

종속적 문헌 추정치를 이용한 모집단 변이 분포의 추정[†]

임 태 진*

Estimating the Population Variability Distribution Using Dependent Estimates from Generic Sources[†]

Tae-Jin Lim*

Abstract

This paper presents a method for estimating the population variability distribution of the failure parameter (failure rate or failure probability) for each failure mode considered in PSA(Probabilistic Safety Assessment). We focus on the utilization of generic estimates from various industry compendia for the estimation. The estimates are complicated statistics of failure data from plants. When the failure data referred in two or more sources are overlapped, dependency occurs among the estimates provided by the sources. This type of problem is first addressed in this paper. We propose methods based on ML-II estimation in Bayesian framework and discuss the characteristics of the proposed estimators. The proposed methods are easy to apply in real field. Numerical examples are also provided.

1. 서론

원자력 발전소(원전)의 재해도(risk)를 정량화 하는 확률론적 안전성 평가(Probabilistic Safety Assessment, PSA)에 있어서 가장 기본

적으로 요구되는 정보는 원전을 구성하는 개별 기기의 신뢰도라고 할 수 있다. 신뢰도는 충분한 시간에 걸쳐 실시된 수명시험 자료로부터 추정하는 것이 바람직하지만, 원자력 발전소와 같이 고 가격 고 신뢰도의 기기(component)들로 구성된 시스템에서는 이러한 수명시험이 현

[†] 본 연구는 1994년 숭실대학교 교내학술연구비 지원에 의하여 연구되었음.

* 숭실대학교 산업공학과

실적으로 불가능하므로 과거의 가동 중 고장이력을 토대로 신뢰도¹⁾를 추정하게 된다. 그러나 가동 연수가 길지 않은 신규 발전소에 있어서는 고장 자료의 희소성으로 인하여 전문가의 경험이나 유사 발전소의 고장 자료에 의존하여 신뢰도를 평가하게 된다. 이러한 유무형의 자료를 일관성 있게 처리하는 도구로써 베이직안 방법이 고려되며, 베이스 정리는 자료와 판단을 결합하여 우리의 인지 상태(state of knowledge)를 일관성 있게 표현하는 가장 근본적인 도구로 쓰인다[11].

원전의 신뢰도 자료는 해당 발전소의 가동 이력에 근거한 발전소 고유 자료(plant specific data)와 유사 발전소의 고장 자료에 근거한 일반 자료(generic data)로 구분된다. 비교적 가동 연수가 짧은 국내 발전소들에 있어서는 일반 자료의 중요성이 매우 크다고 할 수 있다. 일반 자료는 또한 유사 발전소의 고장 자료 및 여러 발전소의 고장 자료를 토대로 추정한 문헌 자료의 두 유형으로 분류된다.

80년대 초부터 일반 발전소 자료를 활용하여 사전(prior) 분포²⁾를 구축하는 베이직안 분석 절차에 대한 연구가 다수 발표되었으나[5,10,11,14,15], 문헌 자료의 활용에 대한 연구는 미미한 실정이다. 문헌 자료는 몇 개의 발전소 자료를 처리하여 얻은 추정치로서, 수집이 용이하여 활용 가치가 크다. 그러나 고려된 발전소 자료가 중복되는 경우 상호 종속성을 갖게 되며 이들의 조건부 결합 분포를 유도하기 어려운 특성을 갖고 있다. 이러한 문헌 자료를 활용하여 체계적으로 사전 분포를 추정하는 방법

론을 제안하는 것이 본 논문의 목적이다.

기존의 신뢰도 자료 처리 방법들[5,10,11,14,15]은 발전소 자료만을 대상으로 하여 자료간의 독립성이 보장되고 자료의 조건부 분포가 쉽게 정의되는 상황에 맞게 개발된 것이므로, 문헌 자료와 같이 상호 종속성을 갖는 통계량 형태의 경우에는 적용되지 않는다. 적절한 처리 방법의 부재로 인하여 과거 PSA 연구에서는 분석자의 주관적인 선별 과정을 통하여 문헌 자료를 처리함으로써 추정의 일관성과 정확성을 보장하기 어려웠다. 개별 기기에 대한 부정확한 신뢰도 추정치가 입력 자료로 쓰일 경우 전체 시스템에 대한 신뢰도 추정치는 실제 값에 비하여 상당히 큰 오차를 갖게 된다. 따라서 정확한 신뢰도 분석을 수행하려면 통계량 형태의 문헌 자료를 처리하는 방법론의 개발이 필요하다.

추정치로 주어지는 문헌 자료는 어느 발전소 자료로부터 어떤 방법을 통해 구해진 것인지 설명되지 않은 경우가 빈번하다. 따라서 이러한 추정 과정에 대한 일관된 가정이 필요하며, 이 가정에 근거하여 문헌 자료(추정치)를 처리하여 발전소 변이 분포를 추정하는 방법을 제안한다.

사용 기호 및 용어

E_0 : 일반적인 공학적 지식 및 가정

E_i : 유사 발전소의 고장 자료

(발전소 $m = 1, \dots, N$)

운전 중 고장 : (x_m, t_m) , t_m 시간 동안 x_m 회 고장

1) PSA에서 기기신뢰도는 지수 수명분포의 가정하에 상수고장률로 정의된다.

2) PSA에서 사전분포는 모수에 대한 주관적 확률분포의 의미보다는 발전소간의 모수 값의 차이에 의해 발생하는 발전소 모집단에서의 변이분포(PVD)의 의미를 갖는다[11]. 따라서 본 연구의 대상은 기존의 점추정 문제와는 다르며 PVD를 추정하는 문제이다.

기동 실패 고장 : (x_m, d_m) , d_m 회 기동 시
도 중 x_m 회 실패

$$\hat{\lambda}_m = \begin{cases} x_m/t_m \text{ or } x_m/d_m, & \text{if } x_m > 0 \\ x_{0.5}^2(1)/(2t_m) \text{ or } x_{0.5}^2(1)/(2d_m), & \\ \text{if } x_m = 0 \end{cases}$$

$y_m = \log(\hat{\lambda}_m)$: 대수 변환된 고장(확)률
추정치

E_2 : 문헌 자료 : 고장(확)률 변이 분포
(PVD) 추정치
(문헌 $i = 1, \dots, B$)

E_3 : 해당 발전소 자료(plant-specific data)

R : 추정에 고려된 모든 발전소의 색인(index) 집합

R_i : 문헌*i*에서 고려된 발전소의 색인 집합

N : 추정에 고려된 모든 발전소 자료의 수.
즉, $N = |R|$.

N_i : 문헌*i*에서 고려된 발전소 자료의 수. 즉,
 $N_i = |R_i|$.

N_{ij} : 문헌*i*와 문헌*j*에서 공통으로 고려된 발전
소 자료의 수. 즉, $N_{ij} = |R_i \cap R_j|$.

i_m : 발전소 m 에서의 고장 모수 ($m = 1, \dots, N$)
운전 중 고장: 고장률 ($0 < i_m$)
기동 실패 고장: 고장 확률 ($0 < i_m < 1$)

E_j 을 베이저안 처리하여 모집단 변이
분포 (Population Variability Distri-
bution, PVD)를 추정한다.

[2단계] 1단계에서 구한 추정치와 여타 문헌
자료 E_2 를 베이저안 처리하여 모든 일
반 자료를 반영하는 PVD를 추정한다.

[3단계] 2단계에서 추정된 분포를 사전 분포로
놓고 발전소 고유 자료 E_3 를 베이저안
처리하여 해당 발전소에서의 고장 모수
의 사후(posterior) 분포를 추정한다.

