

측정불확도의 개념과 GUM

이종화

한국표준과학연구원 화학바이오표준본부 무기분석표준그룹
(접수 2023. 7. 3; 게재확정 2023. 8. 2)

The Concept of Measurement Uncertainty and the GUM

Jong Wha Lee

*Inorganic Metrology Group, Division of Chemical and Biological Metrology,
Korea Research Institute of Standards and Science (KRIS), Gajeong-ro 267, Yuseong-gu, Daejeon 34113, Korea.
E-mail: jongwhalee@kriss.re.kr*

(Received July 3, 2023; Accepted August 2, 2023)

요 약. Guide to the Expression of Uncertainty in Measurement (GUM)가 발간된 1993년부터 측정불확도는 측정학의 핵심 개념으로 여겨져 왔으며 적절한 불확도 평가의 중요성은 꾸준히 높아지고 있다. 하지만 측정불확도의 개념과 GUM에 대해 충분히 정확하고 상세하게 설명하는 국문 자료는 거의 없는 실정이다. 본 총설은 측정불확도 개념과 GUM의 수학적, 역사적, 철학적 배경을 상세하게 설명하며 화학 분야 불확도 평가의 특수성을 논한다.

주제어: 측정불확도, GUM, 측정학

ABSTRACT. Ever since the development of the Guide to the Expression of Uncertainty in Measurement (GUM) in 1993, the concept of measurement uncertainty has been considered a core concept in metrology and the importance of proper uncertainty evaluation has continuously been increasing. Unfortunately, few papers in Korean are available that introduce the concept of measurement uncertainty and the GUM correctly and in sufficient detail. This review describes in detail the mathematical, historical, and philosophical background behind the concept of measurement uncertainty and the GUM, and also discusses some special aspects of uncertainty evaluation in chemical analysis.

Key words: Measurement uncertainty, GUM, Metrology

서 론

모든 측정에는 불확실성이 수반된다. 가령 시료 내 납의 질량 분율이 1 mg/kg으로 측정되었다고 하여 시료 내 납의 질량 분율이 정확히 1 mg/kg이라고 믿는 측정자는 없을 것이다. 통상적으로는 동일한 시료를 재측정하거나 다른 사람이 분석할 경우 결과가 조금은 다를 수 있다고 믿는다. 하지만 만약 충분히 검증된 분석 절차를 사용했다면 측정 결과가 틀리더라도 아주 틀리지는 않았을 것이라고도 믿는다. 즉, 측정 결과들은 다소 불완전하지만 아예 무의미하지는 않다. 측정불확도(measurement uncertainty)는 이러한 측정 결과의 불완전성을 일관되게 표현하기 위해 국제표준화기구(International Organization for Standardization; ISO)가 1993년에 발간한 문서 Guide to the Expression of Uncertainty in Measurement (GUM)¹에서 확립한 개념이다.

GUM 발간 이후 다양한 분야에서 GUM에 따른 측정불

확도 평가의 중요성이 꾸준히 높아지고 있다. 하지만 측정불확도의 기본 개념에 대해 정확하고 상세하게 설명하고 있는 문헌이 부족하여 많은 연구자들이 불확도 평가에 어려움을 겪고 있는 것이 실정이다. 이에 본 논문은 측정불확도의 기본 개념과 배경에 대해 상술하는 것을 목표로 한다. 특히 화학 분야는 시료 전처리(sample pretreatment)와 기질 효과(matrix effect)라는 두 중요한 요소를 불확도 평가시 누락하여 비현실적으로 작은 불확도를 보고하는 경우가 많다. 화학 분야에서 이 문제를 해결하기 위해 활용되는 독특한 불확도 평가 방식에 관해서도 설명한다.

인용 표기 및 용어

측정불확도에 관한 일차 문헌인 GUM은 그 내용이 방대하므로 인용 시에는 GUM의 구체적인 구절을 인용할 필요가 있다. 따라서 본 논문은 GUM의 특정 내용을 인용

할 때 GUM의 구절 번호를 병기한다. 예를 들어 ‘GUM 2.3.5’라는 표기는 GUM의 2.3.5번 구절을 인용하는 것이다.

현재 통용되는 측정불확도 관련 용어의 정의는 GUM의 정의와 2012년 발간된 International Vocabulary of Metrology (VIM)² 제3판의 정의이다. 두 문헌 간 가장 큰 차이는 GUM에서는 평균(average) 등의 단일 대표값을 ‘측정 결과(result of a measurement)’라고 부르며 측정불확도를 측정 결과에 수반되는 부수적인 정보로 보았으나 VIM 제3판에서는 측정을 통해 얻은 총체적인 정보를 ‘측정 결과(measurement result)’로 보며 평균 등의 대표값과 측정불확도로 요약하여 표현할 수 있다고 본다. 즉, VIM 제3판에서는 측정불확도가 측정 결과의 일부이다.³ 이 차이는 본 논문의 목적에는 특기할 만큼 크지 않으므로 본 논문에서는 일관성을 위해 GUM의 용어 정의만을 인용한다. GUM과 VIM 모두 측정학(metrology) 분야에서 널리 인정받는 권위 높은 문서이며 현재 Joint Committee for Guides in Metrology(JCGM)라는 8개 국제 단체의 연합이 개정 및 발간을 담당하고 있다.

개 관

측정불확도에 대한 직관적인 이해를 위해 다음과 같은 간단한 상황을 고려해보자. 한 저울로 두 물체 A와 B의 질량을 측정하여 두 물체의 질량 m_A 와 m_B 가 모두 약 100 g이라는 측정값을 얻었다고 가정하자. 그리고 여러 상황을 고려하여 이 측정값의 가능한 오차를 약 ± 1 g 정도로 추정하였다고 하자. 이때 두 물체의 질량의 합 $m_A + m_B$ 는 얼마로 추정할 수 있을까?

간단한 답은 $m_A + m_B$ 가 약 198 g에서 202 g 사이라는 것이다. m_A 와 m_B 가 모두 약 99 g에서 101 g 사이로 추정되기 때문이다. 물론 더 복잡한 답도 있다. m_A 가 $100 \text{ g} \pm 1 \text{ g}$ 으로 추정된다는 정보는 m_A 가 99 g보다 작거나 101 g보다 큰 값을 취할 여지는 전혀 없는데, 99 g부터 101 g까지라는 구간 사이에서는 특별히 취할 가능성이 높은 값이 없다는 정보라고 풀어 쓸 수 있다. 이 정보는 Fig. 1과 같이 m_A 에 직사각형 형태의 확률 분포(rectangular probability distribution)를 부여함으로써 나타낼 수 있다.

마찬가지로 m_B 에도 동일한 형태의 확률 분포를 부여할 수 있다. 이 두 확률 분포는 수학적으로 합성할 수 있다. 이 두 분포를 수학적으로 합성하면 Fig. 1과 같이 198 g과 202 g을 양 극단으로 하는 삼각형 형태의 확률 분포(triangular probability distribution)를 얻게 된다. 직사각형 분포 두 개를 더했는데 삼각형 분포가 되는 이유는, 1부터 6까지의 값을 무작위로 취하는 주사위를 두 번 던져 그 합을 구하였을 때 2나 12를 얻을 확률보다 6이나 7을 얻을 확률

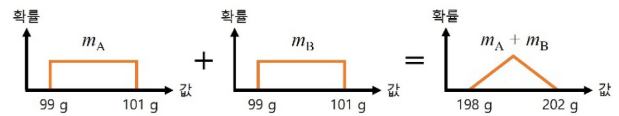


Figure 1. Combining probability distributions.

이 높다는 사실을 떠올려서 직관적으로 이해할 수 있다. 어쨌거나, 이렇게 얻은 삼각형 분포는 단순히 $m_A + m_B$ 가 198 g부터 202 g 사이라는 것보다 더 풍부한 정보를 지니고 있다. 예를 들어 $m_A + m_B$ 가 200 g일 가능성이 가장 높다는 정보를 지니고 있다.

그러므로 사실 $m_A + m_B$ 에 대한 측정 결과를 서로 공유할 때에는 이 최종적인 확률 분포를 공유하는 것이 정보의 소실이 가장 적다. 하지만 이러한 소통 방식은 많은 경우 효율적이지 않다. 많은 경우 더 효율적인 방식은 최종적인 확률 분포에 대한 정보를 축약하여 전달하는 것이다. 확률 분포에 대한 정보를 축약하는 방식은 여러 가지가 있지만 통상적으로 확률 분포를 대표값과 확률 분포의 너비와 관련된 수치로 축약하게 된다. 대표값으로는 기대값(expected value), 중간값(median), 최확값(most probable value) 등이 사용된다. Fig. 1의 간단한 경우에서는 이 셋이 모두 같으므로 구분할 필요가 없다. 확률 분포의 너비와 관련된 수치는 여러 가지가 있으나 실제로 사용되는 방식은 크게 두 가지이다. 첫째는 표준편차(standard deviation)이며 둘째는 확률 분포의 95%를 포괄하는 구간에 관한 수치이다. 이 두 가지 수치를 사용하여 Fig. 1의 삼각형 분포를 축약하여 표현하면 다음과 같다.

- 1) $m_A + m_B = 200.0 \text{ g}$, $u(m_A + m_B) = 0.8 \text{ g}$
- 2) $m_A + m_B = 200.0 \text{ g}$, $U(m_A + m_B) = 1.5 \text{ g}$

후술하겠지만 불확도 평가에서는 ‘표준편차’라는 표현 대신 ‘표준불확도(standard uncertainty; GUM 2.3.1)’라는 표현을 사용하며 통계학에서 주로 사용하는 σ 와 s 라는 기호 대신 u 라는 기호를 사용한다. 또한 95%를 포괄하는 구간에 관한 수치는 ‘확장불확도(expanded uncertainty; GUM 2.3.5)’라고 부르며 U 라는 기호로 나타낸다. 표준불확도와 확장불확도 모두 어떤 측정량이 대표값 외에도 수많은 다른 값을 취할 가능성이 존재한다는 정보를 표현한다. 즉, $m_A + m_B$ 가 약 200 g으로 추정되긴 하는데, 199.5 g일 가능성도 없지 않다는 정보를 간결하게 전달한다. 이 ‘간결한 정보’가 바로 측정불확도이다(GUM D.5.2). 측정불확도는 측정량이 취할 수 있는 값들이 얼마나 넓게 퍼져 있는지(산포도)를 간략하게 표현하는 수치이다. GUM에서의 정의는 다음과 같다(GUM 2.2.3).

“측정량에 합리적으로 부여될 수 있는 값들의 산포도를 나타내는 측정의 결과와 관련된 파라미터(parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand).”

위 정의에서 ‘dispersion’이 종종 ‘분산 특성’으로 번역되어 여러 오해를 일으킨다. ‘분산’은 통계학에서 ‘variance’를 의미하는데, 주로 반복 측정에서 정밀성(precision)의 지표로 사용되는 경우가 많다 보니 불확도를 정밀성과 관계된 지표로 오해하게 된다. 그러나 불확도는 교정(calibration) 등 측정값의 품질에 영향을 줄 수 있는 모든 요소를 종합하여 산출하는 지표이다.

