Applied Radiation and Isotopes 70 (2012) 712–719

Contents lists available at SciVerse ScienceDirect

Applied Radiation and Isotopes

journal homepage: www.elsevier.com/locate/apradiso

Coverage intervals according to MARLAP, Bayesian statistics and the new ISO 11929 for ionising radiation measurements

Dieter Grientschnig [a,*], Ignacio Lira [b]

[a] Chemical Laboratories, Böhler Edelstahl, Mariazeller Str. 25, 8605 Kapfenberg, Austria
[b] Department of Mechanical and Metallurgical Engineering, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile

ARTICLE INFO

Article history:
Received 27 May 2010
Received in revised form
25 July 2011
Accepted 8 December 2011
Available online 16 December 2011

Keywords:
Measurement uncertainty
Bayesian data analysis
Replicate samples
Empirical standard deviation
Extra-Poisson variation
Gaussian distribution

ABSTRACT

The coverage intervals stipulated by ISO 11929 (2010) for estimating the uncertainty from ionising radiation measurements of replicate samples are compared with those of MARLAP (=Multi-Agency Radiological Laboratory Analytical Protocols Manual) and of Bayesian statistics. The latter two intervals agree well despite their different concepts. Whereas for either of them the ratio of the length of the coverage interval and MARLAP's standard uncertainty grows when the number of samples decreases, no such growth arises for the interval mandated by ISO 11929 (2010). It may therefore be too short (e.g. for three samples by a factor of approximately 2).

1. Introduction

A new edition of the international standard ISO 11929 for measurements of ionising radiation has come into force in March 2010, which replaces all eight parts of the preceding ISO 11929 standards series. Despite being mainly concerned with decision thresholds and detection limits, the new standard also includes the calculation of coverage intervals. Only the latter aspect of this standard will be investigated in the present paper. The study will consider the case of extra-Poisson variation determined by replicate measurements, thus contrasting a statistical analysis of low-level radioactivity measurement recently published in this journal by Heisel et al. (2009), which did not accommodate situations of such a kind.

The new ISO 11929 (2010) is claimed to proceed from the principles of Bayesian statistics. It stipulates a procedure for calculating coverage intervals that differs from the one employed in parts 1–4 of the previous ISO 11929 standards series, which are based on conventional (i.e. frequentist) statistics. These parts were in line with the "Multi-Agency Radiological Laboratory Analytical Protocols Manual" (MARLAP, 2004) which in turn, according to its section 19.3, adopts the methods for evaluating measurement uncertainty laid down in the "Guide to the Expression of Uncertainty in Measurement" (JCGM 100, 2008), subsequently simply called the GUM. In view of this equivalence of MARLAP and GUM, mostly references to the former document will be made in Sections 2–6 below, although all these references pertain to the latter as well.

While the GUM (JCGM 100, 2008) mandates the propagation of uncertainties, its recently published Supplement 1 (JCGM 101, 2008) advocates the propagation of probability distributions instead. In the framework of GUM Supplement 1 such distributions are not a frequency distributions but representations of "state of knowledge" or, equivalently, of "degree of belief". In this sense, as shown by Elster et al. (2007), GUM Supplement 1 embodies a Bayesian approach, not one of conventional statistics. By contrast, MARLAP and GUM contain elements of both these branches of statistics. The means advocated in GUM Supplement 1 for the propagation of distributions is the Monte Carlo method. However, as demonstrated in a recent review (Lira and Grientschnig, 2010), for measurement models involving a small number of quantities the Monte Carlo approach can be replaced by analytical calculations followed by numerical integration. Therefore, the procedures of GUM Supplement 1 (JCGM 101, 2008) and of Lira and Grientschnig (2010) are equivalent and will hereinafter be referred to as the calculation "according to Bayesian statistics".

As stated in the introduction of the GUM "it is often necessary to provide an interval about the measurement result that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the quantity subject

* Corresponding author. Tel.: +43 3862 20 36498; fax: +43 3862 20 37430.
E-mail address: dieter.grientschnig@bohler-edelstahl.at (D. Grientschnig).

0969-8043/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apradiso.2011.12.017

to measurement". The GUM requires the best estimate of the measurand to serve as the centre of this interval and the "expanded uncertainty" obtained with a chosen coverage factor to be its half-width. As mentioned in clause 6.2.2 of the GUM, the interval so defined has a certain "coverage probability" or "level of confidence". For this interval neither of the terms "coverage interval" or "confidence interval" nor any other short designation formed by placing a modifier in front of "interval" is introduced in the GUM, whereas the "International Vocabulary of Metrology" (JCGM 200, 2008) does attribute the term "coverage interval" to it. By contrast, the same kind of interval is termed "confidence interval" in the new ISO 11929 (2010). But this designation is not well-chosen because it has a specific meaning in conventional statistics, established by clause 1.28 of ISO 3534-1 (2006), which is not applicable in the context of Bayesian statistics. Therefore, we will adopt the terminology of the "International Vocabulary of Metrology" instead and use the term "coverage interval" throughout, regardless of whether it is determined according to MARLAP, Bayesian statistics or the new edition of ISO 11929 (2010).