이 절차에서 '3단계'는 하나의 모수에 대한
사후 분포를 구하는 것이므로 전형적 베이저안
방법으로 쉽게 해결된다. 기존의 연구는 '1단계'
'와 '3단계'의 처리 절차에 집중되어 '2단계'에
대한 연구는 미미한 실정이다. 본 연구의 대상
은 문헌 자료를 처리하는 '2단계'로서 '1단계'와
의 연관성을 고려하여 다음과 같은 전제 하에
서 수행한다.

첫째, 문헌 자료에서 고려된 발전소 자료들
은 중복 가능하다. 중복된 자료에 대한 상세한
정보는 알 수 없다. 둘째, 각 문헌에서 사용한
PVD 추정 방법은 동일하다. 여기서 두 번째
가정은 성립되지 않는 경우도 있으나, 각 문헌
에서 사용한 추정 방법을 알 수 없는 경우가
대부분이므로 분석의 일관성을 유지하기 위해
요구된다. 상세한 개별 정보가 주어진 자료는
'1단계'에서 처리하여 그 결과를 '2단계'에서 하
나의 문헌 자료로 사용할 수 있다.

해당 발전소의 희소한 고장 자료를 보완하기
위해 유사한 발전소의 고장 자료를 활용하는
체계적인 분석 절차에 대한 첫 연구는 Kap-
lan[10]에 의하여 제안되었다. 그는 유사 발전
소에서의 가동 중 고장 자료를 이용하여 고장
률의 PVD를 추정하고, 이를 다음 단계의 사전

2. 일반(generic) 자료분석절차

유사 발전소의 자료(E_j), 문헌 자료(E_2), 그
리고 발전소 고유 자료(E_3)로 구성된 자료의
처리를 위해 [1]에서 제안한 '3단계 베이저안'
분석 절차는 다음과 같이 요약할 수 있다.

[1단계] 공학적 지식 E_p 과 주어진 발전소 자료

분포로 하여 해당 발전소에서의 사후 분포를 추정하는 '2단계 베이지안' 방법을 제안하였다. 고장률 λ 들의 사전 분포 밀도 함수로서 초모수 μ, σ 를 갖는 대수정규 분포를, 고장 개수의 조건부 pdf로서 포아송분포를 이용하였다. 이후 그는 기동 실패 고장에 대해서도 이 방법을 적용하였다[11]. 이 연구의 가장 큰 특징은 초모수에 대하여 이산화 확률 분포(DPD)를 사용하여 수리적으로 적분이 불가능한 모형을 처리한 점이다. 초모수 (μ, σ)에 대한 2단계 사전 분포를 사용한 점에서 계층적(hierarchical) 베이지안 개념이 사용되었고, 자료로부터 모수의 변이 분포를 추정한 점에서 선형적(empirical) 베이지안 개념도 사용되었다. 그러나 Kaplan [10,11]의 방법은 개별 발전소 자료만을 다룬다는 점에서 제한적이다.

그 후 Mosleh와 Apostolakis[14]는 발전소 자료가 고장모수의 추정치로 주어진 경우에 대하여 Kaplan[10,11]의 절차를 적용하였다. 고장모수 추정치들의 조건부 결합 분포는 대수정규분포로 가정하였다. 추정치의 신빙성을 반영하기 위하여 추정 오차의 사용이 제안되었으나 이 값을 정하는 체계적인 절차가 제시되어 있지 않아 상당한 주관에 개입될 수 있다. 후에 Mosleh[15]는 이 모형을 확장하여 전문가들이 제시한 추정치를 이용하여 전문가간의 변이 분포를 추정하는 베이지안 절차를 제안하였다.

이러한 DPD 접근법은 초모수의 공간을 인위적으로 분할하여 각 구간에 사전 확률을 할당하여야 하는 문제가 있으나, 자료의 수가 충분하지 않은 경우 선형적 사전 분포를 추정할 수 있는 유용한 방법으로 인식되어 여러 PRA 연구에 응용된 바 있다[6,13,17].

이후 Fröner[5]는 감마분포를 사전분포로 하여 발전소 자료(운전 중 고장)를 손쉽게 처리

하여 PVD를 추정하는 방법을 제안하였다. 운전 중 고장 개수의 조건부 분포는 포아송분포이므로 감마분포를 사전분포로 하여 베이지안 처리를 하면 역시 감마 사후 분포를 얻게 된다. 이러한 N개의 pdf를 평균하여 1단계 사후 분포를 구하는 방법이다. 해당 발전소 자료를 처리하는 최종 단계에서도 수치 적분이 필요 없이 사후분포를 계산할 수 있다. 전체적인 구성은 역시 '단계1'과 '단계3'으로 구성된 '이단계 베이지안' 형태이며, 각 단계별 수행 내용도 Kaplan[10,11]의 절차와 유사하다. 이 방법은 1단계 사전분포의 모수를 계산하는 과정이 복잡하고, 개별 pdf들의 단순 평균을 취하여 변이 분포를 추정한 점에서 논리적 설득력이 부족하다.

Vaurio[22]도 역시 감마 사전분포를 이용한 '이단계 베이지안' 방법을 제안하였다. 선형적 베이스 추정 이론에 근거하여 통계적 신뢰구간에 대하여 일관성을 갖는 사전분포를 추정하였다. 그의 방법은 DPD 절차보다 계산이 용이하고 이론적으로 DPD 방법의 약점을 극복한 장점이 있다. 그러나 운전 중 고장에 해당하는 발전소 자료만을 대상으로 하며, 기동 실패 고장 유형에는 적용할 수 없다는 점에서 한계가 있다.

이상의 연구들은 개별 자료의 처리에만 집중되어 있고 문헌 추정치와 같이 요약된 자료의 활용 방법은 제시하지 않고 있다. 이전의 논문 [1,12]에서 기존의 방법론을 보완하고 문헌 자료의 처리를 위해 한 단계를 추가한 '3단계 베이지안' 절차를 이미 제시한 바 있으나, 자료의 종속성과 같은 특성을 반영하는 체계적인 방법론을 제안하지 못하였다. 따라서 본 논문에서는 보다 근본적으로 '단계2'의 처리 방법을 접근하고자 한다.

3. 모집단 변이분포(PVD)의 추정

본 절에서는 문헌 자료간의 종속성을 고려한 PVD의 추정 방법을 제안한다. 문헌 자료들 간의 종속성은 자료 분석자들이 중복되에도 기인한다고 볼 수 있으나, 보다 근본적이고 정량화가 가능한 원인은 각 문헌에서 반영한 발전소 자료들의 중복에 있다. 따라서 종속성을 추정하는데 요구되는 최소한의 자료는 각 문헌에서 처리한 개별 발전소 자료의 수 N_i ($i=1, \dots, B$)와 두 문헌에서 공통적으로 사용한 발전소 자료의 수 N_{ij} 가 된다. PVD를 모수적 분포로 제한하면 모수 추정 문제로 생각할 수 있다. 주어진 모수에 대한 문헌 자료의 조건부 결합 분포가 유도된다면 베이즈 정리를 이용하여 이 문제를 해결할 수 있으나, 문헌 추정치는 발전소 자료들로 구성된 복잡한 통계량이므로 이들의 결합분포가 확정적 형태를 갖는 경우는 드물고 추정치 자체가 확정적 형태(closed form)로 표현되지 않는 모형도 다수 존재한다. 본 연구에서는 확정적 형태의 추정량을 갖게 되는 경우로서 대수정규 모형에 대하여 유형 2 최대우도(ML-II) 방법에 기초하여 문헌 자료를 처리하는 방법을 제안한다.

