution. Otherwise, go to the second stage. This step corresponds to the first stage proposed by Kaplan[2] when there are no generic book data to be updated.

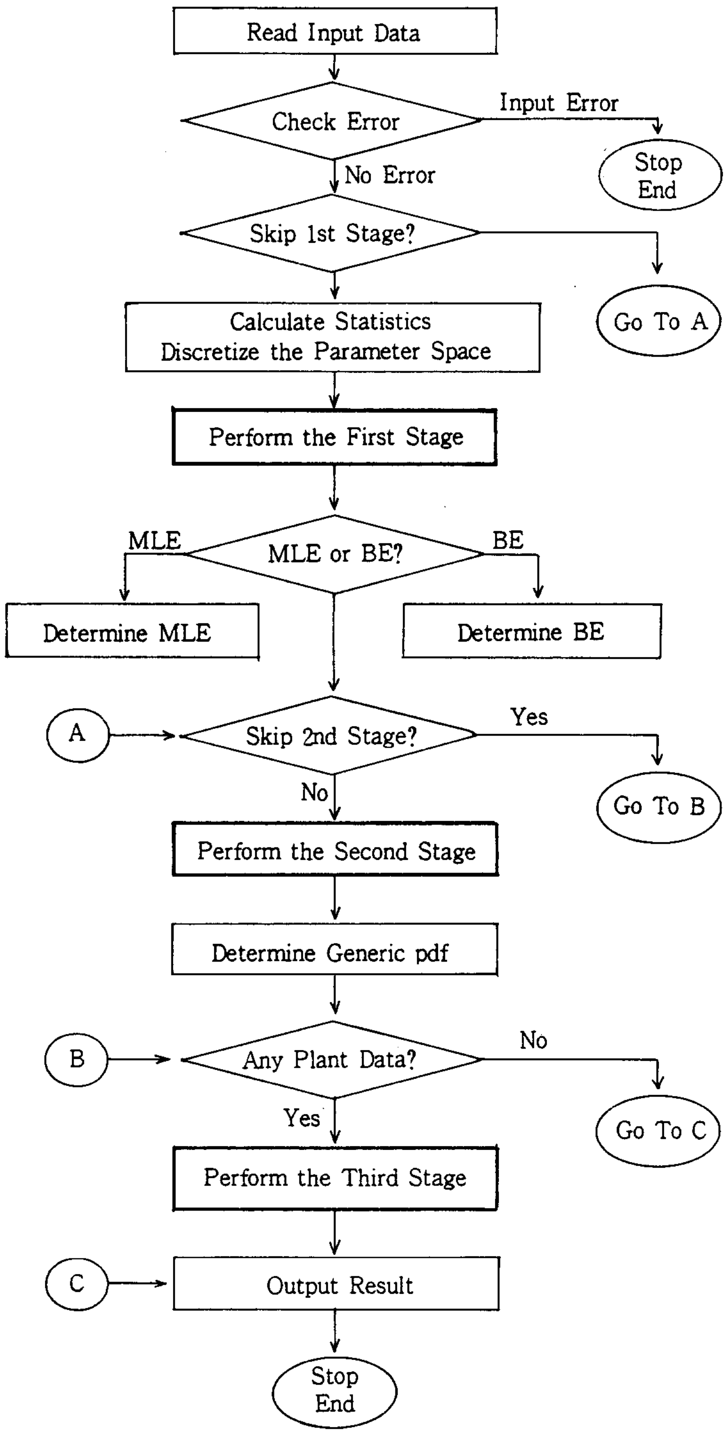

Figure 2. The Flow Diagram of The MPRDP Program

Step 2.

Determine MLE or best estimates from the result of step 1.

4. 1. 3 Second Stage

Step 1.

If there is no book data, go to step 3. Otherwise, convert generic book data into lognormal parameters, μ and σ.

Step 2.

Perform the second stage Bayesian update for the given prior density with the generic book data by using Eq. 13.

Step 3.

Determine the maximum likelihood estimates or the best estimates for from the results of step 2. That is, in this step, the parameters for the third prior distribution are determined.

4. 1. 4 Third Stage

Step 1.

Decide the range of the discretized failure rates which is enough to cover sufficient range, say $\pm 3\sigma$ around μ for the prior distribution. Generate discretized (failure rates) in log-scale or in normal scale.

Step 2.

Perform the third stage Bayesian update with the plant specific data for a given likelihood function (Poisson or binomial) using Eq. 15.

Step 3.

Determine the estimates for the parameters of the third posterior distribution ; mean $E(\Lambda | E_0, E_1, E_2, E_3)$, variance $E(\Lambda^2 | E_0, E_1, E_2, E_3)$, and 5-th, 50-th, 95-th percentiles.

Step 4.

Fit the posterior distribution to a member of a parametric family, using the statistics calculated in step 3. If it is necessary, write out the posterior distribution and fitted posterior distribution in a table for graphic usage.

4. 2 Execution of MPRDP

In order to execute the MPRDP code, an input data file must be provided, whose structure will be explained in section 5. To provide an input file of MPRDP, first define the field starting with character '$' and following key wards(5 characters including '$'). Then type corresponding information on the next line according to the given format.

We do not need all the fields to execute MPRDP since default options are already assigned. Brief description and an example of the input file is provided in figure 3. The execution procedure is as follows.

Step 1.

Construct an input file (for example, sam. dat)

Step 2.

Type 'mprdp'. Then the following question appears on the screen.

'INPUT FILE NAME ?'

Step 3.

Type input file name (for example, sam. dat). Then the following question appears on the screen.

'OUTPUT FILE NAME ?'

Step 4.

Type output file name (for example, sam. out). Then the following question appears on the screen.

'GDB FILE NAME ?'

Step 5.

Type generic data base output file name (for example, sam. gdb). Then the following question appears on the screen.

'PLOT FILE NAME ?'

Step 6.

Type plot file name (for example, sam. plt). Then the program starts and the names of the executed subroutines appear on the screen.

5. INPUT/ OUTPUT

5. 1 Input Data File Structure

The input data file is to be composed of 24 key words, each of which should be started with character '$' followed by 4 characters. Some of them are essential, while others are optional (marked by ✳). An example for input data file is given in figure 3. The whole structure of an input file is as follows.

