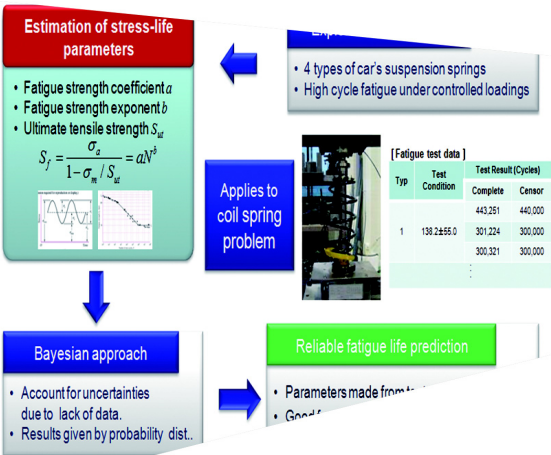


신뢰성분석 및 설계를 위한 베이지안 통계기법 Application of Bayesian Methods in Reliability Problems

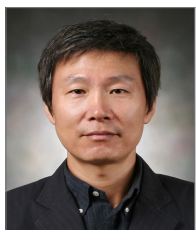


최근 각종 공학문제의 해결을 위해 결정론적(deterministic) 관점 보다는 각종 불확실성을 고려한 확률적(probabilistic) 또는 추계적(stochastic) 관점에서 분석 평가하려는 노력이 증가하고 있다. 이는 신뢰성분석 및 설계 측면에서 특히 중요하게 인식되고 있으며, 이를 구현하는 방법론으로 베이지안 통계기법(Bayesian statistics)이 최근 많은 주목을 받고 있다. 본 글에서는 이러한 기법을 신뢰성평가 및 고장예지에 어떻게 적용하고 있는지 사례를 중심으로 설명하고자 한다.

1. 베이지안 이론¹⁾

현대의 통계이론은 크게 두가지 방향에서 연구·확립되어 왔으며, 이에 따라 확률도 두가지로 나누어 설명하고 있다. 하나는 고전 통계학에서 얘기하는 frequentist적 정의로서 어떤 사상(event)이 발생할 확률을 수많은 반복 결과 발생한 객관적 횟수로 정의한다. 다른 하나는 Bayesian 통계학에서 얘기하는 것으로 확률을 우리가 알고 있는 주관적 믿음의 정도로 정의한다. 이것을 가장 잘 설명하는 예로서 19세기 초 Laplace가 연구했던 토성의 질량 추정문제가 있다. 고전적 관점에서 이것을 추정한다면, 토성의 질량은 측정이 불가능할 뿐이지, 측정할 수만 있다면 아무리 반복해도 동일한 값으로 생각한다. 반면에 Laplace는 이것을 베이지안 관점에서 접근했는데, 이 경우 토성의 질량은 우리가 알고 있는 부족한 관찰정보를 토대로 확률분포로 표현된다. 그 결과 실제로 Laplace는 그림 1과 같은 확률분포를 제시한 바 있다.

베이지안 이론의 핵심은 베이 법칙(Bayes' rule)이다. 이에 의하면 어떤 미지변수의 불확실성은 확률분포로 표현되며, 이는 과거의 경험에 기초한 주관적(subjective) 사전 지식(prior knowledge)과 현재의 데이터에 기반한 객관적(objective) 우도(likelihood)의 곱으로 주어진다. 이를 식으로 표현하면 그림 2(a)와 같으며, 사전분포(prior distribution) $p(\theta)$ 와 데이터 x 의 우도 $L(x|\theta)$ 의 곱에 의해 미지변수 θ 의 사후분포 $p(\theta|x)$ 가 얻어짐을



최 주 호

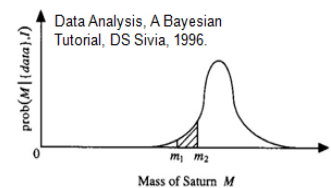
한국항공대학교 항공우주 및 기계공학부 교수



Always single value after many trials ?

