

< 논문 >

함수 근사 모멘트 방법에서 추정된 1~4 차 통계적 모멘트의 수치적 검증

허재성[†] · 곽병만^{*}

(2006년 11월 17일 접수, 2007년 3월 21일 심사완료)

Numerical Verification of the First Four Statistical Moments Estimated by a Function Approximation Moment Method

Jae-Sung Huh and Byung-Man Kwak

Key Words : Uncertainty(불확실성), Random Variable(확률 변수), Statistical Moments(통계적 모멘트), Function Approximation Moment Method(함수 근사 모멘트 방법)

Abstract

This research aims to examine accuracy and efficiency of the first four moments corresponding to mean, standard deviation, skewness, and kurtosis, which are estimated by a function approximation moment method (FAMM). In FAMM, the moments are estimated from an approximating quadratic function of a system response function. The function approximation is performed on a specially selected experimental region for accuracy, and the number of function evaluations is taken equal to that of the unknown coefficients for efficiency. For this purpose, three error-minimizing conditions are utilized and corresponding canonical experimental regions constructed accordingly. An interpolation function is then obtained using a D-optimal design and then the first four moments of it are obtained as the estimates for the system response function. In order to verify accuracy and efficiency of FAMM, several non-linear examples are considered including a polynomial of order 4, an exponential function, and a rational function. The moments calculated from various coefficients of variation show very good accuracy and efficiency in comparison with those from analytic integration or the Monte Carlo simulation and the experimental design technique proposed by Taguchi and updated by D'Errico and Zaino.

1. 서론

공학적 시스템의 성능은 다양한 설계변수 및 시스템인자의 불확실성에 의해 어떤 하나의 값으로 정의되기 보다는 어떤 범위로 파악하는 것이 바람직하다. 즉, 시스템의 통계적 특성을 통해 이해하는 것이 좀 더 현실이라고 할 수 있다. 이러한 특성은 시스템 응답함수(Response function)의 평균,

표준편차와 같은 통계적 모멘트, 더 나아가 그 성능함수의 확률밀도함수를 통해 정의할 수가 있다.

이러한 시스템 성능함수의 통계적 모멘트를 효율적이고 정확하게 추정하는 것은 무엇보다 중요하며, 이를 위한 방법들은 크게 4 가지로 분류-몬테카를로 시뮬레이션 방법, 수치 적분 방법, 실험계획법을 이용한 통계적 해석 방법, 테일러 전개를 이용한 방법-가 가능하다. Evans⁽¹⁾는 구적분법(product quadrature rule)에 의한 통계적 모멘트의 계산이 비효율적임을 보이고, 확률변수가 N 개 존재할 때 $2N^2+1$ 개의 절점으로 구성된 수치 구적분을 고안하였다. Rosenblueth⁽²⁾는 2 점 정보를 이용하여 성능함수의 모멘트를 추정하는 방법을 제안하였다. Zhao 와 Ono⁽³⁾는 비선형 성능함수를

[†] 책임저자, 회원, 자동차부품연구원 신뢰성본부
E-mail : caesarhjs@nate.com
TEL : (041)559-3157 FAX : (041)559-3165

^{*} 한국과학기술원 기계공학과

테일러 급수의 1 차 전개함수로 근사하고 Rosenblatt 변환을 이용하여 비정규 분포를 처리하는 방법을 제안하였다. Taguchi⁽⁴⁾는 3 수준 전조합법을 이용하여 통계적 모멘트를 계산하였으며 D'Errico 와 Zaino⁽⁵⁾는 다구찌가 제안한 수준과 가중치를 수정·보완하여 고 차 모멘트의 정확도를 개선하였다. Seo 와 Kwak⁽⁶⁾은 정규 분포를 따르는 확률변수에 대해서만 적용이 가능하던 기존의 방법을 발전시켜 비정규분포에 대한 수준과 가중치를 제안하였다. 최근에 Lee 와 Kwak⁽⁷⁾은 전조합법을 사용함에 발생하는 방대한 계산량 문제를 반응표면법(response surface method)을 접목하여 해결하고자 하였으며, 최현석⁽⁸⁾은 기존에 제안된 3 수준 실험계획법이 가지는 문제점을 제시하고 이를 극복하기 위한 새로운 수준을 제안하고 5 수준 실험계획법 수립하였다.