(1) $TITLE (A80)(✳ optional)

Input one line of title up to 80 characters

(2) $DISTRIBUTION OPTION (1=POISSON(✳ default), 2=BINOMIAL)

(I5): input IDIST

```
$TITLE
MOV : FAILURE TO OPERATE ON DEMAND
$COMPONENT NAME
MOTOR
OPERATED VALVE
$CODE
 MV
FMODE(FAILURE MODE)
FAILURE TO OPERATE ON DEMAND
$FCODE
 0(Fail to open), C(Fail to close)
$TYPE OF FAILURE
   0
$DISTRIBUTION OPTION(1=POISSON, 2=BINOMIAL)
   2
$ESTUNATIR(1=MLE, 2=BE, 3=BE2)
   1
$SECOND STAGE SKIP(0=NO SKIP, 1=SKIP)
   1
$PRIOR DISTRIBUTION(1=LOGNORMAL, 2=GAMMA)
   1
$CONJUGATE PRIOR FOR LOGNORMAL TO BE USED?(0=NO, 1=YES)
   1
$PARAMETER(MU, SIGMA) INPUT?(0=YES, 1=NO)
   1
$NUMBERS(NUM, NUS, NUZ, NUD, NUD1, NUD2, NUD3, NPS, NUL)
10   5   50   4   0   0   0   0
$PVALUE
   0.05
$GENERIC PLANT DATA
   42.      6725.
   31.      14677
    3.       1505
   60.      11732
   69.      10052
$BOOK DATA
      1
   .003      10.
   .004       0.
   .004      10.
   .0043      3.7
$SOURCE
NUREG/CR-4550
NUREG/CR-1363
OCONEE PRA
SEABROOK PSS
GENERIC PLANTS
```

Figure 3. Example of an Input File

IDIST = 1 for running type failure (Poisson likelihood function)

IDIST = 2 for demand type failure (binomial likelihood function)

(3) $ESTIMATOR (1=MLE(* default), 2=BE)

(I5): input IEST

IEST = 1 for maximum likelihood estimation

IEST = 2 for best estimation

(4) $SKIP THE FIRST STAGE (0=NO SKIP(* default), 1=SKIP)

(I5): input ISKIP

ISKIP = 0 to perform the first stage

ISKIP = 1 to skip the first stage(when only one generic information is available)

(5) $SECOND STAGE SKIP (0=NO SKIP(* default), 1=SKIP)

(I5): input ISND

ISND = 0 to perform the third stage

ISND = 1 to skip the third stage(when plant specific data are not available)

(6) $PRIOR DISTRIBUTION (1=LOGNORMAL(* default), 2=GAMMA)

(I5): input IPRIOR

IPRIOR = 1 for lognormal prior

IRPIOR = 2 for gamma prior

(7) $CONJUGATE PRIOR FOR LOGNORMAL TO BE USED? (0=NO(* default), 1=YES)

(I5): input IEXACT

IEXACT = 0 for numerical integration

IEXACT = 1 for arithmetical integration

(8) $GCON (GAMMA-POISSON CONJUGATE PRIOR?) (0=NO(* default), 1=YES)

(I5): input IGCON

IGCON = 0 for numerical integration

IGCON = 1 for arithmetical integration

(9) $PARAMETER (MUE, SIGMA) INPUT? (0=YES(* default), 1=NO, 2)

(I5): input IPARA

IPARA = 0 when the lognormal parameters are provided

IPARA = 1 when the lognormal parameters are not available

IPARA = 2 when the range of the lognormal parameters are given

(10) $PLOT OPTION (0=NO PLOT(* default), 1=PLOT)

(I5): input IPLT

IPLT = 0 when the plot input file is not necessary

IPLT = 1 when the plot input file is necessary

(11) $NUMBERS (NUM, NUS, NUZ, NUD, NUD1, NUD2, NUD3, NPS, NUL)

(9I5)

Input the following numbers according to the given format

(I5) 1 - 5: NUM = the number of discretized (or)

(I5) 6 -10: NUS = the number of discretized (or)

(I5) 11-15: NUZ = the number of standard normal intervals

(I5) 16-20: NUD = the number of generic plant data

(I5) 21-25: NUD1 = the number of type1 generic book data

(I5) 26-30: NUD2 = the number of type2 generic book data

(I5) 31-35: NUD3 = the number of type3 generic book data

(I5) 35-40: NPS = the number of plant specific data

(I5) 41-45: NUL = the number of discretized

(12) $LAMBDA AXIS INPUT (0=NO INPUT(* default), 1=INPUT)

(I5): input LAMIN

LAMIN = 0 when the range of is not provided

LAMIN = 1 when the range of is provided

(13) $SCALE OPTION (0=LOG SCALE(* default), 1=NORMAL SCALE)

(I5): input ISCALE

ISCALE = 0 to discretize in log scale

ISCALE = 1 to discretize in normal scale

(14) $PVALUE(* default value = 0. 05)

(F10. 0)

input p-value for the confidence interval of

(15) $SPECIFIC DATA (FAILURE, TIME)(* not needed when ISND = 1)

(2F10. 0)

(F10. 0) 1 -10: Input plant specific failure rate, XKS(k)

(F10. 0) 11-20: Input plant specific observation time, TS(k)

(16) $GENERIC DATA (FAILURE, TIME)(* not needed when ISKIP = 1)

(2F10. 0)

(F10. 0) 1 -10: Input generic plant failure rate, XK(m)

(F10. 0) 11-20: Input generic plant observation time, TIME(m)

(17) $BOOK DATA(* not needed when ISND = 1)

(I5): input IBTYPE

IBTYPE = 1 for type1 book data

IBTYPE = 2 for type2 book data

IBTYPE = 3 for type3 book data

(2F10. 0)

if (IBTYPE=1), input median(F10. 0) and error factor(F10. 0): GMED(m) and GEF(m)

if (IBTYPE=2), input mean(F10. 0) and error factor(F10. 0): GMEAN(m) and GEF(m).

