J. Comput. Struct. Eng. Inst. Korea, 30(1)
pp.87~94, February, 2017
https://doi.org/10.7734/COSEIK.2017.30.1.87

pISSN 1229-3059 eISSN 2287-2302
Computational Structural
Engineering Institute of Korea

모수적 · 비모수적 입력모델링 기법을 이용한 신뢰성 해석

강 영 진[1] · 홍 지 민[1] · 임 오 강[1] · 노 유 정[1†]

[1]부산대학교 기계공학부

Reliability Analysis Using Parametric and Nonparametric Input Modeling Methods

Young-Jin Kang[1], Jimin Hong[1], O-Kaung Lim[1] and Yoojeong Noh[1†]

[1]*School of Mechanical Engineering, Pusan Nat'l Univ., Busan, 46241, Korea*

Abstract

Reliability analysis(RA) and Reliability-based design optimization(RBDO) require statistical modeling of input random variables, which is parametrically or nonparametrically determined based on experimental data. For the parametric method, goodness-of-fit (GOF) test and model selection method are widely used, and a sequential statistical modeling method combining the merits of the two methods has been recently proposed. Kernel density estimation(KDE) is often used as a nonparametric method, and it well describes a distribution function when the number of data is small or a density function has multimodal distribution. Although accurate statistical models are needed to obtain accurate RA and RBDO results, accurate statistical modeling is difficult when the number of data is small. In this study, the accuracy of two statistical modeling methods, SSM and KDE, were compared according to the number of data. Through numerical examples, the RA results using the input models modeled by two methods were compared, and appropriate modeling method was proposed according to the number of data.

Keywords : *kernel density estimation, nonparametric statistical modeling, parametric statistical modeling, reliability analysis, sequential statistical modeling*

1. 서 론

자동차, 조선, 항공, 건축 등의 대부분의 공학문제에서는 예측하지 못하는 불확실성(uncertainty)이 내재되어 있다. 최근에 이러한 불확실성을 고려한 통계기반 해석기술의 발달로 인해 다양한 분야에서 시스템의 성능을 통계적으로 해석하고 설계하는 신뢰성 해석, 신뢰성 기반 최적설계(reliability-based design optimization, RBDO) 등에 활발하게 적용되고 있다. 통계적 해석을 위해서는 확률변수에 대한 분포함수의 형태가 정의되어야 하며, 통계적 특성을 나타내는 통계적 모멘트(statistical moments) 또는 통계적 모수(statistical parameter) 등을 사용하여 정량화되어야 한다(Noh *et al.*, 2010). 위와 같이 확률적 분포함수를 이용하여 불확실성을 정량화 및 정의하는 것을 통계모델링(statistical modeling)이라고 하며, 통계모델링 방법은 통계적 모수의 사용 유무에 따라 모수적 · 비모수적 통계모델링 방법으로 구분된다.

대표적인 모수적 기법에는 적합도 검정(goodness of fit test), 모델 선택법(model selection method) 등이 있으며, 최근 모수적 통계모델링 기법으로 적합도 검정과 모델선택법의 장점을 결합한 순차적 통계모델링(sequential statistical modeling, SSM)방법(Kang *et al.*, 2016)이 제안되었다. 비모수적 기법에는 증거이론(evidence theory or Dempster-Shafer theory), 커널밀도추정(kernel density estimation, KDE) 등이 있다. 하지만 기존 연구에서는 모수적 · 비모수적 통계모델링 방법의 정확도(Kang *et al.*, 2016; Noh, 2016) 또는 신뢰성 해석의 측면에서 개별적으로 연구(Lee *et al.*,

† Corresponding author:
Tel: +82-51-510-2308; E-mail: yoonoh@pusan.ac.kr
Received December 27 2016; Revised January 2 2017;
Accepted January 3 2017
©2017 by Computational Structural Engineering Institute of Korea

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(http://creativecommons.org/licenses/by-nc/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

2013; Lim *et al.*, 2012)가 수행되었고, 모수적 · 비모수적 통계모델링의 정확도가 신뢰성 해석결과에 미치는 영향에 대한 연구는 부족하다. 또한 기존 연구에서는 데이터의 개수에 따른 통계모델링 방법의 정확도와 신뢰성 해석에 관한 연구가 충분하지 않아 입력 변수에 따른 제한된 정보나 데이터를 가진 경우에 적합한 통계모델링 방법에 대한 가이드라인이 부족한 실정이다.

본 연구에서는 통계시뮬레이션을 이용해 모수적 밀도추정 방법인 SSM과 비모수적 밀도추정 방법인 KDE를 사용하여 데이터의 개수에 따른 통계모델의 추정 정확도를 비교하였다. 또한 입력모델링기법에 따른 신뢰성 해석의 결과를 분석하기 위해서 수학적 신뢰성 해석 모델인 Lognormal ratio 문제와 수직-수평하중을 받는 외팔보 문제에 대해서 신뢰성 해석을 수행하였다. 입력모델링기법에 따른 추정 정확도와 신뢰성 해석의 결과를 토대로 데이터의 개수 및 시스템의 불확실성 정도에 따라 적합한 입력모델링기법을 제안하였다.

