베이지안 기법을 이용한 소표본 보증데이터 분석 방법 연구

A Study of the Small Sample Warranty Data Analysis Using the Bayesian Approach

김 종 걸* · 성 기 우** · 송 정 무***

Jong-Gurl Kim* · Ki-Woo Sung** · Jung-Moo Song***

Abstract

보증 데이터를 통해 제품의 수명 및 형상모수를 추정할 때 최우추정법과 같은 전통적인 통계 분석방법(Classical Statistical Method)을 많이 사용하였다. 그러나 전통적인 통계 분석방법을 통해 수명과 형상모수의 추정 시 표본의 크기가 작거나 불완전한 경우 추정량의 신뢰성이 떨어진다는 단점이 있고 또 누적된 경험과 과거자료를 충분히 이용하지 못하는 단점도 있다. 이러한 문제점을 해결하기 위해 모수의 사전분포를 가정하는 베이지안(Bayesian) 기법의 적용이 필요하다. 하지만 보증 데이터분석에 있어서 베이지안 기법을 이용한 연구는 아직 미흡한 실정이다.

본 연구에서는 수명분포가 와이블 분포를 갖는 보증데이터를 활용하여 모수 추정의 효율성을 비교 분석하고자 한다. 이를 위해 와이블 분포의 모수가 대수정규분포를 따르는 사전분포를 갖는 베이지안 기법과 전통적 통계기법인 생명표법(Actuarial method)을 활용하여 추정량을 도출하고 비교 분석하였다. 이를 통해 충분한 관측 데이터를 확보할 수 없는 경우에 베이지안 기법을 이용한 보증 데이터 분석방법의 성능을 확인하고자 한다.

Keywords : Warranty Data Analysis, Classical Statistical Method, Bayesian Statistics, Life Data Analysis, 1-Parameter Weibull Model, Weibull-Bayesian Model, Actuarial Method, Interval Censored Data

* 성균관대학교 시스템경영공학과 교수

** 성균관대학교 산업공학과 박사과정

*** 성균관대학교 산업공학과 석사과정

1. 서 론

보증 데이터는 제품의 신뢰도를 산출하기 위한 목적의 데이터로써 일반적으로 보증기간을 갖는 제품에서 고장데이터의 형태로 얻어진다. 보증 데이터는 수집과정에서 제품 고장 수명에 대하여 불완전하고 충분한 관측 데이터를 확보할 수 없는 경우가 빈번하게 발생한다. 특히 실무에서는 빠른 의사결정과 비용문제 때문에 소표본 보증데이터를 분석해야 하는 경우가 많으므로 분석의 어려움이 있다. 이러한 경우 보증 데이터의 불확실성에 대한 분석은 대상에 대한 다양한 정보를 활용해 가용 데이터의 부족으로 인한 분석의 한계를 극복해야 할 필요가 있다.

소표본 데이터의 분석의 필요성과 중요성은 다음과 같다. 정확한 하고 신뢰성 높은 분성을 하기 위해서는 더 많은 데이터의 확보가 필요하다. 그러나 현장에서는 고장 데이터의 수가 곧 비용과 연결되기 때문에 많은 표본을 확보하기 어렵다. 또한 적은 보증데이터로 필드 부품의 정확한 수명을 알 수 있다면 빠른 의사결정 필요시 활용이 가능하다.

보증 데이터를 통해 신뢰도의 정량적인 값을 얻거나 신뢰성수명분야에서 이용되는 기존의 방법은 최우추정법과 같은 전통적인 통계 분석방법을 많이 사용하였다. 하지만 표본의 크기가 작거나 불완전하고 충분한 관측 데이터를 확보할 수 없는 경우 표준오차나 변동계수가 커져서 추정량의 신뢰성이 떨어진다. 또 누적된 경험과 과거자료를 충분히 이용하지 못하는 단점이 있다. 이러한 직접 추정량의 문제점을 해결하기 위한 방안으로 관련 모수의 사전분포를 가정하는 베이지안 통계학(Bayesian Statistics)이 이용되어 왔다.

베이지안 통계학이란 통계학적으로 관심이 있는 모든 것(모수, 결측치, 예측 값 등)을 불확실하지만 이 불확실의 정보가 확률로서 표현된다는 가정에서 시작한다. 이러한 베이지안 통계학은 결과의 해석이 전통적인 통계학 보다 쉽고, 우리의 직관과도 잘 부합한다는 장점을 갖는다. 특히 모수에 대한 사전의 정보가 많은 경우에는 추론을 더 잘 해낼 수 있고 다루는 문제가 복잡한 경우는 전통적인 통계학 보다 쉽게 계산이 가능하다. 그런 장점들에 의해 전통적인 통계학에서 풀 수 없는 아주 복잡한 모형의 분석의 경우 등에 널리 응용되어지고 있다. 하지만 보증 데이터 분석에 있어서 베이지안 기법을 이용한 연구는 미흡한 상태이다.

본 논문의 목적은 보증 조건이 5년 60,000만부인 제품에 대해서 실제 보증 실시 이후 모여진 5년간의 보증 수리 데이터를 통해 소표본 보증데이터 분석 방법 및 활용 방법을 제시하고자 함이며, 데이터 분석에 있어서 표본의 크기가 작거나 충분한 관측 데이터를 확보할 수 없을 때 효율적이고 정확한 모수와 수명 추정이 가능함을 확인하는 것이다. 특히 전통적 통계기법에서 해결하기 어려운 추정의 정확성을 베이지안 기법을 통해서 확인 하고자 한다.