또한 위 정의에서 ‘파라미터(parameter)’를 ‘인수, 인자’라는 의미의 일반적인 용어로 오해하는 경우도 많다. 그러나 이 용어는 “확률 변수의 확률 분포를 묘사하는 양(a quantity used in describing the probability distribution of a random variable)”이라는 의미의 통계적인 용어이다(GUM C.2.7). 즉, 측정불확도는 확률 분포와 밀접하게 관계되어 있는 개념이다.

그런데 측정불확도에 대해 다루는 많은 자료는 측정불확도와 확률 분포 간의 관계를 다루지 않는다. 이는 GUM이 관련 내용을 아주 적극적으로 다루고 있지 않아 그 관계를 정확히 파악하기 어렵기 때문이다. 그 이유를 이해하여야 측정불확도에 대해서 온전히 이해할 수 있는데, 이를 위해서는 GUM이 탄생한 역사적인 배경을 이해할 필요가 있다.

GUM과 측정불확도 개념의 탄생

모든 측정값에는 오차(error)가 있기에 불확실하다는 인식은 새롭지 않다. 라플라스, 가우스를 포함한 많은 수학자들과 과학자들이 측정값의 불확실성을 정량적으로 표현하는 방식을 연구하였으며 이 분야를 오차론(theory of errors)이라고 한다.⁴ 그러나 여러 노력에도 불구하고 1970년대에는 이 ‘오차의 정도’ 혹은 ‘불확실한 정도’를 표현하는 국제적으로 합의된 방식이 없었다.⁵ 이는 국가 측정표준기관(national metrology institute)들이 서로 측정값을 비교하는 데 여러 애로사항들을 유발하였다. 이에 미국 측정표준기관인 National Bureau of Standards [NBS; 현 National Institute of Standards and Technology (NIST)]의 국장이었던 Ernest Ambler는 당시 측정학계에서 높은 권위를 가지고 있던 국제도량형위원회(Comité international des poids et mesures; CIPM)에서 측정에서의 불확실성을 정량적으로 표현하는 합의된 접근법을 개발할 것을 제안하였다(GUM Annex A). 이 제안이 받아들여져 다년 간의 논의

끝에 국제표준화기구에서 1993년 발간한 문서가 바로 GUM이다. 즉, GUM은 측정에서의 불확실성을 정량적으로 표현하는 방식 중 국제적으로 최초로 합의된 방식을 담은 문서라고 볼 수 있으며 그 때문에 측정 분야에서 매우 높은 존중을 받고 있다.

논의를 이어가기 전에 특기할 점은 현재 ‘GUM’이라는 문서가 여러 차례 재출간되었기 때문에 문헌에서 여러 방식으로 인용된다는 점이다. 크게 GUM의 초판본인 1993년 판본, 국제표준화기구에서 재발간한 ISO/IEC Guide 98-3:2008⁶ 판본, 그리고 현재 GUM의 개정을 담당하고 있는 JCGM이 재발간한 JCGM 100:2008¹ 판본이 인용된다. 내용은 동일하므로 모두 같은 문서로 이해하면 된다. 국내에서는 한국표준과학연구원이 발간한 번역본인 ‘측정불확도 표현 지침(KRISS/SP2010-105)’⁷을 일차 문헌으로 오해하는 경우도 종종 있다. 원문은 GUM이다.

몬테 카를로 방법과 불확도 전파의 법칙

중요한 것은 GUM과 측정불확도라는 개념이 1980년대부터 논의되기 시작하여 1993년에 탄생했다는 점이다. 이것이 GUM과 측정불확도라는 개념이 그 토대가 되는 확률 분포에 대한 설명을 전면에 내세우지 못한 이유 중 하나이기 때문이다. 확률 분포라는 개념을 중심에 두고 측정 불확도의 개념을 구성하기 위해서는 Fig. 1과 같이 확률 분포를 합성할 수 있어야 한다. 그러나 Fig. 1과 같이 확률 분포가 간단하게 합성되는 경우는 드물다. 많은 경우 해석학적(analytical)으로 확률 분포를 합성하는 것은 불가능하며 수치적(numerical)으로 합성해야만 한다. 이때 활용되는 것이 컴퓨터를 활용한 몬테 카를로 방법(Monte Carlo method)이다. 그러나 GUM을 만들 당시에는 컴퓨터의 성능이 충분하지 않아 몬테 카를로 방법에 의한 확률 분포 합성을 일반 연구자들이 수행하는 것이 사실상 불가능하였고, GUM은 확률 분포라는 개념이 전면에 나타나지 않는 접근법을 채택하게 되었다.⁸

‘몬테 카를로 방법’은 무작위적인 표본 채취를 활용하여 수치적으로 문제를 해결하는 방식을 말한다. 예를 들어 1부터 6까지의 수가 동일한 확률로 나올 수 있는 주사위를 두 번 던져서 나온 두 눈의 합이 12가 되는 확률을 구하는 문제가 있다고 하자. 간단히 $1/6 \times 1/6 = 1/36$ 이라는 계산을 통해 확률을 구할 수도 있겠지만 컴퓨터를 활용하여 1부터 6까지의 난수(random number)를 두 번 얻고 그 합을 구하는 과정을 수 천 번 반복한 후 합이 12가 된 경우의 비율을 구하는 방법도 있다. 후자의 문제 해결 방식을 몬테 카를로 방법이라고 한다. 이 간단한 경우에는 수치적인 방식이 효과적이지는 않겠지만 이러한 방식이 더 효율적

이거나 현실적으로 유일한 방식이 경우가 많다. 확률 분포의 합성이 이러한 경우이다.

측정량 Y 가 $Y=A/BC$ 라는 관계를 가진다고 하자. 몬테 카를로 방법을 통해서 Y 의 확률 분포를 얻기 위해서는 먼저 A, B, C 에 대해 사전 분포(prior distribution)라고 불리는 확률 분포를 설정해야 한다. 그 후 세 확률 분포에서 확률에 따라 무작위적으로 값을 하나씩 뽑는다. 이를 a, b, c 라고 하자. 이 세 값을 통해 $y=a/bc$ 와 같이 y 값을 하나 구할 수 있다. 이 과정을 수백만 번 반복한 후 얻어진 y 값들의 분포를 정리한 것이 Y 의 확률 분포이다. 이를 사후 분포 (posterior distribution)라고 한다. 이렇게 사후 분포를 얻었다면 측정자는 이 확률 분포를 타인과 그대로 공유할 수도 있고 필요에 따라 표준편차 혹은 95%를 포괄하는 구간 등으로 축약한 후 공유할 수도 있다.

그런데 GUM을 개발하던 당시에는 컴퓨터 자원과 몬테 카를로 방법을 구현하는 적절한 알고리즘에 한계가 있었다. 그래서 GUM은 고전적이지만 확률 분포를 직접적으로 합성하지 않아도 되는 고전적인 방식을 측정불확도 평가 방식의 중심으로 두었다.⁸ 이 방식은 입력량의 확률 분포를 합성하여 측정량의 확률 분포를 구하는 대신, 입력량의 확률 분포의 표준편차만을 합성하여 측정량의 표준편차만을 구하는 방식이다. 이 방식에 사용되는 수식을 오차 전파의 법칙(the law of propagation of errors)이라고 한다. $y=f(x_1, x_2, \dots, x_n)$ 에 대해 오차 전파의 법칙은 식 (1)과 같이 나타낼 수 있다.

$$\sigma^2(y) = \sum_i^n \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma^2(x_i) \quad (1)$$

식 (1)의 의미를 간략하게 설명하자면, 입력량의 표준편차 $\sigma(x_i)$ 를 통해 측정량의 표준편차 $\sigma(y)$ 를 추정할 수 있다는 것이다. 따라서 사실 ‘오차 전파의 법칙’이라는 명칭은 적절하지 않다. 입력량의 오차(측정값과 참값의 차)를 통해 측정량의 오차를 구하는 것이 아니라 입력량의 표준편차를 바탕으로 측정량의 표준편차를 추정하는 식이기 때문이다.⁵ 여기서 표준편차는 어떤 양에 대한 불완전한 지식을 나타내는 지표라고 볼 수 있으므로 측정학에서는 ‘오차 전파의 법칙’이라는 표현 대신 ‘불확도(불확실성) 전파의 법칙(the law of propagation of uncertainties)’이라는 표현을 사용한다. 또한 불확도 평가에서는 표준편차를 σ 로 표기하기보다는 u 로 표기한 후 표준불확도로 부른다. 따라서 식 (1)은 (2)와 같이 입력량의 표준불확도 $u(x_i)$ 를 통해 측정량의 표준불확도 $u(y)$ 를 추정하는 식이 된다 (GUM 5.1.2).

$$u^2(y) = \sum_i^n \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) \quad (2)$$

본 논문에서도 별도의 이유가 없다면 식 (2)와 ‘불확도 전파의 법칙’이라는 명칭을 사용한다. 입력량 간 상관관계(correlation)이 있는 경우 식이 더 복잡해지나(GUM 5.2.2) 여기서는 관련 설명을 생략한다.

참고로 식 (2)는 ‘법칙(law)’이라는 명칭에 걸맞지 않게도 근사식에 불과하다. 이 식의 유도 과정은 굉장히 간단한데, y 에 대해 테일러 급수를 취한 후 고차항을 배제한 것이다(GUM E.3.1). 유도 과정에 아주 특수한 가정이 들어가지 않는 만큼 다양한 경우에 널리 사용할 수 있는 식이지만 항상 정확하지는 않다. 2000년대 후반부터 이 식이 정확하지 않을 것으로 추정될 경우에는 GUM 방식 대신 몬테 카를로 방식을 사용하기를 권장하는 문헌들이 점차 늘고 있으며 몬테 카를로 방식으로 불확도를 평가하는 것도 GUM과 합치되는 방식으로 인정받는다.⁹ 하지만 대부분의 경우 GUM의 근사적인 방식이 실용적인 문제를 일으키지는 않으며, GUM 방식이 불확도 평가 시스템을 구축하기 편리한 측면이 있으므로⁵ 특별한 이유가 없다면 GUM 방식을 우선으로 고려할 것을 권할 만하다.

표준불확도를 얻는 두 가지 방식

결과적으로 현재 GUM에 따른 불확도 평가 방식에서 중심이 되는 수학적 도구는 표준불확도와 표준불확도를 합성하는 식인 불확도 전파의 법칙이다. 복잡한 고민이 필요한 경우가 없는 것은 아니지만 표준불확도의 합성은 일단은 공식을 따라 수행하면 되므로 불확도 평가에서 가장 중요한 것은 결국 입력량의 표준불확도를 구하는 것이다. 입력량의 표준불확도를 구하는 방식은 크게 두 가지이다.