2. 통계모델링 방법

2.1 순차적 통계모델링

순차적 통계모델링인 SSM 방법은 적합도 검정과 모델 선택법을 결합한 방법이다(Kang *et al.*, 2016). 적합도 검정은 후보모델(candidate model)들이 주어진 데이터를 표현하기 적절한지 채택 혹은 기각의 두 가지 조건으로 절대적인 평가가 가능하지만 채택된 모델간의 적합도의 순위를 평가하지 못하고, 모델선택법은 주어진 데이터를 가장 잘 표현하는 순서대로 후보모델의 우선순위를 상대적으로 평가할 수 있지만 절대적인 평가는 할 수 없다. SSM기법은 서로 다른 특성을 가지는 두 가지 기법을 결합하여 서로의 단점을 보완한 방법이다. 즉, 적합도 검정의 절대적인 방법을 통해 선정된 후보모델의 적합성 평가와 부적합한 모델을 1차로 제거하고, 적합성을 만족하는 모델에 대해 모델선택법을 사용하여 주어진 데이터를 잘 표현하는 순위를 매겨서 최우선 순위의 모델을 최종적으로 선택한다. 본 연구에서는 사전 비교연구를 통해 Anderson-Darling(A-D), Chi-squared, Kolmogorov-Smirnov(K-S) 검정 중 가장 정확한 K-S 검정과 Akaike information criterion(AIC), Akaike information criterion correction(AICc), Bayesian information criterion(BIC) 방법 중 가장 정확한 BIC 방법을 SSM 기법에 사용하였다(Kang *et al.*, 2016).

2.1.1 Kolmogorov-Smirnov test

K-S 검정은 주어진 데이터가 후보분포함수로부터 추출되었다는 귀무가설(null hypothesis)을 검정통계량을 사용하여 채택 또는 기각을 하는 가설검정(hypothesis test) 방법이다. K-S 검정 통계량(D_n)은 가정된 분포함수의 누적분포함수(cumulative distribution function, CDF)와 데이터의 경험누적분포함수(empirical cumulative distribution functions, eCDF)사이의 최대 차이를 나타내는 것으로, 만약 통계량이 데이터의 개수(n)와 유의수준(significance level, α)에 의해 결정되는 임계값(D_n^{α}) 보다 작으면 귀무가설은 채택되고, 반대의 경우에는 가설은 기각된다. 일반적으로 임계값은 데이터와 관계없이 정의된 값을 사용하며(Ayyub *et al.*, 2012), K-S 검정 통계량은 다음과 같이 나타낸다.

$$D_n = sup\ |F_n(x) - F(x)| \tag{1}$$

$$F_n = (1/n)\sum_{i=1}^{n} I_{X_i \le x} \tag{2}$$

여기서, $F_n(x)$은 데이터의 경험누적분포함수이고, $F(x)$는 검정에 사용되는 후보모델의 누적분포함수이다. X_i는 확률변수 X의 i번째 데이터 값이고, $I_{X_i \le x}$은 지시함수(indicator function)로서 $X_i \le x$이면 1의 값을 가지고, 다른 경우 0의 값을 가진다.

2.1.2 Bayesian information criterion

BIC는 후보모델들에 주어진 데이터를 사용하여 BIC값을 계산한 후 작은 값을 가지는 순서로 순위를 선정하여 가장 작은 값을 가지는 모델을 가장 적합한 모델로 선정하는 방법이다. BIC는 AIC와 유사하게 최대우도함수(maximum likelihood function)의 값을 사용하나 데이터의 개수와 후보모델의 모수의 개수를 보정계수로 사용하여 AIC의 단점을 보완한 방법이다. BIC값은 다음과 같이 계산된다(Schwarz, 1978).

$$BIC = -2\ln(L) + k\ln(n) \tag{3}$$

$$L = \max\left\{\prod_{i=1}^{n} f(x_i|\boldsymbol{\theta})\right\} \tag{4}$$

여기서, L은 우도함수의 최댓값이고, k는 후보모델에서 모수의 개수, n은 데이터의 개수이다.

2.1.3 순차적 통계모델링의 절차

순차적 통계모델링은 적합도 검정을 사용하여 주어진 데이터를 표현하기 부적합한 모델을 제거하여 후보모델을 축소

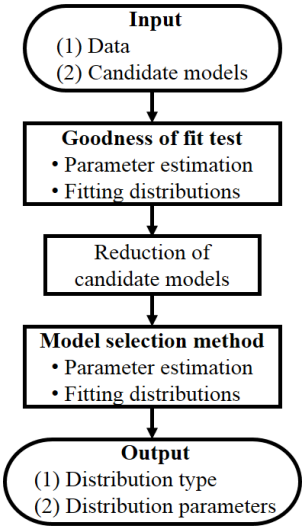

Fig. 1 Process of SSM method

시키고, 축소된 후보모델에 대해 모델선택법을 사용하여 후보모델간의 순위를 선정하여 최우선 순위의 후보모델을 가장 적합한 모델로 선정한다.

Fig. 1은 순차적 통계모델링의 절차를 나타낸다. 먼저 부적합한 후보모델을 제거하기 위해 적합도 검정을 수행하며, 적합도 검정에서 후보모델의 모수는 최대우도추정(maximum likelihood estimation, MLE)방법을 사용하여 추정한다. 순차적 통계모델링 방법은 적합도 검정결과 기각된 분포함수를 제외시켜 후보모델을 축소하고, 축소된 후보모델에 대해 모델선택법을 사용하여 주어진 데이터를 나타내기에 적합한 모델에 대한 우선순위를 정하여 최우선 순위로 선정된 모델을 가장 적합한 모델로 선정한다. 마지막 단계에서는 최종적으로 선정된 모수적 분포함수의 모수를 추정하게 된다.

