

신뢰도 데이터 합성 program의 개발

이광원[†] · 박문희 · 오신규* · 한정민*

호서대학교 안전공학과, *한국가스공사 연구개발원
(2003. 8. 21. 접수 / 2003. 11. 4. 채택)

A Development on Reliability Data Integration Program

Kwang-Won Rhie[†] · Moon-Hi Park · Shin-Kyu Oh* · Jeongmin Han*

Department of Safety Engineering, Hoseo University

*KOGAS R&D Center

(Received August 21, 2003 / Accepted November 4, 2003)

Abstract : Bayes theorem, suggested by the British Mathematician Bayes (18th century), enables the prior estimate of the probability of an event under the condition given by a specific event. This theorem has been frequently used to revise the failure probability of a component or system. 2-stage Bayesian procedure was firstly published by Shultis et al. (1981) and Kaplan (1983), and was further developed based on the studies of Hora & Iman (1990) Papazoglou et al., Pörn(1993).

For a small observed failure number (below 12), the estimated reliability of a system or component is not reliable. In the case in which the reliability data of the corresponding system or component can be found in a generic reliability reference book, however, a reliable estimation of the failure probability can be realized by using Bayes theorem, which jointly makes use of the observed data (specific data) and the data found in reference book (generic data).

Key Words : reliability data, specific data, generic data, 2-stage bayesian procedure

1. 서 론

베이스 정리란 18C 영국 수학자 Bayes가 특수한 사건이 주어진 조건하에서 어떤 사건이 일어날 확률을 산출하는 이론을 정리한 것이며 근래에는 부품이나 시스템의 고장확률 등의 데이터수정에 많이 쓰이고 있다¹⁾. 이후 2단계 베이스이론(2-stage Bayesian procedure)는 Shultis(1981)와 Kaplan(1983)³⁾에 의하여 처음으로 발표 되었으며 Hora & Iman (1990)²⁾, Papazoglou et al.(1991)⁵⁾, Pörn(1993)⁴⁾ 등에 의하여 연구 발전되었다.

관찰되어진 고장수가 적을 때 (대략 12개 이하)는 추정되어지는 신뢰도 수치의 값이 신뢰성을 갖지 못하게 된다. 이때 일반 신뢰도 데이터 book에서 신뢰도 수치가 존재한다면 베이스 이론을 통하여 관찰되어진 데이터(specific 데이터)와 문헌에서 찾은

데이터(generic 데이터)를 동시에 고려 할 수가 있다. 또한 관찰된 데이터가 여러 플랜트의 데이터라면 각 플랜트의 보수정책이나 사용환경이 다르므로 이를 고려해 주어야 하며, 이는 이중적분 등 복잡한 수학문제로 귀결된다. 본 논문에서는 이러한 문제를 일반인들도 사용하기 쉽도록 프로그램을 개발하고자 하였다.

2. 이론적 배경

정량적 평가를 위하여 어떤 부품의 신뢰도 데이터를 수집하던 중 문헌에 나타난 일반적 데이터와 고유 신뢰도자료 등을 분석하여 얻은 specific 데이터 2개를 얻었을 때 이들을 모두 고려한 신뢰도 수치를 얻고자 할 때 베이스 이론을 사용한다. 즉, · 사전적 확률(prior probability): 추가적인 정보가 있기 전의 어떤 사건에 대한 사전적 확률, 즉 문헌 등에 나타난 확률

[†]To whom correspondence should be addressed.
kwrhie@office.hoseo.ac.kr

- 추가정보(evidence): 추가적 정보, 즉 최근 고유 신뢰도자료 등을 분석하여 얻은 확률
 - 사후적확률(posterior probability): 추가적 정보와 사전적 확률을 고려한 사후적 확률
- 이라 하고, P(Y)를 추가정보 즉, 고유 신뢰도자료 등을 분석하여 얻은 확률이라 하고 P(X)가 사전적 확률이라 하면 사후적 확률은

$$P(X|Y) = \frac{P(Y|X) \cdot P(X)}{P(Y)} = \frac{P(Y|X) \cdot P(X)}{\sum P(Y|X) \cdot P(X)} \quad (1)$$

로 표현된다.