ML-II 추정 방법은 모수적 선형적 베이지안(Parametric Empirical Bayesian, PEB) 방법의 일종으로서 모수의 사전 pdf π 가 주어졌을 때 자료 x 의 조건부 pdf $m(x|\pi)$ 를 에 대한 우도 함수로 하여 이를 최대화하는 사전 pdf를 구하는 방법이다. 사전 분포 π 는 초모수를 갖는 모수적 분포로 주어지므로 조건부 pdf를 최대화하는 초모수의 식을 구하는 것이다.

고장 자료는 조건부 독립적이고 고장 모수는 '발전소 변이 분포'라는 공통된 분포를 따르므

로 PEB 방법의 적용이 가능하다. PEB는 공통의 분포를 모수적 분포로 모형화하여 분석하는 방법으로서 초모수의 추정량에 대한 수리적 표현이 가능한 경우에는 문헌 자료와 같은 통계량의 처리에 적합하다. 여기서는 기존 연구에서 고려된 세 가지 자료 모형 중 수리적 해를 구할 수 있는 대수정규 모형에 적용한다.

대수정규 모형은 Mosleh 등[14]이 제안한 모형으로서 개별 발전소의 고장률 추정치가 주어진 경우 다음과 같이 근사적으로 대수정규 분포를 가정한다.

$$\hat{\lambda}_m | \lambda_m, \sigma_m \stackrel{ind}{\sim} LN(\lambda_m, \sigma_m).$$

개별 발전소의 관측치 $\hat{\lambda}_m$ 을 $y_m = \log(\hat{\lambda}_m)$ 으로 대수 변환하면 이들의 조건부 분포는 다음과 같이 독립적으로 정규분포를 따른다.

$$y_m | \theta_m, \sigma_m \stackrel{ind}{\sim} N(\theta_m, \sigma_m^2). \tag{3. 1}$$

위의 식에서 $\theta_m = \log \lambda_m$ 이며 σ_m 은 추정오차를 나타낸다.

이 모형에서는 PVD 역시 공액족인 대수정규분포로 가정되며, 따라서 변환된 모수 θ_m 의 분포는 아래와 같다.

$$\theta_m \stackrel{ind}{\sim} N(\xi, \sigma^2), \quad (m=1, \dots, N). \tag{3. 2}$$

베이즈 정리에 의하여 발전소 자료들의 조건부 결합 분포는 식 (3. 3)과 같이 계산되어 이들은 독립적으로 정규분포를 따름을 알 수 있다. 즉, $y_m \stackrel{ind}{\sim} N(\xi, \sigma^2 + \sigma_m^2)$.

$$m(\vec{y} | \xi, \sigma) = \prod_{m=1}^N \left\{ [2\pi(\sigma^2 + \sigma_m^2)]^{-1/2} \times \exp \left[-\frac{(y_m - \xi)^2}{2(\sigma^2 + \sigma_m^2)} \right] \right\} \tag{3. 3}$$

위의 식을 최대화하는 ML-II 추정치 $\hat{\xi}$ 와 $\hat{\sigma}^2$ 는 식(3.4)의 해로서 구해진다.

$$\begin{cases} \sum_{m=1}^N \frac{(y_m - \hat{\xi})}{\hat{\sigma}^2 + \sigma_m^2} = 0 \\ \sum_{m=1}^N \left[\frac{-1}{\hat{\sigma}^2 + \sigma_m^2} + \frac{(y_m - \hat{\xi})^2}{(\hat{\sigma}^2 + \sigma_m^2)^2} \right] = 0 \end{cases} \quad (3.4)$$

위의 등식에 대한 일반적인 수리 해는 존재하지 않으나, 추정 오차 $\sigma_m^2 = \bar{\sigma}^2$ 로서 모두 동일하다고 가정하면 아래의 해를 얻을 수 있다.

$$\begin{cases} \hat{\xi} = \bar{y} \equiv \frac{1}{N} \sum_{m=1}^N y_m \\ \hat{\sigma}^2 = \max\{0, s^2 - \bar{\sigma}^2\}, s^2 = \frac{1}{N} \sum_{m=1}^M (y_m - \bar{y})^2 \end{cases} \quad (3.5)$$

이와 유사한 과정을 통하여 부분적 자료 R를 사용한 문헌i에서의 ML-II 추정치는 다음과 같이 표현할 수 있다.

$$\begin{cases} \hat{\xi}_i = \bar{y}_i \equiv \frac{1}{N_i} \sum_{m \in R_i} y_m \\ \hat{\sigma}_i^2 = \max\{0, s_i^2 - \bar{\sigma}^2\}, s_i^2 = \frac{1}{N_i} \sum_{m \in R_i} (y_m - \bar{y}_i)^2 \end{cases} \quad (3.6)$$

이제까지 문헌 추정치와 모든 자료를 고려한 추정치와의 관계식을 얻기 위해서 사용된 가정을 정리하면 다음과 같다.

[가정 1] PVD는 대수정규 모형 하에서 ML-II 방법으로 추정한다.

[가정 2] 개별 발전소 자료 y_m 의 추정 오차는 동일한 산포를 갖는다. 즉, $\sigma_m^2 = \bar{\sigma}^2$.

[가정 3] 각 문헌에서 산포 모수의 ML-II 추정치는 양의 값을 갖는다.

먼저 모든 문헌들이 독립적인 경우 문헌 추정치로부터 PVD 모수를 추정하는 관계식은 다음과 같이 정리할 수 있다.

[보조정리 1] 가정 1, 2, 3 하에서 각 문헌에서 고려된 발전소 자료들의 중복이 없는 경우, 부분 자료를 고려한 문헌 추정치 $(\hat{\xi}_i, \hat{\sigma}_i^2); i, \dots, B$ 와 전체 자료를 이용한 ML-II 추정치는 다음의 관계를 갖는다.

$$\hat{\xi} = \frac{1}{N} \sum_{i=1}^B (N_i \hat{\xi}_i) \quad (3.7)$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^B \{N_i (\hat{\sigma}_i^2 + (\hat{\xi} - \hat{\xi}_i)^2)\} \quad (3.8)$$

[증명] 전체 발전소의 집합을 R이라 하면 $R = \cup_{i=1}^B R_i, R_i \cap R_j = \emptyset$ 의 관계가 성립한다. 따라서 (3.7) 식은 아래의 관계를 통하여 자명하다.

$$\begin{aligned} \hat{\xi} = \bar{y} &= \frac{1}{N} \sum_{m=1}^N y_m \\ &= \frac{1}{N} \sum_{i=1}^B \sum_{m \in R_i} y_m = \frac{1}{N} \sum_{i=1}^B (N_i \hat{\xi}_i) \end{aligned}$$

가정 3에 의하여 $\hat{\sigma}_i^2 > 0 (\forall i=1, \dots, B)$ 이므로, $\hat{\sigma}_i^2 = s_i^2 - \bar{\sigma}^2$ 이다. 여기서

$$\begin{aligned} s^2 &= \frac{1}{N} \sum_{m=1}^N (y_m - \bar{y})^2 \\ &= \frac{1}{N} \sum_i \sum_{m \in R_i} [(y_m - \bar{y}_i)^2 + (\bar{y}_i - \bar{y})^2] \\ &= \frac{1}{N} \left[\sum_i N_i (\hat{\sigma}_i^2 + \bar{\sigma}^2) + \sum_i N_i (\bar{y}_i - \bar{y})^2 \right] \end{aligned}$$

따라서

$$\hat{\sigma}^2 = s^2 - \bar{\sigma}^2 = \frac{1}{N} \sum_i N_i \hat{\sigma}_i^2 + \frac{1}{N} \sum_i N_i (\bar{y}_i - \bar{y})^2$$

이 되므로 (3.8) 식을 얻는다. ■

위의 결과에서 식 (3.8)에 의하면 개별자료에 대한 추정 오차의 산포를 모르는 경우에도 모수 추정이 가능하다. 즉, 오차의 산포는 모두 소거되어 관계식에 나타나지 않는다. 문헌 추정치들이 독립적인 경우에는 식 (3.7), (3.8)

과 같은 정확한 계산이 가능하지만, 종속성이 존재하는 경우에는 정보의 손실로 인하여 정확한 관계식을 구할 수 없다. 식 (3. 7)과 (3. 8)을 살펴보면 $\hat{\xi}$ 과 $\hat{\sigma}^2$ 은 각각 $\hat{\xi}_i$ 과 $\hat{\sigma}_i^2 + (\hat{\xi}_i - \hat{\xi})^2$ 들의 선형 조합 형태를 가짐을 알 수 있다. 각각 계수는 표본의 크기 N_i 에 비례하고 이는 추정량의 분산을 최소화하는 역할을 한다. 이에 착안하여 선형 조합 형태의 추정량 중 최소의 분산을 갖는 추정량을 제안하고 그 특성을 다음과 같이 정리한다.