2. 연구영역의 조사

보증 데이터로부터 제품 신뢰도에 관한 유용한 정보를 추출해내기 위해 다양한 통계 기법이 사용되고 있다. 본 챕터에서는 모수 추정에 따른 방식으로 비모수적 방법과 모수적 방법으로 나누어 분석 방법을 분류하였다. 또 본 연구에 활용될 보증 데이터는 와이블 분포를 따르는 것으로 알려진 기계장치에 대한 고장시간과 가동시간을 기록한 보증 데이터를 사용하였는데 이러한 데이터 특성에 맞는 모수 추정 방법을 크게 전통적 접근법(Classical Approach)과 베이지안 접근법(Bayesian Approach)으로 분류하였다.

보증 데이터에서 수명함수의 모수를 추정하는 방법은 크게 두 가지로 구분할 수 있다. 하나는 '모수적 모형(Parametric Model)'을 이용하는 방법이고 하나는 '비모수적 모형(Non-Parametric Model)'을 이용하는 방법이다.

모수적 모형은 수명함수에 대한 일정한 분포를 가정하고 수명자료를 이용해 해당 모델의 모수(Parameter)를 추정하는 접근방식으로 지수분포(Exponential Distribution)와 와이블 분포(Weibull Distribution), 대수정규분포(Lognormal Distribution) 등이 활용되고 있다. 반면 비모수적 모형은 특정한 분포를 가정하지 않고 한 집단의 생존기간을 추정하는 방법으로 수명자료에 대한 구체적인 분포를 가정하는 것이 무리일 때 사용한다. 이는 분포에 대한 가정에 오류 가능성을 줄이고 때로는 추정 시 효율을 높일 수 있는 대안이 된다.

비모수적 방법연구로는 Kaplan과 Meier(1958)가 수명관측의 중단이 임의로 일어나는 임의 중단모형에서 얻은 수명 관측치를 이용하여 신뢰도의 비모수적 추정량을 제안하였고[9], Hu와 Lawless(1998)는 자동차의 주행거리 데이터를 활용해 관측중단시간을 모를 경우를 대비하여 비모수적 추정 방법을 제안하였다[8]. Majeske와 Herrin(1995, 1998)은 차량 오디오 시스템에 대한 보증 데이터를 사례로 들어 제품 및 공정설계 조건에 변동이 있을 시 변동 전후의 시스템을 실증적으로 평가할 수 있는 방법을 제안하였고 보유기간에 따른 생명표(Life Table)와 위험률(Hazard Rate) 플롯을 사용하는 비모수적방법을 개발하였다[11][12].

본 연구에서는 수명이 와이블 분포를 따르는 것으로 알려진 구간관측중단(Interval Censored Data)된 데이터를 가진 기계장치에 대한 고장시간과 가동시간 데이터를 모수 추정에 사용하였고 이러한 보증 데이터 유형에 적용하기 적합한 분석 기법으로 비모수적 신뢰성 분석 방법인 생명표법을 사용하였다. 생명표법은 일정한 구간으로 나누어 구간의 생존확률을 추정하는 방법으로 구간별 고장데이터와 구간 전체에 대한 관측중단 데이터가 존재할 때 사용하는 방법이다. 따라서 복사기기의 복사량과 오토바이의 주행거리등의 분석에는 생명표법이 가장 타당하다[2].

와이블 분포는 수명시간과 관련된 많은 분야에서 응용되는 분포로 분포에 포함된 형상모수(Shape Parameter)를 변화시킴으로서 여러 가지 형태의 분포를 얻을 수 있다. 와이블 분포의 모수추정에 관련된 대표적인 연구는 크게 두 가지 방향으로 접근해왔다. 첫째는 전통적 접근법(Classical Approach)으로 Cohen, Whitten and Ding(1984)는 수정된 최우추정량을 구하였고[4], Lawless(1982)는 변환(log 변환)을 통하여 생존함수

와 관측된 자료를 일차 직선화시킨 후, 최소자승법을 이용하여 와이블 분포와 극단치 분포(Extreme Value Distribution)의 모수를 추정하는 방법을 제안 하였다[7]. 특히 통계적 추론 분야에서 좋은 통계적 성질들을 만족하는 최우추정법에 이용한 방법은 뉴턴-랩슨법(Newton-Raphson Method)등의 비선형방정식(Non-Linear Equation)의 해를 구하는 복잡한 문제이다. 두 번째로 와이블 분포의 모수 추정법으로는 베이지안 접근법(Bayesian Approach)이 있다. 관련된 과거의 자료나 유사한 정보를 사전정보(Prior Information)로 두고 현재자료와 결합하여 사후정보(Posterior Distribution)를 이끌어내는 베이지안 접근법은 많은 장점이 있음에도 불구하고 사후확률에 관련된 계산은 어려운 경우가 많았다. 척도모수와 형상모수를 가지는 와이블분포에 관한 베이지안 추론 문제로 Soland(1968)는 형상모수를 고정한 조건에서 척도모수에 대한 베이즈 추정량을 구하였고[15], Soland(1969)는 척도모수와 형상모수에 대한 베이즈 추정량을 구하였다[14]. 위와 같은 베이즈 추정량을 계산하기 위해 다차원 적분을 해야 하는 경우, 베이즈 추정량을 구하는 문제는 매우 복잡하다. Geman은(1984)은 깁스표본법(Gibbs Sampling Method)을 제안하였다[7]. 깁스표본법은 반복적인 몬테칼로(Iterative Mote Carlo)기법으로, Gelfand와 Smith(1990)는 베이지안 통계계산에 깁스표본법이 매우 유용하게 사용됨을 밝혔다[5]. 특히, 다차원 적분에 매우 유용하다. 그러나 깁스표본법을 적용하기 위해서는 완전 적분 확률분포가 난수(Random Number)발생이 가능한 분포여야 하므로 통계적 적분함수가 아닌 경우에는 이용이 어려운 경우가 많다. 그러나 최근에는 이와 같은 문제에서 Metropolis 알고리즘, adaptive-rejection 표본법 등으로 많은 문제를 해결하고 있다.