본 연구에서는 Huh 등^(9,10)이 최근에 제안하고 초정밀위치결정기구(Nano-positioning system)와 TFT-LCD 유리 기판의 강건최적설계에 적용한 함수 근사 모멘트 방법(Function approximation moment method, FAMM)이 추정하는 1 ~ 4 차 통계적 모멘트의 정확성 및 효율성을 정규분포를 따르는 확률변수로 정의되는 수치 예제를 통해 검증하려 한다.

2. 함수 근사 모멘트 방법

2.1 모멘트의 정의

성능함수에 대한 k 차 통계적 모멘트(Statistical moments)는 다음 수식처럼 각 확률변수의 정의 영역에 대한 적분 형태로 표현된다.

$$E\{g^k\} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} [g(x_1, \dots, x_N) - \mu_g]^k \prod_{i=1}^N \phi_i(x_i) dx_1 \dots dx_N \quad (1)$$

여기서 $g(\mathbf{x})$ 는 시스템의 성능함수이며 $x_1 \sim x_N$ 는 독립 확률변수(Random variable)이고 $\phi_i(x_i)$ 는 i 번째 확률변수가 따르는 확률밀도함수(Probability density function, P.D.F.)이고 μ_g 는 성능함수의 평균이다. 일반적으로 많이 이용되는 모멘트는 4 차 모멘트까지이며 이들은 확률분포의 특성 기술에 이용된다. 1, 2 차 모멘트는 성능함수의 평균(mean) μ 과 분산(variance) σ^2 을 의미한다. 왜도(Skewness) γ_1 (혹은 $\sqrt{\beta_1}$)는 3 차 모멘트를 표준편차(Standard deviation, S.D.)의 3 제곱으로

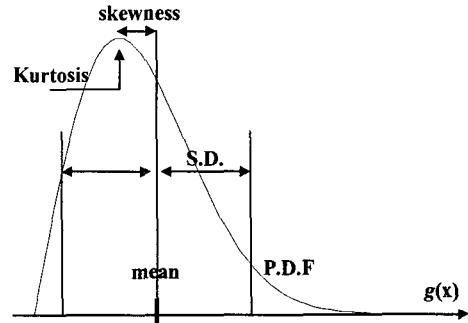


Fig. 1 Definition of probability density function using the first four moments

나눈 값이다. 즉, $\sqrt{\beta_1} = E\{g^3\} / \sigma^3$ 이다. 왜도는 Fig. 1에서 설명하듯이 확률분포의 비대칭 상태를 판단하기 위하여 사용하며, 양수인 경우 왼쪽으로 치우치고 음수인 경우는 반대이며 $\sqrt{\beta_1} = 0$ 이면 정규분포처럼 대칭인 분포를 가진다. 마지막으로 첨도(kurtosis) γ_2 (혹은 β_2)는 4 차 모멘트와 $\beta_2 = E\{g^4\} / \sigma^4$ 관계를 가지며 확률밀도함수의 뾰족한 정도를 표현한다. $\gamma_2 < 0$ 일 때 완첨(platykurtic), $\gamma_2 = 0$ 이면 평첨(mesokurtic), 그리고 $\gamma_2 > 0$ 에서는 급첨(keptokurtic)이라 명명된다.

2.2 함수 근사 모멘트 방법^(9,10)

함수 근사 모멘트 방법(FAMM)은 일반적으로 비명시적 함수(Implicit function)로 표현되는 시스템의 성능함수를 다항식과 같은 명시적 함수(Explicit function)로 가능한 정확하게 근사하여 k 차 통계적 모멘트를 계산하는 방법이다. FAMM에서는 근사함수로 2 차 다항식을 이용한다. 이는 대부분의 시스템 성능함수는 선형이 아닌 비선형성을 가지므로 선형함수로 표현하기에는 부족함이 있기 때문이다. 이 방법에서 핵심 부분은 비명시적 함수를 명시적 함수로 근사할 때 반드시 필요하고 해에 큰 영향을 미치는 실험영역의 체계적 결정 방법이며, 얻어진 실험영역을 정규 실험영역(Canonical experimental region)으로 명명한다. 실험영역이 결정된 이후는 기존의 반응표면법(Response surface method)⁽¹¹⁾의 절차와 동일하게 된다. 즉, 주어진 실험영역에서 실험계획법(Design of experiments⁽¹²⁾)을 활용하여 실험점을 배치하고 다항식의 계수를 최소 자승법(Least square method)으로 추정한다. 마지막으로 얻어진 2 차 다항식의 모멘트는 손쉽게 정확하게 얻어지며 이를 성능함수의 추정치로 활용한다. FAMM의 절차는 아