Not suitable to tackle this problem

Classical approach



Bayesian approach

그림 1 토성질량 추정문제에 대한 베이지안 접근법

나타내고 있다. 이로 인해 베이지안 접근법에서는 객관적 데이터와 주관적 지식을 하나의 프레임워크 내에서 통합 고려하는 것이 가능하다. 그림 2(b)는 θ 의 사전분포가 측정데이터 x 에 의해 사후분포로 업데이트된 것을 보이고 있으며, 새로운 x 가 추가되면 기존의 사후분포는 다시 사전분포가 되면서 새로운 사후분포를 얻게 되는데, 이를 베이지안 업데이트(Bayesian updating)이라고 한다.

베이 법칙은 귀납(inductive)적 추론과정을 가능하게 한다. 이의 반대 과정은 연역(deductive) 과정인데, 이는 원인(cause) 또는 가정(hypothesis)이 주어졌을 때 결과 또는 측정치(observation)를 얻는 순과정(forward process)을 뜻한다. 반면에 귀납 과정은 결과데이터가 주어졌을 때 원인(cause)을 역으로 추정하는 역과정을 뜻하며, 이는 아래의 베이 법칙을 이용하여 구현한다.

$$P(\text{hypothesis} \mid \text{data}) \propto P(\text{data} \mid \text{hypothesis}) \cdot P(\text{hypothesis})$$

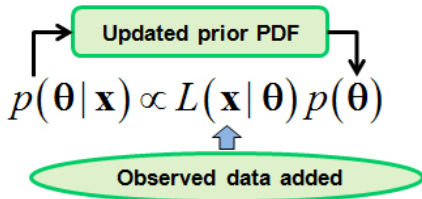
연역과정과 귀납과정을 그림 3에서 비교하였다. 그림 3(a)에서는 모수 θ 가 주어졌을 때 이로부터 모수 θ 를 가지는 정규분포 데이터 X 를 결과로 얻음을 보이고 있고, 그림 3(b)는 이의

역과정으로서 결과 데이터가 있을 때 이로부터 미지 모수 θ 를 추정하는 과정을 보이고 있다.

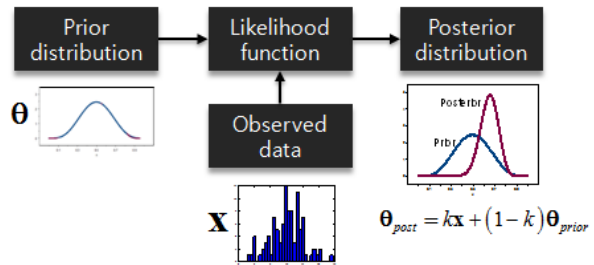
베이지안 이론은 1763년 Thomas Bayes라는 목사가 최초로 제안하였고, 1812년 Laplace에 의해 발견되어 여러 문제에 응용되기 시작하였다. 그러나 베이지안 확률개념의 모호성과 주관성으로 인해 이후 1백년 동안 고전 통계학만 연구되었다. 그러다가 1939년 Jeffrey라는 통계학자에 의해 재발견, 현대적 통계이론이 완성되었다. 그러나 이 역시 실제 응용보다는 이론에 머물렀는데 이유는 연산에 매우 많은 비용이 필요하다는 것이었다. 다행히 1990년대 부터 HW, SW의 발달에 힘입어 베이지안 통계학은 급속한 발전을 맞이하게 되었고, 최근에 의학, 경제학은 물론 각종 공학 분야에서 활발히 연구·적용되고 있다.

2. 동전 던지기 예제²⁾

베이지안 이론을 손쉽게 이해하기 위해 가장 많이 설명되는 것으로 동전 던지기 문제가 있다. 통상 동전 앞면이 나올 확률은 1/2이나, 약간 특이하게 생긴 동전이라면 그 값이 무엇인지는 던져 봐야 알 수 있다. 그러나 이를 위해 무한회의 던지기를 할 수 없는 노릇이라면 제한된 횟수의 결과를 통해 미지의 앞면