첫 번째 방식은 반복 측정을 통해 얻은 데이터를 통계적으로 분석하는 것이다. 대부분의 경우에는 식 (3)과 같이 n 개의 데이터 간 실험 표준편차(experimental standard deviation; GUM B.2.17) s 를 \sqrt{n} 으로 나누어 표준불확도를 구하게 된다(GUM 4.2.3).

$$u = \frac{s}{\sqrt{n}} \quad (3)$$

식 (3)의 수리적인 배경은 중심극한정리(Central Limit Theorem; GUM G.2)이다. 중심극한정리에 따르면 임의의 확률 분포에서 n 개의 표본을 취할 경우 그 평균의 표준편차를 모표준편차 σ 를 n 의 제곱근으로 나눈 값으로 추정할 수 있게 된다. 그런데 많은 경우 모표준편차 σ 를 직접 알 수 없으므로 이를 실험적 추정치 s 로 대체한다.

통계학에서는 s/\sqrt{n} 을 ‘표준 오차(standard error)’라고 부르는 경우가 많다. 하지만 이는 오해의 여지가 큰 명칭

이다. s/\sqrt{n} 은 오차(틀린 정도)를 나타내는 것이 아니라 동일한 방식으로 평균을 다시 얻었을 때 평균이 얼마나 변할 수 있는지를 나타내기 때문이다. 즉, 평균의 불확실성을 표준편차의 형태로 나타낸 것이다. 이 때문에 측정학 문헌에서는 s/\sqrt{n} 을 ‘평균의 표준편차(standard deviation of the mean; GUM B.2.17 NOTE 2)’라고 부른다.

두 번째 방식은 기존 자료나 경험 등에 기반하여 상정한 확률 분포의 표준편차를 구하는 방식이다. 불확도 관련 자료를 보면 $\sqrt{3}$ 이라는 수치가 자주 등장한다. 이는 어떤 양이 취할 수 있는 값의 양 한계치만 알고 있을 경우 직사각형 분포를 상정하게 되는 경우가 많으며, 직사각형 분포의 표준편차는 그 반너비(half width)를 $\sqrt{3}$ 으로 나누는 것과 같기 때문이다. 이를 간략히 증명하면 아래와 같다(GUM 4.4.5).

$-a$ 와 $+a$ 를 양 극단으로 하는 직사각형 형태의 확률밀도 함수(probability density function) $f(x)$ 의 높이는 $1/2a$ 이다. 정의상 확률밀도함수 아래의 면적은 1이어야 하기 때문이다. 따라서 $f(x)$ 는 다음과 같이 쓸 수 있다.

$$f(x) = \frac{1}{2a} \quad (-a \leq x \leq a) \quad (4)$$

확률밀도함수의 표준불확도(표준편차)는 다음과 같이 정의된다.

$$u = \sqrt{\int_{-a}^a (x-\mu)^2 f(x) dx} \quad (5)$$

여기서 μ 는 이 확률 분포의 기대값이며 $\int xf(x)dx$ 로 정의된다. 이 경우에는 0이다. 따라서 $\mu=0$ 으로 두고 적분하면 $u = a/\sqrt{3}$ 을 얻을 수 있다.

같은 방식으로 다른 형태의 분포에 대해서도 표준불확도를 구할 수 있다. 가령 어떤 양의 값이 $-a$ 와 $+a$ 사이로 추정되긴 하는데, 그 중에서는 0일 가능성이 가장 높다고 판단된다면 이 정보를 $-a$ 와 $+a$ 를 양 극단으로 하고 0에 꼭지점을 가지는 삼각형 분포(triangular distribution)로 나타낼 수 있다(GUM F.2.3.3). 이 경우 $u = a/\sqrt{6}$ 이 된다(GUM 4.3.9).

이 방식으로 표준불확도를 구할 때 어떤 확률 분포를 상정할지 선택하는 것은 어느 정도 임의적이며 근사적이다.¹⁰ 가령 위 삼각형 분포에서 양의 값이 $1/2a$ 일 확률은 값이 0일 확률의 정확히 절반이다. 이것이 사실을 정확히 나타내지 않는다고 판단될 경우 확률 분포의 각 부분을 세부적으로 조정하면 그 확률 분포의 표준불확도를 수학적으로 구하는 접근법을 취할 수는 있다.¹⁰ 하지만 통상적으로는 그러한 세부 조정이 유의미한 영향을 줄 가능성은 높지 않다. 불확도 자체의 불확실성은 대부분 중요하지

않기 때문이다. 예를 들어 어떤 물체의 질량을 10% 잘못 알고 있으면 큰일이겠지만 불확도를 10% 잘못 추정하는 것은 실용적으로 크게 중요하지 않을 가능성이 크다. 애초에 위에서 언급한 바와 같이 불확도 전파의 법칙 자체가 근사식이다. 또한 확률 분포를 상정하는 데 사용하는 정보도 어느 정도 불확실할 수밖에 없으므로 대부분의 경우 현실을 적당히 잘 표현할 것으로 기대되는 간단한 형태의 확률 분포를 사용하게 된다(GUM 4.3.9 NOTE 1). 이 때문에 어떤 양이 취할 수 있는 범위만 알고 있을 경우 보통 직사각형 분포를 상정하게 되며 불확도 평가 자료에서는 $\sqrt{3}$ 이 빈번히 등장한다. 온너비(full width)를 기준으로 하면 $2\sqrt{3}$ 으로 나누어야 하므로 $\sqrt{12}$ 도 자주 등장한다.

위에서 설명한 두 가지 방식 중 반복 측정 데이터를 사용한 첫 번째 방식을 A형 불확도 평가(type A evaluation of uncertainty; GUM 2.3.2), 반복 측정 데이터를 사용하지 않은 두 번째 방식을 B형 불확도 평가(type B evaluation of uncertainty; GUM 2.3.3)라고 한다. 즉, 불확도 평가에서는 A형 혹은 B형 평가 방식으로 표준불확도를 얻고 이를 합성하여 최종 불확도를 얻게 된다. 다만 특정 불확도 평가 방식을 항상 A형이나 B형으로 명확하게 나눌 수 있는 것은 아니니 이 구분에 아주 큰 의미를 둘 필요는 없다.¹¹ GUM이 불확도 평가 방식을 크게 두 가지로 나누어서 설명하는 것은 불확도 평가 방식을 어느 정도 정형화하여 이해를 돕기 위한 것이라고 보는 것이 좋다(GUM 3.3.4). 둘 중 어떠한 방식으로 표준불확도를 평가했든, 그 표준불확도는 어떤 양에 대한 측정자의 최선의 지식을 담은 표준편차라는 것은 동일하다(GUM 4.1.6). 후술하겠지만 사실 이것이 GUM의 핵심이다.

확장불확도

표준불확도를 평가하고 합성하여 다음과 같은 측정 결과를 얻었다고 해보자: “이 길이의 최적 추정량은 5.0 m이며 표준불확도는 0.1 m입니다.” 그런데 이 측정 결과는 정확히 무엇을 의미할까? 사실 이 결과의 직관적인 의미를 찾는 것은 쉽지 않다. 이러한 한계를 해결하기 위해 도입된 개념이 확장불확도이다. 확장불확도를 구하면 “이 길이는 95% 확률로 4.8 m와 5.2 m 사이로 추정됩니다”와 같은 표현을 사용할 수 있게 되는데, 이러한 표현은 이 길이가 4.8 m보다 작거나 5.2 m보다 클 가능성이 거의 없다는(적어도 측정자는 그렇게 믿는다는) 상당히 직관적인 의미를 전달한다.

확장불확도를 구하기 위해서는 ‘포함인자(coverage factor; GUM 2.3.6)’라는 것을 구하여 표준불확도에 곱해주면 된다. 확장불확도는 U , 포함 인자는 k 로 표기하므로 식 (6)과

같이 쓸 수 있다(GUM 6.2.1).

$$U = ku \quad (6)$$

복잡한 수학적 배경을 생략하고 설명하자면, 포함 인자는 신뢰의 수준(level of confidence; GUM 6.2.2) p 와 유효자유도(effective degree of freedom; GUM G4.1) v_{eff} 에 의해 완벽하게 결정되므로 실용적으로는 이 두 수치만 구할 수 있으면 확장불확도를 어렵지 않게 얻을 수 있다. 두 수치 중 p 는 역사적인 이유와 실용적인 이유로 대부분 95% (0.95)를 사용하는 것이 관례이므로 실질적으로는 유효자유도 v_{eff} 만 구하면 된다.

유효자유도는 개별 불확도 요소의 자유도를 구한 다음 합성하여 구하게 된다. 개별 불확도 요소의 자유도를 구하는 방식은 불확도 평가 방식에 따라서 나뉜다. 반복 측정을 통해 얻은 표준불확도(A형 불확도 평가)의 경우 대부분 반복 횟수가 n 이었을 때 $n-1$ 이 자유도가 된다(GUM 4.2.6). 피팅(fitting)이나 복잡한 통계적 분석을 통해 얻은 수치는 자유도가 다를 수 있으나 통상적으로 그런 경우에는 자유도를 어떻게 구하면 되는지에 대한 설명이 주어지지 않으므로 자세한 설명은 생략한다.

이전의 경험, 자료를 통해 얻은 표준불확도(B형 불확도 평가)의 경우 불확도 자체의 신뢰성을 평가하여 자유도를 얻을 수 있다(GUM G4.2). 이 방식을 따를 경우, 완벽하게 신뢰할 수 있는 자료를 사용한 경우 자유도가 무한대가 되는 반면, 50%만 신뢰할 수 있는 자료를 사용한 경우 자유도가 2가 된다(GUM G4.2 EXAMPLE). 그런데 실무에서 GUM에서 제시된 이 방식으로 자유도를 얻는 경우는 드물다.⁵ 후술하겠지만 이 방식이 수학적으로 타당하지 않은 방식이기 때문이다. 이 때문에 보통 B형 불확도 평가를 할 때에는 아주 신뢰성이 높은 자료를 사용하며 자유도를 무한대로 두는 것이 일반적이다. 신뢰성이 부족하더라도 표준불확도를 더 크게 하는 것으로 대응하지 자유도를 낮추는 경우는 드물다.

위와 같이 개별 요소의 자유도를 얻었다면 Welch-Satterthwaite 식 (7)을 통해 합성하여 유효자유도를 얻을 수 있다(GUM G4.1).

$$v_{eff} = \frac{u^4(y)}{\sum_i \frac{u_i^4(y)}{v_i}} \quad (7)$$

v_{eff} 을 구했다면 t -분포표(t -table; GUM Table G2)에서 v_{eff} 에 해당하는 값을 읽어서 포함인자를 구할 수 있다. 만약 Microsoft Excel을 보유한 경우 'TINV(0.05, v_{eff})'을 입력하여 간편히 포함인자를 구할 수 있다. 구한 포함인자 식 (6)과 같이 표준불확도에 곱하면 확장불확도가 얻어진다.

위에서 설명한 방식대로 포함인자를 구하지 않고 단순히 $k=2$ 로 가정하는 경우도 많다. 이는 자유도가 아주 작은 경우가 아니라면 k 는 2에 근접하기 때문이다(GUM G6.6). 예를 들어 자유도가 10일 경우 k 는 약 2.2가 되어 k 를 간단히 2라고 가정한 것과 10% 밖에 차이 나지 않는다. 위에서도 언급하였지만 측정값의 상대 오차가 10%라면 문제가 될 경우가 많겠지만 불확도의 오차가 10%인 것이 실용적인 문제로 이어질 가능성은 낮다. 이 때문에 단순히 $k=2$ 를 사용하는 경우가 많다.

