UDC 001.4:389.14:621.317 DOI:

Redefining Standard Measurement Uncertainty

A. Possolo, O. Bodnar

National Institute of Standards and Technology, Gaithersburg, MD 20899-8980, USA; Örebro University, Örebro, Sweden
antonio.possolo@nist.gov; olha.bodnar@oru.se

Abstract

The Guide to the Expression of Uncertainty in Measurement (GUM) defines standard measurement uncertainty as the standard deviation of a probability distribution that describes the uncertainty associated with an estimate of the measurand, and defines expanded uncertainty as a multiple of the standard uncertainty. Monte Carlo methods can produce the expanded uncertainty for 95% coverage as one half of the length of the interval whose endpoints are the 2.5th and 97.5th percentiles of the probability distribution of the estimate of the measurand (when this distribution is approximately symmetrical). This creates an opportunity for a paradox to arise: that the standard uncertainty, defined as a standard deviation, can be larger than the expanded uncertainty. We provide an example involving real measurement data where this paradox arises with high probability, and then offer a new definition of standard uncertainty that agrees numerically with the conventional definition in "normal" cases, but that is still reliable in "abnormal" cases.

Keywords: expanded uncertainty; percentiles; paradox; coverage factor; normal distribution; Dolos distribution.

Received: 10.02.2022 Edited: 17.03.2022 Approved for publication: 24.03.2022

1. Introduction

The *Guide to the Expression of Uncertainty in Measurement* (GUM) [1] defines standard measurement uncertainty (for scalar measurands) as the standard deviation of a probability distribution that describes the uncertainty associated with an estimate of the measurand, and it defines expanded uncertainty as a multiple of the standard uncertainty.

With the advent of Monte Carlo methods for uncertainty propagation, and with the increasing use that is being made of statistical models and methods to characterize measurement uncertainty, it is now fairly common to obtain the expanded uncertainty "directly," not as a multiple of the standard uncertainty.

For example, the expanded uncertainty for 95% coverage can be obtained as one half of the length of the interval whose endpoints are the 2.5th and 97.5th percentiles of the probability distribution of the estimate of the measurand (when this distribution is approximately symmetrical).

This creates an opportunity for a paradox to arise: that the standard uncertainty, defined as a standard deviation, can be larger than the expanded uncertainty, defined in terms of percentiles as exemplified above.

In this contribution we provide an example of a situation involving real measurement data where this paradox arises with very high probability, and then use it as motivation to offer a new definition of standard uncertainty that agrees numerically with the conventional definition in "normal" cases, but that will still be reliable in "abnormal" cases, and that will also resolve the paradox aforementioned.

Our proposal is to redefine standard measurement uncertainty $u(y)$ as half the length of a 68% coverage interval centered at y. Defined in this way, $u(y)$ reproduces the standard deviation when y has a Gaussian probability distribution, and the meaning of $y \pm u(y)$ is independent of the underlying probability distribution.

2. Meaning of Standard Uncertainty

Measurement results are often expressed in the form $y \pm u(y)$, where y denotes the measured value and $u(y)$ denotes the associated standard uncertainty. However, the meaning of such expression of measurement uncertainty is contingent on the probability distribution of the measurand (Fig. 1). For example, if y has a uniform (or rectangular) distribution on interval [0,1], then the interval $y \pm u(y)$ has 55.7% coverage probability.

The standard deviation can provide a misleading expression of the dispersion of a distribution. Consider a discrete distribution Q that we call the Dolos Distribution because, in classical Greek mythology Dolos is the god of trickery, and a master of craftiness.