일반적으로 1단계 베이스이론은 관찰되어지는 모든 부품은 같은 고장률이나 고장확률을 갖는다고 가정되어진다. 하지만 경험적으로 관찰하면 같은 부품일지라도 사용 장소가 틀리면 운전조건이나 보수 정책 등에 의하여 서로 다르게 고장이 일어난다. 이런 관점을 고려한 방법을 2단계 베이스이론(2-stage-bayesian procedure or super population method) 라하며 수학적으로 관심 있는 수치, 즉 사후적 확률을 확률 변수로 취급한다.

각기 다른 plant등에서 사용되어지는 같거나 비슷한 부품의 하위집단(subpopulation)이 존재하고 각각의 하위집단이 k개 존재하며 각 하위집단 당 값 E=(n,T) 나 E=(n,N)이 존재한다고 가정하자. 여기서 n은 고장수를 의미하며 T나 N은 관찰시간이나 요구횟수를 의미한다. 일반성의 상실 없이 추가정보 E1에 대한 관찰을 한다 가정하고 다른 것들을 E2 Ek 라 하자.

고장률이나 고장확률의 불확실성 계산을 위하여 이들은 어떤 미지변수 Ψ 에 의존되는 분포를 갖는 확률변수라 가정한다. 미지변수들은 다음과 같은 베이스 식에 의하여 구하여진다.

$$g(\Psi|E) = \frac{g(\Psi)L(E|\Psi)}{\int_{\Psi} g(\Psi)L(E|\Psi)d\Psi} \quad (2)$$

여기서 g(Ψ)는 선행 분포도함수이며 미지변수의 선행정보를 의미한다. 하위집단들이 서로 독립적이라 가정하자. Likelihood함수를 L(E| Ψ), 우리가 관심 있는 고장률 λ 나 고장확률 p를 x라 표현하면 총확률의 법칙의 적용과 함께 분포밀도함수 g($\lambda|E$)가 주어진다. 다시 한번 총확률의 법칙을 적용하면 변수 x에 대한 분포밀도함수를 구할 수 있으며 이것

은 선행분포함수(prior distribution)로 사용하고 마지막으로 하위집단 E1에 대한 고장률이나 요구횟수 당 고장확률을 결정하기 위한 베이스 식은 다음과 같이 나타나게 된다.

$$f(x|E_1, E) = \frac{L(E_1|x)f(x|E)}{\int_x L(E_1|x)f(x|E)dx} \quad (3)$$

하지만 이 식은 계산에 있어서 단혀있지가 않다. 고장률에 대하여는 감마분포를 그리고 요구횟수 당 고장확률에 대하여는 베타분포를 가정하면 분석적으로 풀 수 있다. 하지만 각 경우에 lognormal 분포를 가정하면 풀지를 못한다. 식 (2)에서 2번의 적분이 요구되는데 이는 지금까지 numerical하게 또는 Monte Carlo방법을 사용하여 풀 수밖에 없다고 알려져 있다.

3. Bayes program

Bayes program은 일반신뢰도 데이터들과 고유경험 데이터들을 기초로 하여 고장확률이나 고장률을 예측하기 위해 windows 환경에서 손쉽게 사용할 수 있도록 개발하였다. 프로그램 "Bayes 2.0"은 문헌에서 2단계 베이스이론 이라고 알려진 방법을 사용하였다.

지금까지 널리 사용되어진 고전적 (또는 1단계) 베이스이론은 관찰되어지는 모든 고장들이 같은 고장률이라는 전제 조건이 필요하지만 2단계 베이스이론은 이런 전제조건이 필요가 없다.

여러 경험상 비슷한 부품들일지라도 설치조건에 따라서 고장현상이 다르게 발생된다는 것이 증명되고 있으며 이런 사항은 하위집단 개념에 반영시킬 수 있다. 하나의 하위집단 안에 묶여지는 부품들을 그들의 성질과 외적설치조건이 거의 차이가 없으며, 같은 고장률이나 고장확률을 갖는다고 전제한다. 관찰되는 전체집단(superpopulation)는 보통 여러 개의 하위집단들로 구성되어 이들 하위집단들은 각각 고장률이나 고장확률은 갖는다고 가정한다.