if (IBTYPE=3), input mean(F10. 0) and standard deviation(F10. 0) ; GMEAN(m) and GSTD(m)

(18) $RAGM (1P2E10. 3)(* needed only when IPARA = 2)

(1PE10. 3): input minimum value of to be discretized

(1PE10. 3): input maximum value of to be discretized

(19) $RAGS (1P2E10. 3)(* needed only when IPARA = 2)

(1PE10. 3): input minimum value of to be discretized

(1PE10. 3): input maximum value of to be discretized

(20) $MUE DISCRETIZATION(* needed only when IPARA = 0)

(5F10. 0): input discretized , XMUE(i), for i = 1,. ,NUM

(21) $SIGMA DISCRETIZATION(* needed only when IPARA = 0)

(5F10. 0): input discretized , SIGMA(j), for j = 1,. ,NUS

(22) $GAMMA PARAMETER INPUT (0=NO INPUT(* default), 1=INPUT)

(I5): input IGAM

IGAM = 0 when gamma parameters are provided

IGAM = 1 when gamma parameters are not provided

(23) $ALPHA DISCRETIZATION(* needed only when IGAM = 1)

(5F10. 0): input discretized , ALPH(j), for j = 1,. ,NUS

(24) $BETA DISCRETIZATION

(5F10. 0): input discretized , BETA(i), for i = 1,. ,NUM

5. 2 Data File Structure for GDB collection

When we collect and process generic data, it is desired to report the input data as well as the result (estimates for parameters of generic failure rate distribution) in a handy format. For this reason, MPRDP provide a routine for reporting a generic data collection form with its results as illustrated in figure 5. The input data structure is as follows.

(1) $COMPONENT NAME (A35): input name of the component

(2) $CODE (A5): input CODE: Component code

(3) $FMODE (A56): input FMODE: Failure mode description

(4) $FCODE (A5): input FCODE: Failure code

(5) $TYPE (I5): input ITYPE: Failure type

ITYPE=0 for running failure

ITYPE=1 for demand failure

ITYPE=3 for standby failure

(6) $SOURCE

(A15): input SOURCE(m) for m=1,. ,NTD+1

NTD = NUD1 + NUD2 + NUD3: total number of generic book data

SOURCE(m): Source of the generic book data

SOURCE(NTD+1) = 'Generic Plants'(* default when NUD 〉 0)

5. 3 Error Checking

When required input data are missing or inconsistent, the program stops automatically and the corresponding error messages appear on the screen.

(1) Unless IPRIOR = 1 or 2, the following error message appears on the screen : 'Prior Distribution Option Error'

(2) Unless IDIST = 1 or 2, the following error message appears on the screen:

'Failure Type Option Error'

(3) Unless IEST = 1 or 2, the following error message appears on the screen:

'Estimator Option Error'

(4) Unless IPARA = 1, 2 or 3, the following error message appears on the screen :

'Lognormal Parameter Option Error'

(5) Unless ISKIP = 0 or 1, the following error message appears on the screen:

'First Stage Skip Option Error'

(6) Unless ISND = 0 or 1, the following error message appears on the screen:

'Second Stage Skip Option Error'

(7) Unless IEXACT = 0 or 1, the following error message appears on the screen :

'Lognormal Conjugate Option Error'

(8) Unless IGCON = 0 or 1, the following error message appears on the screen :

'Gamma-Poisson Conjugate Option Error'

(9) Unless LAMIN = 0 or 1, the following error message appears on the screen :

'Lambda Range Input Option Error'

(10) Unless IPLT = 0 or 1, the following error message appears on the screen :

'Plot File Option Error'

(11) If IGAM not = 0 and IPRIOR = 1, the following error message appears on the screen:

'Gamma Parameter Option Error'

5. 4 Structure of output files

5. 4. 1 General output file

This file contains all the intermediate results during data update.

(1) TITLE

(2) Input Options and Data (A. 1 - A. 7)

(3) Intermediate Results (B. & STAT.)

(4) Result of the first stage update (C. 1 - C. 2)

C. 1. First stage posterior probabilities for discretized parameters

C. 2. Selected parameters and statistics for the first stage posterior distribution

(5) Results of the third stage update (D. 1 - D. 4)

D. 1. Second stage posterior failure rate distribution

D. 2. Statistics for the third stage posterior distribution

D. 3. Fitted posterior distribution and statistics

D. 4. Reference statistics

An example of the output file is in figure 4.

5. 4. 2 Generic data base file

Generic data base file is to be made when one wants to determine generic failure rate distribution from all the available generic sources. GDB file shows all the input generic sources and determined parameters(mean and error factors) in a given generic data collection format. An example of GDB file is in figure 5.

5. 4. 3 Plot file

This file is to provide necessary input for general graphic tools to plot the results of the third stage update. Plot file is composed of five columns.

The first column:discretized (failure rate).

The second column: the second stage posterior cell probabilities.

The third column: their cumulative probabilities.

The fourth column: fitted posterior cell probabilities.

The last column: fitted posterior probability density.