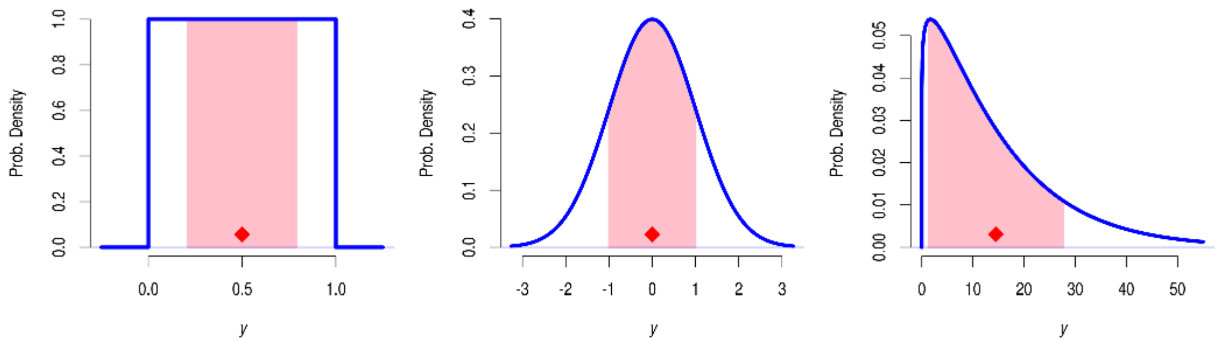

Fig. 1. Different distributions can have the same standard deviation and assign different probabilities to an interval of the form mean plus or minus one standard deviation

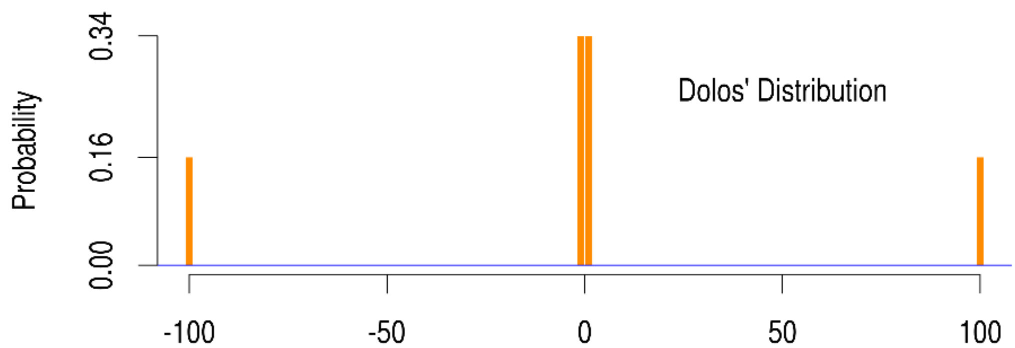

Fig. 2. Dolos's distribution

This distribution has four atoms placed symmetrically around zero, at -100, -1, +1, and +100, and assigns to them the following probabilities. Let $\gamma = 2.5/\sqrt{2}$, and define $p = 1/(2\gamma^2)$ and q=1-p. Now, put $Q(\{-100\}) = Q(\{+100\}) = p$, and $Q(\{-1\}) = Q(\{+1\}) = q$. These probabilities are depicted in Fig. 2.

If Q characterizes the uncertainty associated with a measured value of 0, then the corresponding standard measurement uncertainty is 1 while Q's standard deviation is 56.57. The former is very accurately informative, while the latter is rather misleading.

The standard deviation can be infinite. Consider the situation depicted in Fig. 3, where a horizontal light beam emerges from a small hole in a wall and travels along a 1 m long path at right angles to the wall, towards a flat mirror that oscillates freely around a vertical axis.

When the mirror's surface normal makes an angle A with the beam, its reflection hits the wall at distance $D = \tan(A)$ from the hole (positive to the right of the hole and negative to the left). If A is uniformly (or, rectangularly) distributed between $-\pi/2$ rad and $\pi/2$ rad, then $\Pr(D < d) = \Pr[A < \arctan(d)] = [\arctan(d) + \pi/2]/\pi$, and D's probability density is p_D such that $p_D(d) = 1/[\pi(1+d^2)]$ for $-\infty < d < +\infty$.

As it turns out, both the mean and the standard deviation of D are infinite [2, Page 51], but if the standard uncertainty $u(D)$ is defined as one half of the length of the interval such that $\Pr[-u(D) < D < +u(D)] = 0.68$, as proposed above, then $u(D)$=1.82 m. The infinite standard deviation is useless, while the standard measurement uncertainty tells us that the reflected beam stays within 1.82 m of the exit hole 68% of the time.