본 논문에서는 통계시뮬레이션과 신뢰성 해석을 위해 Birnbaum-Saunders(BS), Exponential, Extreme value, Gamma, Generalized extreme values(GEV), Logisitic (LOG), Log-logistic, Log-normal(LOGN), Nakagami, Normal(NORM), Rayleigh(RAY), t location-scale, Weibull(WBL) 분포함수를 후보모델로 사용하였다.

2.2 커널밀도추정

커널밀도추정(KDE)은 비모수적 밀도추정방법으로서, 통계적 모수 및 모수적 분포함수를 사용하지 않고 오직 주어진 데이터만을 사용하여 입력확률변수의 확률밀도함수를 추정한다. 즉, KDE는 밀도함수의 추정 시, 통계적 모멘트와 모수 및 분포형태와 같은 사전정보가 필요없고 각 데이터에서 생성된 커널함수(kernel function)를 결합하여 밀도함수를 추정한다. KDE는 데이터만을 사용하기 때문에 확률변수가 정의된 모수적 확률밀도함수로 표현되기 어려운 경우 유용한 방법이다. 즉, 데이터의 개수가 극단적으로 적어서 모수적 함수로써 표현이 어렵거나 분포형태가 다봉 분포함수(multimodal distribution function)인 경우에 유용하다(Kang *et al.*, 2016). 추정된 커널밀도함수는 다음과 같다(Silverman, 1986).

$$\hat{f}(x)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right) \tag{5}$$

식 (5)를 재수정 기호 $K_h(t)=(1/h)K(t/h)$를 사용하여 간단하게 나타내면 다음과 같다.

$$\hat{f}_h(x)=\frac{1}{n}\sum_{i=1}^{n}K_h(x-X_i) \tag{6}$$

여기서, X_i는 주어진 데이터이고 $\hat{f}(x)$는 추정된 밀도함수이고 $K(\bullet)$는 커널함수이다. 그리고 h는 커널함수의 대역폭(bandwidth)으로서, 각 커널함수의 폭을 조절하는 평활모수(smoothing parameter)이다. 커널밀도추정에서 커널함수의 종류보다는 최적의 대역폭을 선정하는 것이 정확도에 중요하기 때문에(Wand *et al.*, 1994) 본 연구에서는 최적의 대역폭에 대한 수학적 식에 정의되어 있고 효율성으로 인해 많이 사용되는 가우시안 커널(Gaussian kernel)을 사용하였고, 최적의 대역폭의 선정하는 수식으로 Silverman's rule을 사용하였다(Silverman, 1986; Scott, 1979). 가우시안 커널함수의 수식과 수정된 가우시안 커널함수는 다음과 같다.

$$K(t)=\frac{1}{\sqrt{2\pi}}exp\left\{-\frac{1}{2}t^2\right\} \tag{7}$$

$$K_h(t)=\frac{1}{\sqrt{2\pi}}exp\left\{-\frac{1}{2}\left(\frac{t}{h}\right)^2\right\} \tag{8}$$

그리고 최적의 대역폭을 선정하는 Silverman's rule은 다음과 같다.

$$h^*=\left(\frac{4}{3}\right)^{1/5}\hat{\sigma}n^{-1/5} \tag{9}$$

여기서, $\hat{\sigma}$은 추정된 표준편차로서 본 논문에서는 강건 통계학(robust statistics)을 적용한 표준편차의 강건한 추정치(robust estimate)를 사용하였고 수식은 다음과 같다(Analytical Methods Committee, 1989).

$$\hat{\sigma} = \frac{Median(|X_i - Median(X_i)|)}{0.6745} \tag{10}$$

3. 통계시뮬레이션

SSM과 KDE의 통계모델링 추정 정확도 비교를 위해 분포형태와 모수의 개수가 다른 BS, GEV, LOG, LOGN, NORM, RAY, WBL 분포함수를 실제분포로 가정하였다. Fig. 2는 7개의 모집단의 확률밀도함수(probability density function, PDF)를 나타내며 범례에서 괄호 안의 숫자는 분포함수의 모수의 값을 나타낸다. Table 1은 7개 분포함수의 통계적 모멘트를 나타내었다.

모수적 • 비모수적 기법을 비교하기 위해서는 두 가지 방법 모두의 추정 정확도를 정량화할 수 있는 척도가 필요하다. 본 연구에는 면적척도방법인 교차면적을 이용하여 모델링의 정확도를 평가하였다(Kang *et al.*, 2016). 교차면적은 두 확률밀도함수의 겹치는 면적을 계산하는 것으로 두 분포함수가 완전히 일치하면 1을 가지고 완전히 불일치하면 0을 가진다.