본 논문에서는 다양한 베이지안 기법 중 Weibull-Bayesian 분포를 사용한 베이지안 기법을 사용하고자 한다. 이 방법은 소표본 분석의 불확실성을 줄이기 위해 1-Parameter WEIBULL 분포를 사용하였고, 와이블 분포의 형상모수를 이미 사전에 알고 있는 것을 가정하고 사전분포로 로그정규분포를 사용하였다. 이 방법의 장점은 형상모수에 대한 사전정보만 있다면 더 쉽게 소표본일 때 불확실성을 줄이고 정확한 모수를 추정할 수 있다. 사후분포에 대한 해를 얻기 위해서는 몬테카를로 시뮬레이션 등의 고급기법보다 쉽게 접근이 가능하다. 따라서 본 논문에서는 소표본 보증데이터의 정확한 분석을 위해 Weibull-Bayesian 모델을 제안하였고, 그 성능을 기존 전통적인 분석방법인 생명표법과 비교하였다.

3. 사례 연구

필드에서 수집된 보증데이터는 수명이 와이블 분포를 따르는 것으로 알려진 기계장치에 대한 고장시간과 가동시간을 기록한 데이터를 사용하였다. 이를 위해 와이블 분포의 모수가 대수정규분포를 따르는 사전분포를 갖는 베이지안 기법과 전통적 통계기법 중 비모수적인 분석방법인 생명표법을 활용해 추정량을 비교 하고자 한다.

3.1 데이터수집

보증데이터 분석을 위해 활용된 데이터는 수명이 와이블 분포를 따르는 것으로 알려진 기계장치에 대한 고장시간과 가동시간을 기록한 데이터로 본 기계장치는 5년 60,000매 가동으로 관측 중단된 데이터 이다. 이 보증데이터는 A/S를 통해 접수된 내용으로 판매일, 수리일, 제품번호, 가동시간으로 수집되었다. 총 판매 대수는 40,000대 이며 본 연구 대상이 된 부품에 대한 보증 고장 데이터는 1,435건 이다.

3.2 데이터의 가정

본 연구는 다음과 같은 가정으로 제안하였다.

① 보증기간 내에 발생한 고장은 모두 A/S를 통해 접수된다.
② 제품은 판매일시부터 가동된다.
③ 제품의 고장일시는 수리일시와 동일하다.
④ 분석에 사용된 데이터 값은 모두 참으로 가정한다.

3.3 데이터분석 방법

본 실증 연구에서는 전통적인 분석방법 중 비모수적인 분석방법을 통해 보증기간이 5년까지인 데이터로 분석한 결과를 참으로 보고 베이지안 기법을 이용하여 2년, 1년, 6개월 치로 샘플링한 소표본 데이터를 분석하여 참값과 비교 하였다. 참값에 대한 보증 데이터분석을 비모수적 접근법의 이용으로 한정하였고 고장시간과 가동시간을 모두 고려하기 적절한 생명표법을 활용하여 분석을 진행하였다. 이와 성능을 비교하기 위해서 베이지안 기법을 제안하였고, 이때 사용한 사전분포는 5년 60,000부의 보증 조건을 가진 복사기를 구형, 신형, 최신형으로 제품명을 정하여 각각의 생명표법으로 얻어진 형상 모수 값을 사전분포로 활용하였다. 또한 소표본 데이터 일 때의 성능을 비교하기 위해서 2년, 1년, 6개월로 데이터를 샘플링하고 분석하여 그 성능을 비교하였다. 결과적으로 얼마나 적은 데이터까지 베이지안 기법으로 성능을 유지 할 수 있는지도 살펴보았다. 그리고 분석 결과의 성능을 확인하기 위한 방법은 신뢰성 척도의 하나인 B1과 형상모수를 통해 비교분석하였다. 본 논문에서 활용한 보증데이터분석 기법인 생명표법과 베이지안기법의 분석 절차는 다음과 같다.

3.3.1 보증 데이터 분석 방법론

앞서 2절에서 보증 데이터에서 수명함수의 모수를 추정하는 방법을 모수적 방법과 비모수적 방법으로 분류했다. 이 분류를 토대로 보증 데이터 분석 방법론을 나열하고 본 논문에서 분석에 활용될 데이터에 적합한 분석방법을 찾아본다.