실제적으로 전체집단은 보통 비슷한 종류와 비슷하게 설치된 부품들로 정의된다. 하위집단은 이들 부품들이 서로 다른 설치조건을 가질 때 분류한다. 일반적으로 여러 종류의 부품들의 정리된 부분적인 부품들을 하위집단으로 정의, 사용한다.

수학적 풀이를 가능케 하기 위하여 가정이 필요

하다. 즉 하위집단들의 고장률이나 고장확률들을 확률적 우연변수라 가정하면 미지의 변수를 갖는 어떤 분포함수 즉 *super population* 분포를 갖게 된다. 이러한 분석에서 제일먼저 할 일은 전체집단 분포의 미지변수를 결정하는 일이다. 이 미지변수결정은 베이스이론에 의해 가능하며 이때 결과 값으로 변수의 복합분포, 즉 변수분포를 얻게 된다. 변수결정의 입력 데이터는 관찰되어진 하위집단의 고장들과 일반 신뢰도 데이터가 되며 일반적으로 일반신뢰도 데이터는 *lognormal* 분포모델을 적용하고 *median* 값과 *k-factor* 값을 주는 것이 보통이다. 얻어진 변수분포로부터 궁극적으로는 총확률의 법칙을 사용하여 고장률이나 고장확률의 분포를 얻게 된다.

만약 전체적인 고장률이나 고장확률에 대하여 관심이 있다면 이로서 해답을 얻게 되며 만약 어떤 하나의 하위집단에 관심이 있다면 다시 한 번 이 하위집단에 대하여 베이스 이론을 적용하여 얻을 수 있다. 이 경우 여러 번 같은 데이터를 베이스 이론에 적용하면 틀린 값이 나오므로 관찰되어지는 하위집단은 처음에는 사용하지 말아야 한다.

본 program 은 *visual basic*으로 코드화 하였다. 입력 데이터 파일을 읽어서 여러 개의 출력 데이터 파일을 생성하며 다음과 같은 모델들에 대하여 program화하였다.

- 1) *lognormal* 분포의 전체집단에 대한 고장률의 계산
- 2) *gamma* 분포의 전체집단에 대한 고장률의 계산
- 3) *beta* 분포의 전체집단에 대한 고장확률 계산
- 4) *lognormal* 분포의 전체집단에 대한 고장확률 계산

이들 모델들은 변수가 2개나 3개인 *mixing* 분포에는 사용하지 못한다.

전체집단 분포에서 변수와 고장률이나 고장확률의 값들을 자동으로 주어진 데이터에 의하여 계산되어진다. 빈약한 입력 데이터는 쓸모없는 결과 값을 계산할 수도 있다. 그런 것을 방지하기 위해 입력 데이터에서 모든 변수범위를 미리 줄 수 있도록 하였다. 계산정확성(여러 적분계산 과정에서 발생하는 절단수의 크기와 이에 따른 결과치의 수)는 입력 데이터에서 조절할 수 있다. 즉 4단계 계산의 정확도를 높일 수 있도록 하였다.

계산은 2단계 베이스 이론(*superpopulation* 방법)을 사용하였다. 이때 일반적으로 많이 사용되어지는

Ignorance-Prior 방법을 사용하였다. 즉 구형분포를 사용하여 계산하였다. 단 한 가지 예외는 *lognormal* 분포모델에서의 변수분포 계산과정이며 이때에는 수학적 계산상의 이유로 *k-factor*의 값을 계산하기 위하여 *prior* 값은 입력 데이터에서 줄 수 있다.

