

2-Bayes 이론을 이용한 데이터 처리방법에 관한 연구

이만수 · 이광원[†] · 김태훈 · 윤익근* · 오영달 · 서두현

호서대학교 안전보건학과 · *한국가스공사 연구개발원

(2009. 8. 12. 접수 / 2009. 12. 9. 채택)

A Study on the Processing Method of Reliability Database using 2-Bayes Theory

M. S. Lee · K. W. Rhee[†] · T. H. Kim

I. K. Yoon* · Y. D. Oh · D. H. Seo

Department of Safety and Health Engineering Hoseo University

*Korea Gas Corporation R&D Division

(Received August 12, 2009 / Accepted December 9, 2009)

Abstract : The safety assessment for facility industry is now being periodically performed in Korea. For the purpose of scientific safety management, QRA(Quantitative Risk Assessment) is also being performed, and reliability data of the facilities is essential to perform the assessment. The necessary reliability data for QRA have been generally announced the values in other process industries, which results in the drop of risk reliability. The most appropriate method is to perform a direct reliability analysis towards the facilities undergoing safety assessment. In this study, the distinction between homogeneous sample estimation and multi-sample estimation of reliability data clarify using 2-Bayes theory.

Key Words : database, reliability, quantitative analysis, failure rate, MTTF

1. 서론

설비 및 시스템의 정량적 안전성 평가를 수행하기 위해서는 시간당 또는 가동회수당 부품들의 고장률, 고장도, 평균수명 등의 신뢰도 데이터가 필요하다. 이러한 신뢰도 데이터는 통계적으로 정확하게 측정되어야 정량적 안전성 평가의 신뢰성 있는 분석결과 값이 도출될 수 있다. 때때로 관찰되는 사고/고장수의 부족으로 충분한 신뢰도 데이터를 수집하기 위해서는 여러 집단을 동시에 관찰하여야만 되는 경우가 발생된다. 이때 동종의 부품이라도 다른 집단에 있는 경우 다른 신뢰도 특성을 갖게 되는데 그 이유는 각 지역마다 환경과 운전 및 보수지침 등이 다르기 때문이다. 하지만 근래에 이르러 각 기업들은 비용 대 효율을 극대화하기 위해 표준화된 운전지침 및 보수지침 등을 수립하여 운영을 하고 있기 때문에 비록 집단이 다르더라도

동일한 부품이나 설비를 사용하면 그에 대한 신뢰도 데이터 역시 비슷한 양상을 보일 것이라 사료된다. 이에 본 연구에서는 2-Bayes 이론을 이용한 다중표본 추정 방법과 균일 표본 추정방법으로 구한 신뢰도데이터에 어떠한 차이점을 보이며 그 양상은 어떤지 알아보려고 한다.

2. 다중표본 신뢰도의 합성 방법

여러 집단의 신뢰도를 합성하는 방법은 크게 4가지로 나눌 수 있다.

1) 모든 집단에 균일한 가중치를 부여하여 계산하는 방법

신뢰도 데이터군이 n 개이고 이들 데이터군의 신뢰도가 모두 같은 경우에는 다음 식 (1)과 같이 평균수명(MTTF: Mean Time to Failure)을 계산할 수 있다.

[†] To whom correspondence should be addressed.
kwrhie@hoseo.edu

$$MTTF = \frac{1}{k} \sum_{i=1}^k MTTF_i = \frac{1}{k} \sum_{i=1}^n \frac{r_i}{t_i} \quad (1)$$

$MTTF_i$: i 번째 그룹의 $MTTF$

With k : 데이터 그룹의 수

n : 관찰부품수

t_i : i 번째 부품의 관찰시간

r_i : i 번째 부품의 고장갯수

2) 데이터 군이 모두 유사한 보수와 운전 정책으로 운전 되었고 부품 역시 동일하다고 가정된 경우 식 (2)와 같이 계산할 수 있다.

$$MTTF = \frac{\sum_{i=1}^n r_i}{\sum_{i=1}^n t_i} \quad (2)$$

3) 1-step Bayes 이론

Bayes 방법은 한 집단(population)을 구성하는 모든 부품들이 관찰하는 특성에 대하여 차이가 없어야 한다는 전제조건을 필요로 한다. 이는 한 집단의 모든 성분이 같은 고장률/확률을 갖게 된다는 의미이다. 하지만 경험상으로 보면 비슷한 부품들일지라도 다른 plant에서 운영된다면 이런 전제조건을 적용하기에는 무리가 따른다. 비슷한 동종의 부품이 다른 보수정책이나 운전조건에서 사용되어 지는 경우 이들 신뢰도 데이터를 합산하는 경우에는 단순합산의 경우 통계학적으로 무리가 따른다. 이런 이유로 1-step 또는 2-step Bayes 방법(super population method)이 사용되기 시작하였다.