[보조정리 2] 가정 1, 2, 3 하에서 각 문헌에서 고려된 발전소 자료들이 중복되는 경우, 문헌 추정치들을 이용한 위치모수의 최소분산 선형 불편추정량(Minimum variance linear unbiased estimator)은 식 (3. 9)로 계산되며, 추정량의 분산은 식 (3. 10)과 같다.

$$\hat{\xi}_L = (\mathbf{1}^T \Sigma_0^{-1} \hat{\xi}) / (\mathbf{1}^T \Sigma_0^{-1} \mathbf{1}), \quad (3. 9)$$

여기서 $\mathbf{1} \equiv (\mathbf{1}, \dots, \mathbf{1})^T$, $\hat{\xi} = (\hat{\xi}_1, \hat{\xi}_2, \dots, \hat{\xi}_B)^T$,

$$\Sigma_0 = \begin{pmatrix} 1/N_1 & N_{12}/(N_1N_2) & \dots & N_{1B}/(N_1N_B) \\ N_{12}/(N_1N_2) & 1/N_2 & \dots & N_{2B}/(N_2N_B) \\ \dots & \dots & \dots & \dots \\ N_{1B}/(N_1N_B) & N_{2B}/(N_2N_B) & \dots & 1/N_B \end{pmatrix}$$

$$\text{Var}(\hat{\xi}_L) = (\sigma^2 + \bar{\sigma}^2) / (\mathbf{1}^T \Sigma_0^{-1} \mathbf{1}) \quad (3.10)$$

단, $N_i = N_j = N_{ij}$ 인 문헌 i 와 문헌 j 가 존재하면 Σ_0 은 역행렬이 존재하지 않으며 이때는 두 문헌간에 완전한 종속 관계가 있으므로 한 문헌의 자료를 제거하여 처리한다.

[증명]

식 (3. 6)으로부터 문헌 모수 추정량들의 분산과 공분산은 다음과 같이 계산된다.

$$\text{Var}(\hat{\xi}_i) = (\sigma^2 + \bar{\sigma}^2) / N_i,$$

$$\text{Cov}(\hat{\xi}_i, \hat{\xi}_j) = (\sigma^2 + \bar{\sigma}^2) N_{ij} / (N_i N_j), \quad \forall i \neq j$$

따라서 $\text{Var}(\hat{\xi}) = (\sigma^2 + \bar{\sigma}^2) \Sigma_0$ 로 표현된다. 위치모수의 선형 추정량을 $\hat{\xi}_L = \sum_{i=1}^B p_i \hat{\xi}_i$ 라하면 불편 추정량이 되기 위한 조건은 $\sum_{i=1}^B p_i = 1$ 이다. 따라서 아래의 비선형 문제를 풀면 최소분산불편 추정량을 구할 수 있고 그 해는 식 (3. 9)와 같다.

$$\text{Minimize } \text{Var}(\sum_{i=1}^B p_i \hat{\xi}_i) \text{ sub.to } \sum_{i=1}^B p_i = 1$$

단, $N_i = N_j = N_{ij}$ 인 문헌 i 와 문헌 j 가 존재하면 Σ_0 의 i 열과 j 열이 같아지므로 역행렬이 존재하지 않는다. ■

위의 정리에서 식 (3. 9)는 다음과 같은 선형 모형에서의 일반화된 최소 제곱 추정치 (Generalized Least Square Estimate, GLSE)의 식과 동일하다. 따라서 Gauss-Markov 정리에 의해서도 최량 선형 불편추정량(Best linear unbiased estimator, BLUE)임을 알 수 있다.

$$\hat{\xi} = \xi \mathbf{1} + u, \quad u \sim (0, \Sigma_0)$$

다음으로 산포 모수의 추정량 $\hat{\sigma}^2$ 을 $\hat{\sigma}_i^2 + (\bar{y}_i - \bar{y})^2$ 들의 선형 조합 형태로 제한하면 다음과 같은 정리를 얻는다.

[정리 1] 가정 1, 2, 3 하에서 각 문헌에서 고려된 발전소 자료들이 중복되는 경우, 산포 모수에 대한 $w_i = \hat{\sigma}_i^2 + (\bar{y}_i - \bar{y})^2$ 들의 선형 결합 추정량 중 최소의 분산을 갖는 추정량은 식 (3. 11)과 같이 구해진다. 또한 이 추정량의 분산은 식 (3. 12)와 같다.

$$\hat{\sigma}_L^2 = (\mathbf{1}^T \Sigma_0^{-1} \mathbf{w}) / (\mathbf{1}^T \Sigma_0^{-1} \mathbf{1}), \quad (3.11)$$

여기서 $\mathbf{1} \equiv (1, \dots, 1)$, $\mathbf{w} = (w_1, w_2, \dots, w_B)^T$.

$$\begin{aligned} \text{Var}(\hat{\sigma}_L^2) &= 2(\sigma^2 + \bar{\sigma}^2)^2 \\ &\times \{(1 - 2/N)/(\mathbf{1}^T \Sigma_0^{-1} \mathbf{1}) + 1/N^2\} \end{aligned} \quad (3.12)$$

[증명]

다음과 같은 선형 추정량을 고려하자.

$$\hat{\sigma}_L^2 = \sum_{i=1}^B q_i \{ \hat{\alpha}_i^2 + (\bar{y}_i - \bar{y})^2 \}, \quad \sum_{i=1}^B q_i = 1.$$

여기서 $s_L^2 = \sigma_L^2 + \bar{\sigma}^2$ 로 정의하면 아래의 식을 얻는다.

$$\begin{aligned} s_L^2 &= \sum_{i=1}^B q_i \{ s_i^2 + (\bar{y}_i - \bar{y})^2 \} \\ &= \sum_{i=1}^B \{ (q_i/N_i) \sum_{m \in R_i} (y_m - \bar{y})^2 \} \end{aligned}$$

따라서 위 식의 분산을 최소화하면 $\hat{\alpha}_i^2$ 의 분산도 최소화되므로 아래와 같은 비선형 문제의 해를 구하면 된다.

$$\text{Minimize } \text{Var}(\sum_{i=1}^B q_i v_i) \text{ sub. to } \sum_{i=1}^B q_i = 1,$$

$$\text{여기서 } v_i = (1/N_i) \sum_{m \in R_i} (y_m - \bar{y})^2$$

다음의 과정을 통하여 $s_L^2 \equiv \sum_{i=1}^B q_i v_i$ 의 분산을 구한다.