[표 1] 보증데이터 분석 방법론[1]

구분	방법론	특 징	분석의 타당성
모수	모수적 방법	- 모집단이 특정 분포(Exponential, Weibull, Log-normal Distribution 등)를 따르고 있다고 가정 - 모집단이 특정분포를 따르지 않을 경우 분석할 수 없음	X
비모수	생명표법 (Actuarial-Method)	- 일정한 구간으로 나누어 구간의 생존확률 추정 - 구간별 고장데이터와 구간 전체에 대한 관측중단 데이터가 존재할 때 사용 - 관측중단 발생 구간을 조정된 고장 비율로 사용	◎
	카플란-마이어 (Kaplan-Meier)	- 관측 중단된 시점마다 수명확률을 구하여 누적의 합으로 누적수명을 추정 - 표본수가 50이하일 때 유용한 방법	X
	다회 관측중단 데이터 분석	- 구간별 고장데이터와 우측 관측중단된 데이터가 혼합되어 있는 경우 사용 - 사용기간별 분석 시 사용되는 방법으로 주행거리분포 분석에서는 사용되지 않음	X
	랭크법 (Rank Method)	- 관측된 고장데이터로 추이 파악 가능 - 초기고장과 이상치 구별 가능 - 관측되는 모든 데이터에 신뢰도가 측정되므로 분석 및 관리에 어려움이 있음	○

각각의 기법을 정리해본 결과 인쇄부수를 사용하여 보증 데이터를 분석하기 위한 타당한 방법론으로 생명표법, 랭크법이 적절하다고 할 수 있다. 이 중 좀 더 분석의 타당성이 높은 생명표법으로 분석을 실시한다. 생명표법을 통해 고장 및 관측중단 데이터를 정해진 인쇄부수 구간(예: 5,000 인쇄부수 간격)별로 정리하여 분석이 가능해진다. 또 데이터 량이 방대한 보증 데이터를 요약하여 나타내므로 데이터 분석, 유지, 관리가 용이하다. 구간별 데이터의 수량을 이용하므로 초기고장이나 이상치의 영향에 대해 상대적으로 덜 민감(Robust)하다는 장점도 있다.

3.3.2 생명표법

생명표법은 Cutler와 Ederer가 1958년에 발표한 방법으로 고장 및 관측중단 데이터를 정해진 가동시간의 구간별로 정리하여 그 개수를 이용한 분석방법이다. 5년 60,000부의 보증조건에서 얻어진 신형의 보증 데이터를 가지고 신뢰도를 추정하는 방법으로는 다음과 같다.

① 구간데이터를 정리한다. : m개 구간, 최소 제품 수 = n

구간 i : $(t_{i-1}, t_i], t_0 = 0$

n_i : 구간 i초 t_{i-1}서 가동 중인 제품 수(Risk set), $n_0 = n$

c_i : 구간 i에서 관측 중단된(Censored) 제품 수, $c_0 = 0$

d_i : 구간 i에서의 고장 제품 수, $d_0 = 0$

$n_i = n_{i-1} = d_{i-1} - c_{i-1}, i = 1,2,..,m$

여기서 구간 i의 관측중단 수 c_i는 보고그룹의 중단 수 c'_i와 비보고 그룹의 중단 수 c''_i의합으로 구성된다. 비보고 그룹의 중단 수 c''_i는 가동시간 분포와 총비보고 제품 수 M으로부터 구한다.

$c''_i = M \times G(t_i) - G(t_{i-1})$

[표 2] 구간 절단된 구간데이터

순위 i	t_i	d_i	c_i	$d_i + c_i$	n_i	$n''_i = n_i - \frac{c_i}{2}$	$p_i = 1 - \frac{d_i}{n''_i}$	$F(t_i)$
1	5	4	17	21	153332	153323.6	0.99997	0.00003
2	10	13	206	219	153311	153207.9	0.99992	0.00011
3	15	15	787	802	153092	152698.3	0.99990	0.00021
4	20	16	1923	1939	152290	151328.3	0.99989	0.00031
5	25	19	3723	3742	150351	148489.3	0.99987	0.00044
6	30	31	6200	6231	146609	143509.0	0.99978	0.00066
7	35	42	9233	9275	140378	135761.4	0.99969	0.00097
8	40	22	12527	12549	131103	124839.2	0.99982	0.00114
9	45	30	15611	15641	118554	110748.4	0.99973	0.00141
10	50	40	17905	17945	102913	93960.6	0.99957	0.00184
11	55	29	18866	18895	84968	75535.2	0.99962	0.00222
12	60	92	65981	66073	66073	33082.7	0.99722	0.00500

② 구간 i에서의 고장비율을 산출한다.

구간 i 고장비율 = (구간 고장 수 d_i) / (구간 초 가동 제품 수 n'_i). c_i개 관측중단은 구간 전체에 걸쳐 발생한다.