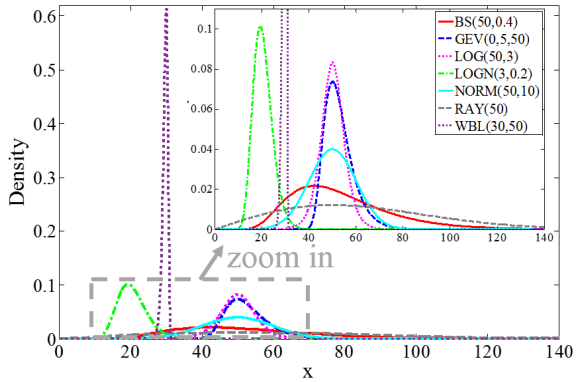

Fig. 2 PDFs of assumed populations

Table 1 Statistical moments of true distributions

Distribution	Mean	Standard dev.	Skewness	Kurtosis
BS(50,0.4)	53.93	21.85	1.19	5.34
GEV(0,5,50)	52.90	6.41	1.14	5.38
LOG(50,3)	50.02	5.44	0	4.12
LOGN(3,0.2)	20.50	4.14	0.61	3.63
NORM(50,10)	50.02	9.98	0	3
RAY(50)	62.62	32.79	0.64	3.24
WBL(30,50)	29.67	0.75	-1.02	4.87

3.1 추정 정확도의 비교

가정된 7개의 실제 모델로부터 데이터의 개수(n)를 3, 5, 7, 10, 20, 30, 50, 100, 200, 300까지 증가시켜가면서 각 1,000개의 데이터 세트를 생성하였다. 생성된 데이터 세트에 대해 SSM과 KDE를 사용하여 밀도함수를 추정하고 추정된 밀도함수와 모집단의 확률밀도함수간의 교차면적을 계산하여 데이터의 개수에 따른 SSM과 KDE의 통계모델링 정확도를 검증하였다. Fig. 3~5는 1,000개의 데이터 세트에 대해 SSM과 KDE를 사용하여 추정된 밀도함수와 모집단과의 평균 교차면적을 y축에 표시하고, x축에 데이터의 개수를 로그-스케일로 나타내었다.

Fig. 3은 모수의 개수가 1개인 RAY 분포에 대한 교차면적을 나타내었다. 샘플 수가 적은 경우($n \le 5$) KDE가 SSM에 비해 실제 모델의 추정 정확도가 더 높은데, 그 이유는

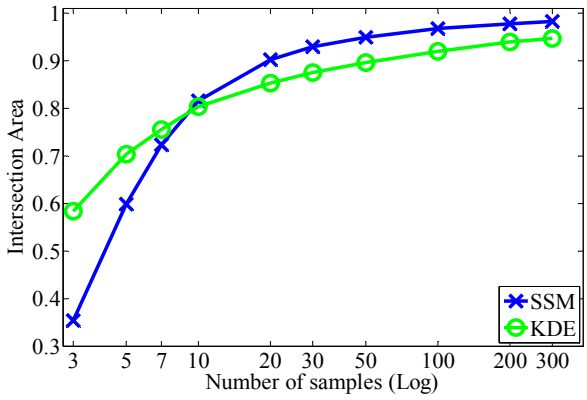

Fig. 3 Averaged intersection areas of RAY dist.

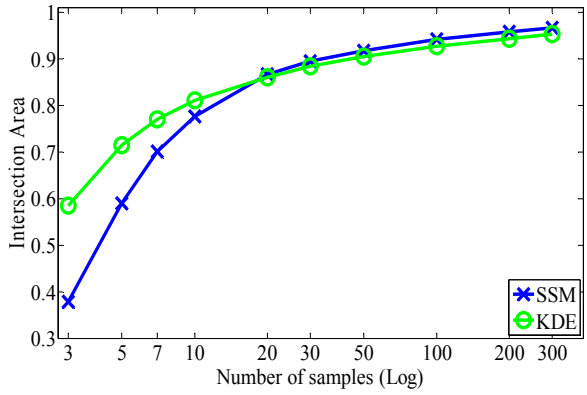

Fig. 4 Averaged intersection areas of NORM dist.

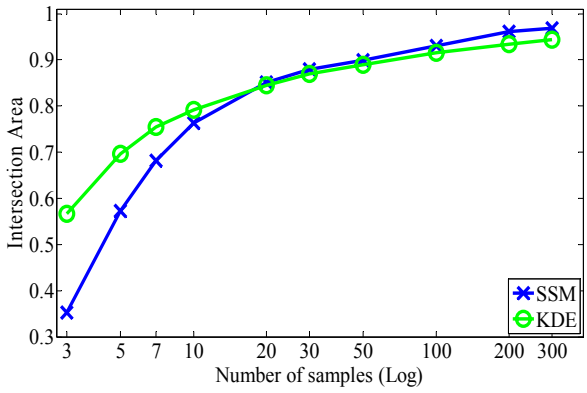

Fig. 5 Averaged intersection areas of GEV dist.

SSM은 샘플 수가 적은 경우 잘못된 모델을 선택하는 경우가 많기 때문에 샘플로부터 직접 모델링하는 KDE 방법에 비해 추정 정확도가 낮다. 샘플 수가 증가함에 따라 두 방법의 추정 정확도는 증가하지만, KDE는 비모수적인 방법이므로 실제 모델로 수렴하는 속도가 느린 반면, SSM은 정확한 모델을 선택하는 횟수가 증가하므로 실제 모델로 더 빨리 수렴하게 된다.