정량적 평가를 위하여 어떤 부품의 신뢰도 data를 수집하던 중 문헌에 나타난 일반적 data와 직접 얻은 specific data 등 2개를 갖고 있으며 이들을 모두 고려한 신뢰도 수치를 얻고자 할 때 이 이론을 사용한다. 사전적 확률, 추가정보, 사후적 확률은 다음과 같다²⁾.

prior probability : 추가적인 정보가 있기 전의 어떤 사건에 대한 사전적 확률, 즉 문헌 등에 나타난 확률

evidence : 추가적인 정보, 즉 최근 얻은 확률

posterior probability : 사후적 확률, 추가적인 정보와 사전적 확률을 고려한 확률

$P(Y)$ 를 evidence 즉, 현장의 실제사고 등을 통하여 얻은 확률이라 하고 $P(X)$ 가 사전적 확률이라 하

면 사후적 확률은 식 (3)처럼 표현된다.

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

예를 들어 어떤 부품의 평균수명이 A문헌에는 10개월, B문헌에서 5개월로 나와 있는 경우 이 부품의 평균수명 적용은 두 문헌의 신뢰도가 같다고 전제하면 $(10 \times 0.5) + (5 \times 0.5) = 7.5$ (개월)로 하게 된다. 만약 3개월 동안 이 부품을 사용하였는데 무사고이었다는 경험치를 반영시키려고 한다면 두 문헌의 신뢰도는 다음과 같다.

i) 각 문헌에서 3개월 동안 무고장일 확률은 다음과 같다.

$$P(Y|X=10) = e^{-\lambda t} = e^{-\frac{3}{10}} = 0.7408$$

$$P(Y|X=5) = e^{-\frac{3}{5}} = 0.5488$$

ii) 각 문헌의 가중치는 다음과 같이 수정된다.

$$P(X=10|Y) = \frac{(0.7408 \times 0.5)}{(0.7408 \times 0.5) + (0.5488 \times 0.5)} = 0.5744$$

$$P(X=5|Y) = \frac{(0.5488 \times 0.5)}{(0.7408 \times 0.5) + (0.5488 \times 0.5)} = 0.4256$$

iii) 결국 수정된 평균수명은 다음과 같이 계산된다. $(10 \times 0.5744) + (5 \times 0.4256) = 7.8717$ (개월)

일반적으로 one stage Bayes 이론은 관찰되어지는 모든 부품은 같은 고장률이나 고장확률을 갖는다고 가정한다. 하지만 경험적으로 관찰하면 같은 부품일지라도 사용 장소가 틀리면 운전조건이나 보수정책 등에 의하여 서로 다르게 고장이 일어난다. 이런 관점을 고려한 방법을 2-stage- Bayesian procedure 또는 super population method라 하며 수학적으로 관심 있는 수치, 즉 사후적 확률을 확률 변수로 취급한다.

4) 2-step Bayes 이론

이 이론은 신뢰도 측정의 체계적인 평가를 위해 여러 개의 집단을 동시에 관찰하여야 하는 경우 다음과 같은 단계가 필요하다.

- 비슷한 부품들에 대한 집단의 정의 및 부품들의 속성수집
- 통계적 측정의 기초자료로 쓰일 수 있는 운전

경험과 고장 사건들의 수집
 · 각각 다른 집단에서 나타나는 고장 형태의 차이를 반영할 수 있는 통계적 분석 방법 적용

i) 추정량의 가정

- k개의 다른 샘플이 존재한다. 즉, 서로 상이한 운전 정책과 환경을 갖는 동일 부품의 데이터군이 k개 존재 한다.
- i번째의 데이터군에서 우리는 $i = 1, 2, \dots, n$ 에 대한 총 관찰시간 τ 동안에 r_i 개의 고장을 관찰한다.
- i번째 샘플은 $i = 1, 2, \dots, n$ 에 대하여 각각 일정한 고장률 λ_i 를 갖고 있다.
- 다른 운전 및 환경적 조건들 때문에, 고장률 λ_i 는 k개의 데이터군 사이에서 다를 수 있다.

k개의 데이터군 간의 고장률 변동은 고장률이 확률밀도함수 $\pi(\lambda)$ 에 의해 주어진 어떤 분포를 갖는 확률변수라고 가정함으로써 모형화 된다¹⁾.