An example of plot file is in figure 6.

```
MOV:FAILURE TO OPERATE ON DEMAND

A.1 THE FIRST STAGE PRIOR DISTRIBUTION
      --> UNIFORM PRIOR FOR THE PARAMETER SPACE
A.2 THE GENERIC (PRIOR) DISTRIBUTION
      --> LOGNORMAL PRIOR DISTRIBUTION
A.3 FAILURE TYPE
      --> DEMAND TYPE FAILURE (BINOMIAL LIKELIHOOD FUNCTION)
A.4 ANALYSIS OPTIONS
    (1) METHOD OF ESTIMATION
        --> MAXIMUM LIKELIHOOD ESTIMATION
    (2) PARAMETER INPUT
        --> NO INPUT FOR LOGNORMAL PARAMETERS
    (3) SKIP THE FIRST STAGE ?                      :NO
    (4) SKIP THE SECOND STAGE ?                     :YES
    (5) CONJUGATE PRIOR FOR LOGNORMAL UPDATE ?      :YES
    (6) GAMMA-POISSON CONJUGATE PRIOR ?             :NO
    (7) IS THE RANGE OF LAMBDA GIVEN ?              :NO
    (8) IS PLOT FILE GENERATED ?                    :NO
A.5 INPUT DATA NUMBERS
    (1) THE NUMBER OF DISCRETIZED MUE IS            10
    (2) THE NUMBER OF DISCRETIZEFD SIGMA IS          5
    (3) THE NUMBER OF DICRETIZED Z IS               50
    (4) THE NUMBER OF GENERIC DATA IS                5
    (5) THE NUMBER OF TYPE1 BOOK DATA IS             4
    (6) THE NUMBER OF TYPE2 BOOK DATA IS             0
    (7) THE NUMBER OF TYPE3 BOOK DATA IS             0
    (8) THE NUMBER OF PLANT SPECIFIC DATA IS         0
    (9) THE NUMBER OF DISCRETIZED LAMBDA IS          0
A.7 GENERIC DATA
    THE GENERIC PLANT DATA
          FAILURES        TIME/DEMAND
    ---------------------------------
     1      42.0            6725.0
     2      31.0           14677.0
     3       3.0            1505.0
     4      60.0           11732.0
     5      69.0           10052.0
    ---------------------------------
    THE TYPE1 GENERIC BOOK DATA
              MEAN         ERROR FACTOR
    ---------------------------------
     1      3.000E-03      1.000E+01
     2      4.000E-03      0.000E+00
     3      4.000E-03      1.000E+01
     4      4.300E-03      3.700E+00
    ---------------------------------

B. INTERMEDIATE RESULTS

  (1) LOGNORMAL PARAMETERS INDUCED FROM GENERIC BOOK DATA
                 MU             SIGMA
          ---------------------------
             1.126E-03      1.400E+00
             2.426E-03      1.400E+00
             1.502E-03      1.400E+00
             3.134E-03      7.954E-01
          ---------------------------

  CASE: NO PARAMERS ARE GIVEN

  STAT.1 THE RANGE OF GENERIC PLANT DATA
  MEAN          LOWER         UPPER         EF            SIG
  ----------------------------------------------------------------
  6.2454E-03    4.5006E-03    8.4421E-03    1.3696E+00    1.9119E-01
  2.1121E-03    1.4348E-03    2.9981E-03    1.4455E+00    2.2399E-01
  1.9934E-03    4.1078E-04    5.8292E-03    3.7670E+00    8.0625E-01
  5.1142E-03    3.8202E-03    6.5042E-03    1.3048E+00    1.6175E-01
  6.8643E-03    5.2447E-03    8.5951E-03    1.2802E+00    1.5015E-01
   STAT.2 THE RANGE OF PARAMETERS
     MINIMUM MUE FROM DATA          1.1261E-03
     MAXIMUM MUE FROM DATA          6.8643E-03
     MINIMUM SIGMA FROM DATA        1.5015E-01
     MAXIMUM SIGMA FROM DATA        1.3999E+00
   (1) AUTOMATICALLY DISCRETIZED MUE
   1.126E-03   1.377E-03   1.683E-03   2.057E-03   2.515E-03
   3.074E-03   3.758E-03   4.594E-03   5.615E-03   6.864E-03
   (2) AUTOMATICALLY DISCRETIZED SIGMA
   .15015      .46258      .77501      1.08744     1.39987
   STAT: THE CONFIDENCE INTERVAL OF GENERIC PLANT FAILURE RATES
   1. SAMPLE MEAN FAILURE RATE IS    4.466E-03
   2. POOLED SAMPLE MEAN IS            4.587E-03
   3. 95 % CONFIDENCE INTERVAL IS (3.959E-03 ,5.239E-03)
   DATA( 1)  ( 4.200E+01 6.725E+03 6.245E-03) MAY BE AN OUTLIER
   DATA( 2)  ( 3.100E+01 1.468E+04 2.112E-03) MAY BE AN OUTLIER
   DATA( 3)  ( 3.000E+00 1.505E+03 1.993E-03) MAY BE AN OUTLIER
   DATA( 5)  ( 6.900E+01 1.005E+04 6.864E-03) MAY BE AN OUTLIER
    DISCRETIZED Z
  -5.250E+00  -4.750E+00  -4.250E+00  -3.750E+00  -3.250E+00
  -2.469E+00  -1.786E+00  -1.532E+00  -1.353E+00  -1.210E+00
  -1.089E+00  -9.818E-01  -8.850E-01  -7.958E-01  -7.127E-01
  -6.342E-01  -5.594E-01  -4.876E-01  -4.183E-01  -3.510E-01
  -2.852E-01  -2.206E-01  -1.570E-01  -9.393E-02  -3.127E-01

  3.127E-02   9.393E-02   1.570E-01   2.206E-01   2.852E-01
  3.510E-01   4.183E-01   4.876E-01   5.594E-01   6.342E-01
  7.127E-01   7.958E-01   8.850E-01   9.818E-01   1.089E+00
  1.210E+00   1.353E+00   1.532E+00   1.786E+00   2.469E+00