먼저

$$\text{Var}(y_m - \bar{y})^2 = 2(\sigma^2 + \bar{\sigma}^2)^2 (N-1)^2 / N^2,$$

$$\text{Cov}((y_m - \bar{y})^2, (y_k - \bar{y})^2) = 2(\sigma^2 + \bar{\sigma}^2)^2 / N^2,$$

$$\forall m \neq k.$$

다음으로

$$\text{Var}(\sum_{m \in R_i} (y_m - \bar{y})^2) = 2(\sigma^2 + \bar{\sigma}^2)^2$$

$$\times N_i (1 - 2/N + N_i/N^2),$$

$$\text{Cov}(\sum_{m \in R_i} (y_m - \bar{y})^2, \sum_{m \in R_j} (y_m - \bar{y})^2)$$

$$= 2(\sigma^2 + \bar{\sigma}^2)^2 \{(1 - 2/N)N_i + N_i N_j / N^2\}.$$

$\mathbf{v} = (v_1, v_2, \dots, v_B)^T$ 라 하면

$$\text{Var}(\mathbf{v}) = 2(\sigma^2 + \bar{\sigma}^2)^2 \Sigma_v, \text{ 여기서}$$

$\Sigma_v = (1 - 2/N)\Sigma_0 + (\mathbf{1} \cdot \mathbf{1}^T)/N^2$ 으로 계산된다.

따라서

$$\text{Var}(s_L^2) = \text{Var}(\mathbf{q}^T \mathbf{v})$$

$$= 2(\sigma^2 + \bar{\sigma}^2) \{ (1 - 2/N) \mathbf{q}^T \Sigma_0 \mathbf{q} + 1/N^2 \}$$

따라서 이 식을 최소화하는 계수 벡터는 보조 정리 2의 계수 벡터와 동일하므로 식 (3.11)을 얻는다. 이를 위의 식에 대입하면 이때의 분산은 식 (3.12)와 같다. ■

앞의 정리를 적용할 때 $\hat{\xi}_L = \bar{y}_i$ 이지만, 일반적으로 $\hat{\xi}_L$ 은 \bar{y} 와 같지 않으므로 오차 제곱의 추정량 w_i 에 \bar{y} 대신 $\hat{\xi}_L$ 을 사용하면 최소 분산을 가짐을 보장할 수는 없다. 그러나 다음과 같이 문헌간에 독립성이나 포함 관계가 성립하는 경우에는 $\hat{\xi}_L = \bar{y}$ 이 성립하여 최소 분산이 보장된다.

[정리 2] 식 (3.13)의 조건이 성립하면 $\hat{\xi}_L = \bar{y}$ 이다. 단, $N_i = N_j = N_{ij}$ 인 문헌 i 와 문헌 j 가 존재하면 Σ_v 의 역행렬이 존재하지 않으며 이때는 두 문헌간에 완전한 종속 관계가 있으므로 한 문헌의 자료를 제거하여 Σ_0 를 재구성한다.

$$R_i \cap R_j = \emptyset \text{ or } R_i \cap R_j = R_i, \quad \forall i \neq j \text{ s.t. } N_i \leq N_j \quad (3.13)$$

[증명]

조건이 만족되면 임의의 두 문헌간에 독립 관계 혹은 포함 관계가 성립한다. 따라서 일반성을 잃지 않고 문헌의 순서를 재배열하여 Σ_0 행렬을 다음과 같이 구성할 수 있다.

$$\Sigma_c = \begin{pmatrix} \Sigma_1 & 0 & \cdots & 0 \\ 0 & \Sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \Sigma_k \end{pmatrix},$$

여기서 K 는 독립적 문헌 그룹의 수를 나타내며 $1 \leq K \leq B$.

그룹 k 에 속한 문헌의 수를 b_k 라하고 k 그룹까지의 누적 수를 $b(k)$ 라 하면,

$$\sum_{k=1}^K b_k = B, \quad b(0) = 0, \quad b(k) = \sum_{r=1}^k b_r, \quad (1 \leq k \leq K-1).$$

역시 일반성을 잃지 않고 각 그룹에 속한 문헌을 크기 순서로 재배열하여 Σ_k 를 다음과 같이 표현할 수 있다. 단, $N_{b(k-1)+1} < N_{b(k-1)+2} < \cdots < N_{b(k)}$. (등호가 성립하는 두 문헌은 동일한 것이므로 하나를 제거함.)

$$\Sigma_k = \begin{pmatrix} 1/N_{b(k-1)+1} & 1/N_{b(k-1)+2} & \cdots & 1/N_{b(k)} \\ 1/N_{b(k-1)+2} & 1/N_{b(k-1)+3} & \cdots & 1/N_{b(k)} \\ \vdots & \vdots & \ddots & \vdots \\ 1/N_{b(k)} & 1/N_{b(k)} & \cdots & 1/N_{b(k)} \end{pmatrix}$$

Σ_0 의 $b(k-1)+i$ 열을 단위벡터 1로 치환한 행렬을 $\Sigma_c^{(b(k-1)+i)}$ 라하고, Σ_k 의 i 열을 단위벡터로 치환한 행렬을 $\Sigma_k^{(i)}$ 라 하면 다음과 같은 관계를 얻는다.

$$|\Sigma_0| = |\Sigma_1| \cdots |\Sigma_k|,$$

$$|\Sigma_0^{(b(k-1)+i)}| = |\Sigma_1| \cdots |\Sigma_k^{(i)}| \cdots |\Sigma_k|,$$

Cramer의 법칙을 이용하여 $b(k-1)+i$ 번째 가중 계수를 구하면

$$D_{b(k-1)+i} = (|\Sigma_0^{(b(k-1)+i)}| / |\Sigma_0|) (1^T \Sigma_0^{-1} 1)^{-1}$$

$$= (|\Sigma_k^{(i)}| / |\Sigma_k|) (1^T \Sigma_k^{-1} 1)^{-1}.$$

여기서 $\Sigma_k^{(i)}$ 의 구조를 살펴보면 마지막 열은 동일한 원소 $(1/N_{b(k)})$ 로 구성되어 있다. $\Sigma_k^{(i)}$ 는 Σ_k 의 i 열을 벡터 1로 대치한 행렬이므로 다음의 식을 얻는다.

$$|\Sigma_k^{(i)}| = \begin{cases} 0 & 1 \leq i < b_k, \\ N_{b(k)} / |\Sigma_k|, & i = b_k \end{cases}$$

따라서

$$D_{b(k-1)+i} = \begin{cases} 0, & 1 \leq i < b_k, \\ N_{b(k)} / (1^T \Sigma_0^{-1} 1) \\ = N_{b(k)} / \sum_{k=1}^K N_{b(k)}, & i = b_k \end{cases}$$

$$\hat{\xi}_i = \sum_{k=1}^K \sum_{i=1}^{b_k} D_{b(k-1)+i} \hat{\xi}_{b(k-1)+i}$$

$$= \sum_{k=1}^K D_{b(k)} \hat{\xi}_{b(k)}$$

$$= \sum_{k=1}^K N_{b(k)} \hat{\xi}_{b(k)} / \sum_{k=1}^K N_{b(k)}$$

위의 식은 독립적 문헌 그룹 내에서 타 문헌을 포함하는 대표적 문헌 추정치들을 식 (3.7)과 같이 선형 조합한 것이므로 보조정리 1에 의하여 전체 평균과 같다. ■

4. 예제

4.1 단일 중복의 영향 (동일한 크기의 표본)

각각 20개씩의 표본을 이용하여 추정된 세 개의 문헌 자료가 주어지고 두 문헌간에만 자료의 중복이 있는 경우 그 영향을 살펴본다. 위치모수 추정치는 편의상 $(\hat{\xi}_1, \hat{\xi}_2, \hat{\xi}_3) = (1, 2, 3)$ 으로 가정하고 산포 모수 추정치는 다음과 같이 세 가지 경우를 고려하였다.