평균 고장률

$$\theta = \int_0^{\infty} \lambda \cdot \pi(\lambda) d\lambda \quad (4)$$

분산

$$\sigma = \int_0^{\infty} (\lambda - \hat{\theta})^2 \cdot \pi(\lambda) d\lambda \quad (5)$$

ii) 다중 표본에서의 분석절차

- ① 데이터를 모두 합쳐서 평균고장률 θ 의 초기 추정치 $\hat{\theta}_1$ 을 계산한다.

$$\hat{\theta}_1 = \frac{\text{Total No. of failures}}{\text{Total time in service}} = \frac{\sum_{i=1}^n r_i}{\sum_{i=1}^n \tau_i} \quad (6)$$

- ② 다음 식 (7), (8), (9)와 같이 표준 편차와 분산을 계산한다.

$$S_1 = \sum_{i=1}^k \tau_i \quad S_2 = \sum_{i=1}^k \tau_i^2 \quad (7)(8)$$

$$V = \sum_{i=1}^k \frac{(r_i - \hat{\theta}_1 \cdot \tau_i)^2}{\tau_i} = \sum_{i=1}^k \frac{r_i}{\tau_i} - \hat{\theta}_1^2 \cdot S_1 \quad (9)$$

- ③ 표본들 간의 변동을 나타내는 분산 σ^2 에 대한 추정치는 다음 식 (10)과 같이 계산된다.

$$V = \sum_{i=1}^k \frac{(r_i - \hat{\theta}_1 \cdot \tau_i)^2}{\tau_i} = \sum_{i=1}^k \frac{r_i}{\tau_i} - \hat{\theta}_1^2 \cdot S_1 \quad (10)$$

0보다 클 때, 그렇지 않으면 0.

- ④ 평균고장률 θ 의 최종 추정치 θ^* 은 다음과 같이 계산된다. 즉 θ^* 이 최종 추정된 고장률이다.

$$\theta^* = \frac{1}{\sum_{i=1}^k \frac{1}{\hat{\theta}_1 + \hat{\sigma}^2}} \times \sum_{i=1}^k \left(\frac{1}{\hat{\theta}_1 + \hat{\sigma}^2} \times \frac{r_i}{\tau_i} \right) \quad (11)$$

3. 사례연구

본 장에서는 2장에서 설명한 4가지의 방법을 실제 가스설비 주요부품에 적용하여 계산을 하여보고 차이점을 고찰 하였다.

유사한 가스 설비 3곳에 대한 기본적인 신뢰도 데이터 및 위의 4가지 방법으로 계산한 결과는 다음 표와 같다. 이때 A, B, C집단의 관찰시간 및 운전시간(단위는 백만시간)은 다음 Table 1과 같다.

Table 1. Calendar time and operation time of LNG pump & air compressor

		(단위 : 1,000,000시간)		
		A	B	C
LNG Pump	Calendar Time	1.8912	3.0439	0.9189
	Operational Time	0.782	0.6592	0.2999
Air compressor	Calendar Time	0.4693	0.6132	0.1885
	Operational Time	0.2750	0.3616	0.0785

A, B, C 그룹은 LNG 설비로서 가동년수는 A, B, C그룹 순이며, 이중 규모는 B그룹이 가장 크다. 신뢰도 데이터 계산에는 단순교체는 제외 하였다. 작동시간은 B그룹이 큰 이유는부품수가 많아 가장 오래된 A그룹보다 부품의 총 가동시간이 길기 때문이다.

신뢰도 계산 결과를 살펴보면 다음과 같다.

- Method 1 : 각 Group의 고장 개수 및 운전시간의 분포와 상관없이 절대적인 고장개수 및 운전시간에 영향을 받음.

Table 2. MTTF of LNG pump

(단위 : 1,000,000시간)

Failure Mode	Number of Failure			Calendar time				Operational time			
	A Group	B Group	C Group	Method 1	Method 2	Method 3	Method 4	Method 1	Method 2	Method 3	Method 4
소음진동	2	227	2	25.94	39.46	3.81	26.08	117.85	132.67	10.59	118.03
운전지연	0	92	5	11.89	16.57	7.28	12.02	52.07	55.71	21.12	52.22
구조결함	2	74	7	11	14.18	8.89	11.12	46.05	47.67	27.00	46.17
외부누출(P)	0	48	11	9.25	10.08	10.44	9.25	36.5	33.89	35.87	36.43
작동불량	4	45	9	8.9	9.91	10.59	8.97	34.46	33.31	34.41	34.47
표시불량	0	35	3	4.92	6.49	5.02	5.05	21.03	21.82	14.42	21.15
내부누출	0	12	0	1.31	2.05	3.73	1.55	6.07	6.89	11.32	6.28
오작동	0	3	1	0.69	0.68	2.30	0.68	2.63	2.3	6.67	2.3
전송불량	0	0	2	0.73	0.34	2.06	0.56	2.73	1.72	4.62	1.72
오알람	0	1	2	0.84	0.51	1.62	0.51	2.22	1.15	6.30	1.66
외부누출(U)	0	1	0	0.11	0.17	1.61	0.17	0.51	0.57	4.60	0.57
합계	8	538	42	75.75	100.45	51.12	75.69	322.12	337.71	155.02	322.25