래와 같으며, 다음 절에서 이 방법의 핵심단계인 1 과 2 단계를 상세히 기술한다.

1 단계: N 개의 확률변수에서 i 번째 변수에 대한 1 차원 문제에 대해 아래에 기술된 세 가지의 예러 최소화 조건에 의해 서로 다른 3 개의 점을 얻는다.

2 단계: 각 확률변수에서 결정된 3 개의 점으로부터 N 직면체(N -orthotope)를 정의하고 이를 정규 실험영역(Canonical experimental region)이라 한다.

3 단계: 얻어진 영역에서 $(N + 1)(N + 2)/2$ 의 실험점을 가지는 실험계획법을 실시하고 근사함수 $\bar{g}(x)$ 의 미지수를 결정한다.

4 단계: 근사함수인 $\bar{g}(x)$ 로부터 계산된 4 차 모멘트를 성능함수 $g(x)$ 의 추정치로 인정한다.

2.2.1 1 단계: 1 차원 문제에서 예러 최소화

FAMM 에서는 근사함수인 2 차 다항식이 그 이하의 차수를 가지는 다항식뿐만 아니라 3 차 다항식의 1 ~ 4 차 모멘트까지도 가능한 정확하게 추정할 수 있는 2 차 다항식의 계수를 찾고자 한다. 그 이유는 2 차 다항식보다는 3 차 다항식이 일반적인 성능함수를 근사 하는데 유리하다고 판단되어지기 때문이다. 이 때 2 차 다항식 계수는 3 개의 다른 점에 의해 결정이 된다. 즉 3 개 점의 위치를 최적화하여 2 차 다항식으로 3 차 다항식의 모멘트도 가능한 정확히 추정하려는 것이다. 그러므로 성능함수인 $g(x)$ 와 근사함수인 $\bar{g}(x)$ 는 각각 식 (2)과 (3)처럼 3 차 다항식과 2 차 다항식으로 가정한다.

$$g(x) = ax^3 + bx^2 + cx + d \tag{2}$$

$$\bar{g}(x) = \bar{a}x^2 + \bar{b}x + \bar{c} \tag{3}$$

우선 2 차 다항식으로 3 차 다항식을 근사하기 위하여 임의의 3 개 점 (l, m, n)에서 계산된 함수 값을 2 차 다항식도 만족해야 하므로 식 (4)에서 (6)이 얻어진다. 이 연립 방정식을 풀면 2 차 다항식의 계수가 식 (7)처럼 3 차 다항식의 계수와 세 점의 좌표로 표현된다.

$$g(l) = al^3 + bl^2 + cl + d = \bar{a}l^2 + \bar{b}l + \bar{c} \tag{4}$$

$$g(m) = am^3 + bm^2 + cm + d = \bar{a}m^2 + \bar{b}m + \bar{c} \tag{5}$$

$$g(n) = an^3 + bn^2 + cn + d = \bar{a}n^2 + \bar{b}n + \bar{c} \tag{6}$$

$$\bar{g}(x) = (b + a(l + m + n))x^2 + (c - a(l \cdot m + l \cdot n + m \cdot n))x + (d + a \cdot l \cdot m \cdot n) \tag{7}$$

얻어진 근사함수인 2 차 다항식으로 계산되는 1 ~ 4 차 모멘트가 본래 함수인 3 차 다항식의 모멘트와 가능한 일치할 수 있도록 근사 함수의 계수를 결정하여야 한다. FAMM 에서는 이는 달성하기 위하여 아래의 3 가지 조건을 제시하고, 이 조건들은 2 차 다항식을 정의하는 세 점의 좌표(l, m, n)을 결정하게 된다.