[경우 1] $(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_3^2) = (1, 2, 3)$

[경우 2] $(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_3^2) = (2, 2, 2)$

[경우 3] $(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_3^2) = (3, 2, 1)$

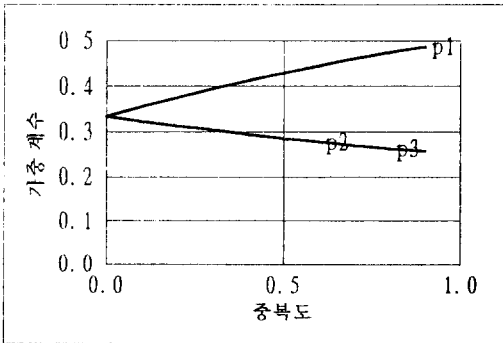
이에 대하여 계산한 결과는 <표 1>과 같다.

표에서 유효 표본 크기는 $N_e = 1 / (1^T \Sigma_0^{-1} 1)$ 로서 추정치의 정밀도(precision)를 나타내며 자료의 중복이 증가함에 따라 단조감소함을 알 수 있다. $N_2 = N_3 = N_{2,3}$ 인 경우에는 Σ_0 의 역

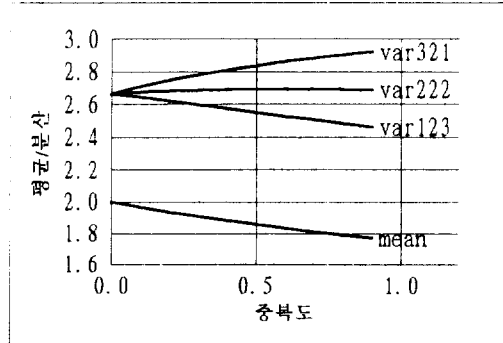
행렬이 존재하지 않으나, 문헌2와 3이 동일한 것으로 간주되므로 유효 표본 크기는 40이 된다. [그림 1]을 보면 독립적인 문헌1에 대한 가중 계수(p_1)는 중복의 증가에 따라 증가하고 종속적인 문헌2와 3의 가중 계수($p_2=p_3$)는 중복의 증가에 따라 감소한다. 따라서 [그림 2]에서 보는 바와 같이 각 모수의 추정치는 중복의 증가에 따라 상대적 독립성이 커지는 문헌1의 추정치 쪽으로 접근함을 알 수 있다.

<표 1> 단일 중복의 영향 ($N_1 = N_2 = N_3 = 20, N_{12} = N_{13} = 0, N_{23} = 2r$)

r	N_e	p_1	$p_2=p_3$	$\hat{\xi}$ (1,2,3)	$\hat{\sigma}_1^2$ (1,2,3)	$\hat{\sigma}_2^2$ (2,2,2)	$\hat{\sigma}_3^2$ (3,2,1)
0	60.00	.3333	.3333	2.00	2.6667	2.6667	2.6667
1	56.36	.3548	.3226	1.97	2.6441	2.6764	2.7086
2	53.33	.3750	.3125	1.94	2.6211	2.6836	2.7461
3	50.77	.3939	.3030	1.91	2.5978	2.6887	2.7796
4	48.57	.4118	.2941	1.88	2.5744	2.6920	2.8097
5	46.66	.4286	.2857	1.86	2.5510	2.6939	2.8367
6	45.00	.4444	.2778	1.83	2.5278	2.6944	2.8611
7	43.53	.4595	.2703	1.81	2.5048	2.6939	2.8831
8	42.22	.4737	.2632	1.79	2.4820	2.6925	2.9031
9	41.05	.4872	.2564	1.77	2.4596	2.6903	2.9211
10	40.00	-	-	-	-	-	-



[그림 1] 단일중복에 의한 가중 계수의 변화



[그림 2] 단일중복에 의한 추정치의 변화

4.2 다중 중복의 영향 (상이한 크기의 표본)

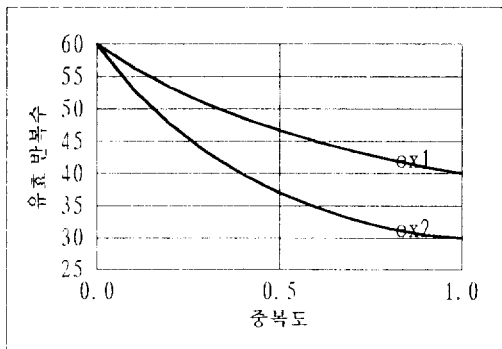
각각 10, 20, 30개씩의 표본을 이용하여 추정 한 세개의 문헌 자료가 주어지고 세 문헌간에 자료의 중복이 있는 경우 그 영향을 살펴본다. 문헌1과 타 문헌과의 중복자료 수는 0에서 10 까지, 문헌2와 문헌3과의 중복자료 수는 0에서 20까지 변화시켜 보았다. 앞에서와 마찬가지로 위치모수 추정치는 편의상 $(\hat{\xi}_1, \hat{\xi}_2, \hat{\xi}_3) = (1, 2, 3)$ 으로 가정하고 산포모수 추정치는 (1, 2, 3), (2, 2, 2), (3, 2, 1) 세 가지 경우를 고려하였다. 이에 대하여 계산한 결과는 <표 2>와 같다.

이 경우에도 유효 표본 크기는 자료의 중복

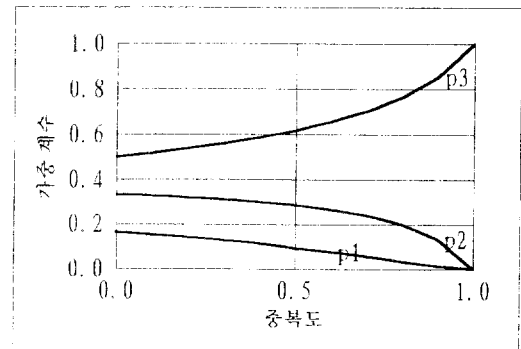
이 증가함에 따라 $60(=N_1+N_2+N_3)$ 에서 30 $(=N_3)$ 까지 단조 감소하고 있다. [그림 3]을 보면 중복성의 증가에 따라 유효 반복수는 대수적으로 감소함을 알 수 있다. [그림 4]를 보면 중복의 증가에 따른 독립적 자료의 감소로 인하여 문헌1과 문헌2의 가중 계수(p_1, p_2)는 단조 감소하여 0으로 수렴하는 반면 문헌3의 가중 계수(p_3)는 1로 수렴한다. 즉, $N_{12} = N_{13} = 10, N_{23} = 20$ 인 경우에는 완전한 종속관계가 성립하여 모든 정보는 문헌3에 귀속되므로 이때의 가중 계수는 (0,0,1)인 것이다. 따라서 [그림 5]에 나타난 바와 같이 PVD 모수의 추정치도 문헌3의 추정값으로 수렴하는 특성을 보인다.

