2000 Proceedings of the Autumn Conference, Korean Statistical Society

측정 및 분석 불확도 계산 프로그램(PUMA)의 제작 알고리즘

우 진춘, 문동민, 백홍범*

한국표준과학연구원 물질량연구부, *KISCOM.

1 서론.

측정 및 분석 결과의 신뢰도에 관한 정량적 표현의 방법으로서 오차나 오차 표기 방법을 오랫동안 사용하여 왔다. 그러나 오차를 아무리 잘 평가하여 보정하더라도 표현된 결과가 정확한가 하는데는 의문이 남아 있었다. 이러한 문제점을 해결하고, 측정된 결과의 품질을 하나의 통일된 척도(불확도)로 표기하고자, 국제표준화기구(ISO)가 BIPM, IEC, IFCC, IUPAC, IUPAP, OIML 등의 국제기구와 합동으로 측정 불확도 표현 지침서(Guide to the Expression of Uncertainty in Measurement, GUM)[1]를 1993년에 발행하였다.

GUM에 의한 불확도의 정의와 계산 방법이 종래에 측정 및 분석 과학 분야에서 사용된 오차 표현 방법에 비해 보다 과학적이고 엄격한 방법이므로 이 분야에 종사하는 측정/ 분석 담당자, 품질관리 책임자들은 이 지침서를 적용하고자 노력하고 있는 현실이다. 그러나 실제의 측정 및 분석 분야에 적용해본 결과, 여러 문제점과 어려움이 발견되어 이에 대한 해결 노력이 확대되고있다[2,3]. 여기에서는 이러한 노력의 하나로써 제작하여 시험 운영하고 있는 Program for Uncertainty-calculation in Measurements and Analysis(PUMA)의 알고리즘에 대하여 기술하고자 한다.

2 본론

2-1 GUM에 의한 불확도 계산 방법

먼저, 한국표준과학연구원에서 번역하여 발간한 측정불확도 표현 지침서에 따르면, 정의된 "uncertainty"를 "불확도"로 번역하여 사용하기로 하고 있으며, 다음과 같이 정의하고 있다. 불확도는 **"측정결과에 관련하여, 측정량을 합리적으로 추정한 값의 분산 특성을 나타내는 파라미터"**이다. 이와 같이 정의된 불확도 계산 방법은 크게 4 가지 단계로 구분되며, 그 과정은 다음과 같다.

1) 요인별 관계 모델의 설정

최종 결과의 추정값, y와 이것을 구하기 위하여, 관계된 여러 요인별 입력 추정값($x_1, x_2, \ldots x_n$)에 의한 관계 모델을 식으로 구한다.

$$y = f(x_1, x_2, \ldots, x_n) \qquad \langle 1 \rangle$$

2) 요인별 측정값과 표준불확도(standard uncertainty, $u(x_i)$)의 계산

요인별 입력값을 각각 구한 후, 유형에 따라 각 요인의 표준불확도를 구하는데, 이것을 구하는 방법에는 두 가지가 있다. 첫째로, 여러 번 측정하여 구한 평균값을 해당 요인의 측정값으로 사용할 경우, 평균값의 표준편차(표준오차)를 표준불확도로 정의하며, 이러한 유형으로 구한 것을 A형 불확도라 한다. 둘째로, 해당 요인의 측정값이 선험적인 자료, 관계된 물질과 기기의 성질, 제작자의 규

격, 교정 보고서의 자료, 편람에 나타난 자료로부터 구해진다면, 표준불확도 $u(x_i)$ 는 해당 측정값의 변동성에 관하여 얻을 수 있는 모든 정보와 근거한 과학적 판단에 의해 평가되며, 이러한 유형으로 구한 것을 B형 불확도라 한다.

3) 합성표준불확도(combined standard uncertainty, $u_c(y)$)의 계산

최종 결과, y에 대한 합성표준불확도는 설정된 관계식을 이용한 불확도(오차)의 전파 방법에 의해 계산된다. 이때 합성표준불확도의 계산식은 두 개 이상의 변수들 간에 상관성이 없거나 각각의 요인별 표준불확도 값이 측정값과 비교하여 상대적으로 작을 때 식 <3>과 같이 표현된다.

$$u_c^2(y) = \sum(\partial f/\partial x_i)^2 \cdot u^2(x_i) \quad \text{<3>}$$

4) 확장불확도(expanded uncertainty, U)

표현하고자하는 신뢰수준에 상당하는 포함인자를 곱하여 확장불확도를 구한다.

$$U = k \cdot u_c \quad \text{<4>}$$

자유도와 신뢰수준에 따른 t 값을 구하여 k 값으로 이용하며, 최종 결과의 자유도가 클 때, 95 % 신뢰수준의 값으로서 k = 2를 이용한다.

이와 같은 방법에 의한 불확도 계산 방법이 역추정 과정에 이용하기에는 문제점이 있기 때문에, PUMA에서는 역추정에서도 이들 과정의 적용이 용이하도록 Matrix 계산에 의해 선형대수학적으로 처리하였다[4-6].