그런데 화학 분석에서는 $k=2$ 로 가정하는 방식이 적절하지 않은 경우가 많다. 화학 분석에서는 시료 간 무작위적인 변동(random variation)이 가장 큰 불확도 요소인 경우가 많기 때문이다. 이런 경우에서 시료의 수가 충분하지 않다면 $k=2$ 라는 가정은 확장불확도의 과소평가로 이어질 수 있다. 가령 시료 수가 3이며 시료 간 차이가 가장 큰 불확도 요소인 경우 유효자유도는 약 2가 되는데, 이 경우 적절한 k 는 약 4로 단순히 k 를 2로 가정한 것과 두 배나 다르다. 따라서 적어도 한 번은 위의 방식으로 포함인자를 구하여 $k=2$ 라는 가정이 유효한지를 살펴보는 것이 바람직하다.

확장불확도에 대해 한 가지 특기할 점은 확장불확도를 구하는 GUM 방식에 여러 가지 수학적 문제가 있다는 점이다. GUM 방식의 수학적 배경은 t -분포(t -distribution)라는 것인데, t -분포를 활용하기 위해서는 표본을 취하는 원분포가 정규 분포(normal distribution)여야 한다. 그런데 이 조건이 충족되지 않는 경우도 많다. 더 큰 문제는 자유도라는 개념은 원래 B형 불확도 평가를 통해 얻은 데이터에는 적용될 수 없는 개념이라는 점이다. 이러한 여러 문제로 인해 현재 GUM의 확장불확도 산출 방식은 많은 비판의 대상이 되고 있다. 위에서 표준불확도를 산출하는 방식도 어느 정도 근사적이라고 설명하였다. 그러나 이 한계는 많은 경우 실용적인 문제로 이어지지 않으며 개념적인 문제는 아닌 데 반해 확장불확도를 구하는 방식은 때때로 실용적인 문제로 이어지며 개념 자체의 유효성이 의문시된다.

이러한 문제는 몬테 카를로 방법을 사용하는 것으로 해결할 수 있다.⁸ 사실 “95% 신뢰의 수준에서의 확장불확도”란 “측정량의 확률 분포의 95%를 차지하는 구간”이라는 의미이다. GUM의 “확장불확도”의 정의를 살펴보면 아래와 같다(GUM 2.3.5).

“측정량에 합리적으로 부여할 수 있는 값들의 분포의 상당 부분을 포함할 것으로 기대될 수 있는 측정의 결과와 관련된 구간을 나타내는 양(quantity defining an interval about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that

could reasonably be attributed to the measurand).”

GUM에서의 문제는 측정량의 확률 분포를 직접 구하지 않기 때문에 발생한다. 측정량의 확률 분포를 직접 구할 수 없어서 그 형태에 대해 가정을 하는 과정에서 문제가 발생하는 것이다. 따라서 확률 분포를 직접 구하는 몬테 카를로 방식에서는 이 문제가 상당 부분 해결된다. 확률 분포가 있다면 95% 신뢰의 수준에서의 구간은 손쉽게 계산할 수 있기 때문이다.

이러한 장점을 고려하여 JCGM은 2014년에 몬테 카를로 방법을 기본으로 하도록 GUM을 개정할 것을 제안하였다.⁸ 그러나 최종적으로 이 제안은 기각되었다. GUM에 개선할 점이 없는 것은 아니나 GUM의 한계가 실용적인 문제로 이어지는 경우는 드물며, 이제야 GUM이 자리를 잡아가고 있는 상황에서 전면 개정을 하는 것은 관련 분야에 큰 혼란을 일으킬 것이라는 관련 기관의 의견들 때문이었다.^{8,12} JCGM은 이러한 반대 의견들이 아직까지 GUM 방식과 몬테 카를로 방식에 대한 이해의 저변에 높지 않기 때문으로 판단, 관련 설명 자료를 적극적으로 제작하여 배포한 후 미래에 개정을 다시 추진하기로 계획하였다.⁸ JCGM이 2008년부터 발간하고 있는 JCGM 10x 시리즈 문서들이 이러한 문서들이다. 이 중에서 몬테 카를로 방법의 적용을 설명하는 JCGM 101:2008은 많은 분야에서 적극적으로 인용되고 사용되고 있다.⁹

개인적으로 확장불확도는 어디까지나 직관적인 이해를 돕기 위해 제공되는 성격이 강하므로 어느 정도의 문제는 실용적으로 용인될 수 있다고 생각한다. 대부분의 수학적 작업에 사용되는 것은 결국 확장불확도가 아닌 표준불확도다. 가령 교정성적서나 표준물질 인증서에는 통상적으로 확장불확도가 주어지나 사용자들은 결국 이를 표준불확도로 변환하여 사용한다. 측정값을 서로 비교할 때도 대부분 표준불확도가 사용된다.

따라서 일단은 GUM 방식으로 확장불확도를 구하되, GUM 방식이 실용적인 문제로 이어질 것으로 기대되는 경우에는 몬테 카를로 방식의 사용을 고려해보는 것이 합리적이라고 하겠다. 이는 NIST의 내부 권장사항과도 같다.¹² 이러한 경우로는 먼저 순수한 물질의 순도를 구하거나 검출 한계 근처에서 측정을 수행하는 경우 등 양이 취할 수 있는 값에 물리적인 한계가 있는 경우가 있다. 이런 경우 몬테 카를로 방법 대신 GUM 방식을 사용할 경우 순도가 100% 이상으로 나타나거나 농도가 음수로 나타나는 등 물리적으로 불가능한 결과를 얻을 수 있다.¹³ 또한 자유도가 작은 경우나 정규 분포(normal distribution)에서 크게 벗어난 형태의 확률 분포를 합성할 때에도 몬테 카를로 방식이 유용하다(GUM G.2.3). NIST가 제공하는 웹페이지에서 몬테 카를로 방법을 사용한 불확도 평가를 수행해볼

수 있다.¹⁴

간단한 예시

간단한 예시를 통해 지금까지 설명한 개념들이 어떻게 적용될 수 있는지를 설명한다. 한 직각삼각형의 빗변의 길이 a 를 다른 두 변의 길이 b, c 를 통해 추정한다고 해보자. 피타고라스 정리에 의해 a 는 다음과 같다.

$$a = \sqrt{b^2 + c^2} \tag{8}$$

참고 자료를 통해 b 를 2.9에서 3.1사이로 추정할 수 있다고 하자. 이 경우 b 의 최적 추정치(best estimate)는 3이 되며 $u(b) = 0.1/\sqrt{3} = 0.058$ 이 된다. 또한 c 를 반복 측정을 통해 구하였으며 이 때 데이터가 4.02, 3.98, 4.04, 3.97, 3.99였다고 하자. 이 경우 c 의 최적 추정치는 4가 되며 $u(c) = s(c)/\sqrt{5} = 0.013$ 이 된다.

식 (8)에 불확도 전파의 법칙을 적용하면 $u(a)$ 는 다음과 같다.

$$u^2(a) = \left(\frac{\partial a}{\partial b}\right)^2 u^2(b) + \left(\frac{\partial a}{\partial c}\right)^2 u^2(c) = \frac{b^2}{b^2 + c^2} u^2(b) + \frac{c^2}{b^2 + c^2} u^2(c) \tag{9}$$

값을 대입하면 $u^2(a) = 0.0013$, $u(a) = 0.036$ 을 얻을 수 있다. 만약 참고 자료 b 가 아주 믿을 만하여 자유도를 무한대로 둘 수 있다면 $v_{eff} = 568$ 가 되어 95% 신뢰의 수준에서의 $k = 1.97$ 이 된다. 근사적으로 $k = 2$ 를 사용하면 $U(a) = 0.072$ 이다. 만약 몬테 카를로 방식을 사용할 경우 $u(a) = 0.036$, $U(a) = 0.063$ 가 얻어진다. 확장불확도만 조금 다른데, 이는 본 예시에서 정규 분포와 그 형태가 많이 다른 직사각형 분포가 최종 불확도에 많이 기여하여 최종 확률 분포도 정규 분포에서 벗어난 형태를 지니기 때문이다(Fig. 2). 실용적인 관점에서는 차이가 아주 크지는 않다.

이 예시에서 볼 수 있듯이 불확도 평가는 복잡하거나 개념적으로 어렵지 않으며 통계에 대한 아주 전문적인 지식이 필요하지도 않다. 불확도 평가를 위해서는 Fig. 3과 같이 1) 측정량을 정하고, 2) 측정 전략을 수립하여 식(모델 식)으로 나타내고, 3) 불확도 요인을 파악하여 표준불확도의 형태로 나타내고, 4) 표준불확도를 합성하여 합성 표준불확도를 얻고, 5) 포함인자를 구하여 확장불확도를 얻으면 된다. 만약 몬테 카를로 방법을 사용한다면 불확도 요인을 파악하여 확률 분포의 형태로 나타내고 몬테 카를로 방법으로 합성한 후 최종 확률 분포의 형태를 분석하여 표준불확도나 확장불확도를 얻게 된다. 이 중에서 가장 어려운 것은 측정 전략을 수립하고 그 측정 방식에서 어떠한 요인이 불확실성을 유발하는지를 파악한 후 정량적으로 표현하는 것이다. 위의 예시에서 불확도 평가가 아주 간단했던 이유도 위 예시가 이 핵심적인 부분을 간

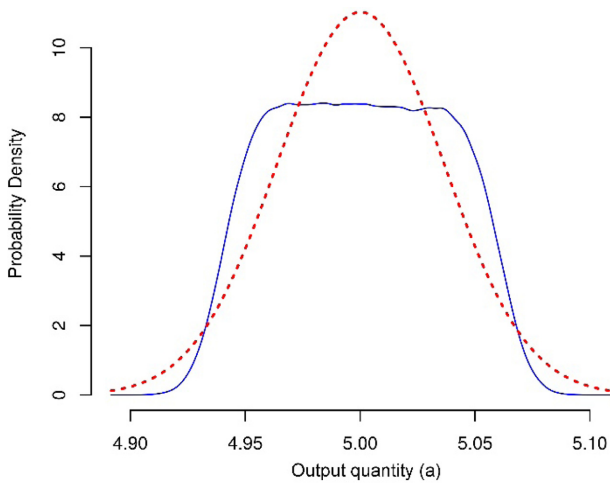


Figure 2. A comparison of a posterior distribution obtained using the Monte Carlo method implemented in the NIST Uncertainty Machine (solid blue)¹⁴ and the corresponding normal distribution (red dash).

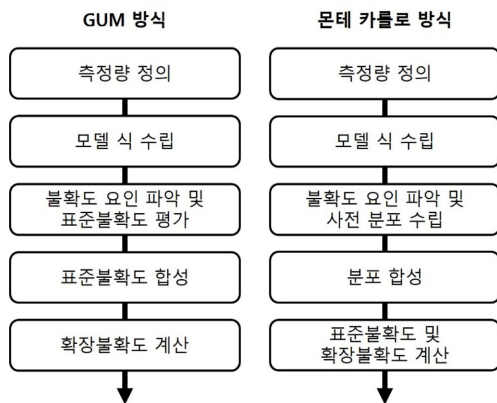


Figure 3. A flow chart of uncertainty evaluation using the GUM approach and the Monte Carlo approach.