NORM 분포가 실제 모델일 경우에도 RAY 분포와 마찬가지로 샘플 수가 적은 경우 KDE 방법이 SSM보다 실제 모델의 추정 정확도가 더 높다. 단, RAY 분포의 경우와 다른 점은 $n \leq 20$까지도 KDE의 정확도가 SSM보다 더 높다. 그 이유는 Fig. 2와 같이 RAY 분포가 다른 분포 모델과 형태가 상이하여 SSM을 사용할 경우 정확한 모델을 쉽게 찾을 수 있는 반면 NORM은 유사한 모델이 많으므로 실제 모델을 찾는데 더 많은 수의 샘플을 필요로 한다. 또한, NORM 분포가 실제 모델인 경우 두 방법의 수렴도가 유사한데, 그 이유는 NORM 분포는 RAY 분포에 비해 대칭성이 있고 분산도 적어 KDE로 모델링하기가 용이하기 때문이다.

Fig. 5는 모수의 개수가 3인 GEV 분포에 대한 교차면적을 나타낸다. GEV 분포는 NORM분포와 유사하게, $n \leq 20$까지 KDE보다 SSM의 추정 정확도가 더 높다. NORM 분포와 GEV 분포 결과가 유사한 이유는 논문에서 사용된 GEV 분포는 Fig. 2에서 나타낸 바와 같이 NORM 분포보다 첨도가 크고 왜도도 있으나 전반적으로 분포 형태의 비선형성이 낮아 KDE 방법으로도 쉽게 통계 모델을 얻을 수 있기 때문이다.

RAY, NORM, GEV분포함수 외에 다른 분포함수들에서도 3가지 분포함수들과 유사한 경향이 확인되었다. Fig. 6은 데이터에 개수에 따라 7개의 분포함수에 대한 교차면적을 나타낸 그림이다. BS, LOG와 LOGN 분포는 NORM과 GEV 분포와 유사하여 SSM이 높은 추정 정확도를 갖기 위해서 충분한 수의 샘플을 필요로 한다. RAY 분포를 제외하고 대부분의 분포 모델에서 $n \leq 10$에서는 KDE가 SSM보다 정확하며, $n \geq 20$일 때 SSM의 추정 정확도가 KDE에 비해 더 높아지게 된다. $n \geq 50$인 경우 SSM의 교차면적은 모든 분포함수에 대해 0.9 이상을 가지며 매우 높은 추정 정확도를 보인다.

4. 신뢰성 해석

모수적·비모수적 입력모델링 기법에 따른 신뢰성 해석의 결과를 비교하기 위해서 수학적 모델과 간단한 외팔보 문제에 대해 SSM과 KDE로부터 추정된 분포함수를 이용하여 신뢰성 해석을 수행하였다.

4.1 입력모델링에 따른 절차

입력모델링 기법에 따른 신뢰성 해석의 결과를 비교하기 위해서 문제에서 정의된 입력변수에 대한 모집단의 확률변수의 분포함수로부터 데이터의 개수를 3, 5, 7, 10, 20, 30, 50, 100, 200, 300까지 증가시켜 가면서 각 1,000개의 데이터 세트를 추출하였으며, 각각의 데이터 세트에서 추출된 샘플을 이용하여 통계모델링을 수행하였다. 그리고 통계모델링 기법을 사용하지 않은 경우와 비교를 위해서 추출된 샘플을 균일분포함수(uniform distribution function, UNI)로 가정하였고,

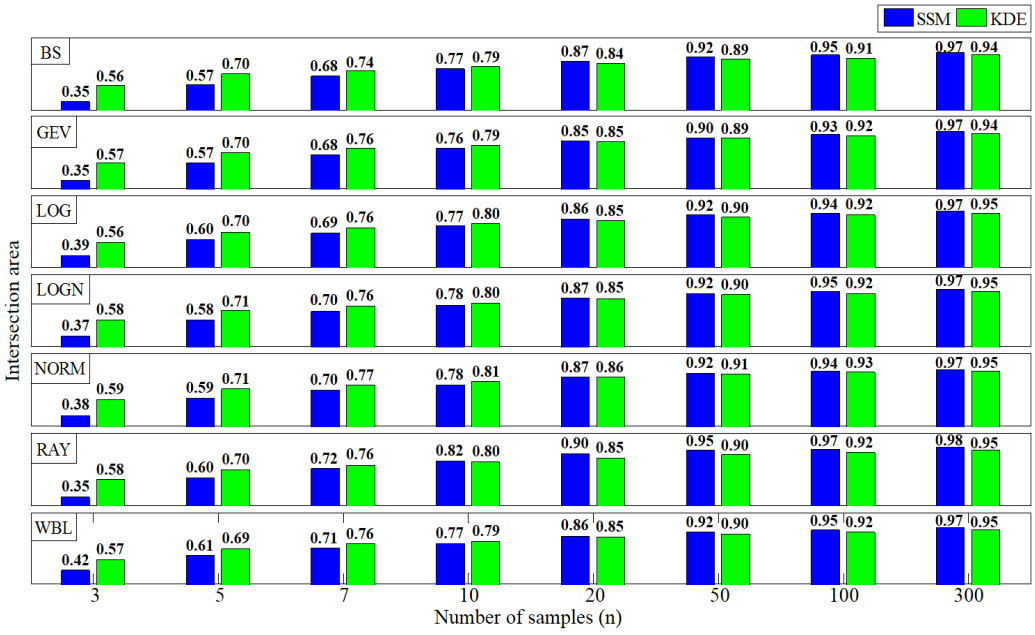

Fig. 6 Intersection areas according to sample sizes

모수는 최대우도추정방법을 이용하여 데이터로부터 계산하였다. 신뢰성 해석 방법인 몬테카를로 시뮬레이션(Monte Carlo simulation, MCS)을 수행하기 위해 SSM, KDE, UNI으로 추정된 입력변수의 분포함수로부터 100,000개의 시뮬레이션 샘플을 추출하고, 한계상태함수(limit-state function)를 이용하여 100,000개의 샘플에 대한 파손확률(probability of failure)을 계산하였다. 마지막으로 SSM, KDE, UNI을 통해 계산된 파손확률은 실제 분포함수의 파손확률과 비교 분석하였다.