3. Paradox of the U's

Consider the following measurement results: w_1=34.3 mg/kg, w_2=32.9 mg/kg, and w_3=31.9 mg/kg, with expanded uncertainties (for 95% coverage) U_1=13.1 mg/kg, U_2=8.8 mg/kg, and U_3=5.1 mg/kg, each based on a single degree of freedom. Since the expanded uncertainties all are based on the same number of degrees of freedom, and all have the same coverage probability, the weighted average of the measured values is $a = (w_1/U_1^2 + w_2/U_2^2 + w_3/U_3^2)/(1/U_1^2 + 1/U_2^2 + 1/U_3^2)$. These results are mutually consistent, and their weighted average is 32.4 mg/kg.

To evaluate $u(a)$ and $U_{95\%}(a)$ we model each measured value as an outcome of a rescaled and shifted Student's t random variable with 1 degree of freedom, and interpret the corresponding expanded uncertainty

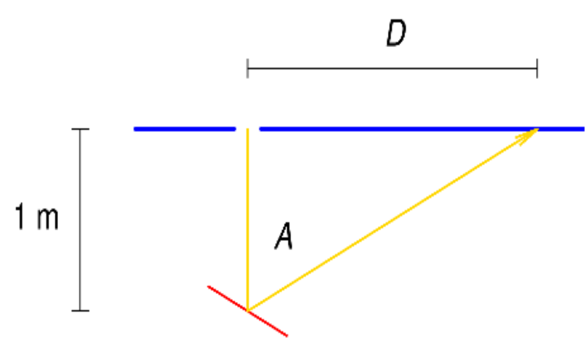

Fig. 3. Reflection of light beam by oscillation mirror

as being half the length of the interval whose endpoints are the 2.5th and 97.5th percentiles of this distribution.

The following R [3] code implements the propagation of these distributions according with the GUM Supplement 1 [4] to obtain a sample of size K from the probability distribution of a, and to derive the associated standard and expanded uncertainties from this sample.

```
  x = c(34.3, 32.9, 31.9); U = c(13.1, 8.8, 5.1);
nu = c( 1, 1, 1)
  w = (1/U^2)/sum(1/U^2)
  a = sum(w*x)
  K = 1e5; aB = numeric(K)
  sigma = U/qt(0.975, nu)
  for (k in 1:K)
  {
      xB = x + sigma*rt(n, df=nu)
      aB[k] = sum(w*xB)
  }
  ua = sd(aB);  U95a = diff(quantile(aB,
probs=c(0.025, 0.975)))/2
  names(U95a) = NULL;  c("u(a)"=ua,
"U95(a)"=U95a)
```

The standard deviation of the weighted average in a Monte Carlo sample of size 1000 happened to be 27 mg/kg, and $U_{95\%}(a)$ = 6.5 mg/kg (half the difference between the 97.5th and 2.5th percentiles of the Monte Carlo sample). Therefore, for this particular Monte Carlo sample, is the standard uncertainty is defined as the standard deviation, then it will be larger than the expanded uncertainty for 95% coverage. Fig. 4 shows the corresponding probability density estimate. This paradoxical behavior is not a fluke. In fact, with K=1000, $u(a)$ will be greater than $U_{95\%}(a)$ with probability 96.1%. And this is so only because K is finite, because a actually has infinite standard deviation, while the expanded uncertainty, as interpreted here, is finite and meaningful.

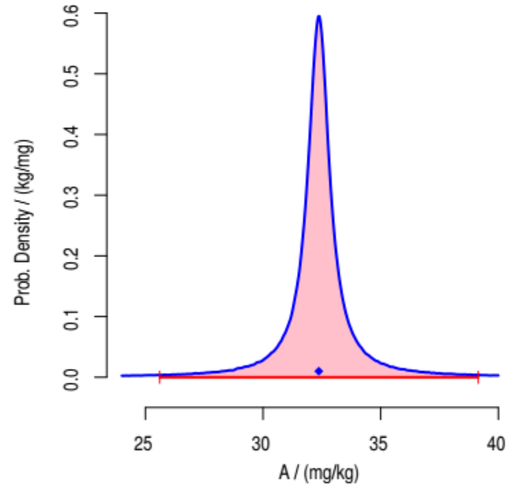

Fig. 4. Monte Carlo probability distribution of weighted average *A*

Since, in this case, the weighted average A has a symmetric distribution, if we define $u(A)$ as half the difference between 82nd and 16th percentiles of A's distribution, then we obtain $u(A)$ =1 mg/kg and $U_{95\%}(A)$ =6.5 mg/kg, which resolves the paradox.