조건 1: 세 개의 실험 중에서 하나를 x 의 확률밀도함수에서 극대에 해당하는 곳에 위치 시킨다.

$$\text{조건 2: } \int_{-\infty}^{\infty} g(x) \cdot \phi(x) dx = \int_{-\infty}^{\infty} \bar{g}(x) \cdot \phi(x) dx \tag{8}$$

$$\text{조건 3: } \text{Minimize } E = \frac{1}{2} \left(\frac{\mu_\varepsilon}{\mu_{\min}} + \frac{s_\varepsilon}{s_{\min}} \right), \tag{9}$$

여기서 예러함수인 ε 와 $\mu_\varepsilon, s_\varepsilon$ 은 아래와 같이 정의되며, μ_{\min} 와 s_{\min} 은 각각 예러함수의 두 모멘트 최소값이다.

$$\varepsilon = |g(x) - \bar{g}(x)| \tag{10}$$

$$\mu_\varepsilon = \int_{-\infty}^{\infty} \varepsilon(x) \cdot \phi(x) dx \tag{11}$$

$$s_\varepsilon^2 = \int_{-\infty}^{\infty} \varepsilon(x)^2 \cdot \phi(x) dx \tag{12}$$

정규 분포인 경우는 위 세 조건을 모두 만족하는 서로 다른 세 점의 위치 (l, m, n)은 $(m - 1.38184\sigma, m, m + 1.38184\sigma)$ 이며, 비정규 분포에도 적용이 가능하다. 자세한 유도과정은 참고문헌 (9,10)에 기술되어 있다.

2.2.2 2 단계: 정규 실험영역 결정

앞 절에서 결정된 각 축에 대응되는 3 개의 서로 다른 점은 정규 실험영역을 결정하는데 사용된다. 각 확률변수에 대응되는 3 개 점의 중앙 점(m)은 정규 실험영역의 중심을 정의하며 다른 두 점은 영역의 크기를 나타내게 된다. Fig. 2 는 확률변수가 2 개인 경우의 정규 실험영역의 예이다.

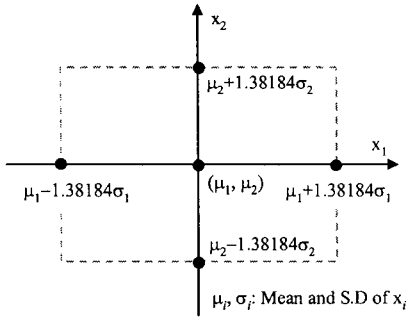


Fig. 2 Optimized positions of each variable and canonical experimental region in case of two random variables

Table 1 Mean of random variables

	μ_{x_1}	μ_{x_2}	μ_{x_3}
Mean	1.0	1.0	1.0

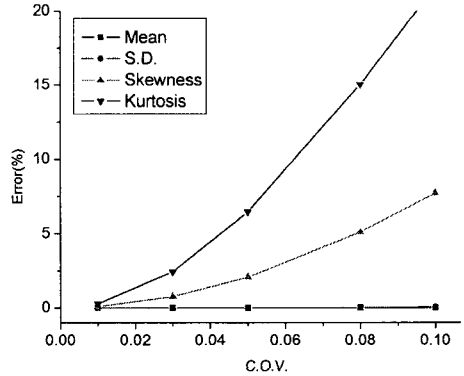
3. 검증 예제

3.1 고차 다항식

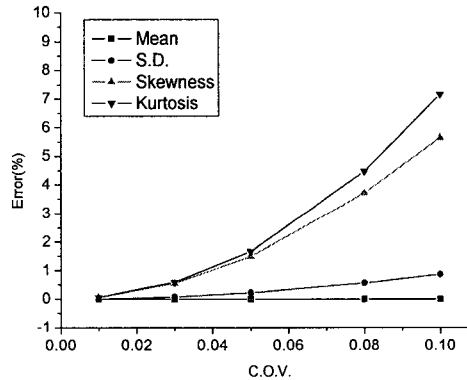
FAMM 에서 얻어지는 통계적 모멘트를 수치적으로 검증하기 위하여 먼저 식 (13)처럼 정의되는 4 차 다항식에 적용하며 각 확률변수는 정규 분포를 따르며 평균은 Table 1 과 같다. 참고로 이 함수는 Zhao 와 Ono⁽³⁾의 검증예제로 사용된 문제이다.