략하게만 다룬 관념적인 예시이기 때문이다. 각 분야에서 측정 전략을 수립하고 불확실성을 유발하는 요소를 파악하는 방법을 익히는 데 왕도는 없으며 해당 분야의 문헌을 토대로 일반적으로 통용되는 방식을 모방해본 후 조금씩 비판적으로 개선해보는 것이 최선이다. 추후 다른 논문에서 구체적인 예를 다룰 계획이다. 현존하는 자료로는 교정 분야에서는 EA-4/02,¹⁵ 화학 분석에서는 Quantifying Uncertainty in Analytical Measurement (QUAM)¹⁶가 유용하다.

위 예시를 토대로 몇 가지 개념적인 설명을 더하면 다음과 같다. 불확도 전파의 법칙에 따라 합성하여 최종적으로 얻은 표준불확도를 합성표준불확도(combined standard uncertainty; GUM 2.3.4)라고 하며 기호 u_c 로 나타낸다. 일상적으로는 최종 불확도(final uncertainty), 총 불확도(total uncertainty), 전체 불확도(overall uncertainty) 등의 표현도 통용되며 간단히 표준불확도라고 부르는 경우도 많다.

불확도 전파의 법칙에서의 편미분 항 $\partial f/\partial x_i$ 을 감도 계수(sensitivity coefficient; GUM 5.1.3)라고 한다. 이 편미분 항이 클수록 관련 표준불확도의 작은 변화도 최종 불확도에 큰 영향을 주게 되기 때문이다.

어떤 항의 감도 계수와 표준불확도를 곱하여 얻은 수치 $u_i = (\partial f/\partial x_i)u(x_i)$ 가 i 요소에 의한 불확도이다. u_i^2/u_c^2 을 열거하여 어떤 요소가 합성표준불확도에 영향을 많이 미치는지를 편리하게 파악해볼 수 있으며 이를 불확도 총괄표(uncertainty budget)라고 한다.¹⁷ 이렇게 구한 수치를 모두 더하면 1 (100%)이 되어 각 요소가 최종 불확도에 얼마나 기여하는지를 한 눈에 파악할 수 있기 때문이다. 예를 들어 위의 예시에서는 이 수치가 b 와 c 에 대해 각각 약 92%와 8%가 되는데 이를 통해 추후 불확도를 더 낮추고 싶을 경우 b 에 의한 불확도를 우선적으로 낮추어야 한다는 것을 알 수 있게 된다. 현재 국내에서는 이렇게 계산된 수치를 ‘기여율(contribution)’이라고 부르며 이 계산 방식이 정형화되어 있다.

그런데 이러한 계산 방식에 한계가 없는 것은 아니다. 예를 들어 위의 방식처럼 제공된 수치를 합성표준불확도의 제공으로 나누어 계산한 수치는 총합이 100%가 되어 매우 직관적이라는 장점은 있으나 합성표준불확도는 결국 제공근을 취하여 구하게 되므로 실제 영향과는 다소 차이가 있다. 그리고 화학 분야에서는 불확도 평가를 할 때 반복성(repeatability; GUM B.2.15)에 기여하는 요소를 별도로 종합하여 평가하는 경우가 많은데,¹⁶⁻¹⁷ 이 경우 위 방식으로 얻은 수치는 해석에 주의를 요한다. 이러한 이유들 때문인지 국외에서는 ‘기여도(contribution)’를 평가할 때 특정 방식을 사용하라는 규정을 찾기 어려우며 단순히 u_i 를 정리하여 열거하는 경우도 많다.¹⁶ 개인적으로는 각 요소가 최종불확도에 미치는 영향을 파악한다는 목적에 부합하는 방식이라면 모두 용인될 수 있다고 생각한다.

화학 분야 측정의 특수성과 하향식 불확도 평가

지금까지 설명한 방식과 같이 개별 불확도 요소를 파악하고 평가한 후 합성하여 최종 불확도를 얻는 방식을 상향식 불확도 평가(bottom-up evaluation of uncertainty)라고 하며 가장 보편적인 불확도 평가 방식이다. 화학 분야에서는 하향식 불확도 평가(top-down evaluation of uncertainty)라는 또 다른 불확도 평가 방식이 있다.¹⁸ 많은 경우 일반 시험 분석에서는 이 방식이 현실적인 불확도 평가 방식이다. 개별 불확도 요소를 찾고 합리적으로 평가하는 것이 쉽지 않기 때문이다.

1990년대 초부터 화학 측정의 품질 향상을 위한 여러 국제적인 노력이 있었다. 당시 핵심적인 논제 중 하나가

과연 화학 측정이 ‘측정’인지 여부였다. 이 논의는 화학 측정에도 물리 측정과 동일한 개념을 적용할 수 있다는 결론에 다다르며 마무리되었다.¹⁹⁻²⁰

하지만 화학 측정과 물리 측정 사이 개념적인 차이는 크지 않더라도 실무적인 차이는 크다. 이를 명확히 보여주는 것이 바로 유효성 검토용 기질 인증표준물질(matrix certified reference material)의 존재다. 물리 분야에서는 측정 장비를 교정해주거나 분동 등 사용자가 교정에 직접 사용할 수 있는 도구를 제공함으로써 측정 표준을 보급한다. 화학 분야에서도 ‘표준 가스’ 혹은 ‘표준 용액’ 등으로 불리는 인증표준물질이 동일한 역할을 한다. 그런데 화학 분야에서는 기질 인증표준물질이라는 또 다른 측정 표준도 존재한다. 플라스틱 속 유해 물질의 양이나 식품 내 영양 성분의 양 등 실제 시료에 가까운 물질에 대해 인증값(certified value)이 있는 물질을 기질 인증표준물질이라고 한다.²¹ 이러한 물질은 화학 분야에서는 기기 교정만으로 측정 결과의 품질을 확보하기 어렵기 때문에 개발된 것이다. 화학 분석에서는 R^2 값이 1에 가깝도록 교정 곡선(calibration curve)을 잘 얻었더라도 시료 전처리 등 다른 단계에서 문제가 발생하면 소용이 없으므로 교정용 측정 표준을 보급하는 것만으로 측정 결과의 품질을 보증하기 어렵다. 그래서 실제 시료에 가까운 물질을 보급함으로써 측정자가 전체적인 측정 과정의 유효성을 검토할 수 있게 돕는다. 이러한 유형의 측정 표준은 화학 분야에만 존재한다.

GUM방식에 따라 통상적인 화학 분석에서 불확도 요소로 볼 수 있는 것을 꼽아보면 사용한 표준물질의 불확실성, 질량 측정 및 부피 측정에서의 불확실성, 기기 분석에서의 불확실성 등이 있다. 대부분의 경우 이 요소들로 인한 상대불확도는 도합 1% 남짓인 경우가 많고 5%를 넘는 경우는 드물다. 이와 모순되게도 서로 다른 화학 분석 시험 기관의 측정 결과는 통상적으로 약 50% 정도의 상대적인 편차를 보인다.²²

화학 분석에서의 불확실성은 주로 시료 전처리나 기질 효과 등 모델 식에 드러나지 않는 요소에 의해 발생한다. 화학 분석에서는 직선 형태의 교정 곡선을 그린 후 그 역함수를 활용하여 정량을 하는 경우가 많은데 이러한 계산 방식에서는 시료 전처리와 기질 효과로 인한 오차가 고려되지 않으므로 교정 곡선의 불확도만 고려하면 비현실적으로 작은 불확도를 얻게 된다. 사실 이 요소들을 식에 넣어서 고려하는 것이 불가능하지는 않으나¹⁶ 이 요소들로 인한 불확실성을 합리적으로 평가하기 어렵다는 문제점이 남는다. 이 요소들로 인한 불확실성은 잘 정립된 이론과 관계식을 정확히 따르지 않으며, 측정자가 완벽하게 제어하거나 이해하기 어렵고 실험 조건의 미묘한 차이가 다른 결과를 낼 수 있기 때문이다.

이러한 문제에 대응하여 측정자가 취할 수 있는 접근법은 크게 두 가지로 요약할 수 있다. 첫째는 철저한 유효성 검토를 통해 전처리나 기질 효과로 인한 문제를 거의 완전히 제거하는 것이다. 이러한 접근법은 주로 국가 측정표준 기관이 인증표준물질을 개발할 때 사용한다. 일반 시험 기관에서도 이러한 접근법을 취하는 것이 불가능하지는 않지만 현실적이기는 어렵다. 화학 분석에서는 성분마다 측정법이 달라야 할뿐더러 같은 성분도 서로 다른 기질에 포함되어 있거나 농도가 다르면 측정법을 별도로 최적화해야 하는 경우가 많은데 그 모든 경우에 대해 유효성 검토를 완벽하게 수행하는 데 필요한 자원이 상당하기 때문이다.

따라서 유효성 검토 데이터를 활용한 하향식 불확도 평가라는 두 번째 접근법이 보다 현실적이라고 평가받는다.¹⁶ 이 방식은 인증표준물질 사용 결과나 숙련도 시험(proficiency testing) 참여 결과 등을 바탕으로 전체적인 불확도를 한번에 평가하는 방식이다.²³ 예를 들어 측정 대상과 유사한 인증표준물질을 여러 차례, 장기간 활용하여 인증값 대비 $\pm 20\%$ 정도의 상대 오차를 얻었다면 이 자료를 활용하여 불확도 평가를 할 수 있다. 만약 여러 차례 숙련도 시험에 참석하여 만족스러운 결과를 얻었다면 스스로의 측정 능력이 적어도 일반적인 시험 기관의 수준 정도는 된다고 판단하여 호르위츠 공식(Horwitz equation)²²에 따라 불확도 평가를 하거나 참여한 숙련도 시험의 시험 기관 간 재현성(interlaboratory reproducibility)을 바탕으로 불확도 평가를 하는 방식이다.

하향식 불확도 평가 방식을 사용할 때 주의할 점은 세 가지이다. 첫째, 이 방식은 측정 절차를 일정 수준 이상으로 최적화한 후에 사용해야 된다는 점이다. 가령 회수율(recovery)이 30%에서 90%까지 크게 변하는데 이를 해결하지 않고 단순히 편차의 범위를 불확도에 산입하는 것은 적절하지 않다. 적어도 인증값 대비 상대 오차가 20% 보다는 작아지도록 측정법을 최적화할 필요가 있다. 둘째, 화학 분석의 대상이 워낙 방대하다 보니 필요와 정확히 일치하는 인증표준물질이나 숙련도 시험 프로그램이 없을 수 있다는 점이다. 이 경우 최대한 유사한 물질이나 프로그램을 활용하거나 유사한 시료에 농도를 알고 있는 표준 물질을 첨가한 후 회수율을 확인한 자료를 활용하는 것을 고려해볼 수 있다. 마지막은 아직까지 국내에서는 하향식 불확도 평가 방식에 대한 인식이 부족하다는 점이다. 이 때문에 하향식 불확도 평가 방식을 사용할 경우 시험 기관 공인(creditation) 과정에서 여러 애로사항이 생길 수 있으므로 이 접근법을 뒷받침할 수 있는 근거 자료를 준비해 둘 필요가 있다. 예를 들어 식품 분석 분야에서는 국제적인 식품 규격을 제정하는 Codex Alimentarius의 지침에서 하향식 불확도 평가에 관한 권장 사항을 확인할 수 있다.²³

GUM의 혁신

이제 불확도 평가 실무에서 이해하여야 하는 기본 개념들은 모두 설명하였다. 아래에서는 완결성을 위해 보다 관념적인 설명을 덧붙인다.