  3.250E+00   3.750E+00   4.250E+00   4.750E+00   5.250E+00

       CELL PROBABILITY
     2.867E-07   3.111E-06   2.827E-05   2.010E-04   1.117E-03
     2.493E-02   2.493E-02   2.493E-02   2.493E-02   2.493E-02
     2.493E-02   2.493E-02   2.493E-02   2.493E-02   2.493E-02
     2.493E-02   2.493E-02   2.493E-02   2.493E-02   2.493E-02
     2.493E-02   2.493E-02   2.493E-02   2.493E-02   2.493E-02
     2.493E-02   2.493E-02   2.493E-02   2.493E-02   2.493E-02
     2.493E-02   2.493E-02   2.493E-02   2.493E-02   2.493E-02
     2.493E-02   2.493E-02   2.493E-02   2.493E-02   2.493E-02
     2.493E-02   2.493E-02   2.493E-02   2.493E-02   2.493E-02
     1.117E-03   2.010E-04   2.827E-05   3.111E-06   2.867E-07
NUM      XMUE        PMUE                PEVID
                                          SIGMA
                           1.50E-01  4.63E-01  7.75E-01  1.09E+00  1.40E+00
 1   1.13E-03  2.76E-03
                           9.71E-41  1.19E-08  1.21E-04  1.09E-03  1.55E-03
 2   1.38E-03  5.92E-03
                           3.91E-31  2.13E-06  2.18E-04  2.99E-03  2.71E-03
 3   1.68E-03  1.50E-02
                           7.37E-23  1.30E-04  4.20E-03  6.34E-03  4.30E-03
 4   2.06E-03  3.40E-02
                           5.88E-16  2.52E-03  1.37E-02  1.16E-02  6.18E-03
 5   2.51E-03  6.82E-02
                           1.75E-10  9.01E-03  3.32E-02  1.79E-02  8.03E-03
 6   3.07E-03  1.49E-01
                           1.62E-06  5.77E-02  5.84E-02  2.36E-02  9.45E-03
 7   3.76E-03  2.55E-01
                           4.20E-04  1.43E-01  7.53E-02  2.65E-02  1.01E-02
 8   4.59E-03  2.53E-01
                           2.49E-03  1.44E-01  7.13E-02  2.52E-02  9.71E-03
 9   5.62E-03  1.48E-01
                           2.80E-04  6.94E-02  4.35E-02  2.04E-02  8.48E-03
10   6.86E-03  6.80E-02
                           5.14E-07  2.31E-02  2.51E-02  1.40E-02  6.69E-03
  C.2 SELECTED PARAMETERS FOR THE FIRST STAGE
   MLE FOR (MUE,SIGMA) IS : (   4.5936E-03      4.6288E-01)
   GENERIC MEAN WOULD BE   :     5.1123E-03
   GENERIC ERROR FACTOR    :     2.1403E+00
   GENERIC  5-TH QUANTILE :      2.3886E-03
   GENERIC 95-TH QUANTILE :      1.0942E-02
   BE FOR (MLE,SIGMA) IS   : (   4.1613E-03      7.2150E-01)
   GENERIC MEAN WOULD BE   :     5.3984E-03
   GENERIC ERROR FACTOR    :     3.2788E+00
   GENERIC  5-TH QUANTILE :      1.6475E-03

   GENERIC 95-TH QUANTILE :    1.7689E-02

    C. RESULTS OF THE FIRST STAGE UPDATE

  C.1 THE FIRST STAGE POSTERIOR DISTRIBUTION
NUM      XMUE        PMUE                PEVID
                                          SIGMA
                           1.50E-01  4.63E-01  7.75E-01  1.09E+00  1.40E+00
 1   1.13E-03  2.34E-02
                           9.27E-04  4.16E-03  7.27E-03  6.55E-03  4.52E-03
 2   1.38E-03  4.15E-02
                           4.02E-03  1.02E-02  1.24E-02  9.19E-03  5.70E-03
 3   1.68E-03  7.07E-02
                           1.30E-02  2.07E-02  1.85E-02  1.18E-02  6.73E-03
 4   2.06E-03  1.11E-01
                           3.14E-02  3.44E-02  2.43E-02  1.38E-02  7.44E-03
 5   2.51E-03  1.54E-01
                           5.65E-02  4.68E-02  2.80E-02  1.48E-02  7.71E-03
 6   3.07E-03  1.79E-01
                           7.61E-02  5.23E-02  2.84E-02  1.45E-02  7.49E-03
 7   3.76E-03  1.69E-01
                           7.62E-02  4.79E-02  2.53E-02  1.30E-02  6.82E-03
 8   4.59E-03  1.29E-01
                           5.70E-02  3.61E-02  1.96E-02  1.06E-02  5.81E-03
 9   5.62E-03  8.03E-02
                           3.18E-02  2.23E-02  1.36E-02  7.95E-03  4.65E-03
10   6.86E-03  4.16E-02
                           1.32E-02  1.13E-02  8.20E-03  5.45E-03  3.48E-03
  C.2 SELECTED PARAMETERS FOR THE FIRST STAGE
   MLE FOR (MUE,SIGMA) IS : (   3.7577E-03      1.5015E-01)
   GENERIC MEAN WOULD BE   :     3.8003E-03
   GENERIC ERROR FACTOR    :     1.2802E+00
   GENERIC  5-TH QUANTILE :      2.9686E-03
   GENERIC 95-TH QUANTILE :      4.8651E-03
   BE FOR (MLE,SIGMA) IS   : (   3.3342E-03      5.3179E-01)
   GENERIC MEAN WOULD BE   :     3.8406E-03
   GENERIC ERROR FACTOR    :     2.3984E+00
   GENERIC  5-TH QUANTILE :      1.6013E-03
   GENERIC 95-TH QUANTILE :      9.2111E-03
   BE FOR (MLE,SIGMA) IS   : (   2.9970E-03      7.6248E-01)
   GENERIC MEAN WOULD BE   :     4.0081E-03
   GENERIC ERROR FACTOR    :     3.5053E+00
   GENERIC  5-TH QUANTILE :      8.5500E-04

   GENERIC 95-TH QUANTILE :      1.0506E-02
```