a) c_i개 관측중단이 구간 초에서 발생한다고 가정할 때 : $n'_i = n_{i\_}c_i$

b) c_i개 관측중단이 구간 끝에서 발생한다고 가정할 때 : $n'_i = n_i$

▸ 조정된 $n'_i = n_i - \frac{c_i}{2}$

▸ 구간 i 고장비율 $= \frac{d_i}{n_i} = \frac{d_i}{n_i - \frac{c_i}{2}}$

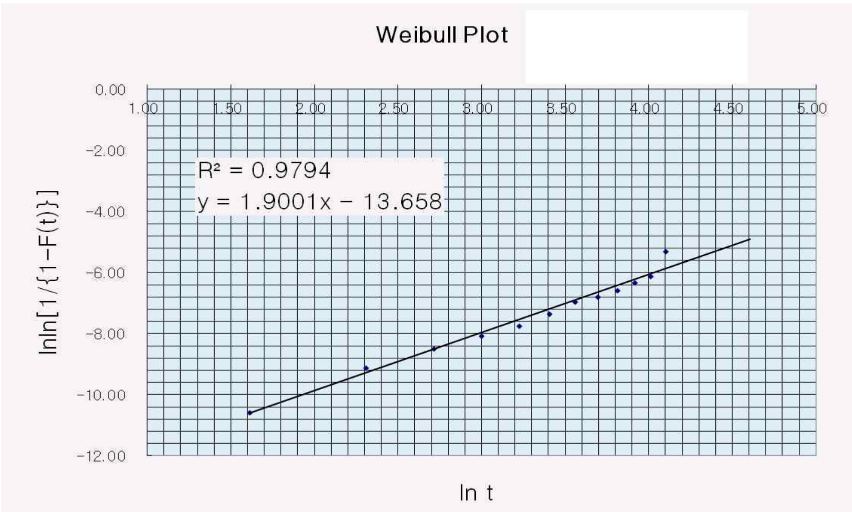
③ 신뢰도와 불 신뢰도를 추정(생명표에 의한 방법)한다. 구간 i의 끝 t_i에서의 신뢰도와 불 신뢰도 추정은 다음과 같다.

$$\widehat{R(t_i)} = \prod_{j=1}^{i}(1 - \frac{d_i}{n'_j}),\ i = 1,2,..,m$$

$$\widehat{F(t_i)} = 1 - \widehat{R(t_i)} = 1 - \prod_{j=1}^{i}(1 - \frac{d_j}{n'_j}),\ i = 1,2,..,m$$

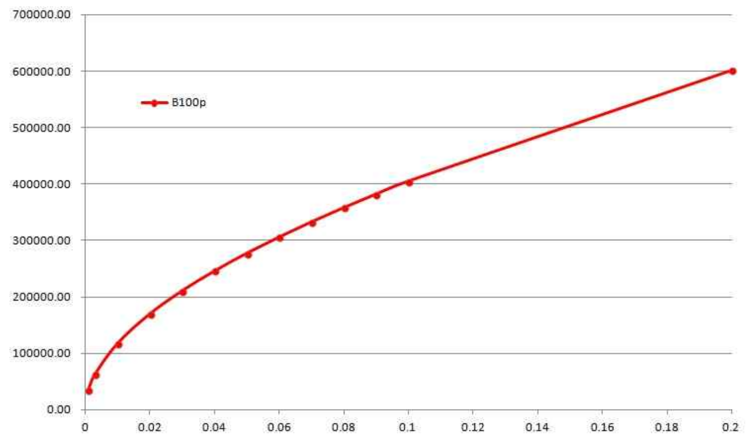
④ Weibull Plot을 작성하고, 형상모수와 척도모수를 추정한다.

형상모수 β : 1.9001
척도모수 θ : 1323.510



[그림 1] 생명표법을 통한 Weibull Plot

⑤ 추정된 수명분포로부터 B_{100p}수명을 추정한다.



[그림 2] 생명표법을 통한 B_{100p}수명 추정

[표 3] 생명표법을 통한 B_{100p}수명 추정

Percentiles	B100p수명
0.01	117,570
0.02	169,779
0.03	210,730
0.04	245,842
0.05	277,234
0.06	305,997
0.07	332,781
0.08	358,013
0.09	381,989
0.1	404,924

3.3.3 베이지안 기법

통계학적인 관점으로부터 작은 샘플 사이즈로 너무 적게 관측된 고장들을 다룰 때 추정의 불확실성은 증가하고 의미 없음에 가까운 폭이 넓은 신뢰한계를 만들어 낸다. 작은 샘플 사이즈로부터 야기된 불확실성을 줄이기 위하여 1-Parameter Weibull 분포가 사용되어졌다[16]. 특히 형상 모수에 대한 사전 정보가 이용 가능할 때 1-Parameter Weibull은 와이블 분포의 형상 모수가 알려졌다는 것을 가정한다. 왜냐하면 그것은 오직 척도 모수가 추정되어지기 때문이다. 이렇게 하여 불확실성은 크게 줄어든다. 하지만, 만약 추정된 β의 형상 모수가 참값과 거리가 멀다면 추정된 결과는 부정확할 것이다. 이 챕터에서는 베이지안 모델에 대해서 논의한다. 1-Parameter Weibull과 같은 상수 값을 사용하는 것 대신에 β의 형상모수에 대한 분포가 고려되었다. β의 분포는 보증 데이터, 전문가의 의견, 이전의 시험(Prior Tests), 발달 시험(Developmental Testing)등을 통해서 얻어질 수 있고 사전분포로 불린다.