The objective of this study is to compare for a particular example of activity measurement, presented in Section 2, the coverage intervals that ensue from the three approaches at hand. Sections 3, 4 and 5 are devoted to dealing with the example according to MARLAP, Bayesian statistics and the new ISO 11929 (2010), respectively. The article concludes with a discussion of the results and a summary.

2. Example of specific ^{137}Cs activity measurement

2.1. Original data

The following example derives from a scenario considered by Michel and Kirchhoff (1999) of measuring the specific ^{137}Cs activity of a batch of waste matter that despite reduction to small pieces and mechanical homogenisation is still fairly inhomogeneous with respect to activity. The average activity per mass of this waste shall be determined on the basis of several samples taken from different spots of the batch. Up to ten samples are assumed to be collected and subjected to a single-channel gross counting measurement with a counting time t_G=3600 s for each sample. By "counting time" we refer to the measuring system's live time, the uncertainty of which is assumed to be negligible. The gross counting results adopted from Michel and Kirchhoff (1999) are

$$n_{G,i} = (847,523,867,720,778,636,881,672,1102,756), \quad i = 1,\dots,10. \tag{1a}$$

In order to study the effect of the number of samples on the width of the coverage interval of the resulting specific activity, either all 10 counting results are accounted for or just part of them from the beginning of the series, i.e. $n_{G,i}$, i=1,...,K, with K set to 3, 4, 5, 6, 8 or 10. The means $\overline{n}_G$ and empirical standard deviations s_G for these six numbers K of individual counting results are displayed in Table 1, which along with Table 2 also shows the coverage intervals ensuing from these data. Given that the standard deviations exceed by far the square roots of the means $\overline{n}_G$, the scatter of the results cannot be solely due to Poisson counting statistics but must originate predominantly from other causes, such as random effects of sampling and sample treatment. Therefore it seems reasonable to assume, as Michel and Kirchhoff (1999) did, that the gross counts originate from a Gaussian distribution of unknown variance. The quantity estimated by means of the gross counts in conjunction with the corresponding counting time t_G is the gross count rate G.

Following the scenario of Michel and Kirchhoff (1999), the gross effect measurements are complemented with counting the radiation of 10 different blank samples, allotting a counting time of 7200 s for each of them. The standard deviation of these 10 counting results agrees well with the square root of the mean counts observed, so that Poisson counting statistics can be assumed to apply. Thus, for the sake of simplicity we consider only the sum n_B=2561 of the counts for the 10 blank samples and their total counting time t_B=72,000 s. The fact that this number of blank counts is supposed to be drawn from a Poisson distribution enables us to forgo the empirical determination of the variance suggested by Michel and Kirchhoff (1999). The quantity inferable from the information described is the blank count rate B.

The third quantity required is the calibration factor F, obtained by dividing the portion of the count rate originating from transformations of ^{137}Cs in the sample through its specific ^{137}Cs activity. This factor equals the product of the efficiency, i.e. the fraction of transformations of ^{137}Cs leading to counts, and the mass of the sample. However, due to attenuation effects within the sample, its mass may have a bearing on the efficiency. Therefore, a fixed sample mass is considered, so that it can be fused with the respective efficiency to the calibration factor F. The estimate f of this factor is supposed to originate from an earlier calibration utilising a sample of known specific ^{137}Cs activity, whose mass of 0.800 kg concurs with that of the waste samples at hand. Adopting from Michel and Kirchhoff (1999) the efficiency of 0.017 obtained in this way leads to f=0.0136 s^{-1} Bq^{-1} kg. Instead of taking this as the perfectly known value of the factor F we accommodate its uncertainty by assuming that only its lower and upper limits, $f-\Delta f$ and $f+\Delta f$, are given, where Δf= 0.0003 s^{-1} Bq^{-1} kg, and that all values within these limits are equally likely. This corresponds to adopting a rectangular probability distribution for F.

2.2. Modified data

The original data for the example are such that the uncertainties arising from the blank measurement and the calibration

Table 1
Limits of the 95% coverage interval of the specific ^{137}Cs activity according to MARLAP for the original data, based on K samples.

Gross count data			Result and uncertainty			Coverage interval	
K	$\overline{n}_G$	s_G	a	$u(a)$	$\nu_{A,eff}$	a_L	a_U
3	745.7	193.1	12.61	2.28	2.02	2.89	22.34
4	739.3	158.2	12.48	1.62	3.06	7.38	17.59
5	747.0	138.1	12.64	1.27	4.15	9.16	16.13
6	728.5	131.6	12.26	1.11	5.23	9.45	15.08
8	740.5	126.4	12.51	0.93	7.48	10.34	14.68
10	778.2	159.4	13.28	1.04	9.54	10.94	15.62

Specific activity given in Bq/kg.

Table 2
Limits of the 95% coverage intervals of the specific ^{137}Cs activity according to Bayesian statistics and to the new ISO 11929 for the original data, based on K samples.

K	Bayesian statistics			New ISO 11929	
	c	α_L	α_U	$a^{\triangleleft}$	$a^{\triangleright}$
3	1.0158	4.11	21.13	8.14	17.09
4	1.0023	7.43	17.55	9.30	15.67
5	1.0003	9.14	16.16	10.15	15.14
6	1.0001	9.43	15.10	10.09	14.44
8	1.0000	10.33	14.70	10.69	14.33
10	1.0000	10.93	15.64	11.23	15.33

Specific activity given in Bq/kg.