많은 문헌에서 GUM이 새롭고 혁신적이라고 말한다. 그런데 지금까지 설명한 내용들을 돌이켜보면 새롭고 혁신적인 측면은 보이지 않는다. 가령 GUM 방식의 중심이 되는 중심극한정리, 불확도 전파의 법칙, t -분포 등은 GUM 훨씬 이전에 확립된 수학적 도구들이다.^{24,25} 그렇다면 GUM에서 새로운 것은 도대체 무엇일까? 왜 GUM 이전까지는 오래 전부터 알려진 수학적 도구들을 엮어내지 못했을까?

이 의문에 대한 답은 굉장히 현학적이다. GUM의 혁신은 수학적 것이 아니다. 바로 확률과 측정에 대한 비전통적인 시각을 도입한 것이다. 확률에 대해서는 베이즈주의(Bayesianism)²⁶를 도입하였으며 측정에 대해서는 도구주의(instrumentalism)²⁷⁻²⁸를 도입하였다. 국내에는 이에 관해 정확히 설명하는 자료가 전무하다시피 하여 GUM이 왜 새롭고 혁신적인 문헌인지를 정확히 이해하는 것이 쉽지 않다. 아래에서는 이 주제에 관해 간략히 다루며 GUM의 의의를 설명한다.

빈도주의자 통계와 베이즈 통계

‘확률(probability)’이라는 용어는 분야를 막론하고 빈번히 사용된다. 그런데 막상 ‘확률’이 무엇을 의미하는지를 설명하기는 쉽지 않다. 가령 ‘내일 비가 올 확률이 90%다’라는 말에서 ‘90%’라는 표현이 무엇을 의미하는지를 설명하는 것은 생각보다 어렵다. 이 문구의 의미는 크게 두 가지로 해석할 수 있다(GUM C.2.1 NOTE).

첫째, ‘내일과 같은 환경이 무한히 주어질 경우 그 중 9/10이 비가 올 것이다.’

둘째, ‘나는 절대로 일어나지 않을 것이라 믿는 사건에는 0이라는 수를, 반드시 일어날 것이라고 믿는 사건에는 1이라는 수를 부여하는데, 내일 비가 올 것이라는 사건에 대해 내가 가진 최선의 정보를 확률의 공리(axioms of probability)에 따라 정리해보면 0.9라는 수를 매길 수 있다.’

이 중에서 첫 번째 시각은 ‘확률’을 무한한 반복에서 어떤 사건이 일어나는 상대적인 빈도(relative frequency)로 본다. 이러한 시각을 확률에 대한 빈도주의자 해석(frequentist interpretation of probability)이라고 한다. 두 번째 시각은 ‘확률’을 어떤 사건이 일어날 것이라는 믿음의 정도(degree of belief)라고 본다. 이러한 시각을 확률에 대한 베이즈 해석(Bayesian interpretation of probability)이라고 한다.^{10,26,29} 여기서 ‘베이즈’는 베이즈 정리(Bayes' rule)에 관해 연구한

수학자 토머스 베이즈(Thomas Bayes)의 이름을 따온 것이다.

GUM이 작성되던 당시만 하더라도 베이즈주의적 시각은 타당하지 않은 것으로 여겨졌다.^{26,30} 수학과 과학은 객관적이어야 하므로 ‘믿음의 정도’와 같은 주관적인 개념은 용인될 수 없다는 것이 논지였다. 그러나 현대에는 이와 같은 시각에는 여러 오해가 있다는 견해 하에 베이즈주의적 시각이 자연스럽게 받아들여지고 있다.

베이즈 통계학(Bayesian statistics)은 기존 확률을 베이즈 정리와 추가로 얻어지는 정보를 이용하여 꾸준히 갱신하는 방식의 통계학을 의미한다. 이 부분에서는 논란이 될 부분이 없으나 가장 큰 문제는 ‘최초의 추측’에 있었다. 베이즈 통계학은 기존 확률을 갱신하는 통계학이므로 ‘기존 확률’이 반드시 있어야 한다. 그런데 그렇다면 ‘최초의 확률’은 어떻게 얻어야 할까? 예를 들어 빨간 공과 파란 공이 일정한 비율로 들어 있는 상자에서 공을 하나씩 꺼내며 그 결과를 확인하여 원래 상자에 들어 있는 빨간 공과 파란 공의 비율을 구한다고 할 때, 베이즈 통계학을 사용하기 위해서는 빨간 공과 파란 공의 비율을 최초로 한번 임의로 추측할 필요가 있다. 베이즈 통계학자들은 이 최초의 추측에 사용할 수 있는 충분한 자료가 없다면 모든 확률을 같다고 두는 등 임의로 추측하는 것도 가능하다고 주장하였다. 이러한 추측은 개인적인 믿음에 기반한 것이므로 베이즈 통계학은 주관적인(subjective) 통계학이라고 여겨졌다.

그러나 이는 오해였다. 최초의 추측은 임의적이고 주관적일 수 있으나, 자료가 충분히 축적되면 결국 베이즈 통계학의 결과물은 최초의 추측보다 자료의 영향을 더 크게 받으므로 주관적이지만 한 결과로 볼 수 없기 때문이다. 게다가 빈도 자료를 기반으로 한 빈도주의자 통계학에서도 자료를 수집하고 평가하는 데 임의의 판단이 필요한 경우가 많다. 사실 대부분의 학문의 핵심은 전문가의 판단(expert judgement)이다. 가령 전문가들은 실험을 어떻게 수행해야 하는지, 데이터를 어떻게 해석해야 하는지에 대한 나름의 직관을 갖추고 있는데, 왜 그러한 원칙을 따라야 하는지에 대한 순환적이지 않고 반증불가능한 근거를 별도로 갖추지는 못한 경우가 많다(사실 철학적으로 이것은 불가능하다). 그럼에도 수많은 자료보다 이러한 직관이 훨씬 중요한 경우가 많다. 특히 자료가 불충분한 경우에는 가용한 경험과 자료를 바탕으로 최선을 다해 임의로 판단을 할 수밖에 없다.¹⁰ 결국 문제는 부족한 자료에 있는 것이지, 자료가 부족할 때 최선을 다해 판단을 내리는 것 자체가 문제는 아닌 것이다.

GUM 이전에는 이러한 시각은 타당하지 않은 시각이었다. 측정을 수행할 때에는 반드시 반복 측정을 기반으로 해야만 했다.³¹ GUM의 용어를 사용하자면 “A형 불확도

평가”만 가능했던 것이다. 물론 그 한계가 잘 알려져 있었기에 반복 측정을 통해 ‘우연 오차(random error)’를 추정하고 자료, 경험에 의거하여 ‘계통 오차(systematic error)’를 추정하여 합성하는 경우도 있었다.³² 그러나 이러한 합성 방식을 뒷받침할 논리적인 근거는 없었다.⁵ 가령 GUM 이전에는 교정 성적서나 기기 사양명세서(specification)에 기재된 오차 범위와 허용 한계치를 반복 측정에서의 표준편차와 합성할 논리적인 근거가 없었다. GUM은 이 두 가지 정보가 모두 동등하다고 선언하고 당연히 합성할 수 있는 것으로 보았다. 즉, “B형 불확도 평가”도 가능하다고 선언하였다. 이것이 GUM의 첫 번째 혁신이다. GUM을 보면 곳곳에서 이러한 새로운 시각을 설명하고 정당화하고 있다(GUM E.3.6 등). 당시에는 이 시각이 새롭고 도전적인 시각이었기 때문이다.

이러한 철학적 논의에서 실무적인 함의를 찾아보자면 측정불확도 평가에서는 충분한 자료를 사용하는 것이 중요하다라는 것과, 단 하나의 옳은 측정불확도 평가 방식은 존재하지 않는다는 점이다. GUM 방식은 데이터에 대한 통계적 처리에만 초점이 맞추어져 있던 종래의 방식에 비해 측정자에게 자료를 선택, 해석, 활용하는 데 훨씬 많은 자유를 제공하지만 동시에 측정자에게 막중한 책임도 부여한다.⁵ 만약 교정 성적서, 기기 사양명세서 등의 데이터가 불확도에 크게 기여하는데 이 데이터를 믿기 어렵다고 해보자. 그렇다면 타 기관에서 교정을 한 번 더 받거나 기기의 사양을 직접 확인하는 등 추가적인 자료로 신뢰성을 확보하여야 한다. 다른 예로 반복 측정을 통해 얻은 여러 데이터 중 일부가 다른 데이터와 다른 것처럼 보인다고 해보자. 그렇다면 가용한 모든 자료와 경험을 사용하여 그러한 차이가 발생한 원인을 파악하거나 추가 실험을 수행하는 것이 최우선이다. 이제 단순히 통계학적인 기법을 기계적으로 적용한 후 결과의 품질에 대한 책임을 통계학적인 도구에 미룰 수 있는 시대는 지났다. GUM에 따른 불확도 평가의 체계에서는 충분하고 믿을만한 자료를 갖추어 최종 결과물의 신뢰성을 확보하는 책임이 측정자에게 있다.³³ 측정에서의 판단은 어느 정도 주관적이고 임의적일 수밖에 없다. 그러나 그런 판단을 타 전문가들도 받아들일 수 있을 만큼 충분한 자료와 근거를 토대로 수행하여 다수의 주관적인 객체가 널리 받아들일 수 있는 상호주관적(intersubjective)인 결과를 내어놓아야 한다.^{26,34}

다음 GUM의 구절이 이러한 논지를 잘 담은 것으로 평가받으며 자주 인용된다(GUM 3.4.8).

“본 지침이 측정에서의 불확실성을 평가하는 체계를 제공하고는 있으나, 본 지침은 비판적 사고, 지적 정직성, 전문적인 기량을 대체할 수 없다. 측정불확도 평가는 단순 반복 작업도, 순수히 수학적인 작업도 아니며 측정량과

측정에 관한 상세한 지식에 의존한다. 평가된 불확도의 품질과 유용성은 궁극적으로 그 값을 얻는 데 기여한 자의 이해 수준, 비판적 분석 및 정직성에 달려 있다(Although this Guide provides a framework for assessing uncertainty, it cannot substitute for critical thinking, intellectual honesty and professional skill. The evaluation of uncertainty is neither a routine task nor a purely mathematical one; it depends on detailed knowledge of the nature of the measurand and of the measurement. The quality and utility of the uncertainty quoted for the result of a measurement therefore ultimately depend on the understanding, critical analysis, and integrity of those who contribute to the assignment of its value).”