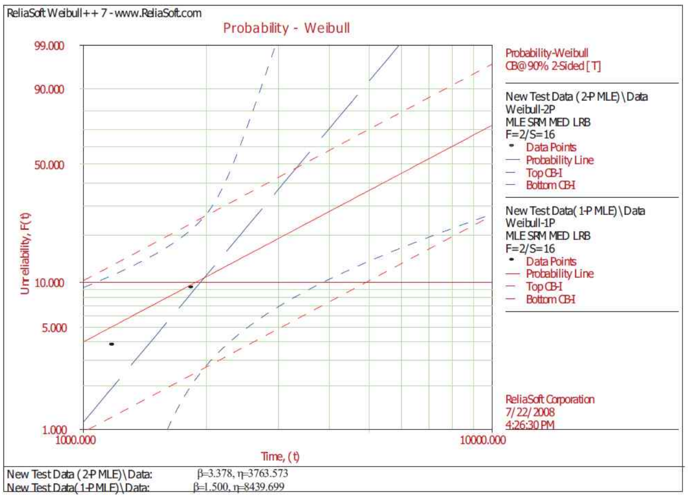
3.3.3.1 1-Parameter Weibull Model

1-Parameter Weibull은 와이블 분포의 형상모수가 알려졌다는 가정하기 때문에 작은 샘플 사이즈로부터 야기된 불확실성을 줄이는데 효과적이다. 1-Parameter Weibull은 사실 2-Parameter Weibull 분포의 특별한 예이다. 만약 추정된 형상모수 β가 참값에 가깝다면 2-Paremeter Weibull을 사용한 모수의 추정보다 정확도가 더 높을 것이다. 2-Parameter Weibull 분포의 확률밀도 함수는 다음과 같다.;

$$f(t,\beta,\eta)=\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}e^{-\left(\frac{t}{\eta}\right)^{\beta}} \qquad (1)$$

여기서,
$f(\ )$: 확률밀도함수(Probability Density Function)
β : 형상 모수(The Shape Parameter)
η : 척도 모수(The Scale Parameter)

만약 β의 값이 주어진다면 2-Parameter Weibull은 오직 η가 추정되는 것이 필요한 1-Parameter Model이 된다.



[그림3] Probability Plot for 1-Parameter and 2-Parameter Weibull[3]

[그림3]은 1-Parameter Weibull 과 2-Parameter Weibull의 비교를 목적으로 예를 든 그림으로. 위의 분석으로부터 1-Parameter Weibull Model 이 β의 한쪽으로의 치우침을 줄이고 모델의 불확실성을 줄이기 위해 더 좋은 선택으로 보인다. 하지만, 1-Parameter Weilbull을 적용한 추정은 β가 알려져 있어야 하고 정확해야 한다.

3.3.3.2 Weibull-Bayesian Model

베이지안 이론은 모델의 모수에 대한 사전 정보를 이용하는 것이 가능할 때 많은 분야에서 응용되어 졌다. 베이지안 기법은 거의 고장이 관측되지 않거나 사전 정보를 이용 가능할 때 신뢰성 공학 분야에서 당연한 선택이 되었다. 예를 들어, Martz, Waller and Fickas은 컴퍼넌트와 2명체계(binomial system)의 직렬 시스템에 대한 신뢰도를 추정하기 위해 베이지안 기법을 사용하였다[13]. Willits, Dietz and Moore는 아주 작은 샘플을 가진 직렬 시스템의 신뢰도를 추정하기 위하여 베이즈 몬테 카를로 접근법을 사용하였다[17].

Weibull-Bayesian Model(Wei-Bayesian)은 와이블 분포의 특성과 베이지안 통계학의 개념을 결합한 것으로 베이즈 기법을 사용한 모형의 모수를 추정하는 것에 대한 접근이다. 일반적으로 만약 형상 모수와 척도 모수에 대한 사전 분포가 모두 알려져 있다면 베이즈 공식은 다음과 같이 쓸 수 있다.:

$$g(\beta,\eta \mid data) = \frac{P(data \mid \beta,\eta)\varphi(\beta,\eta)}{\int_0^\infty \int_0^\infty P(data \mid \beta,\eta)\varphi(\beta,\eta)\,d\beta d\eta} \qquad (2)$$

만약 형상 모수와 척도 모수에 대한 사전 분포가 독립 이라면 식(2)는 다음과 같이 수정 될 수 있다.:

$$g(\beta,\eta \mid data) = \frac{P(data \mid \beta,\eta)\varphi(\beta)\varphi(\eta)}{\int_0^\infty \int_0^\infty P(data \mid \beta,\eta)\varphi(\beta)\varphi(\eta)\,d\beta d\eta} \qquad (3)$$

여기서 :

$g(\beta,\eta \mid data)$은 β와 η에 대한 접합후분포(Joint Posterior Distribution)이다.

$P(data \mid \beta,\eta)$는 주어진 β와 η에 대해서 데이터를 얻기 위한 우도 이다.

$\varphi(\beta)$는 형상 모수 β에 대한 사전 분포이다.

$\varphi(\eta)$는 척도 모수 η에 대한 사전 분포이다.

식(2) 또는 (3)로부터 β와 η에 대한 주변의사후분포(Marginal Posterior Distribution)는 다음과 같다.

$$g(\beta \mid data) = \int_0^\infty g(\beta,\eta \mid data)\,d\eta \qquad (4)$$

$$g(\eta \mid data) = \int_0^\infty g(\beta,\eta \mid data)\,d\beta$$

Weibull-Bayesian 모형의 적용에 있어서 η는 대개 밀도 함수(Density Function) $\varphi(\eta)=1/\eta$을 가진 무정보사전분포를 따르는데 이를 가지고 추정을 실시한다. 이것은 제프리 사전(밀도함수)라고 불리 우며 η에 대한 로그 변환(Logarithmic Transformation)을 수행함으로써 구해진다. 구체적으로 말하면, η가 항상 양수이기 때문에 그것은 추정되어질 수 있다. $\ln(\eta)$는 $U(-\infty,+\infty)$범위의 유니폼 분포(Uniform Distribution)를 따른다. 유니폼 사전분포는 사전정보를 얻을 수 없다는 의미를 포함한다. Jeffreys의 이론에 적용해보면 다음과 같다[6]. "일반적으로 대략적인 사전정보가 없는 것은 Fisher정보의 제곱근에 대해 비례하여 얻어진다."