4.2 Lognormal ratio 문제

첫 번째 예제는 Lognormal ratio 문제이다(Eldred *et al.*, 2007). 이 문제는 두 확률변수의 비율에 관한 간단한 수학문제로서 확률변수 x_1, x_2는 모두 대수정규분포 LOGN(1,0.5)을 따르며 이 분포의 평균과 표준편차는 각각 3.08, 1.6, 왜도와 첨도는 각각 1.75, 8.9, 변동계수(coefficient of variation, COV)는 0.5로서 왼쪽으로 많이 치우치면서 산포와 첨도가 매우 큰 분포형태를 가지며, 독립변수로 가정하였다.

본 문제에서 한계상태함수는 다음과 같이 정의된다.

$$g(\boldsymbol{x}) = R - \frac{x_1}{x_2} \tag{11}$$

여기서, R은 파손에 대한 임계값으로 0.3(확정론적 변수)을 사용하였다. 파손은 한계상태함수가 0보다 클 때 발생하는 것으로 가정하였다.

Fig. 7은 본 문제에서 1,000개의 데이터 세트에 대한 입력모델링에 따른 파손확률의 최솟값, 평균값, 최댓값을 나타내었으며, 수평막대가 최소-최대값을 의미하고 ×, ○, * 표시(marker)가 SSM, KDE, UNI의 평균 파손확률을 나타낸다. 범례에서 SSM과 KDE는 각각의 방법으로 입력변수를 모델링

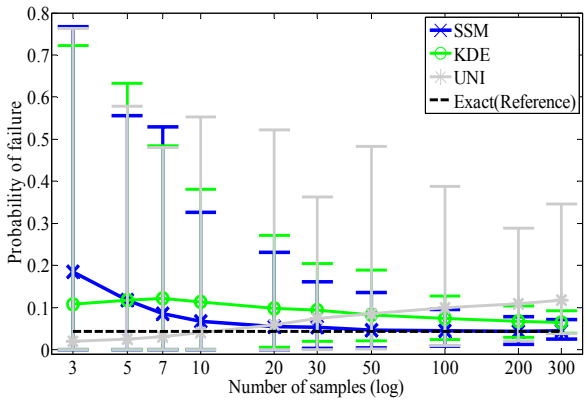

Fig. 7 Probabilities of failure in lognormal ratio

하여 신뢰성 해석을 수행한 결과이고 UNI는 균일분포로 모델링하여 신뢰성 해석을 수행한 결과이다. Exact는 실제 분포함수에 대한 파손확률이며, 4.4%로 계산되었다.

UNI 분포로 가정한 경우에는 데이터의 개수가 증가함에 따라 최소-최대값의 구간의 변동이 심하였고 파손확률의 평균이 점점 증가하였다. 예를 들어, $n \le 10$인 경우에는 UNI 분포를 사용하여 계산된 파손확률은 실제 파손확률값보다 낮았고, $n \ge 20$인 경우에는 실제 파손확률값보다 높게 추정하여 부정확하고 불안정한 결과를 보였다. 그 이유는 샘플 수가 적은 경우 UNI 분포의 폭이 좁아 파손확률이 0에 더 가까워지는 반면, 증가함에 따라 UNI 분포의 꼬리 부분이 두꺼워지면서 파손확률이 큰 값으로 치우치게 되면서 평균값 역시 실제 파손확률값보다 점점 더 커지게 된다.

반면 KDE를 이용한 신뢰성 해석 결과의 상・하한은 UNI 분포에 비해 안정적으로 실제 파손확률값 근처로 수렴하는 것을 알 수 있으며, 평균적으로 실제 파손확률값보다 높게 계산된다. 그 이유는 KDE는 모수적 분포모델에 비해 각 샘플 주변으로 가우시안 커널 함수를 생성하기 때문에 SSM에 비해 분포의 꼬리부분이 두꺼워져(heavy tail) 더 보수적으로 파손확률이 계산되기 때문이다. 반면 SSM을 사용할 경우 특정한 모델로 수렴하기 때문에 데이터의 개수가 증가함에 따라 평균값과 최소-최대값의 구간이 실제 파손확률에 빠르게 수렴하게 된다. 두 방법 모두 $n \le 10$에서는 최소-최대값의 구간이 커서 데이터의 품질에 민감한 것을 확인할 수 있으며 파손확률의 오차가 큰 것을 확인할 수 있다. 하지만 SSM은 $n=3$에서는 정확도가 매우 낮으며 $n=10$까지는 데이터의 개수에 매우 민감하고, 최소-최대값의 구간이 KDE보다 대체적으로 크게 나오며 불안정한 결과를 보인다.