확률이 무엇인가에 대한 시각 차이로 인한 여파는 아직도 남아 있다. 예를 들어 위에서 불확도 평가에서는 표준편차 대신 표준불확도라는 용어를 사용한다고 설명하였다. 사실 이러한 별도의 용어를 사용하는 이유도 GUM을 제작하던 당시에는 통계학자들이 B형 불확도 평가로 얻어진 표준편차를 표준편차로 인정할 수 없었기 때문이다.³⁵ 비슷한 이유로 GUM에서는 ‘신뢰구간(confidence interval)’, ‘신뢰수준(confidence level)’, ‘t 값(t-value)’이라는 용어 대신 ‘포함구간(coverage interval)’, ‘신뢰의 수준’, ‘포함인자’ 등의 표현을 사용한다(GUM 6.2.2). 이러한 용어 상 차이는 불확도를 처음 접하는 사람들이 불확도 평가의 수학적 배경을 정확히 파악하고 이해하기 어렵게 만든다. 베イズ주의에 대한 반감이 많이 사라진 2000년대부터는 굳이 표준불확도라는 표현을 사용하지 않고 표준편차라는 표현을 사용하는 문헌도 자주 보인다.

2000년대부터는 베イズ 통계학의 도구를 사용하여 불확도를 평가하는 방법에 대한 연구들이 활발하게 출판되고 있다. 이러한 경향에 주의할 점도 있다. 베イズ 통계학의 도구를 통해 대부분의 문제를 해결할 수 있다는, 베イズ 통계학을 전가의 보도처럼 사용하는 경향도 보인다는 점이다. 이러한 경향은 통계적 도구를 기계적으로 적용하기 전에 측정 전문가의 면밀한 검토와 판단이 있어야 한다는 GUM의 혁신에 역행하는 것과 다름 없으니 주의할 만하다.

참값과 오차

GUM의 또 다른 혁신은 더 현학적이다. 측정에 대한 전통적인 시각은 ‘참값(true value)’에 대한 정보를 얻는 것을 측정의 목표라고 말한다.^{36,37} 이러한 시각은 각 양에 인간의 인식과 무관하게 ‘옳은 값’이 정해져 있으며 이를 최선을 다해 추정하는 것이 측정의 의의라는 시각이다. 그런데 이 ‘옳은 값’은 알 수 없다. 이는 측정을 통해서만 추정할 수 있는데 어떤 측정도 완벽하지 않기 때문이다. 예를 들어

어떤 책상의 길이를 최선을 다해서 추정한다 한들, 그렇게 얻은 측정값이 정말 참값과 일치하는지는 알 수 없다. 측정자가 알지 못하는 오류가 있지 않다는 것을 증명하는 것이 불가능하기 때문이다. 이러한 인식론적(epistemological) 문제는 많은 개념적인 문제를 낳았다. 측정의 목표는 참값인데, 결국 참값은 알 수 없다. 사실 측정값이 참값과 얼마나 떨어져 있는지(오차)조차도 알 수 없다. 그렇다면 측정은 모두 무의미한 것이 아닐까? 측정 결과의 신뢰성은 대체 어떻게 평가해야 할까?

GUM은 이러한 개념적 문제를 해결했다고 말하며 참값, 오차, 불확도라는 개념에 관하여 많은 분량을 할애하여 설명한다(GUM 2.2.4, Annex D, E.5.1 등). 하지만 GUM이 제시한 해결책이 정확히 무엇인지를 설명하는 자료는 소수이며 그나마도 난해하다.^{11,38} 많은 자료는 ‘참값과 오차를 알 수 없으므로 참값과 오차에 대해 언급하지 말자’를 GUM이 제시한 해결책으로 본다.³⁹ 이러한 해석에 따라 GUM 이후 측정 관련 문헌에서는 ‘참값’과 ‘오차’라는 용어가 일거에 사라졌다. 측정학계에서 널리 존중 받는 문헌에서 이 두 용어를 쓰지 말 것을 권하니, 이 용어를 썼다가 측정과 불확도에 대해 잘 알지 못하는 것으로 낙인이 찍히는 것을 두려워했기 때문이다.³⁸

그런데 2000년대 후반부터 변화가 보이기 시작한다. 측정의 궁극적인 목표는 참값인데, 참값을 알 수 없다는 문제는 참값이라는 표현을 금지시키는 것으로 해결될 수 없다는 주장 때문이다.^{11,38} 이러한 주장이 점차 받아들여지기 시작하여 이제 많은 연구자들이 참값과 오차라는 표현을 사용하고 있으며 JCGM과 같은 측정학계에서 가장 권위가 높은 조직의 문서에서도 참값과 오차라는 표현이 거리낌 없이 활용되고 있다.

개인적으로는 두 가지 시각 모두에 오류가 있다고 생각한다. 무언가를 알 수 없다는 문제를 그 문제에 대해 언급하지 않음으로써 해결할 수 없는 것은 맞다. 그러나 그 문제에 대해 다시 언급하는 것이 문제를 해결해주는 것은 아니다. 참값과 오차라는 개념을 다시 사용하자고 주장하는 문헌들은 참값과 오차를 결코 알 수 없다는 인식론적 문제에 대한 어떠한 명확한 해결책도 제시하지 않는다.^{11,38}

아래는 이 주제에 관한 저자의 사견이며 아직까지 널리 받아들여지는 시각은 아니므로 받아들이는 데 주의를 요한다. 관심 있는 독자들께서는 최근 저자가 이 문제에 관해 정리한 논문을 참조하여 직접 판단을 내리시기를 권한다.⁴⁰ 저자의 사견은 이 모든 개념적인 혼란이 GUM의 주장을 잘못 이해했기 때문에 발생했다는 것이며 GUM의 원래 논지가 ‘참값과 오차라는 개념 자체가 유효하지 않다’라는 것이다. 즉, 참값과 오차라는 표현을 사용하지 말자는 것이 아니라 그 개념들 자체가 유효하지 않다는 것

이 논지라고 생각한다.

위에서 설명한 참값 중심의 시각을 ‘실재론(realism)’이라고 한다.^{28,36-37} 실재론은 넓게는 어떤 이론 내 대상들이 실제로 존재한다는 시각을 말하는데,²⁷ 측정에서는 인간의 인식과 무관한 ‘옳은 값’들이 존재하며 이 값들을 추정하는 것이 측정의 궁극적인 목표라는 시각을 말한다. 그러나 전혀 다른 시각도 있다. 이 시각은 경험론(empiricism), 도구주의(instrumentalism), 조작주의(operationalism), 실용주의(pragmatism) 등으로 부를 수 있는데²⁷ 일단 여기서는 “도구주의”라는 표현을 사용한다.²⁸ 도구주의는 어떤 이론 내 대상들은 관념적인 도구(instrument)에 불과하며 그것들이 실제로 존재하는지 여부는 알 수도 없고 중요하지도 않다는 시각이다.^{27,41} 측정에서는 측정의 대상이 되는 양들이 인간의 인식과 무관하게 존재하는게 아니라, 인간이 편의에 따라 발명한 관념적인 도구라는 시각을 말한다.^{28,42} 따라서 인간의 인식과 무관한 ‘참값’이란 존재하지 않으며 측정이란 인간이 서로 합의한 방식대로 어떤 대상에 대한 정보를 얻고 표현하는 행위가 된다.^{28,32}

Churchill Eisenhart는 NIST 통계학 부서의 창설자로, 측정학과 통계학에서 다양한 기여를 하였다. 그의 논문 “Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems”는 GUM 이전까지 측정에서의 불확실성을 평가하는 데 있어 교과서적인 위치에 있었는데,⁵ 그의 논문에서 ‘참값’에 관한 도구주의적인 시각을 찾아볼 수 있다.³²

“이러한 논의는 즉시 다음 질문을 유발한다: 무언가의 특정 성질의 크기의 ‘참값’이란 대체 어떻게 정의되는가? 결국 어떤 양의 크기의 ‘참값’이란 그 크기의 측정에 관해 전문가들이 합의한 *모범적인 측정법*에 의해 정의된다 (This immediately raises the question: Just how is the “true value” of the magnitude of a particular property of something defined? In the final analysis, the “true value” of the magnitude or a quantity is defined by agreement among experts on an *exemplar method* for the measurement of its magnitude).”

“전술한 바에서 명확하듯이, 특정 양의 크기에 대한 ‘참값’은 이 양의 크기를 필요로 하는 목적과 밀접하게 연관되어 있으며, 결국 그러한 목적과 독립적으로 의미 있게 혹은 유용하게 정의될 수 없다. 이 사실이 과학과 공학에서 더 널리 인지되면서 ‘참값’이라는 전통적인 용어가 사라지기를 바라며, ‘목표값’ 등의 적절한 용어로 대체되어 측정자가 특정 목적을 위해 얻고 싶어하는 값을 명확히 하고 우리가 그 값을 필요로 하는 목적을 초월하여 독립적으로 존재한다는 인상을 주지 않게 되기를 바란다 (Indeed, as is evident from the foregoing, the “true value” of the magnitude of a particular quantity is intimately linked to

the purpose for which a value of the magnitude of this quantity is needed, and its “true value” cannot, in the final analysis, be defined meaningfully and usefully in isolation from these needs. Therefore, as this fact becomes more widely recognized, in science and engineering, I hope that the traditional term “true value” will be discarded in measurement theory and practice, and replaced by some more appropriate term such as “target value” that conveys the idea of being the value that one would like to obtain for the purpose in hand, without any implication that it is some sort of permanent constant preexisting and transcending any use that we may have for it).”

실재론과 도구주의에 대해 더 상세히 설명하는 것은 본 논문의 목적을 벗어나지만 독자의 이해를 돕기 위해 간략하게 비유하면 다음과 같다. 상대성 이론, 열역학 법칙 등 현대 과학에서 최상위의 권위를 가지는 이론에 대한 과학자들의 통상적인 견해는 이 이론들이 반드시 참이라고 확신을 할 수는 없고 미래에 틀린 것으로 밝혀질 가능성이 없는 것은 아니지만 아마도 참에 아주 근접할 것이라는 견해이다. 이러한 견해를 실재론적 견해라고 볼 수 있다. 반면 도구주의적인 견해는 어떤 이론이 참이냐는 의문 자체가 무의미하며 중요한 것은 그 이론을 토대로 얻어지는 결과물이 유용한지 여부라는 견해라고 할 수 있다.^{27,41}

많은 실험 과학자들에게 도구주의적 시각은 너무 유보적인 시각으로 보이며 철학적 논의에 특별히 관심이 있는 경우가 아니라면 그 논거를 이해하는 것도 쉽지 않다. 따라서 역사적으로 많은 실험 과학자들은 실재론적인 시각을 견지했다.⁴³⁻⁴⁵ 사실 대부분의 철학자들도 실재론자에 속한다.⁴⁶ 반면에 측정에 관해 연구하던 수학자, 이론가, 철학자들은 부분적으로나마 도구주의적인 시각을 견지한 경우가 많다.^{43-45,47-48} 저자의 사견으로 이는 어느 정도 필연적인 결과이다. 사실 ‘상대성 이론이 미래에 틀릴 수도 있다는 가능성’은 흥미로운 사고 주제이지만 반드시 해결해야 할 문제는 아니다. 아무리 열심히 측정을 수행하여도 참값을 알 수 없다는 개념적 문제 또한 실무에서 문제를 일으키는 경우는 드물다. 그러나 측정이라는 행위 자체의 개념적인 토대를 연구하는 입장에서는 이는 반드시 해결해야 하는 문제이기에 도구주의적 시각을 부분적으로나마 견지하며 문제를 해결하려 노력하게 되었다고 생각한다.