$$g = x_1^2 x_2^2 + 2x_3^4 \quad (13)$$

FAMM 의 적용범위 및 정확성을 확인하기 위해 변동계수(Coefficient of variation, C.O.V.)를 증가시키면서 정 해(Analytic solution)와 비교하여 오차를 산출한다. 또한 기존 방법과의 비교는 Taguchi⁽⁴⁾가 제안하고 D'errico 와 Zaino⁽⁵⁾가 수정·보완한 방법을 활용한다. Fig. 3 의 (a)와 (b)는 두 방법으로 부터 얻은 모멘트의 오차를 나타낸다. 두 방법 모두 평균과 표준편차는 변동계수가 10%까지 1% 미만의 에러를 가지며 정확히 따라가고 있으나, 고차 모멘트 즉 왜도와 첨도는 변동계수가 5%를 넘어가면서 오차가 급격히 증가하기 시작한다. FAMM 에서는 근사함수가 가지는 오차가 고차 모멘트로 갈수록 점점 더 증폭되기 때문이고, 또한 3 가지 에러 최소화 조건에서 에러의 1, 2 차



(a) Errors of the method proposed by D'Errico and Zaino



(b) Errors of FAMM

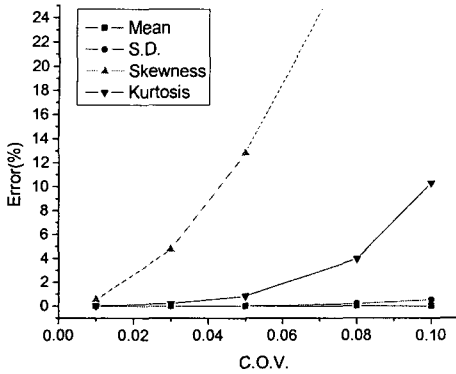
Fig. 3 Results calculated from two methods: Polynomial of order 4

Table 2 Mean of random variables

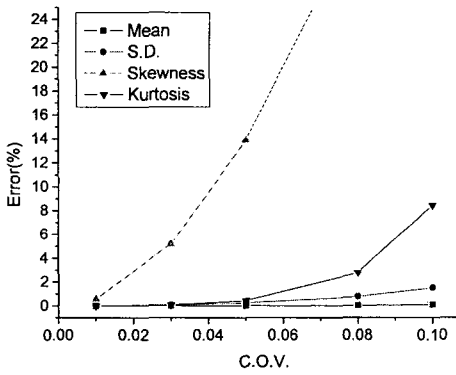
	μ_{x_1}	μ_{x_2}
Mean	1.0	1.0

모멘트만이 포함되어 있기 때문이라 판단된다.

여기서 주목할 부분은 FAMM 이 D'Errico 와 Zaino 의 방법보다 고차 모멘트에서 상대적으로 오차가 작은 것을 Fig. 3 에서 확인할 수가 있다. 또한 D'Errico 와 Zaino 가 제안하는 방법은 3^N 전조합법을 활용하여 27 번의 함수 계산을 필요로 하지만 FAMM 은 10 번의 함수 계산만을 요구한다. 그 이유는 앞 절의 절차에서 설명하였듯이 FAMM 은 2 차 보간 다항식 (Quadratic interpolating polynomial)을 사용하므로 다항식 계수의 개수만큼의 실험만을 요구하기 때문이다.



(a) Errors of the method proposed by D'Errico and Zaino



(b) Errors of FMM

Fig. 4 Results calculated from two methods: exponential function

3.2 지수 함수

다음 검증 예제는 식 (14)처럼 정의되는 지수함수이며 모멘트 계산은 Table 2의 평균점에서 이루어지며 확률변수는 정규 분포를 따른다고 가정한다. 앞의 예제와 동일하게 2 가지 방법의 결과와 정 해가 비교되며 그 결과는 Fig. 4 처럼 요약된다.