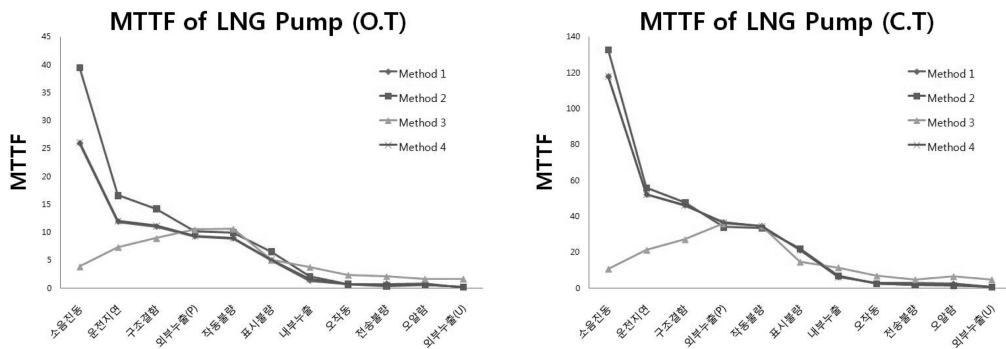


Fig. 1. MTTF of LNG pump.

- Method 2 : Method 4와 유사한 값을 보이거나 한 Group에 고장률이 치우쳐져 있는 경우 불안한 값을 보임.
- Method 3 : Group A와 B의 고장률에 의존도가

- 크기 때문에 Group C의 고장률이 아무리 높게 나와도 일정 이상으로는 높아지지 않음.
- Method 4 : 각 Group의 분산을 고려한 고장률의 분배를 통해 일정한 고장률 값을 가짐.

Table 3. MTTF of air compressor

(단위 : 1,000,000시간)

Failure Mode	Number of Failure			Calendar time				Operational time			
	A Group	B Group	C Group	Method 1	Method 2	Method 3	Method 4	Method 1	Method 2	Method 3	Method 4
운전지연	0	32	2	20.93	26.19	17.64	21.39	37.99	47.54	37.33	38.97
표시불량	0	24	2	16.58	20.03	17.41	17.01	30.62	36.36	36.05	31.36
작동불량	0	18	3	15.09	16.18	20.09	15.2	29.33	29.37	39.35	28.64
구조결함	0	2	12	22.31	10.79	7.94	21.6	52.78	19.58	13.95	51.07
소음진동	0	9	1	6.66	7.7	12.10	7.36	12.54	13.98	23.29	13.98
오알람	0	9	0	4.89	6.93	14.91	5.79	8.3	12.59	27.65	10.8
외부누출(P)	2	4	1	5.29	5.39	9.39	5.39	10.36	9.79	16.97	9.79
외부누출(U)	2	3	1	4.74	4.62	8.59	4.62	9.44	8.39	15.27	8.39
오작동	0	2	1	2.86	2.31	8.29	2.31	6.09	4.2	14.49	4.2
전송불량	0	0	1	1.77	0.77	8.68	0.77	4.25	1.4	15.21	2.03
합계	4	103	24	101.12	100.92	104.33	100.85	201.69	183.18	204.28	199.87

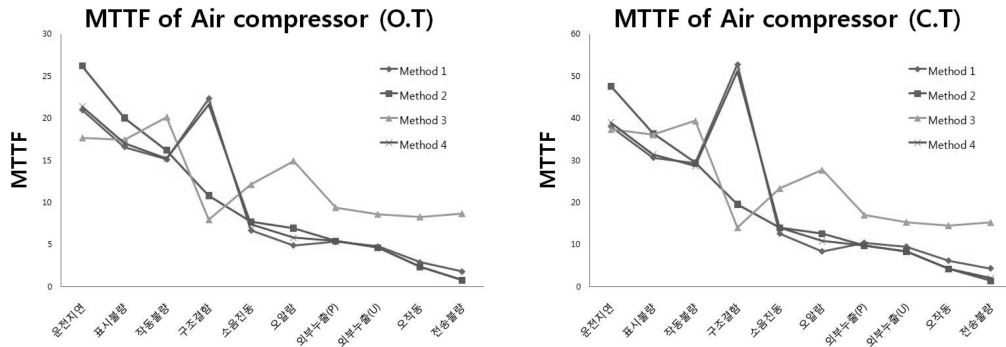


Fig. 2. MTTF of air compressor.