예제에서 UNI 분포는 데이터로부터 추정된 분포 모수를 이용하여 모델링되었으나, 실제 문제에서는 변수에 대한 상・하한만 주어질 수도 있다. 예를 들어 실제 분포를 모르지만 정확한 평균값을 기준으로 95%의 백분위 값에 해당되는 상・하한을 정확하게 알고 있고, 이로부터 UNI 분포를 모델링하여 신뢰성 분석을 한다고 가정하자. 이러한 경우, 파손확률은 11.4%로 보수적으로 계산되나 실제 파손확률값인 4.4%와 차이가 크므로 적합하지 않다. 또한 이 역시 상・하한에 대한 정보를 정확히 아는 경우에 계산된 값이고, 실제로 잘못된 상・하한을 사용하는 경우 신뢰성 해석 결과 역시 부정확할 수 있으므로 입력 변수에 대한 대푯값 및 상・하한에 관한 정확한 정보가 필요하다.

4.3 외팔보 문제

두 번째 예제는 균일한 단면을 가진 외팔보에 관한 예제이며

Fig. 8에 나타내었다(Eldred *et al.*, 2007). 탄성계수(E), 수평하중(X), 그리고 수직하중(Y)는 모두 정규분포를 따른다고 가정하였으며, $E \sim N(2.9\times10^7, 1.45\times10^6)$[psi], $X \sim N(500,100)$[lb], $Y \sim N(1000,100)$[lb]를 가진다. 모두 정규분포를 따르므로 모집단의 왜도는 모두 0이고 첨도는 3이다. 각 변수의 COV는 각각 0.05, 0.2, 0.1이므로 Lognormal ratio 문제에 비해 산포가 작다. 외팔보 문제에서 한계상태함수는 다음과 같이 정의된다.

$$g_D(E,X,Y)=\frac{4L^3}{Ewt}\sqrt{\left(\frac{Y}{t^2}\right)^2+\left(\frac{X}{w^2}\right)^2} \tag{12}$$

여기서, L, w, t는 매개 변수로서 각각 100in., 2.5in., 3.5in.로 선정하였고, D_0는 허용 처짐량이며 2.15in.로 선정하였다.

Fig. 9는 외팔보 문제에 대해 입력모델링에 따른 파손확률의 최솟값, 평균값, 최댓값을 나타내며, 실제 분포함수를 사용한 파손확률은 6.91%로 계산되었다. UNI분포로 가정한 경우에는 첫 번째 예제처럼 데이터의 개수가 증가함에 따라 최소-최대값의 구간의 변동이 심하고 파손확률이 점점 증가하면서 신뢰성 해석의 결과가 부정확하고 불안정한 결과를 산출하였다.

반면 KDE를 사용한 경우에는 첫 번째 문제처럼 데이터의 개수에 상관없이 보수적인 결과를 보였으며, 샘플 수가 증가함에 따라 SSM과 같이 실제 파손확률값에 수렴함을 확인 할수 있다. Lognormal ratio 문제에 비해 보 문제에서 KDE는 실제 파손확률값에 더 빨리 수렴한다. 그 이유는 KDE가 비대칭 분포함수(LOGN)에 비해 대칭성을 가지는 분포함수(NORM)에 대해서는 높은 추정 정확도를 보이기 때문이다. 하지만 여전히 꼬리가 두껍게 예측되는 특성 때문에 SSM에 의한 결과 보다 파손확률이 크게 예측된다.

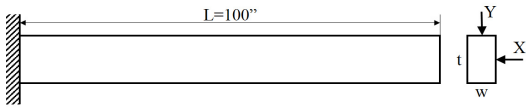

Fig. 8 Cantilever beam problem

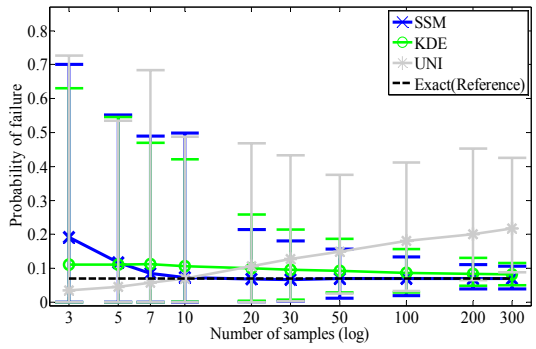

Fig. 9 Probabilities of failure in the cantilever beam

두 방법 모두 $n \le 10$인 경우에서는 최소-최대값의 구간이 크지만, $n \ge 20$인 경우에서는 최소-최대값의 구간이 매우 감소하였다. 특히, SSM의 경우 Lognormal ratio 문제와 다르게 $n=10$에서 최소-최대값의 구간이 매우 크다가 $n=20$에서 갑자기 감소한다. 데이터의 개수가 10개인 경우 최소-최대값의 구간의 크기가 매우 큰 이유는 SSM은 분포형태가 대칭이고 비선형성이 적은 정규분포와 같은 분포에 대해서는 데이터의 개수가 적은 경우 유사한 형태의 분포함수를 선정하게 되는 경우가 많다. 분포함수의 추정 정확도(교차면적)의 관점에서는 유사한 형태를 사용하여도 정확도에 거의 변화가 없지만 꼬리부분이 중요한 신뢰성 해석에서는 그 차이가 커진다(유사모델들은 서로 꼬리 부분의 두께에서 차이가 발생한다). 하지만 데이터의 개수가 충분해지면 유사모델을 선정하는 비율이 낮아지게 되고 꼬리부분의 통계모델링의 결과도 우수해져서 파손확률의 최소-최대값의 구간이 급격히 감소된다.