입력 데이터는 주어진 형식에 의해 주게 되며 결과는 다음과 같이 출력 데이터에 주어진다. 즉 평균 값, 표준편차 5% 백분율, 50% 백분율, 95% 백분율 등의 수치 값들과 그래프 값(횡좌표, *prior* 값, *posterior* 값) 그리고 고장률이나 고장확률을 표현해주는 분포곡선(보통 21개의 좌표 값을 계산하여 이은)을 주게 된다.

a. 결과표

결과표에는 분석의 중요 결과 값들이 표현된다. 결과 값으로서 평균값, 표준편차, 5%, 50%, 95% 백분율들을 포함한다. 또한 *text* 형식으로 부품명과 형식 고장유형, 그리고 특정 하위집단을 관찰한 경우는 이 하위집단의 이름을 포함한다.

b. 계산점

계산과정에서 적분 등을 할 경우 21곳의 계산 점을 계산하게 되며 이때 계산되어지는 계산 점에서의 경험적 분포도함수 값들을 수록한다. 또한 경험적 분포함수의 평균값과 표준편차들도 주어지게 된다.

c. Plot 데이터

prior 값과 *posterior* 값들을 포함하는 데이터다. 이 데이터에는 각 줄에 3개의 값들 즉 고장률이나 고장확률 값과 *prior* 결과 값의 밀도, *posterior* 결과 값의 밀도들을 포함한다. 이 데이터는 많은 *graphic program*에 입력 데이터로 활용될 수 있으며 이로서 분포함수의 그림을 얻어낼 수 있다.

d. 입력 데이터

파일의 입력부를 클릭 한 후 다음과 같이 입력 데이터를 입력한다(Fig. 1~3참조).

- a) 프로젝트명 : *text type*으로 32자 이내
- b) 저장할 file 이름 : 저장되는 파일명, 같은 이름으로 결과 데이터가 생성된다. (*.*kur*, *.*lan*) *text type*으로 32자 이내
- c) 관찰부품의 명칭
- d) 고장유형 : *text type*으로 32자 이내 (빈칸없이 입력)

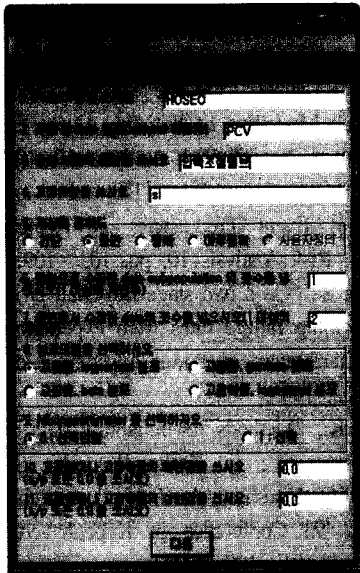


Fig. 1. Input data window

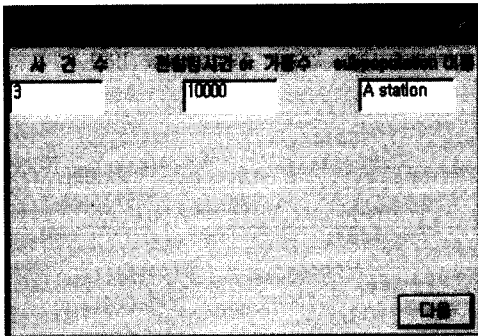


Fig. 2. Experience data

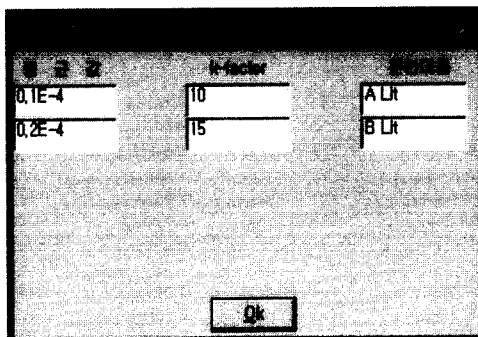


Fig. 3. Literature data

- e) 계산의 정확도 : option 값으로 하나만 선택가능
- f) 경험으로 수집된 data 하위집단의 개수 : 1 이상의 자연수

- g) 문헌에서 수집된 데이터의 개수 : 1 이상의 자연수
- h) 분포모델
- i) Mixing parameter
- j) 고장률이나 고장확률의 하한값 (λ/p 또는 default 값인 0)
- k) 고장률이나 고장확률의 상한값 (λ/p 또는 default 값인 0)