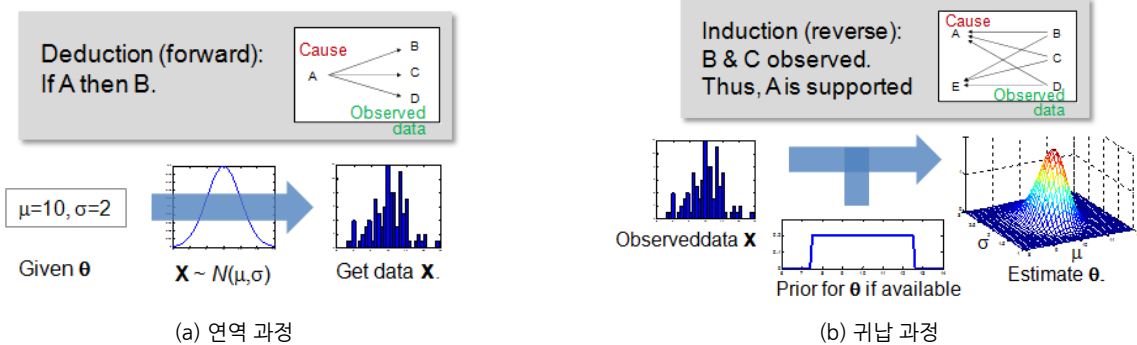


(a) 베이 법칙



(b) 사전분포가 데이터에 의해 사후분포로 업데이트

그림 2 베이 법칙 및 베이지안 업데이트



(a) 연역 과정

(b) 귀납 과정

그림 3 연역과정과 귀납과정

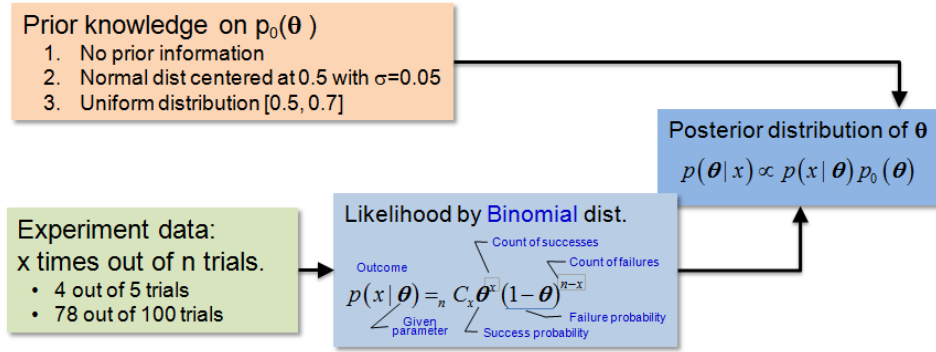


그림 4 동전 던지기 문제의 베이지안 접근법

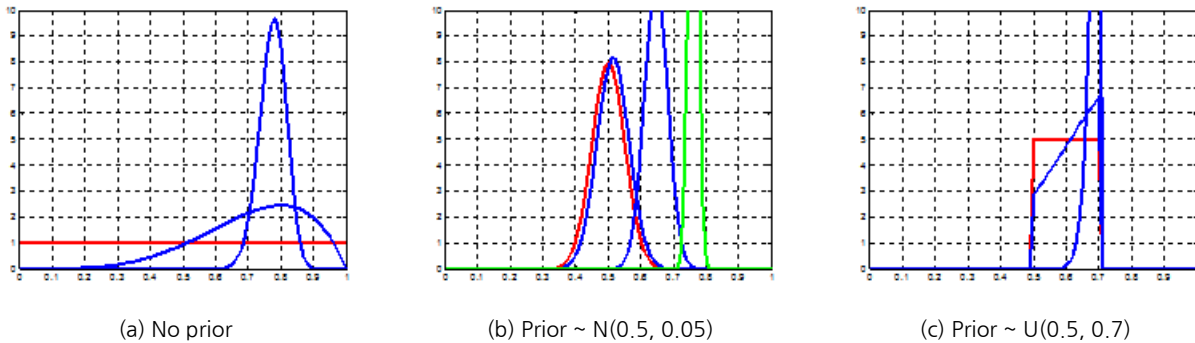


그림 5 동전 던지기 문제의 베이지안 추론해

확률을 추정할 수 밖에 없다. 이것을 베이지안 관점에서 사후 확률분포로 추정하고자 한다. 이를 위한 과정을 그림 4에 보였다.