<표 2> 다중 중복의 영향 ((N_1, N_2, N_3)=(10,20,30), $N_{12} = N_{13} = r$, $N_{23} = 2r$)

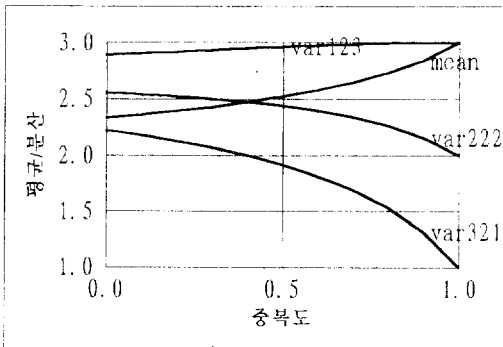
r	N_e	p_1	p_2	p_3	$\hat{\xi}_{1.}$ (1,2,3)	$\hat{\sigma}_{L^2}$ (1,2,3)	$\hat{\sigma}_{L^2}$ (2,2,2)	$\hat{\sigma}_{L^2}$ (3,2,1)
0	60.00	.1667	.3333	.5000	2.3333	2.8889	2.5556	2.2222
1	53.01	.1550	.3273	.5177	2.3627	2.9039	2.5412	2.1785
2	47.60	.1422	.3200	.5378	2.3956	2.9191	2.5235	2.1280
3	43.32	.1281	.3111	.5608	2.4327	2.9344	2.5017	2.0690
4	39.87	.1125	.3000	.5875	2.4750	2.9494	2.4744	1.9994
5	37.06	.0952	.2857	.6190	2.5238	2.9637	2.4399	1.9161
6	34.77	.0762	.2667	.6571	2.5810	2.9768	2.3958	1.8149
7	32.92	.0553	.2400	.7046	2.6492	2.9877	2.3385	1.6893
8	31.47	.0333	.2000	.7667	2.7333	2.9956	2.2622	1.5289
9	30.44	.0121	.1333	.8545	2.8424	2.9994	2.1570	1.3146
10	30.00	.0000	.0000	1.000	3.0000	3.0000	2.0000	1.0000



[그림 3] 중복에 의한 유효 반복수의 변화



[그림 4] 다중 중복에 의한 가중 계수의 변화



[그림 5] 다중 중복에 의한 추정치의 변화

4.3 모터 구동 밸브(MOV, Motor Operated Valve) 기동 고장

모터 구동 밸브의 기동 고장 확률에 대한 문헌 자료를 수집하여 <표 3>에 정리하였다. 각

문헌 자료에서 사용한 표본의 수와 중복된 표본의 수는 정확히 파악되지는 않으나, 일 예로 표와 같이 가정하였다. 문헌 자료는 평균 추정치와 error factor(EF) 추정치로서 대수정규모형 하에서 다음과 같이 변환된다.

$$\hat{\sigma}_i = \log(EF_i / z_{0.95})$$

$$\hat{\xi}_i = \log(\text{mean}_i) - \hat{\sigma}_i^2 / 2$$

중간 결과로서 독립성 가정 하에서의 계수 $p_i(D)$ 와 중속성 가정 하에서 계산된 계수 $p_i(D)$ 를 첨부하였고, 계산 결과를 표 하단에 실었다. 자료의 중속성을 가정한 경우 유효 표본크기는 230에서 108.4로 크게 줄었으나, 자료들이 비교적 균일한 양상을 보임으로 인하여 PVD의 모수 추정치에는 큰 변화가 없었다.

<표 3>

모터 구동 밸브 기동 고장 자료

자료원	출판년월	평균 (10 ⁻³ /rv)	EF	$\hat{\xi}_i$	$\hat{\sigma}_i^2$	N_i	$N_j(i < j)$	$p_i(D)$	$p_i(D)$
WASH-1400	75.10	1.25	3.0	-6.91	0.45	10	-5,5,5,1, 0, 5, 5, 0	0.0435	0.0022
German RS	79	3.6	7.0	-6.33	1.0	12	-5,5,5,2, 0, 5, 5, 0	0.0522	0.0254
Sizewell PSA	82.06	1.4	10.0	-7.55	1.6	12	-5,5,5,2, 0, 5, 5, 0	0.0522	0.0254
NC-1363	82.10	4.2	10.0	-6.45	1.6	30	-5,5,5,2, 5,10,15,0	0.1304	0.0962
Seabrook PSS	83.12	4.3	3.7	-5.77	0.3	10	-5,5,5,5, 2, 2, 2, 0	0.0435	0.0467
IEEE 500	84.01	4.0	10.0	-6.50	1.6	40	-5,5,5,5, 5, 10, 15, 0	0.1739	0.2225
Oconee PRA	84.06	4.0	10.0	-5.94	1.6	50	-5,5,5,5, 5, 5, 20, 1	0.2174	0.2346
NC-4550	87.12	3.0	10.0	-6.79	1.6	60	-5,5,5,5, 5, 5, 5, 0	0.2609	0.2963
6 Plants		3.8	1.9	-5.65	0.5	6		0.0261	0.0506
독립적 추정		4.2	10.0	-6.46	1.96	230			
중속적 추정		4.5	10.0	-6.40	1.97	108.4			

4.4 솔레노이드 구동 밸브(Solenoid Operated Valve) 기동 고장

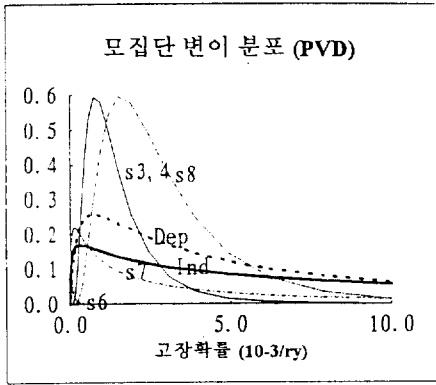
솔레노이드 구동 밸브의 기동 고장 확률에 대한 문헌 자료를 수집하여 <표 4>에 정리하였다. 앞의 예와의 큰 차이점은 분포의 위치모수 및 산포 모수 추정치가 자료에 따라 큰 차이를 보인다는 점이다.

표와 같이 종속성을 높게 가정한 결과 유효 표본크기는 270에서 77.6으로 크게 줄었다. 특히 NC-2728 자료[4]는 NREP DB[18]를 그대로 인용한 것으로 판명되어 분석에서 제외하였다. 종속적 추정시의 가중 계수 $p_i(D)$ 중에 음수가 나오는 경우도 있는데, 그 의미는 이 자료가 여러 번 반영되었으므로 과도하게 중복된

정보를 제외시키는 것으로 해석할 수 있다. [그림 6]에는 가중 계수 $p_i(I)$ 가 큰 문헌들의 PVD와 독립성 가정 하에서 추정된 PVD (Ind로 표기), 종속성을 고려하여 추정된 PVD (Dep로 표기)를 도식하였다. 이 경우 PVD (Ind)가 s6(Oconee PRA)의 영향을 받고 있음을 알 수 있다. [그림 7]에는 가중 계수 $p_i(D)$ 가 큰 문헌들의 PVD와 PVD(Ind), PVD(Dep)를 도식하였다. 이 경우 PVD(Dep)는 s6의 영향을 벗어나 s7, s8, s9의 특성을 반영하고 있음을 알 수 있다. 종속성의 가정 하에서 산포모수의 추정 값이 크게 감소하였는데, 그 이유는 타 자료와의 차이가 큰 Oconee PRA 자료[17]에 대한 가중치가 소멸되고 마지막 세 문헌에 가중치가 집중되었기 때문인 것으로 판단된다.