$$g = -e^{-(x_1-7)} - x_2 + 9 \tag{14}$$

앞의 예제와 유사하게 FMM 과 D'Errico 와 Zaino 가 제안한 방법으로 계산된 평균과 표준편차는 변동계수가 증가하더라도 2% 미만의 미미한 오차를 가진다. 그러나 왜도와 첨도는 변동계수가 3%이상부터 상당히 큰 오차를 가짐을 Fig. 4는 보여주고 있다. 특히 왜도의 경우는 두 방법 모두에서 변동계수가 10%인 경우 40%가 훨씬 넘는

Table 3 Mean of random variables

	μ_{x_1}	μ_{x_2}	μ_{x_3}	μ_{x_4}
Mean	0.5	360	2.26×10^{-3}	4.0

큰 오차를 가진다. 앞 예제와 다르게 첨도보다 왜도가 더 큰 에러를 가지는 이유는 왜도의 값이 첨도에 비해 작아서 상대적으로 작은 차이도 큰 오차로 표현되기 때문이다. 또한, FMM 은 기본적으로 성능함수를 2 차 다항식으로 근사하기 때문에 비선형성이 큰 지수함수의 경우는 이 예제처럼 상대적으로 큰 오차를 가지게 된다. 이는 FMM 의 적용 한계이며 이러한 한계는 정확도가 우수한 것으로 알려진 D'Errico 와 Zaino 방법의 결과에서도 유사하게 확인된다.

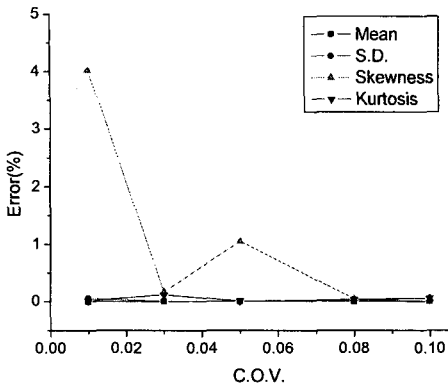
이러한 결과를 얻기 위해 필요한 함수 계산은 FMM 이 6 번을 필요로 하는 반면에, D'Errico 와 Zaino 가 제안한 방법은 9 번의 함수 계산을 요구하므로 FMM 이 더 효율적이다.

3.3 유리함수

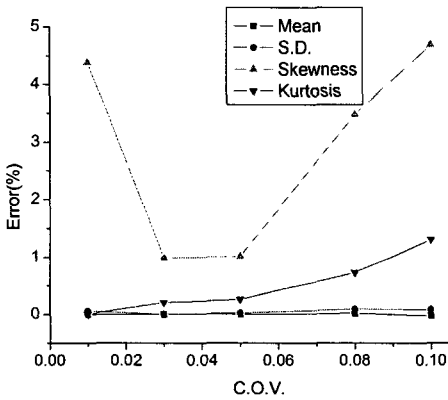
마지막 예제는 식 (15)처럼 표현되는 유리함수이며 모멘트의 계산은 Table 3의 지점에서 수행되며 역시 확률변수가 모두 정규분포를 따른다. 앞의 두 예제와 다른 점은 주어진 함수의 정 해를 얻는 것이 매우 힘들어 몬테카를로 시뮬레이션 방법(Monte Carlo simulation, MCS)를 사용하였으며 MCS의 해가 가지는 해의 산포를 감소시키기 위해 사용된 샘플의 수는 500 만이다.

$$g = x_1 x_2 x_3 - \frac{x_1 x_2^2 x_3^4}{x_4} - x_1 \tag{15}$$

두 방법은 변동계수가 10%의 범위에서 최대 5%정도의 오차 범위에서 정확하게 500 만 번의 MCS 결과와 일치한다. 그러나 Fig. 5에서 알 수 있듯이 앞의 예제와는 다른 양상을 보여주고 있다. 즉, 500 만 번의 샘플을 발생하여 얻은 MCS의 해 조차도 정 해가 아니므로 이를 기준으로 얻은 오차는 Fig. 5처럼 변동계수의 증가에 따라 증가하는 것이 아니라 어느 범위 내에서 불규칙한 양상을 나타내고 있다. 그리고 두 방법에서 모두 지수함수의 예와 유사하게 왜도가 다른 모멘트에 비해 상대적으로 큰 오차를 가지고 있다.