일단 사후분포가 방정식(3)에서 얻어졌다면 시간에 따른 고장의 확률밀도함수는 다음과 같이 계산이 된다.

$$f(t|data)=\int_0^\infty\int_0^\infty f(t,\beta,\eta)g(\beta,\eta\mid data)d\eta d\beta \tag{5}$$

Weibull-Bayesian Model 에서는 전통적 통계학에서와 같이 점 추정 보다는 주어진 시간에서의 신뢰성과 고장 비율에서 BX수명과 같은 모든 모수 모형과 신뢰성 메트릭스에 대한 분포(The Posterior PDF)가 얻어진다. 그러므로 만약 점 추정이 필요하다면 사전 확률밀도함수가 계산되는 것이 필요하다.

4. 비교분석 및 결론

4.1 비교분석

신형/최신형 복사기로부터 전체 5년 보증데이터를 통해서 다음과 같이 형상 모수, 척도 모수, B1 수명을 구하였다. 분석방법은 생명표법을 통해 추정 값을 와이블 플롯에 작성하고, 형상 모수와 척도 모수를 추정한다. 이때 5년 데이터 분석을 통한 결과는 다음과 같다.

[표 4] 생명표법을 통한 B_{100p}수명 추정

	생명표법 5년		생명표법 2년		생명표법 1년		생명표법 6개월	
	형상 모수	B1	형상 모수	B1	형상 모수	B1	형상 모수	B1
신형	1.9364	33020.69	1.4896	46425.22	2.1934	32103.83	0.6701	497203.7
최신형	1.9001	117570.6	1.6065	179146.6	1.3794	311763.1	1.4038	341013.2

생명표법을 통해 5년 데이터를 분석한 결과 형상모수 1.9364, B1수명 33,020매로 분석 되었다. 이제 베이지안 기법을 사용하기 위해 와이블 분포 형상 모수의 사전분포를 구하였다. 이때 사용한 사전분포는 신제품 또는 구형제품 같은 동일한 사양을 가지고 있는 제품의 5년 구형(1.9001), 신형(1.9364), 최신형(1.944)의 형상 모수 값을 사전분포로 활용하였다. Weibull-Bayesian 분석은 ReliaSoft사의 Weibull++7 프로그램을 통해을 실시하였다. 무정보 사전 분포는 η에 대해 사용되어 진다. η에 대해 무정보 사전 분포를 사용하는 것은 η에 대한 어떠한 사전의 정보도 우리는 가지고 있지 않음을 의미한다. β모수는 [표5]에서 주어진 과거의 값을 가지고 있다.

[표 5] 각 제품에서 획득한 형상모수 값

모델명	수집된 형상모수(β) 값
구형	1.9001
신형	1.9346
최신형	1.944

β에 대한 사전분포는 [표5] 의 데이터로부터 결정되어진다. Bayesian-Weibull 모형에 위 데이터를 적합 시키기 위하여 베타의 사전 분포가 측정되는 것이 필요하다. 이전의 테스트의 β값에 근거하여 베타에 대한 사전 분포는 μ=0.65583, σ=0.01394를 가진 로그노말 분포가 되도록 결정되었다. 이 방법은 ReliaSoft사의 와이블++7 프로그램 데이터 입력시트에 [표5]의 값이 입력한 후 회귀분석법(PRX, Rank Regression on X)에 근거하여 분석함으로써 얻어졌다.

[표 6] 표본크기별(기간별) 베이지안 분석결과 및 생명표법 분석결과

분석구분	척도	신형	최신형	비교자료	척도	신형	최신형
생명표법\ (5년 고장 데이터)	형상 모수	1.9364	1.9001				
	B_1	33020.69	117570.6				
베이지안 (2년 고장 데이터)	형상 모수	1.87075	1.889544	생명표법 (2년 고장 데이터)	형상 모수	1.4896	1.6065
	B_1	48247.24	152986.5		B_1	46425.22	179146.6
베이지안 (1년 고장 데이터)	형상 모수	1.877575	1.916643	생명표법 (1년 고장 데이터)	형상 모수	2.1934	1.3794
	B_1	52879.1	172160.4		B_1	32103.83	311763.1
베이지안 (6개월 고장 데이터)	형상 모수	1.906541	1.922784	생명표법 (6개월 고장 데이터)	형상 모수	0.6701	1.4038
	B_1	62575.91	179729.4		B_1	497203.7	341013.2

보증서비스를 통해 얻어진 고장보고 수에 따라 전통적인 분석방법인 생명표법과 베이지안 기법의 차이를 비교하기 위해 고장데이터 수집기간(고장보고 수)을 다르게 하여 비교분석을 진행하였다. 총 판매 수는 40,000대로 고정하였으며 고장보고 기간을 2년, 1년, 6개월 으로 나누어 비교를 진행하였다. 고장보고기간에 따른 비교 결과를 통하여 고장보고기간이 길수록 베이지안 기법을 이용한 분석방법의 B1수명이 차이가 줄어들고 있음을 확인 할 수 있으나 B1수명 정확도가 떨어짐을 확인하였다. 또한 형상모수의 추정은 6개월 데이터로도 5년 고장데이터와 거의 같은 추정의 정확도를 보여주고 있다. 이제 같은 기간을 분석한 생명표법과 베이지안기법의 성능을 비교해보면 2년 고장데이터에 대한 분석결과는 형상모수, B1수명에 크게 차이가 없음을 확인하였으나, 1년/6개월 고장데이터의 경우는 베이지안 기법을 이용한 분석방법의 성능이 더 좋음을 확인할 수 있었다.