2-2 불확도 산출을 위한 OLS의 재검토

OLS(Ordinary Least Square method)에서 일반적으로 적용되는 통계처리 방법과 실질적으로 측정/분석하는 과정에서의 조건이 상이하기 때문에 좀 더 일반화된 불확도 처리 방법을 도입하기 위하여 반응값(y_i)과 표준값(x_i)에 각각 오차항을 하나씩 더 삽입하여 관계 모델을 구하였다.

$$y_i + \varepsilon_{y_i} = f(x_i + \varepsilon_{x_i}) + \tau_i \quad \text{<5>}$$

이것을 정리하면, 식 <6>과 같은 식을 얻을 수 있다.

$$y_i \approx f(x_i) + \frac{df}{dx}\varepsilon_{x_i} + \varepsilon_{y_i} + \tau_i \quad \text{<6>}$$

이 식에서 총 오차량은, 다음과 같이 표현된다.

$$\frac{df}{dx}\varepsilon_{x_i} + \varepsilon_{y_i} + \tau_i = \varepsilon_i^* \quad \text{<7>}$$

또한, 최소제곱법이라는 것이 다음 식 <8>의 Q값을 최소화하는 과정이며, 두 종류의 오차 모델에서 실질적으로 오차량이 서로 같기 때문에, OLS 방법에 의해서 구한 관계식의 계수와 여기에서 설명하는 방법에 따라 최소제곱법으로 계산된 계수가 서로 같다.

$$Q = \sum_{i=1}^{n}\left(\frac{df}{dx}\varepsilon_{x_i} + \varepsilon_{y_i} + \tau_i\right)^2 = \sum_{i=1}^{n}(\varepsilon_i^{*})^2 = \sum_{i=1}^{n}\{y_i - f(x_i)\}^2 \qquad \text{<8>}$$

따라서, 회귀식을 n차 식으로서, 식 <9>와 같이 설정하였다면, 이 식의 계수들은 OLS에 의해 구할 수 있을 것이며 이들의 상관 관계도 역시 구할 수 있다.

$$y = f(x) = a_0 + a_1x + \cdot\cdot\cdot\cdot\cdot\cdot\cdot + a_nx^n \qquad \text{<9>}$$

식 <5>의 원형적인 모델식과 식 <9>를 근거로 하여 불확도 계산을 위한 관계 모델을 설정하면 식 <10>과 같다.

$$y_o = a_0 + a_1x_o + \cdot\cdot\cdot\cdot\cdot\cdot\cdot + a_kx_o^k + \tau \qquad \text{<10>}$$

여기에서 τ의 값과 표준불확도를 구하는 것이 중요한데, 그 값은 0이라는 것을 알 수 있고, 표준불확도는 다음과 같이 구할 수 있다. 실질적으로, 총분산 $\sigma^2\left(\frac{df}{dx}\varepsilon_x + \varepsilon_y + \tau\right)$, $\frac{df}{dx}\varepsilon_x$, ε_y 및 τ가 독립적이라고 볼 때, 다음과 같이 표시할 수 있기 때문이다.

$$\sigma^2\left(\frac{df}{dx}\varepsilon_x + \varepsilon_y + \tau\right) = \sigma^2(\varepsilon^{*}) = \sigma^2(\varepsilon_y) + \sigma^2\left(\frac{df}{dx}\varepsilon_x\right) + \sigma^2(\tau) \qquad \text{<11>}$$

여기서 각각의 분산은 표준값과 반응값의 표준불확도 값으로부터 구할 수 있으며, 이를 이용하여 τ의 분산을 계산할 수 식 <12>와 같이 구할 수 있다.

$$s^2(\tau) = \left(\frac{\sum_{i=1}^{n}(y_i - y_i^0)^2}{n-k-1}\right) - \left(\frac{\sum_{i=1}^{n}\{u^2(y_i)\times\nu_{y_i}\}}{\sum_{i=1}^{n}\nu_{y_i}}\right) - \left(\frac{\sum_{i=1}^{n}\left\{\left(\frac{df}{dx}\right)^2 u^2(x_i)\times\nu_{x_i}\right\}}{\sum_{i=1}^{n}\nu_{x_i}}\right) \qquad \text{<12>}$$

여기서 $u(y_i)$는 표준시료 i의 표준값의 표준불확도

$u(x_i)$는 표준시료 i의 반응값의 표준불확도

ν_{y_i}는 표준시료 i의 표준값의 자유도

ν_{x_i}는 표준시료 i의 반응값의 자유도

3. 결론

측정 및 분석 결과의 불확실성을 계산하는 과정을 GUM에 의한 알고리즘에 기초하여 재정립하였다. 이러한 알고리즘은 PUMA라는 컴퓨터 프로그램으로 구성하여 시범 운영하고 있으며, 여기에서는 이에 대한 기본 알고리즘을 기술하였다.

현재, PUMA를 측정/분석 과정에 적용해도 몇 가지 해결해야할 문제점이 발견되었고, 그것은 다음과 같다.

1) 한 시료에 대하여 측정값과 불확도를 구하는 과정을 여러 번 반복하는 경우
2) 고차식 및 비선형의 검정곡선을 이용한 분석 결과의 추정 및 역추정의 경우
3) 오차가 존재한다는 사실은 알고 있으나, 이것의 불확실성을 정량화하기 어려운 경우
4) 측정과 병행하여 품질관리를 수행하는 경우

이들의 대부분은 측정과 통계적인 방법[4-6]을 병행하여 연구함으로서 해결될 수 있을 것으로 기대되며, 특히, 본 학회의 회원들과 함께 논의되어 해결되기를 기대한다.

참고문헌

1. ISO(1993), “Guide to the Expression of Uncertainty in Measurements,” ISO, Switzerland.
2. NIST Technical Note 1297(1993), “Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results,” NIST, USA.
3. EURACHEM(1995), “Quantifying Uncertainty in Analytical Measurements”
4. DIN 1319-4, "Grundlagan der Messtechnik", DIN, Deuchland
5. W. Bremer and W. Haesselbarth(1997), "Controlling uncertainty in calibration", Analytica Chmica Acta, Vol 348, page 61.
6. NIST Special Publication 500-244(2000), "An Interpretation of the guide to the Expression of Uncertainty in Measurements", NIST, USA.