본 예제에서 앞면확률의 참값이 0.78이라고 가정하자. 그러나 이를 모른다고 하고 유한한 데이터에 기반하여 추정하자. 만일 앞면확률에 대한 사전지식이 있다면 이를 활용하는 것이 도움 되는데, 이를 위해 세가지를 고려한다. 첫째는 사전지식이 없는 경우, 둘째는 대체로 앞면확률이 평균 0.5, 표준편차 0.05의 정규 분포일 것이라고 생각하는 경우, 셋째는 앞면확률이 적어도 0.5와 0.7 사이에서만 존재한다고 믿는 경우이다. 측정데이터는 두가지가 있다. 하나는 적은 데이터로서 5회 실시하여 4회가 앞면이 나왔고, 다른 하나는 많은 데이터로서 100회 실시하여 78회 앞면이 나왔다.

이에 대해 베이지안 이론을 적용하면 미지의 앞면확률에 대해 그림 5와 같은 결과를 얻는다. 그림 5(a)는 사전지식이 없는 경우로 처음에는 0과 1사이의 균일분포로 주어지며, 적은 데이터를 적용하면 최대치가 0.8 근방에 있는 넓게 퍼진 확률분포가, 많은 데이터를 적용하면 0.78 근방에서 더욱 뾰족해진 확률분포가 얻어진다. 즉 측정데이터가 많아 질수록 참해인 0.78 근방으로 분포가 집중되며, 이것이 미지확률에 대한 우리 믿음의

정도를 나타내고 있다. 그림 5(b)에서는 사전지식이 0.5 근방의 정규분포로 표현된 경우로서, 데이터가 증가하면서 확률분포가 우측으로 이동하고 있으나 5(a)에 비해 많이 왼쪽에 위치하고 있는데 이는 잘못된 사전정보 탓이다. 매우 많은 데이터(천회 시도)를 적용하고 나서야 비로소 참해에 근접한 결과를 얻는다. 그림 5(c)에서는 사전지식을 0.5~0.7 근방에만 국한하는 바람에 데이터를 아무리 많이 적용해도 0.7 벽을 넘지 못하고 있다. 이로부터 사전정보의 올바른 적용이 변수의 추정에 큰 영향을 미침을 알 수 있다.

이렇게 하여 앞면확률을 사후분포로 추정하고 나면, 이를 가지고 여러 가지 시나리오에 대한 사후예측(Posterior prediction)을 할 수 있다. 예를 들어 θ 의 사후분포가 그림 5(a)의 세 번째 것이라고 하자. 이를 바탕으로 새로 5회 동전 던지기를 한 결과 모두 앞면일 확률은 얼마인지 예측하고 싶다. 만일 참해 $\theta = 0.78$ 을 알고 있다고 한다면, 그 결과는 $0.78^5 = 0.289$ 로서 단일값이 될 것이다. 그러나 정답을 모른채 θ 에 대해 그림 5(a)와 같은 확률분포로 알고 있다면, 이 때는 이 분포를 따르는 θ 의 샘플을 1만개 생성한 후 각 θ 에 대해 모두 앞면 나올 확률인 0.78^5 를 일일이 구하면, 이렇게 구해진 1만개의 샘플이 사후 예측된 확률분포가 되는 것이다. 이를 그림 6에 보였다.