(a) Errors of the method proposed by D'Errico and Zaino



(b) Errors of FAMM

Fig. 5 Results calculated from two methods: Rational function

4. 결론

본 연구에서는 함수 근사 모멘트 방법(FAMM)에서 추정되는 1~4 차 통계적 모멘트의 정확성과 효율성을 3 개의 수치 예제를 통해 검증하였다. 정확도 측면에서는 평균과 표준편차의 경우는 변동계수가 10%까지도 FAMM 은 정해 혹은 몬테카를로 시뮬레이션 결과와 거의 일치한다. 반면에 고차 모멘트인 왜도와 첨도는 고차 다항식과 지수 함수에서 변동계수가 5%부터 오차가 현저하게 증가하는 모습을 보인다. 이는 정확성이 상당히 우수한 것으로 알려진 D'Errico 와 Zaino 방법에서도 유사한 특징을 보였다. 즉, 이는 두 방법의 적용 범위를 벗어난 것이라 보는 것이 타당할 것이다. FAMM 은 모멘트를 추정하는데 있어서 필요한 함수 호출은 $(N+1)(N+2)/2$ 의 횟수를 필요로 하므로 상당히 효율적인 방법이라 할 수 있다.

대부분의 공학적 문제는 많은 노력과 시간을 요구하는 유한요소해석과 같은 CAE 기술을 통해 그 성능함수의 값이 얻어진다. 한 걸음 더 나아가 이러한 통계적 특성을 기초로 하여 설계를 하고자 하는 강건 최적설계(Roust optimal design)나 신뢰성 기반 최적설계(Reliability-based optimal design)는 방대한 계산량을 요구한다. 이 검증 예제들에서 보여주듯이 FAMM 가 가지는 정확성과 효율성은 이러한 문제의 적용에서 좋은 결과를 보여주리라 예상된다.

참고문헌

- (1) Evans, D. H., 1975, "Statistical Tolerancing: The State of the Art. Part II: Methods for Estimating Moments," *Journal of Quality Technology*, Vol. 7, No. 1, pp. 1~12.
- (2) Rosenblueth, E., 1981, "Two-Point Estimate in Probabilities," *Applied Mathematics Modeling*, Vol. 5, pp. 329~335.
- (3) Zhao, Y. G. and Ono, T., 2000, "New Point Estimates for Probability Moments," *Journal of Engineering Mechanics*, Vol. 30, No. 4, pp. 433~436.
- (4) Taguchi, G., 1978, "Performance Analysis Design," *International Journal of Production Research*, Vol. 16, pp. 176~188.
- (5) D'Errico, J. R., and Zaino Jr., N. A., 1988, "Statistical Tolerancing Using a Modification of Taguchi's Method," *Technometrics*, Vol. 30, No. 4, pp. 397~405.
- (6) Seo, H. K., and Kwak, B. M., 2002, "Efficient Statistical Tolerance Analysis for General Distribution Using Three Point Information," *International Journal for Production Research*, pp. 931~944.
- (7) Lee, S. H., and Kwak, B. M., 2006, "Response Surface Augmented Moment Method for Efficient Reliability Analysis," *Structural Safety*, Vol. 28, No. 3, pp. 261~272.
- (8) Choi, H. S., 2005, "Moment Based Reliability Analysis for General Distributions Using Multi-Level DOE," Master thesis, KAIST.
- (9) Huh, J. S., Kim, K. H., Kang, D. W., Gweon, D. G., and Kwak, B. M., 2006, "Performance Evaluation of Precision Nanopositioning Devices Caused by Uncertainties Due to Tolerances Using Function Approximation Moment Method," *Review of Scientific Instruments*, Published online 19 January 2006.
- (10) Huh, J. S., Jung, B. C., Lee, T. Y., and Kwak, B. M., 2006, "A Study on the Robust Optimal Supporting Positions of TFT-LCD Glass Panel," *Transactions of the KSME A*, Vol. 30, No. 8, pp. 1001~1007.
- (11) Myers, R. H., and Montgomery, D. C., 1995, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley & Sons.
- (12) Montgomery, D. C., 1997, *Design and Analysis of Experiments*, John Wiley & Sons.