입력부 폼을 이용하여 만들어진 입력 데이터 file 의 1번째 줄은 프로젝트명, 2번째 줄은 파일이름, 3, 4번째는 자동으로 생성되는 결과파일, 5번째 줄은 관찰부품, 6번째 줄은 고장유형, 7번째 줄은 차례대로 계산의 정확도, 경험으로 수집된 데이터 하위집단의 개수, 문헌에서 수집된 데이터의 개수, 분포모델, mixing parameter, 고장률이나 고장확률의 하한값, 고장률이나 고장확률의 상한값 들이며 8번째 줄은 bayes 2.0에서 꼭 필요한 데이터이다. 9번째 줄은 차례대로 경험 데이터의 사건 수, 관찰된 시간 또는 가동 수, 하위집단 이름, 10번째줄은 차례대로 평균값, k-factor, 문헌이름 이다.

e. Output 데이터

출력 데이터는 크게 결과 확인부의 계산, plot 데이터 및 파일부의 결과확인이다. 파일 이름은 입력부 폼에 입력한 파일이름.bat를 선택한다.

결과확인부에는 a)분석일련번호 b)부품유형 c)고장유형 d)분석종류 e)평균값, 표준편차, 5% 백분율, 50% 백분율, 95% 백분율 등을 계산하여 보여준다 (Fig. 4 참조).

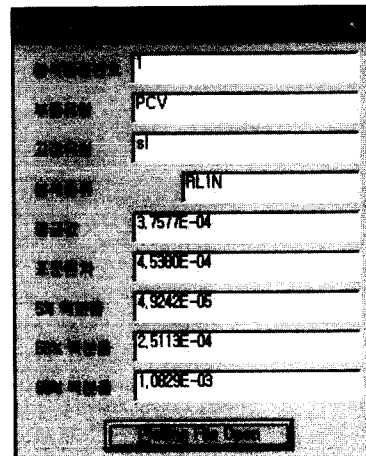


Fig. 4. Result table

또한 plot 데이터는 'test.lan' file에 저장되어진다. 이 file은 header 첫줄에 주어진 주석문이 쓰여진다. 그 다음 줄에는 각각의 부품유형, 고장유형, 분석유형이 차례로 써지게 된다. 각 구간에서는 시작점과 끝점에서의 고장률/확률과 구간 내에서 사용된 분포밀도 값 등이 주어진다.

4. 결 론

본 논문에서는 2단계 베이스이론을 사용하여 여러 공장이나 현장의 고유 신뢰도데이터와 문헌들로부터 구하여지는 신뢰도데이터의 합성을 일반인들도 손쉽게 사용할 수 있는 프로그램을 개발하였다. 지금까지는 복잡한 이중적분 등의 문제로 보통 1단계 베이스이론을 사용하였고, 이에 따라 경험데이터들은 수집 장소의 보수정책차이 등을 고려할 수가 없었으나 본 논문에서 소개한 프로그램은 이들을 고려할 수 있게 하여준다. 또한 신뢰도데이터의 분포 모양 등도 관찰할 수 있도록 하여 불확실성의 계산 등에도 유용하게 사용할 수 있도록 하였으며, 수명기가 보통 갖고 있는 여러 가지 분포함수(lognormal, gamma, beta분포)들의 고장률이나 고장확률을계산할 수 있도록 하여 사용자의 편의를 도모하였다.

참고문헌

- 1) Hofer, E. et al., "On the Solution Approach for Bayesian Modelling of Initiating Event Frequencies and Failure Rates", Risk Analysis 17, 1997.
- 2) Hora, S.C & Iman, R.L., "Bayesian Modelling of Initiating Event Frequencies at Nuclear Power Plants", Risk Analysis 10, pp. 103~109, 1990.
- 3) Kaplan, S., "On a two stage Bayesian procedure for determeining failure rates from experimental data", IEEE Transactions on Power Apparatus and Systems, PAS-102, 1983.
- 4) Poern, K., "Methods for reliability data analysis-A new perspective, in: Kafka, P.,J.(eds), Safety and Reliability Assesment-An Integral Approach", Proc. ESREL '93 Conference in Kopenhagen, Elsevier 1993, pp. 101~108, 1993.
- 5) Papazoglou, I. et al., "Bayes - A Computer Code for Data Anaysis under Lack of Knowledge and Population Variability", National Center for Scientific Research Demokritos, Athen, June 1991.