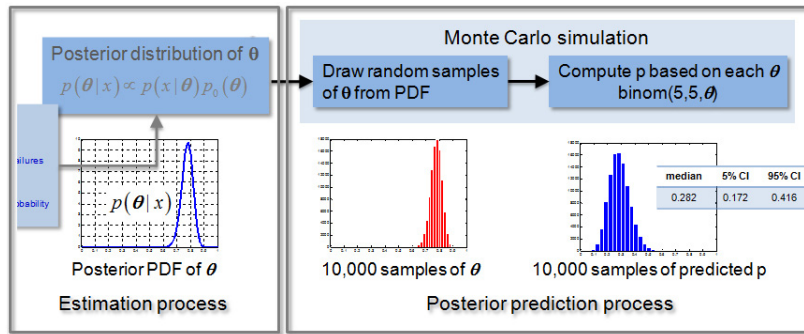


그림 6 사후 예측분포의 과정

3. 케이스 스터디 : 자동차 서스펜션 스프링의 피로수명예측^{3),4)}

서스펜션 스프링의 피로해석은 자동차 안전을 위해 매우 중요한 역할을 한다. 그러나 피로물성의 불확실성으로 인해 해석결과를 신뢰하기 어렵고, 따라서 확인을 위해 결국 많은 비용과 시간이 소모되는 수명시험을 수행한다. 본 연구에서는 이러한 시험 데이터를 가지고 베이지안 접근법을 활용하여 물성치를 역추정

하는 과정을 소개한다. 이렇게 추정된 물성을 가지고 해석을 수행한다면 그 결과는 더욱 신뢰할 수 있을 것이다. 한편 이러한 수명시험은 어차피 개발과정에서 항상 거치는 과정이기에 데이터는 지속적으로 축적된다. 따라서 이것이 축적될수록 해석은 더욱 믿음직스러울 것이다. 이를 그림 7에 보였다. 스프링의 피로해석에는 반복하중 하에서 대변위 및 접촉이 수반된 비선형해석을 통한 응력해석이 포함된다. 전체적인 피로해석 과정과 응력해석을 위한 스프링 모델, 해석결과 최대응력 지점 등

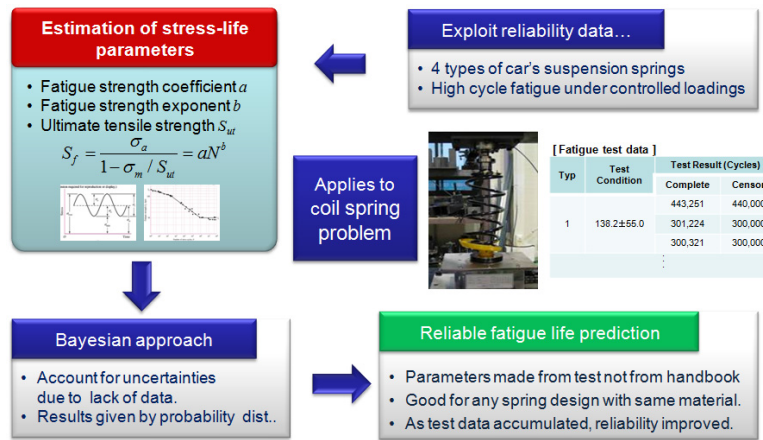


그림 7 스프링 물성추정을 위한 베이지안 추론과정

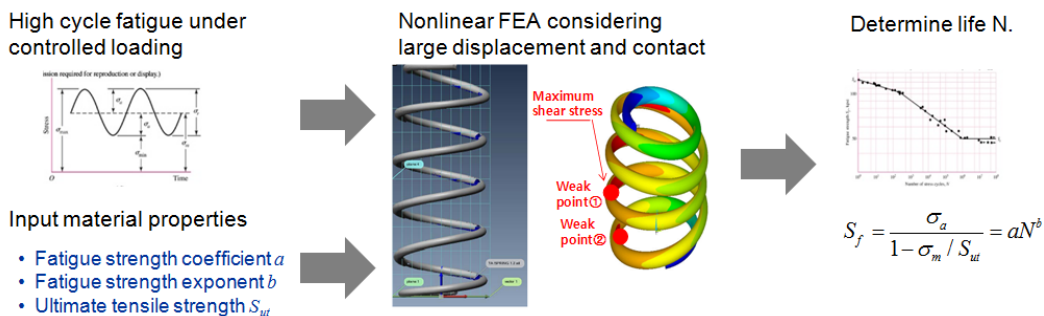


그림 8 스프링 피로해석의 과정

Spring type	Test condition	Mean stress (MPa)	Alternating stress (MPa)	Result (cycles)	
				Complete data	Censored data
A	138.2±55.0	1142.97	290.12	1.1923	1.1287
				1.0958	1.0954
				1.0955	1.0954
				1.1025	1.1432
Total 65 data Case 1 : Using 14 data Case 2 : Using 33 data (adding 17 more data) Case 3 : Using 65 data (adding 32 more data)				0.9917	1.1134
⋮					
M	146.8±57.1	1288.98	339.15	1.1015	1.1010
				1.1011	1.1010
N	154.4±58.3	1127.14	330.43	1.1142	1.1136

그림 9 스프링 수명시험 결과

을 그림 8에 보였다.