1-step Bayes 이론을 이용한 Method 3의 경우 고장수가 관찰되지 않은 경우 불안정한 값을 가지게 되며 Method 1, 3은 Method 4를 이용한 고장률값보다 편차가 심한 편이고, 거의 대부분의 경우에는 2-step Bayes 이론을 이용한 Method 4의 고장률값이 가장 신뢰도가 높은 것으로 확인된다.

즉, 2-step Bayes 이론은 분산에 의한 계산으로 편차에 대한 영향이 적어 데이터의 신뢰도가 가장 높게 나오므로 2-step Bayes 이론에 의한 추정값을 구하는 것이 가장 좋은 방법임을 확인 할 수 있다.

4. 결론

본 연구에서는 다중표본에 대한 신뢰도 데이터 합성 방법 4가지에 대하여 고찰 하였다. 다중표본의 고장률이 비슷한 경우에는 대체적으로 4가지 방법 모두 비슷한 값을 보여주고 있으나 고장률에 차이를 보이는 데이터군은 차이가 많이 남을 알 수 있다.

- 1) 동일 가중치를 부여한 방법의 경우에는 현실적으로 고장률이 너무 크거나 작은 집단이 존재하는 경우에는 이에 영향을 많이 받을 수밖에 없으며 대체적으로 추정 고장률보다 큰 값을 보여준다.
- 2) 동일 집단이라는 가정하의 추정치는 작은 고장률이 존재하는 경우, 이의 영향을 받아 추정 고장률보다 작은 값을 보여준다.
- 3) 고장률의 차이가 많이 나는 경우에는 동일 가중치 부여 방법으로 계산한 값은 같은 부품이라는 가정하의 값에 비하여 크게 나옴을 확인 하였고 대체적으로 2-step Bayes 이론을 적용한 값은 그사이에서 값을 보여준다.
- 4) 1-step 방식의 계산은 고장수가 관찰 되지 않은 경우 불안정한 값을 보여준다.

5) 현장에서 수집되어지는 데이터의 합성 방법은 이론적으로 각기 사용 되어져야 하는 범주가 다르며, 알맞게 사용되어져야 한다. 가중치를 모두 같게 부여하는 방법과 다중표본을 하나의 표본으로 관찰하는 방법은 추정량 보다 크거나 작으며, 모든 경우에 비교적 2-step Bayes 이론에 의한 값이 가장 신뢰도가 높은 것으로 확인 되었다.

본 연구는 관찰 기간이 충분히 큰 경우라는 가정 하에서 시행하였으며 추후 관찰기간이 서로 다른 경우의 신뢰도 데이터 추정에 대한 고찰이 필요하다고 사료 된다.

참고문헌

- 1) OREDA, "Offshore Reliability Data Handbook. 4th edition", DNV Technica, 2002.
- 2) H. J. Wingender, 5th EuReDATA "Reliability Data Collection and Use in Risk and Availability Assessment", 1986.
- 3) Henri Procaccia, Spyros P. Arsenis, Patrick Aufort, Preface by G.Volta. "European Industry Reliability Data Bank", 1998.
- 4) Center for Chemical Process Safety of the American Institute of Chemical Engineers, "Guidelines for process equipment reliability data with data tables", 1989.
- 5) C. David, "Selected Statistical Paper of Sir David Cox", Cambridge, pp. 489~495, 2006.
- 6) Y. M. Zhang; Z. Yang, "Reliability-based Sensitivity Analysis of Vehicle Components with Non-normal Distribution Parameters", International journal of automotive technology. Vol. 10, No. 2, pp. 181~194, 2009. 3.
- 7) G. I. Schueller, "Efficient Monte Carlo Simulation

- Procedures in Structural Uncertainty and Reliability Analysis : Recent Advances”, Structural Engineering and Mechanics : An International Journal. Vol. 32, No. 1, pp. 1~20, 2009. 5.
- 8) Allen Zielnik, Kurt Scott, “Durability, Reliability & Financial Risk in PV”, Inter PV. Vol. 1, No. 3, pp. 60~62, 2009. 5.
- 9) M. Muzzammil, N. A. Siddiqui, A. F. Siddiqui, “Reliability Considerations in Bridge Pier Scouring”, Structural Engineering and Mechanics : An International Journal. Vol. 28, No. 1, pp. 1~18, 2008. 1.