Figure 4. Example of an Output File

GENERIC COMPONENT FAILURE DATA PREPARATION SHEET FOR ASR PRA
SHEET NO. 2

COMPONENT NAME : MOTOR OPERATED VALVE CODE : MV
(BOUNDARY) :
FAILURE MODE : FAILURE REMAIN OPEN
CODE : T(Transfer Closed), P(Plugged)
TYPE : 3 / 1(*)

FAILURE DATA

MEAN	EF	SOURCE	REMARKS
1.3000E−07	3.0000E+00	NUREG /CR−4550	
5.7000E−08	1.0000E+01	NUREG /CR−1363	
2.0000E−07	1.0000E+01	OCONEE PRA	
9.3000E−08	5.0000E+00	SEABROOK PSS	
2.0000E−07	1.0000E+01	NUREG /CR−2815	
1.0000E−07	3.0000E+00	IREP NUREG 272	
1.2000E−07	7.2000E+00	WASH−1400	
5.3000E−08	1.0000E+01	ZION PSS	
1.6000E−07	1.0000E+01	SHOREHAM PRA	
1.6485E−07	5.4939E+10	GENERIC PLANTS	4 Plants 0 1890000 0 3220000 0 1429000 1 817399

VALUE CALCUALTED : MEAN=9.5709E−08, EF = 3.6182E+00

MEAN	EF	CORREL. NO(**)	REMARKS

RATIONALE :

(*) Type 0 : demand(/d), 1 : run(/hr), 3 : stand by(/hr)
(**) This correlation Number is used for Uncertainty Analysis in KIRAP

Figure 5. GDB Output File : MOV failure to remain open

4.655E-04	1.353E-07	1.353E-07	4.064E-28	4.064E-28	1.634E-10
6.546E-04	2.789E-07	4.142E-07	7.723E-27	8.128E-27	1.217E-09
9.187E-04	7.986E-07	1.213E-06	2.906E-25	2.987E-25	9.039E-09
1.289E-03	2.185E-06	3.398E-06	8.616E-24	8.195E-24	5.983E-08
1.809E-03	5.710E-06	9.108E-06	2.549E-22	2.638E-22	3.531E-07
2.539E-03	1.426E-05	2.336E-05	7.530E-21	7.794E-21	1.857E-06
3.564E-03	3.400E-05	5.736E-05	2.222E-19	2.300E-19	8.712E-06
5.001E-03	7.745E-05	1.348E-04	6.548E-18	6.778E-18	3.643E-05
7.109E-03	1.686E-04	3.034E-04	1.925E-16	1.993E-16	1.358E-04
9.850E-03	3.505E-04	6.538E-04	5.639E-15	5.839E-15	4.514E-04
1.382E-02	6.961E-04	1.350E-03	1.645E-13	1.703E-13	1.338E-03
1.940E-02	1.321E-03	2.670E-03	4.767E-12	4.937E-12	3.535E-03
2.722E-02	2.393E-03	5.064E-03	1.369E-10	1.418E-10	8.327E-03
3.821E-02	4.144E-03	9.208E-03	3.882E-09	4.024E-09	1.749E-02
5.362E-02	6.854E-03	1.606E-02	1.082E-07	1.122E-07	3.276E-02
7.525E-02	1.083E-02	2.689E-02	2.940E-06	3.052E-06	5.470E-02
1.506E-01	1.635E-02	4.324E-02	7.715E-05	8.021E-05	8.145E-02
1.482E-01	2.357E-02	6.681E-02	1.928E-03	2.008E-03	1.081E-01
2.080E-01	3.246E-02	9.927E-02	4.498E-02	4.699E-02	1.280E-01
2.919E-01	4.272E-02	1.420E-01	9.530E-01	1.000E+00	1.351E-01
4.096E-01	5.369E-02	1.957E-01	0.000E+00	1.000E+00	1.271E-01
5.749E-01	6.448E-02	2.602E-01	0.000E+00	1.000E+00	1.067E-01
8.068E-01	7.396E-02	3.341E-01	0.000E+00	1.000E+00	7.983E-02
1.132E+00	8.104E-02	4.152E-01	0.000E+00	1.000E+00	5.326E-02
1.589E+00	8.484E-02	5.000E-01	0.000E+00	1.000E+00	3.168E-02
2.230E+00	8.484E-02	8.848E-01	0.000E+00	1.000E+00	1.680E-02
3.129E+00	8.104E-02	6.659E-01	0.000E+00	1.000E+00	7.945E-03
4.329E+00	7.396E-02	7.398E-01	0.000E+00	1.000E+00	3.350E-03
6.163E+00	6.448E-02	8.043E-01	0.000E+00	1.000E+00	1.259E-03
8.650E+00	5.369E-02	8.580-01	0.000E+00	1.000E+00	4.221E-04
1.214E+01	4.272E-02	9.007E-02	0.000E+00	1.000E+00	1.262E-04
1.704E+01	3.246E-02	9.332-01	0.000E+00	1.000E+00	3.361E-05
2.391E+01	2.357E-02	9.568E-01	0.000E+00	1.000+00	7.985E-06
3.355E+01	1.635E-02	9.731E-01	0.000E+00	1.000E+00	1.691E-06
4.709E+01	1.083E-02	9,839E-01	0.000E+00	1.000E+00	3.193E-07
6.608E+01	6.854E-03	9.908E-01	0.000E+00	1.000E+00	5.375E-08
9.273E+01	4.144E-03	9.949E-01	0.000E+00	1.000E+00	9.288E-09
1.301E+02	2.393E-03	9.973E-01	0.000E+00	1.000E+00	0.000E+00
1.826E+02	1.321E-03	9,987E-01	0.000E+00	1.000E+00	0.000E+00
2.563E+02	6.961E-04	9.993E-01	0.000E+00	1.000E+00	0.000E+00
3.597E+02	3.505E-04	9,997E-01	0.000E+00	1.000E+00	0.000E+00
5.048E+02	1.686E-04	9.999E-01	9.000E+00	1.000E+00	0.000E+00
7.085E+02	7.745E-05	9.999E-01	0.000E+00	1.000E+00	0.000E+00
9.942E+02	3.400E-05	1.000E+00	0.000E+00	1.000E+00	0.000E+00
1.395E+03	1.426-05	1.000E+00	0.000E+00	1.000E+00	0.000E+00
1.958E+03	5.710E-06	1.000E+00	0.000E+00	1.000E+00	0.000E+00
2.748E+03	2.185E-06	1.000E+00	0.000E+00	1.000E+00	0.000E+00
3.857E+03	7.986E-07	1.000E+00	0.000E+00	1.000E+00	0.000E+00
5.412E+03	2.789E-07	1.000E+00	0.000E+00	1.000E+00	0.000E+00
7.956E+03	1.353E-07	0.000E+00	0.000E+00	1.000E+00	0.000E+00

Figure 6. Example of a Plot File

6. APPLICATION

6. 1 MOV Failure To Operate on Demand

Ref. [7] provides generic parameter estimates for the failure rate distributions of various components as well as their sources. The first example is for the case of 'MOV failure to operate on demand' mode. The source data are updated to compare with the point estimate proposed by Ref. [6] and to provide error factor associated with the estimation. Table 1 shows the source data. The corresponding input data file is also given in figure 3.