스프링 수명시험 데이터는 해당 회사에서 2년간 총 14종에 대해 수행한 65개가 확보되었으며, 이를 수집 시점에 따라 세 케이스로 구분하였고, 그림 9에 보였다. 이들은 모두 동일재료, 동일 열처리를 거친 스프링으로, 크기나 작용하중만 다를 뿐이다. 그림에서 complete는 파손이 발생한 것을, censor는 해당 사이클 까지 파손되지 않아 종료된 것을 뜻한다. 사이클의 단위는 보안 상 임의 값으로 나누었다. 이러한 데이터에 기반하여 미지의 피로관련 물성 및 시험오차를 베이지안 추론기법에 의거하여 확률분포로 추정된 결과가 그림 10에 있다. 여기에 있는 확률

분포는 측정된 수명데이터에 기반하여 추정된 물성치를 확률 분포로 나타낸 것이다. 수명 데이터가 적은 case1에서 많은 case3으로 가면서 물성치의 확률분포가 뾰족해지고 있으며, 데이터가 계속 쌓이면 그 신뢰도가 더욱 향상될 것임을 알 수 있다. 특이한 것은 이렇게 추정된 물성치가 문헌에서 제시된 값(그림에서 deterministic value로 표시)과 큰 차이를 보인다는 것이다. 즉 피로해석을 수행할 때 단순히 문헌값을 가져와서 해를 구하는 것이 실제 해를 구하는데 도움이 되지 않음을 말한다.

이렇게 추정된 물성치는 확률분포로 표현되어 있다. 이 물성치를 다시 14종의 스프링에 적용하여 피로해석을 수행하면,

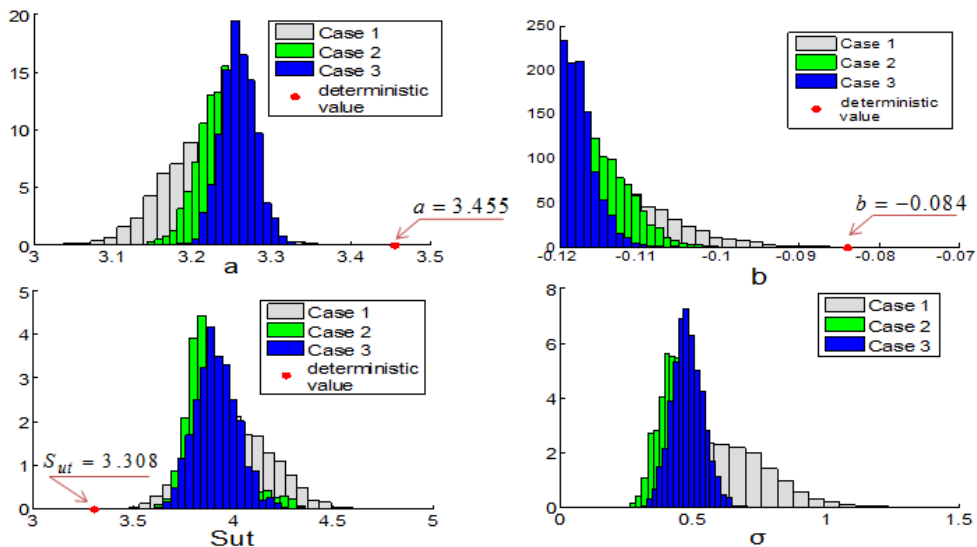
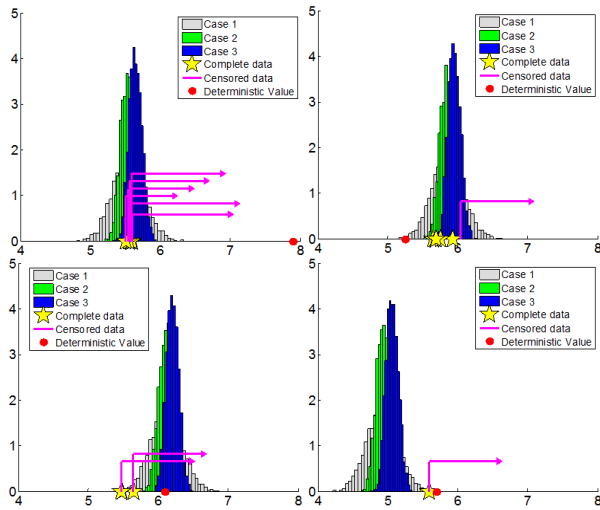