GUM은 참값과 오차라는 무의미한 용어를 쓰지 말 것을 누차 권하고 있다. 저자가 보기에 이 주장은 참값과 오차라는 개념 자체가 타당하지 않다는 도구주의적 주장이다. 하지만 많은 연구자들이 이 주장을 ‘확실히 모르니까 언급하지 말아야 한다’라는 불가지론적(agnostic)인 입장으로 해석하다가 만족하지 못하고 실재론적인 시각으로 항하게 된다.^{11,49} 또한 GUM은 참값과 오차에 대해 “오차와

불확도는 구분해야 한다”(GUM 3.2.2 NOTE 2), “참값은 여러 개일 수 있다”(GUM B.2.3 NOTE 3), “참값이라는 표현에서 ‘참’이라는 단어는 동어반복이다”(GUM D.3.5)와 같은 서로 다른 주장을 불분명하게 섞어서 제시하며 개념적인 혼란을 가중시킨다.

논의의 완결성을 위해 현학적인 설명을 덧붙였지만 사실 대부분의 독자들이 알아야 하는 내용은 간결하다. 첫째, GUM 발간 초기 많은 사람들이 참값과 오차라는 표현을 피했다는 점이다. 둘째, 이제는 참값과 오차라는 표현을 사용해야만 한다는 시각이 더 우세이며 많은 권위 있는 문서에서 이 표현들을 거리낌 없이 사용하고 있다는 점이다. 저자도 참값과 오차라는 표현을 사용하는 데 많은 이점이 있다고 생각한다. 참값과 오차라는 표현을 사용하면 측정불확도에 대해 ‘참값의 가능한 범위에 대한 표현’, ‘오차 범위에 대한 현대적인 표현’ 등과 같은 간결한 설명이 가능해진다. GUM도 이러한 표현이 불가능하지 않다고 말한다(GUM 2.2.4). 사실 참값과 오차라는 표현을 사용할 때 개념적인 문제가 발생하는 경우는 소수이다.

다만 측정이란 행위의 개념적 기반에 대해 관심이 많은 독자라면 왜 참값과 오차라는 개념이 문제가 될 수 있는지, 참값과 오차라는 개념이 없으면 측정의 목표는 무엇인지 등에 대해 다루는 참고문헌을 확인해보기를 권한다.⁴⁰ 참값과 오차라는 개념의 문제를 알아야 GUM이 ‘측정불확도’라는 용어를 위와 같이 간결하게 정의하지 않았는지를 이해할 수 있다.⁵⁰⁻⁵¹

안타까운 것은 문헌에서 이 개념적인 문제에 대해 명확히 다루는 경우가 거의 없다는 것이다. 이 주제에 관한 논의가 대부분 GUM 제작과 관련된 소수의 집단 내부에서 이루어지며 공개된 문헌에서는 논의가 거의 이루어지지 않고 있어 논란의 여지가 존재한다는 것 자체를 인지하지 못하는 연구자들이 대부분이고 논의를 통해 서로의 시각을 개선해나갈 기회도 놓치고 있다.³¹ 향후 이 분야의 발전을 위해 보다 공개적이고 투명한 논의가 이루어지기를 희망해본다.

결론

본 논문은 측정불확도의 개념과 그 배경을 상술하였다. 본 논문을 통해 저자는 측정불확도 평가에서는 수학과 통계에 대한 전문적인 지식보다는 측정의 본질에 관한 개념적인 이해와 특정 분야 측정에 대한 기술적인 전문성이 더 중요하다는 것을 보여주려 노력하였다. 이러한 노력이 국내 측정불확도 평가의 이해 증진과 저변 확대에 일부나마 기여하기를 기대한다.

Acknowledgments. 본 논문은 한국표준과학연구원의 기관고유사업 “무기분석 측정표준 기반 확립(과제 번호 23011061)” 및 “무기분석 표준물질 개발(과제 번호 23011062)”의 지원으로 작성되었다.

REFERENCES

- JCGM 100:2008, *Evaluation of Measurement Data — Guide to the Expression of Uncertainty in Measurement*, BIPM: Sèvres, 2008.
- JCGM 200:2012, *International Vocabulary of Metrology — Basic and General Concepts and Associated Terms (VIM)*, BIPM: Sèvres, 2012.
- De Bièvre, P., *Accred. Qual. Assur.* **2007**, *12*, 279.
- Sheynin, O. B., *Arch. Hist. Exact Sci.* **1979**, *20*, 21.
- Kacker, R.; Sommer, K.-D.; Kessel, R., *Metrologia* **2007**, *44*, 513.
- ISO/IEC Guide 98-3:2008, *Uncertainty of Measurement — Part 3: Guide to the Expression of Uncertainty in Measurement (GUM:1995)*, ISO: Geneva, 2008.
- 한국표준과학연구원, *측정불확도 표현 지침*, 한국표준과학연구원:대전, 2010.
- Bich, W.; Cox, M.; Michotte, C., *Metrologia* **2016**, *53*, S149.
- JCGM 101:2008, *Evaluation of Measurement Data — Supplement 1 to the “Guide to the Expression of Uncertainty in Measurement” — Propagation of Distributions Using a Monte Carlo method*, BIPM: Sèvres, 2008.
- O'Hagan, A., *Metrologia* **2014**, *51*, S237.
- Ehrlich, C., *Metrologia* **2014**, *51*, S145.
- Possolo, A., *Metrologia* **2016**, *53*, S17.
- Westwood, S.; Lippa, K.; Shimuzu, Y.; Lalerle, B.; Saito, T.; Diewer, D.; Dai, X.; Davies, S.; Ricci, M.; Baldan, A.; Lang, B.; Sarge, S.; Wang, H.; Pratt, K.; Josephs, R.; Mariassy, M.; Pfeifer, D.; Warren, J.; Brenser, W.; Ellison, S.; Toman, B.; Nelson, M.; Huang, T.; Fajgelj, A.; Gören, A.; Mackay, L.; Wielgosz, R., *Pure Appl. Chem.* **2023**, *95*, 1.
- Lafarge, T.; Possolo, A., *NCSLI Meas.* **2015**, *10*, 20.
- EA-4/02 M:2022, *Evaluation of the Uncertainty of Measurement in Calibration*, European Accreditation, 2022.
- EURACHEM/CITAC Guide CG 4, *Quantifying Uncertainty in Analytical Measurement*, Third edition, Eurachem/CITAC, 2012.
- Lee, J. W., *Metrologia* **2022**, *59*, 065010.
- L. R. Ellison, S., *Analyst* **1998**, *123*, 1387.
- De Bièvre, P., *Accred. Qual. Assur.* **2012**, *17*, 231.
- Williams, A., *Accred. Qual. Assur.* **2001**, *6*, 73.
- ISO Guide 30:2015, *Reference Materials — Selected Terms and Definitions*, ISO: Geneva, 2015.
- Horwitz, W.; Albert, R., *J. AOAC Int.* **2019**, *89*, 1095.
- CXG 54-2004, *Guidelines on Measurement Uncertainty*, Codex Alimentarius Commission: Rome, 2023.
- Stoudt, S.; Pintar, A.; Possolo, A., *Metrologia* **2021**, *58*, 015014.
- Student, *Biometrika* **1908**, *6*, 1.
- 이영의, *베이지주의*, 한국문화사:서울, 2015.
- Chakravarty, A., Scientific Realism. In *The Stanford Encyclopedia of Philosophy (Summer 2017 Edition)*, Zalta, E. N., Ed. 2017.
- Mari, L.; Giordani, A., Modeling Measurement: Error and Uncertainty. In *Error and Uncertainty in Scientific Practice*, Petersen, A. C.; Hon, G.; Boumans, M., Eds. Pickering & Chatto: London, 2014.
- Hájek, A., Interpretations of Probability. In *The Stanford Encyclopedia of Philosophy (Fall 2019 Edition)*, Zalta, E. N., Ed. 2019.
- 샤론 버치 맥그레인, *불멸의 이론 - 베이지 정리는 어떻게 250년 동안 불확실한 세상을 지배하였는가*, 휴먼사이언스:서울, 2013.
- Kacker, R. N., *Measurement* **2018**, *127*, 525.
- Eisenhart, C., *J. Res. Natl. Bur. Stand. C* **1963**, *67*, 161.
- De Bièvre, P., *Accred. Qual. Assur.* **2008**, *13*, 177.
- Mari, L.; Carbone, P.; Petri, D., *IEEE Trans. Instrum. Meas.* **2012**, *61*, 2107.
- BIPM, *Report of the BIPM Working Group on the Statement of Uncertainties to the Comité International des Poids et Mesures*, BIPM: Sèvres, 1980.
- Mari, L., *Measurement* **2003**, *34*, 17.
- Mari, L., *Measurement* **2005**, *38*, 259.
- Bich, W., *IEEE Trans. Instrum. Meas.* **2012**, *61*, 2153.
- Alexandrov, Y. I., *Fresenius J. Anal. Chem.* **2001**, *370*, 690.
- Lee, J. W.; Hwang, E.; Kacker, R. N., *Accred. Qual. Assur.* **2022**, *27*, 235.
- Legg, C.; Hookway, C., Pragmatism. In *The Stanford Encyclopedia of Philosophy (Summer 2021 Edition)*, Zalta, E. N., Ed. 2021.
- 장하석, *온도계의 철학*. 동아시아: 서울, 2013.
- De Boer, J., *Metrologia* **1995**, *31*, 405.
- De Courtenay, N., The Double Interpretation of the Equation of Physics and the Quest for Common Meanings. In *Standardization in Measurement: Philosophical, Historical and Sociological Issues*, Schlaudt, O.; Huber, L., Eds. Pickering & Chatto: London, 2015.
- Silsbee, F. B., *J. Res. Natl. Bur. Stand. C* **1962**, *66C*, 137.
- Monton, B.; Mohler, C., Constructive Empiricism. In *The Stanford Encyclopedia of Philosophy (Summer 2021 Edition)*, Zalta, E. N., Ed. 2021.
- Rossi, G. B., *Measurement* **2007**, *40*, 545.
- Sherry, D., *Stud. Hist. Philos. Sci. A* **2011**, *42*, 509.
- Ehrlich, C.; Dybkaer, R.; Wöger, W., *Accred. Qual. Assur.* **2007**, *12*, 201.
- Mari, L., *Metrologia* **2014**, *52*, R1.
- Mari, L.; Narduzzi, C.; Nordin, G.; Trapmann, S., *Measurement* **2020**, *152*, 107397.