4. Conclusions

The meaning of $u(y)$ as a standard deviation, and the coverage of an interval like $y \pm u(y)$, are strongly dependent on the underlying probability distribution. The standard uncertainty, defined as standard deviation of the probability distribution of the measurand, can be (much) larger than expanded uncertainty, as illustrated in the paradox of the U's. Furthermore, the standard deviation can be a misleading indication of the dispersion of a distribution, or it can be infinite, while the redefined standard uncertainty continues to be informative and meaningful even in the rather abnormal situations described in section 2.

Therefore, we propose that standard measurement uncertainty $u(y)$ be redefined as half the length of a 68% coverage interval centered at y. Defined in this way, $u(y)$ reproduces the standard deviation when y has a Gaussian probability distribution, and the meaning of $y \pm u(y)$ is independent of the underlying probability distribution.

Перевизначення стандартної невизначеності вимірювання

А. Поссоло, О. Боднар

National Institute of Standards and Technology, Gaithersburg, MD 20899-8980, USA; Örebro University, Örebro, Sweden
antonio.possolo@nist.gov; olha.bodnar@oru.se

Анотація

GUM визначає стандартну невизначеність вимірювання як стандартне відхилення розподілу ймовірностей, що описує невизначеність, пов'язану з оцінкою вимірюваної величини, і визначає розширену невизначеність як кратну стандартній невизначеності. Методи Монте-Карло можуть оцінити розширену невизначеність як половину довжини 95% інтервалу невизначеності, кінцевими точками якого є 2,5-й і 97,5-й перцентилі розподілу ймовірностей оцінки вимірюваної величини (коли цей розподіл є приблизно симетричним). Це створює можливість для виникнення парадоксу: стандартна невизначеність, визначена як стандартне відхилення, може бути більшою за розширену невизначеність. Ми надаємо приклад із реальними даними вимірювань, де цей парадокс виникає з високою ймовірністю, а потім пропонуємо нове визначення стандартної невизначеності, яке чисельно узгоджується зі звичайним визначенням у випадку нормального розподілу, і залишається достовірним також для інших розподілів.

Ключові слова: стандартне відхилення; перцентилі розподілу; парадокс; коефіцієнт покриття; нормальний розподіл; розподіл Долоса.

Переопределение стандартной неопределенности измерения

А. Поссоло, О. Боднар

National Institute of Standards and Technology, Gaithersburg, MD 20899-8980, USA; Örebro University, Örebro, Sweden
antonio.possolo@nist.gov; olha.bodnar@oru.se

Аннотация

GUM определяет стандартную неопределенность измерения как стандартное отклонение распределения вероятностей, описывающее неопределенность, связанную с оценкой измеряемой величины, и определяет расширенную неопределенность как кратную стандартной неопределенности. Методы Монте-Карло могут оценить расширенную неопределенность как половину длины 95% интервала неопределенности, конечными точками которого являются 2,5-й и 97,5-й процентили распределения вероятностей оценки измеряемой величины (когда это распределение является примерно симметричным). Это создает возможность для возникновения парадокса: стандартная неопределенность, определенная как стандартное отклонение, может быть больше расширенной неопределенности. Мы приводим пример с реальными данными измерений, где этот парадокс возникает с высокой вероятностью, а затем предлагаем новое определение стандартной неопределенности, численно согласующееся с обычным определением в случае нормального распределения, и остается достоверным также для других распределений.

Ключевые слова: стандартное отклонение; процентили распределения; парадокс; коэффициент покрытия; нормальное распределение; распределение Долоса.

References

1. JCGM 100:2008. Evaluation of measurement data – Guide to the expression of uncertainty in measurement. JCGM, 2008. 120 p. Available at: www.bipm.org/en/publications/guides/gum.html
2. Feller W. An Introduction to Probability Theory and Its Applications. Vol. II. 2nd edition. John Wiley & Sons, New York, 1971. 683 p.
3. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria, 2021. Available at: https://www.R-project.org/
4. 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. JCGM, 2008. 90 p.