그림 10 스프링 피로물성 추정결과



(a) 대표적 스프링의 피로수명 확률분포

Spring type	Result of fatigue life test (cycles)				B ₁₀ Life
	Complete data		Censored data		
A	1.1293	1.0955	1.1287	1.0954	1.1196
	1.0958	1.0955	1.0954	1.0954	
B	1.1025	0.9917	1.1432	1.1113	1.0535
	1.1119	1.1298	1.1113		
	1.1294	1.1134			
	1.1116				
	1.1862	1.1514	1.2083		
- B ₁₀ life is 10 th percentile value of the life distribution. 10 percent of the data may fail before B ₁₀ life. - 4 out of 65 test data (red part) fails earlier than the expected B ₁₀ life. Favorable number according to the definition.					
	1.1097	1.1020	1.1088	1.1010	
G	1.1190		1.1182		0.8904
H	1.1365	1.1362	1.2000	1.2000	1.2133
I	1.1142		1.1136		1.0216
J	1.1121		1.1113		1.0192
K			1.1204	1.1204	1.4979
L	1.1291		1.1287		1.1771
M	1.1539		1.2000	1.2009	1.1088
N	1.1734		1.2000		1.0380

(b) 예측된 B₁₀수명을 실제 수명시험결과로 검증.

그림 11 추정된 물성을 이용하여 스프링 피로해석을 수행한 결과

그 결과 얻어지는 수명 역시 확률분포로 표현된다. 이것과 실제 수명시험에서 얻은 시험값을 비교함으로써 추정된 수명에 대한 검증할 수 있다. 그림 11에 그 결과를 수록하였다. 그림 11(a)에는 14종의 스프링 중 네 종에 대한 수명분포와 실제 시험결과를 함께 표시하였는데, 위의 두 스프링은 예측된 확률분포 내에 실제 시험결과(노란 별표)가 잘 위치한 반면, 아래 두 스프링은 실제 수명이 예측된 확률분포보다 더 짧거나 긴 결과를 보이고 있다. 문헌값을 통해 구해진 수명(붉은 점)과도 비교했는데 확률분포의 중앙에 잘 위치한 경우도 있지만 아예 일치하지 않거나 꼬리부분에 위치하기도 하여 일관되지 못하였다. 그림 11(b)는 예측된 확률분포로부터 B₁₀수명을 구하고 이를 시험결과와 비교한 표이다. 대부분 B₁₀수명 보다 실제 수명이 더 큰 반면 일부 그렇지 않은 경우(붉은 색칠한 칸)도 있었으나, 전체 시험데이터 대비 10% 수준에 불과하였다.

4. 참고문헌

1. Sivia, Devinder S. Data analysis : a Bayesian tutorial. Oxford university press, 1996.
2. Gelman Andrew et al. Bayesian data analysis. Vol. 2. London: Chapman & Hall/CRC, 2014.
3. 허찬영 등, “베이지안 접근법을 이용한 스프링 피로 수명 파라미터의 역 추정”, 대한기계학회 논문집 A 35권 4호 (2011): 393-400.
4. Dawn An, Joo-Ho Choi, “Improved MCMC method for parameter estimation based on marginal probability density function.” Journal of Mechanical Science and Technology 27.6 (2013) : 1771-1779.

[섭외: 이진학 편집위원]