4.2 결 론

본 논문에서는 구간중도절단된 와이블 분포에서 모수들에 대한 정보 사전 분포를 가정한 후, 모수들의 사후분포를 통하여 모수를 추정하는 Weibull-Bayesian을 제시하였다. 이를 통해 소표본 보증데이터 분석방법에 대한 제안을 하고자 하였다. 또한 시뮬레이션을 통하여 보증데이터 신뢰성분석의 전통적 통계적 개념의 모수 추정방법인 생명표법과 베이지안 기법의 성능을 비교해 보았다.

첫째, 베이지안 기법을 이용한 보증데이터분석방법은 전통적인 분석보다 소표본일 때 더 정확한 추정을 보여주었다. 특히 1년 미만 고장데이터에는 성능이 좋다.

둘째, 베이지안 기법을 이용한 보증데이터분석방법은 형상모수 추정에 대한 정확도는 높은 편이나, 수명척도에 추정에서는 분산이 큼을 확인하였다.

셋째, 베이지안 기법을 이용한 보증데이터분석 방법을 통해 신제품을 출시 후 해당 부품의 보증기간이 길지 않더라도 형상모수와 수명을 전통적인 통계분석방법 보다 정확하게 추정이 가능하다.

넷째, 다음 연구과제는 베이지안 기법을 이용한 분석방법에서 수명척도의 분산을 줄이는 방법에 대한 연구가 필요하다.

5. 참 고 문 헌

[1] 김종걸, 북미 10년 10만 마일 보증 통계 분석, 성균관대학교 산학협력단, 산학협동 연구과제 최종보고서, 2012

[2] 김혜미, 보증데이터를 이용한 신뢰성 분석과 복사기 부품 사례 연구, 성균관대학교, 석사학위논문, 2012.

[3] Alexander Aron and Huairui Guo., “Improving the 1-Parameter Weibull: ABayesian

Approach", IEEE 2009 Proceedings Annual Reliability and Maintainability Symposium, 2009.
[4] Cohen, A. C., Whitten, B. J. and Ding, Y., "Modified Moment Estimation for the Three-parameter Weibull Distribution". J. Qual. Tech. vol.16, pp.159-167, 1984.
[5] Gelfan d, A. and Smith, A.F. M., "Sampling Based Approaches to Calculating Marginal Densities", Journal of the American Statistical Association, Vol.85, pp.398-409, 1990.
[6] Gelman. A, Carlin. J, Stern. H and Rubin, D, Bayesian Data Analysis, 2nd Edition, New York, 2004.
[7] Geman, S. and D. Geman., "Stochastic Relaxation, Gibbs Distribution and Bayesian Restoration of Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.6, pp.721-741, 1984.
[8] Hu, X. J., Lawless, J. F. and Suzuki, K., "Nonparametric Estimation of a Lifetime Distribution When Censoring Times Are Missing", Technometrics, Vol.40, pp.3-13, 1998.
[9] Kaplan, E. L. and Meier, P., "Non-parametric Estimation From Incomplete Sample", Journal of the American statistical Association, Vol.53, pp.457-481, 1957.
[10] Lawless, J. F., Statistical Models and Methods for Lifetime Data. Wiley, New York. 1982.
[11] Majeske, K. D. and Herrin, G. D., "Assessing Mixture-model Goodness-of-fit with an Application to Automobile Warranty Data", Proceedings of Annual Reliability and Maintainability Symposium, pp.378-383, 1995.
[12] Majeske, K. D. and Herrin, G. D., "Determining Warranty Benefits for Automobile Design Changes", Proceedings of Annual Reliability and Maintainability Symposium, pp.94-99, 1998.
[13] Martz, H. F., Waller, R. A., and Fickas, E. T., "Bayesian Reliability Analysis of Series Systems of Binomial Subsystems and Components", Technometrics, vol.30, pp.143-154, 1998.
[14] Soland, R. M., "Bayesian Analysis of the Weibull Process with Unknown Scale and Shape Parameters" IEEE Transactions on Reliability, vol.18(4), pp.181-184, 1969.
[15] Soland, R. M., "Bayesian Analysis of the Weibull Process with UnKnown Scale Parameter and Its Application to Acceptance Sampling", IEEE Transactions on Reliability, pp.84-90, 1968
[16] Shepherd. J. D., "V-belt Reliability: A Statistical Study of Large Sample Size Fatigue Tests", Society of Automotive Engineers (SAE), 1980.
[17] Willits, C. J. Dietz, D. C.,Moore A. H., "Series-system Reliability-Estimation Using Very Small Binomial Samples", IEEE Trans. Reliability, vol.46, no.2, pp.296-302, 1997.