<표 4> 솔레노이드 구동 밸브 기동 고장 자료

자료원	출판년월	평균 (10 ⁻³ /ry)	EF	$\hat{\xi}_i$	$\hat{\sigma}^2$	N_i	$N_{ij}(i < j)$	$p_i(I)$	$p_i(D)$
WASH-1400	75.10	1.0	3.0	-7.13	0.45	10	5,5,5,1, 5, 5, 5, 0	0.0370	0.0334
German RS	79	13.0	20.0	-6.00	3.32	12	5,5,5,2, 5, 5, 5, 0	0.0444	0.0641
NREP DB	82.06	1.0	3.0	-7.13	0.45	40	5,5,40,5, 15, 35, 35, 0	0.1481	-0.0619
NC-2728	83.01	1.0	3.0	-7.13	0.45	40	5,5,5,2, 15, 35, 35, 0	0.1481	-
Seabrook PSS	83.12	2.4	10.0	-7.01	1.96	10	5,5,5,5, 5, 5, 5, 0	0.0370	0.0516
Oconee PRA	84.06	0.013	12.7	-12.44	2.39	40	5,5,5,5, 5, 30, 35, 0	0.1481	-0.0184
NC-2815	85.08	0.73	20.0	-8.88	3.32	50	5,5,5,5, 5, 5, 40, 0	0.1852	0.2720
NC-4550	87.12	2.0	3.0	-6.44	0.45	60	5,5,5,5, 5, 5, 5, 0	0.2222	0.5565
Swedish NPP	-	0.26	5.0	-8.73	0.15	8		0.0296	0.1030
독립적 추정		4.2	48.2	-8.08	5.55	270			
종속적 추정		4.5	13.0	-7.21	2.13	77.6			



[그림 6] 독립성 가정 하에서의 주요 문헌 추정치와 PVD

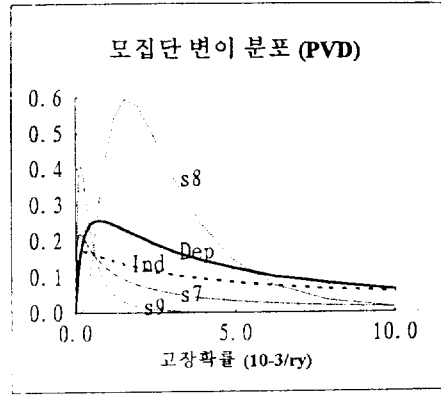


그림 7 종속성을 고려한 경우의 주요 문헌 추정치와 PVD

5. 결론 및 토의

본 연구에서는 종속적인 문헌 추정치들을 사용하여 PVD를 추정하는 방법을 제안하였다. PVD의 유형은 PSA 연구에서 가정되는 대수 정규분포로 제한하고 ML-II 방법으로 종속적 추정치들을 활용하는 관계식을 유도하였다. 추정 오차를 최소화하는 위치 모수의 추정치의 식 (3. 9)는 Winkler[24]가 점추정 문제에서 제안한 식과 동일하게 계산되었다. 그러나 산포 모수의 추정 식은 전혀 다른 양상을 보인다. 기존의 점추정 문제에서는 추정 오차로서의 산포를 고려하므로 자료가 많아질 수록 산포도 감소하게 되나, 분포 추정 문제에서의 산포는 자료의 증가에 따라 그 추정의 정밀도가 증가할 뿐이지 추정값 자체가 감소하지는 않는다. 또한 식 (3. 8)과 (3. 11)에서 보듯이 산포 모수 추정치뿐만 아니라 위치 모수 추정치의 산포도 PVD의 산포 추정에 영향을 미친다.

본 연구에서 사용한 가정은 몇 가지 제한성을 갖고 있다. 대수정규 모형은 PSA 연구에서

인정되고 있으므로 큰 문제가 없으나, ML-II 방법은 PVD를 모수적 분포로 제한하므로 정보의 손실이 발생할 수 있다. 이에 대한 대안으로 후보 분포들의 우도(likelihood)를 반영한 평균적 분포 추정 방법을 생각할 수 있다. 또한 개별 발전소 자료의 추정 오차가 동일한 산포를 갖는다는 가정도 제한적이다. 문헌별로 추정의 정밀도를 고려하여 표본의 크기를 조정하는 방법을 통해 식 (3. 7), (3. 9), (3. 11)에서 보듯이 추정의 정밀도 자체를 추정할 필요는 없다는 것을 시사하고 있다.

몇 가지 예제를 통한 추정 결과를 보면 문헌 추정치간의 종속성은 최종 추정 결과에 큰 영향을 미칠 수 있음을 알 수 있다. 종속성의 증가에 따라 유효한 표본의 수가 감소하여 추정의 정밀도는 감소되며, 상대적으로 독립성이 큰 자료의 비중이 증가함을 알 수 있다. PSA에서 사용되는 일반 문헌 자료에 대해서 적용해 본 결과, 자료간에 상당한 종속성이 잠재해 있으며 이러한 종속성을 정확히 반영할 필요가 있음을 알 수 있다.

참 고 문 헌

- [1] 임태진, "3단계 베이저안 처리 절차 및 신뢰도 자료 처리 코드 개발," 『경영과학』, 제11권, 제2호 (1994), pp. 1-27.
- [2] *ALWR PRA KAG, Advanced Light Water Reactor Utility Requirements Documents*, Chapter 1, Appendix A, PRA Key Assumptions and Groundrules, EPRI, 1991.
- [3] Bari, Buslik and Cho, NUREG/CR-2815, *Probabilistic Safety Analysis Procedures Guide*, BNL, 1985.
- [4] Carlson, NUREG/CR-2728, *Interlim Reliability Evaluation Program Procedure Guide*, SNL, 1983.
- [5] Fröhner, F. H., "Analytic Bayesian Solution of the Two-Stage Poisson-Type Problem in Probabilistic Risk Analysis," *Risk Analysis*, Vol. 5, No. 3 (1985), pp. 217-234.
- [6] Garrick, *PLG-0300, Seabrook Station Probabilistic Safety Study*, PL&G, Public Service Co. of New Hampshire and Yankee Atomic Electric Company, 1983.
- [7] *German Risk Study*, GRS, FRG, 1979.
- [8] Hubble and Miller, NUREG/CR-1363, *Data Summaries of Licensee Event Reports of Valves at U. S. Commercial Nuclear Power Plants*, rev. 1, EG&G, 1982.
- [9] *IEEE Std 500. IEEE Guide to the Collection and Presentation of Electrical, Electronic, Sensing Component and Mechanical Equipment Reliability Data for Nuclear Power Generating Stations*, IEEE, 1984.
- [10] 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.
- [11] Kaplan, S., "On the Use of Data and Judgment in Probabilistic Risk and Safety Analysis," *Nuclear Eng. & Design*, Vol. 93 (1986), pp. 123-134.
- [12] 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.
- [13] *Midland Nuclear Plant Probabilistic Risk Assessment*, Pickard, Lowe, and Garrick, Inc., 1985.
- [14] 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.
- [15] Mosleh, A., "Bayesian Modeling of Expert-to-expert Variability and Dependence in Estimating Rare Event Frequencies," *Reliability Eng. and System Safety*, Vol. 38 (1992), pp. 47-57.
- [16] NUREG/CR-4550, *Analysis of Core Damage Frequency from Internal Events: Methodology Guidelines*, SNL, 1987.

- [17] *Oconee PRA : A Probabilistic Risk Assessment of Oconee Unit 3*, NS-AC-60, Duke Power Co., 1984.
- [18] Oswald, Gentillon, Matthews and Meachum, *NREP D/B, Generic Data Base for Data and Models Chapter of the National Reliability Evaluation Program (NREP) Guide*, INEL, 1982.
- [19] *PWR/RX 312 Sizewell 'B' PWR Pre-Construction Safety Report, Component Failure Data for PWR System Reliability Assessment*, NNC, UK, 1982.
- [20] *RKS 85-25, Reliability Data Book for Components in Swedish Nuclear Power Plants*, RKS, SKI, 1985.
- [21] Tomic and Lederman, *IAEA-TEC-DOC-478, Component Reliability Data for Use in PSA*, IAEA, 1988.
- [22] Vaurio, J.K., "On Analytic Empirical Bayes Estimation of Failure Rates", *Risk Analysis*, Vol.7, No3 (1987), pp. 329-338.
- [23] *WASH-1400, Reactor Safety Study: An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants*, USNRC, 1975.
- [24] Winkler, R.L., "Commbining Probability Distributions Form Dependent Information Sources", *Management Science*. Vol.27, No.4(1981), pp479-488.