Table 1. Generic data for MOV demand failure

Plants	No. of failures	No. of demands
Oconee	42	6,725
Zion	31	14,677
Indian Point	3	1,505
Millstone	60	11,732
PWR X	69	10,052
Source	Failure rate	Error factor
NUREG /CR－4550	3.0E－3	10
NUREG /CR－1363	4.0E－3	－
Oconee PRA	4.0E－3	10
Seabrook PSS	4.3E－3	3.7

The generic failure rate distribution was calculated by the computer code MPRDP. The ranges of the parameters were calculated to be (1.1E-3, 6.5E-3) for the median and (0.32, 1.4) for the dispersion parameter . The resulting distribution has a 5th and 95th percentile of 6.8E-4 /d and 1.1E-2 /d, respectively, with a mean value of 4.0E-3 /d and error factor of 4. For this example, we can observe the followings. First, the estimated value 2.83E-3 /d for the median and 0.84 for the dispersion parameter are in the automatic ranges calculated above, so the automatic parameter generation is satisfactory. Second, the estimated mean value coincides with that proposed by Ref. [7]. Third, this procedure provides additional valuable information, the error factor associated with the estimation, which was not available in Ref. [7].

6. 2 MDP Failure To Start on Demand

Ref. [3] provides a generic data set for the failure frequency of standby motor-driven pumps to start on demand. We chose this data set to compare our result with those given by Ref. [3]. The data are shown in table 2. The resulting distribution has a 5th and 95th percentile of 9.27E-5/d and 1.22E-2/d, respectively, with a mean value of 3.21E-3/d and associated error factor of 11.5. Our mean value estimate is slightly (about 2.5%) lower than that of Ref. [3], while our range factor is significantly (about 40%) larger than that of Ref. [3].

The MLE of failure rate for the plant data was calculated to be 3.75E-3/d, which is much higher than the overall mean value estimate. Thus we can see that the effect of plant data on the mean value estimate was less significant in our procedure than in the procedure of Ref. [3]. The reason for the range factor associated with our estimate to be larger than that of Ref. [3] can be also explained. Our procedure updated the plant data first to propose a set of generic parameter estimates, so it actually used 4 sets of book data whose estimated failure rates vary by factor of 38 in the second update. On the other hand, the procedure proposed by Ref. [3] updates 3 sets of book data and 8 sets of plant data at the same time.

The main issue in this example is the interpretation of available sources. Ref. [3] regarded each source as the experience from a plant, whereas we believe each source is composed of several plants' experience.

Table 2. Generic data for MOV demand failure

Plants	No. of failures	No. of demands
Plnat 1 MDP	3	3,140
Plnat 2 MDP	7	793
Plnat 3 MDP	2	800
Plnat 4 CSP	0	828
Plnat 4 RHRP	1	860
Plnat 5 LPIP	0	223
Plnat 5 HPIP	1	530
Plnat 5 CSP	3	140
Source	Failure Rate	Error Factor
Source 1 Pump	1.0E−3	5
Source 2 SBP	3.5E−3	3
Source 3 MDP	2.0E−3	10

6. 3 Artificial Examples

Two examples for generic book data updates were selected to examine the characteristics of the MPRDP. Suppose we have generic data for a failure rate estimate as in table 3. The mean failure rate estimates vary by factor of 100, while their associated error factors are estimated to be the same.

Table 3. Artificial generic data (I)

Sources	Failure Rate	Error Factor
Source 1	1.0E−3	5
Source 2	1.0E−4	5
Source 3	1.0E−5	5

In this case, the MLE for the median and error factor can be calculated by the equation (6.2. 51) and (6.2.52) of Ref. [9]. They are 6.2E-5 for the median and 14.0 for the error factor resulting in 2.2E-4 for the mean. The MLE calculated by the MPRDP are 5.9E-5 for the median and 13.8 for the error factor resulting in 2.1E-4 for the mean, and they can be exactly estimated if more refined parameter space is adopted. However, the best estimates are calculated to be 5.8E-5 for the median and 17.2 for the error factor resulting in 2.6E-4 for the mean. The discrepancy is due to the dependency of the best estimates on the discretized parameter spaces.

Table 4 illustrates another generic data collection. In this case the MLE for the error factor does not locate in the feasible reason, i. e. , it can not be defined. However, the best estimates calculated by the MPRDP was 5.3E-4 for the median and 5.3 for the error factor resulting in 8. 8E-4 for the mean, which seem to be reasonable considering that the parameter spaces were generated based on only three data.

Table 4. Artificial generic data (II)

Sources	Failure Rate	Error Factor
Source 1	1.0E−3	3
Source 2	1.0E−3	5
Source 3	1.0E−3	10

6. 4 Construction of a Generic Reliability Database

EPRI URD[7] and IAEA-TECDOC-478[8] provide various generic data sources composed of generic plant data and generic book type data. As a pilot study, we developed a generic reliability database by updating data provided by these sources by using the procedure developed in this study. Almost 30 generic data sources were reviewed, and available data were collected and screened from them. We processed reliability data for about 100 safety related components frequently modeled in PSA. The resulting parameter estimates were reported in three forms; (1) generic data collection work sheets, (2) distributions showing the range of input data as well as the resulting distribution, (3) a summary table which can be used as an input deck for future PSA works. The details are included in Ref. [4].

As an illustrations, figure 5 shows an GDB collection format for 'MOV failure to remain open,' respectively. Figure 7 shows the ranges of input data and the result of first two stage updates in terms of mean failure rate and 90% confidence intervals. The confidence interval for the result shows good agreement to the input data.

7. CONCLUSION AND FUTURE DIRECTION

The 'three-stage' Bayesian procedure to treat generic and plant-specific data is proposed. We embodied the procedure in the MPRDP code which can combine broad range of generic data and plant specific data to determine a reasonable failure rate distribution. The weight of the generic book data was re-investigated, and a new procedure treating separately generic plant data and book data was developed. Most discretization functions are automated, so that it provides a consistent result.

We also developed a reliability data base for about 100 safety related components from almost 30 data sources. It can be directly used as an input to PSA works. After reviewing the range of input and output data, we concluded that the updates are satisfactory, since most of the resulting distributions seem to reasonably cover the input ranges.