5. 결 론

본 논문에서는 모수적·비모수적 입력모델링 기법에 따른 특성을 비교하기 위해서 모수적 모델링 방법인 순차적 통계모델링 방법과 비모수적 방법인 커널밀도추정을 이용하여 입력변수의 모델링의 정확성을 비교 분석하였고, 두 모델링 기법을 이용한 두 가지 신뢰성 해석 예제를 통하여서 입력모델링기법에 따른 신뢰성 해석의 결과를 비교 분석하였다. 결과는 다음과 같이 요약할 수 있다.

1) 데이터 수가 10개 이하인 경우 대체적으로 KDE의 추정 정확도가 높지만 데이터 수가 20개 이상인 경우에는 SSM의 정확도가 더욱 높다. 특히 분포 모델의 형태가 타분포와 상이한 모델인 경우(RAY) 데이터 수가 적은 경우에도 SSM의 정확도가 높았다.
2) KDE를 이용한 신뢰성 해석은 데이터의 개수에 상관없이 SSM에 비해 보수적으로 파손확률을 예측하였으며 UNI 분포로 가정한 경우보다 정확성과 안정성이 우수하였다.
3) 데이터 수가 증가함에 따라 SSM이 KDE 보다 실제 파손확률값에 더 빨리 수렴하는 경향을 보였다.

향후 연구계획은 데이터의 개수에 관계없이 적용할 수 있는 통합된 통계모델링 방법을 개발할 예정이며, 기존의 모수적·비모수적 통계모델링 방법 및 개발된 방법을 실제 공학예제에 적용하여 통계모델링의 정확도가 신뢰성 해석 및 신뢰성 기반 최적설계(RBDO)에 미치는 영향에 대해 연구를 수행할 예정이다.

감사의 글

이 논문은 부산대학교 기본연구지원사업(2년)에 의하여 연구되었음.

References

Analytical Methods Committee (1989) Robust Statistics-how NOT to Reject Outliers. Part 1. Basic Concepts, *Analyst*, 114(12), pp.1693~1697.

Ayyub, B.M., McCuen, R.H. (2012) *Probability Statistics, and Reliability for Engineers and Scientists*, CRC Press, Florida.

Eldred, M.S., Agarwal, H., Perez, V.M., Wojtkiewicz Jr. S.F., Renaud, J.E. (2007) Investigation of Reliability Method Formulations in DAKOTA/UQ, *Struct. & Infrastruct. Eng.*, 3(3), pp.199~213.

Kang, Y.J., Lim, O.K., Noh, Y. (2016) Sequential Statistical Modeling Method for Distribution Type Identification, *Struct. Multidisc. Optim.*, online first.

Lee, T.H., Choi, J.S., Lim, W.C., Cho, S.G., Lee, M., Hong, S. (2013) Nonparametric Reliability Analysis for Design of a Mechanical System Working on an Inaccessible Area, *10th World Congress on Structural and Multidisciplinary Optimization*, Orlando, Florida, USA.

Lim, W., Lee, T.H. (2012) Reliability based Design Optimization using Akaike Information Criterion for Discrete Information, *Trans. Korean Soc. Mech. Eng., A*, 36(8), pp.921~927.

Noh, Y. (2016) A Comparison Study on Statistical Modeling Methods, *J. Korea Acad.-Industrial co. Soc.*, 17(5), pp.645~665.

Noh, Y., Choi, K.K., Lee, I. (2010) Identification of Marginal and Joint CDFs using Bayesian Method for RBDO, *Struct. Multidisc. Optim.*, 40(1), pp.35~51.

Schwarz, G. (1978) Estimating the Dimension of a Model, *Ann. Statistics*, 6(2), pp.461~464.

Scott, D.W. (1979) On Optimal and Data-based Histograms, *Biom.*, 66(3), pp.605~610.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*, 26, CRC Press, London.

Wand, M.P., Jones, M.C. (1994) *Kernel Smoothing*, CRC Press, London.

요 지

신뢰성 해석 및 신뢰성기반 최적설계는 불확실성을 고려한 확률변수를 입력 값으로 요구하며, 확률변수는 모수적·비모수적 통계모델링 방법을 사용하여 확률분포함수의 형태로 정량화 된다. 신뢰성 해석과 같은 통계적 해석은 입력되는 확률분포함수의 특성이 결과값에 영향을 미치게 되며, 확률분포함수는 통계모델링 방법에 따라 다른 형태를 가지게 된다. 본 연구에서는 모수적 통계모델링 방법인 순차적 통계모델링 방법과 비모수적 방법인 커널밀도추정을 사용하여 데이터의 개수에 따른 통계모델링의 결과를 분석하였다. 또한 수치예제를 통해 두 가지 기법에 따른 신뢰성 해석의 결과를 분석하였고, 데이터의 개수에 따른 적절한 기법을 제안하였다.

핵심용어 : 커널밀도추정, 모수적 통계모델링, 비모수적 통계모델링, 신뢰성 해석, 순차적 통계모델링