Further study will be done in two phases. The first study will focus on the development of a procedure to treat the dependencies among various generic sources. The second phase will be on enhancing the user interface to have easier data input and output.

Reliability data analyses performed so far used to mainly depend on analysts' subjective judgments. While experts' judgments have provided key roles in PSAs, the need for a consistent and systematic procedure to overcome personal bias is still growing. The proposed program is believed to be a good substitute for experts' judgments in reliability data analyses.

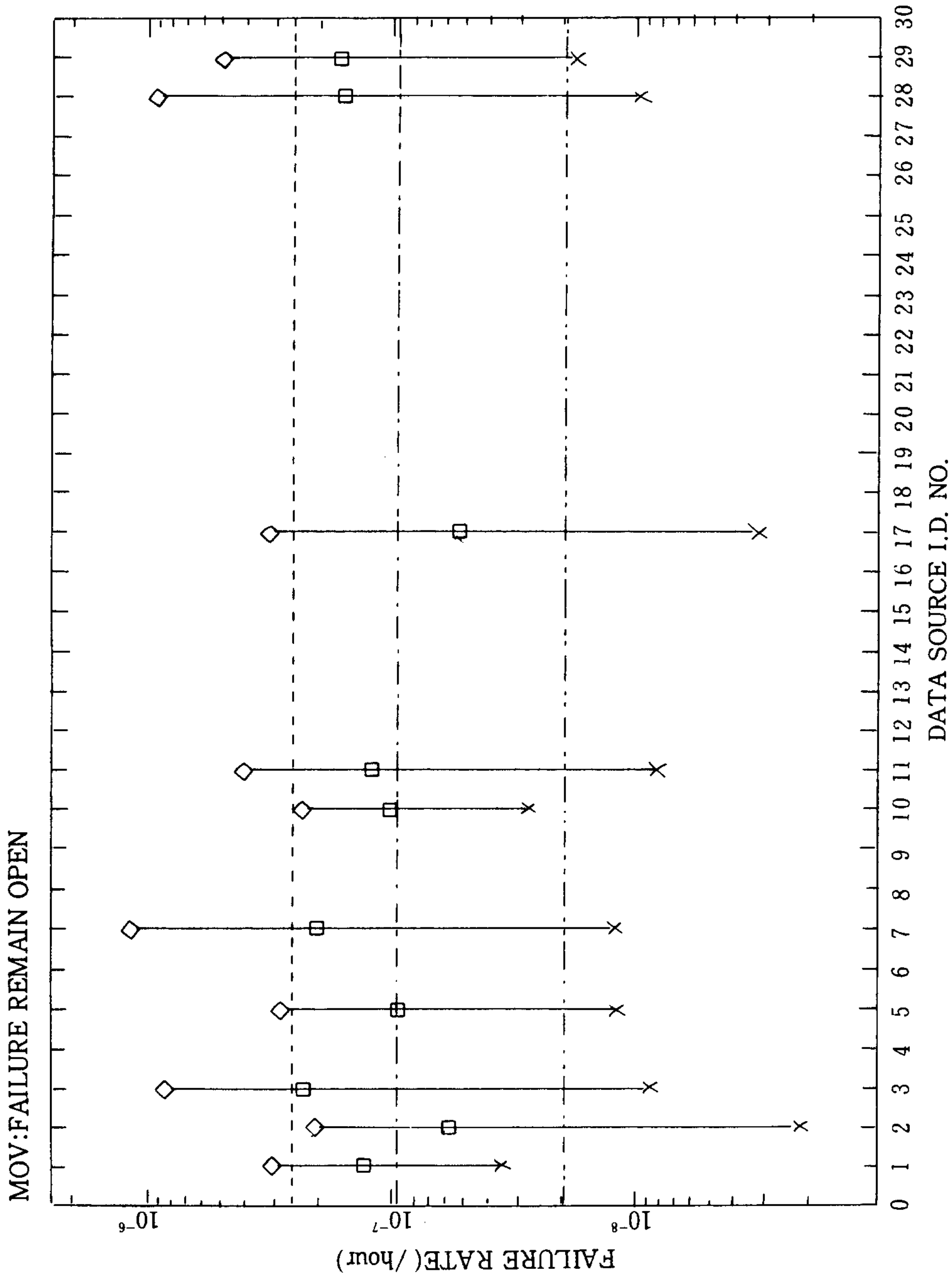

Figure 7. Range of Input Data and Result(MOV failure to remain open)

REFERENCES

1. Lim, T. J. and C. K. Park, "Enhancement of the Bayesian procedure for developing a reliability data base," *Proc. of PSA Int'l Topical Meeting, Florida* (1993)
2. Kaplan, S. , "On a Two-Stage Bayesian Procedure for determining Failure Rates from Experimental Data," *IEEE Trans. on Power Apparatus and Systems, Vol. PAS-102*, No. 1 (1983), pp. 195-202.
3. Mosleh, A. and G. Apostolakis, "The development of a Generic Data Base for Failure Rates," *ANS/ENS Int'l Topical Meeting on Probabilistic Safety Methods and Applications* (1985)
4. Lim, T. J. , C. K. Park, J. H. Park, T. W. Kim, "Survey and Comparison of Generic Component Reliability Data and Establishment of Preliminary Generic Database for USE in PSA," *KAERI/TR-364/93* (1993)
5. Midland Nuclear Plant Probabilistic Risk Assessment, Pickard, Lowe, and Garrick, Inc. (1985)
6. Oconee PRA : A Probabilistic Risk Assessment of Oconee Unit 3, NSAC-60, Duke Power Co. (1984)
7. Advanced Light Water Reactor Requirements Document, Vol. II, Utility Requirements for Evolutionary Plants, Appx. A to Chapter 1. PRA Key Assumptions and Groundrules, EPRI (1990)
8. Component Reliability Data for Use in PSA, IAEA-TECDOC-478, IAEA (1988)
9. Seabrook Station Probabilistic Safety Study, Public Service Co. of New Hampshire and Yankee Atomic Electric Company (1983)