BAYESIAN UNCERTAINTIES IN MEASUREMENT AND CHARACTERISTIC LIMITS: ISO 11929 AND BEYOND

Rolf Michel
Center for Radiation Protection and Radioecology, Leibniz University Hannover
Herrenhäuser Str. 2, D-30419 Hannover, Germany, Fax: +49-511-762-3008,
michel@zsr.uni-hannover.de

Abstract: The quantification of measurement uncertainties and of characteristic limits, such as the decision threshold, detection limit, and limits of a confidence or coverage interval, is an essential task in metrology. The ISO Guide to the Expression of Uncertainty in Measurement (GUM), its recent Supplement1: “Propagation of distributions using a Monte Carlo method”, and the standard ISO 11929 “Determination of the detection limit and decision threshold for ionizing radiation measurements” provide an internationally standardized methodology for this task. Based on Bayesian statistics and a Bayesian theory of measurement uncertainty, characteristic limits can be calculated taking into account all sources of uncertainty. This approach consists of the complete evaluation of a measurement according to the GUM and the succeeding determination of the characteristic limits by using the standard uncertainty obtained from the evaluation. This procedure is described here for several particular models of evaluation. It is, however, generally applicable to a large variety of measurement tasks. It is proposed for the revision of those parts of ISO 11929 which are still based on conventional statistics and, therefore, do not allow to take completely into account all the components of measurement uncertainty in the calculation of the characteristic limits. The approach used in ISO 11929-7 and in the revised ISO/FDIS 11929 is consistently applicable to the extension of the GUM Supplement 1, which makes use of Monte Carlo techniques to quantify measurement uncertainties.

1 Introduction

Measurement uncertainties and characteristic limits, i.e. the decision threshold, the detection limit and the limits of the confidence or coverage interval, are essential ingredients of quality control in all fields of environmental monitoring and assessment. A result of a measurement without a statement about its associated uncertainty is worthless since it does not to allow quantifying potential environmental hazards or to demonstrate compliance of practices with legal requirements. More basically, any scientific statement has to be made with incomplete knowledge and, therefore, its uncertainty must be assessed.

Due to the fact that measurement uncertainties and characteristic limits are increasingly referred to in legislation and legal regulations, internationally standardized procedures are needed to quantify measurement uncertainties and characteristic limits. To satisfy these needs with respect to measurement uncertainties, ISO published the ISO Guide to the Expression of Uncertainty in Measurement (GUM) [1] in 1993 which allows a standardized quantification of measurement uncertainties. The ISO methodology is widely accepted and has been adopted also by national and international bodies; e.g. [2, 3].

Characteristic limits are capable to deal with a general problem in nuclear and other analytical techniques as well as generally in metrology, namely that a measurand is usually to be determined in the presence of a background or a blank. From this fact three questions arise: (1) Is there a contribution of the sample analyzed among the signals measured? (2) Is the analytical method used suited to perform the measurement task? (3) Which range of true values may be reasonably attributed to the measurand given the measured results?

These three questions can be answered by the concept of characteristic limits. The decision threshold allows a decision to be made for a measurement on whether or not, for instance, radiation of a possibly radioactive sample is present. The detection limit allows a decision on whether or not the measurement procedure intended for application to the measurement meets the requirements to be fulfilled and is therefore appropriate for the measurement purpose. Finally, the limits of the confidence or – as it is now preferred - coverage interval enclose with a specified probability the true value of the measurand to be measured.

The standard series ISO 11929 [4 - 11] provide characteristic limits for a diversity of application fields in nuclear radiation measurements. The methodology of ISO 11929 is, moreover, applicable to most other metrological problems. However, as the development of the GUM and of characteristic limits proceeded widely independently, also ISO standards on the capability of detection exist, e.g. ISO 11843-1 [12], which are not compatible with the terminology or methodology of the GUM [1], the recent GUM Supplement 1 [13], the VIM [14], and ISO 11929 [4 - 11].

One particular problem arises from the fact that the statistical foundation of the GUM was not clear at the time when it was published. It was a mixture of Bayesian and frequentist or conventional statistics which have completely different understanding of the term “probability”. In the GUM Supplement 1 [13] it is now explicitly stated that the GUM methodology can only be derived from Bayesian statistics (compare chapter 3). The Bayesian foundation of the GUM [15 - 16] took considerable time and also developed asynchronously to the GUM [1].

The transition from conventional to Bayesian statistics also affected ISO 11929. The older Parts 1 to 4 of ISO 11929 [4 - 7] were based on conventional statistics, while the modern Parts 5 to 8 of ISO 11929 [8 - 11] are already Bayesian. Therefore, Parts 1 to 4 of ISO 11929 need urgently a revision based on the common, already laid statistical foundation of Part 7 of ISO 11929 [10].

This paper outlines the general foundation of the Bayesian measurement uncertainties and characteristic limits and gives exemplarily some applications to measurements of ionizing radiation. Further information on the revision of ISO 11929, which is presently underway and which will combine all Parts of ISO 11929 in one document, may be found elsewhere [17, 18]. In contrast to the earlier publications [17, 18], this paper makes use of a new set of symbols in accordance with a decision of Working Group 14 “Radioactivity Measurements” of ISO TC 85 SC 2 for the revision ISO 11929. Finally, the paper gives some information how to deal beyond ISO 11929 with situations where Monte Carlo methods are applied according to the GUM Supplement 1 [13].

2 Uncertainty in Measurement

The starting point of any measurement is the definition of the quantity Y to be measured and for which the characteristic limits are to be determined. This measurand♣ is, for instance, the concentration of an element or an activity of radionuclides in a sample. The measurand is connected to input quantities X_i ($i = 1,...,n$) which originate from measurements or from other sources of information by a model of evaluation. Examples of input quantities are net peak areas from γ-spectra, efficiency data of a detector, sample masses, and chemical yields. The model of evaluation is a mathematical relationship:

♣ For the proper use of the metrological terms see ref. [14].

$$Y = G(X_1, X_2, ..., X_n) \tag{1}$$

Note, that the model function G needs not necessarily to be explicitly available. The model may also be given in form of a computer code.

Measurements yield estimates x_i of the true values of the input quantities X_i. The estimates x_i have standard uncertainties $u(x_i)$ associated with them. The evaluation analysis yields an estimate y of the measurand Y using y as an estimate of the true value $\tilde{y}$ of Y in equation 1 and one obtains

$$y = G(x_1, x_2, ..., x_n) \tag{2}$$

If the input quantities X_i are not correlated, the standard uncertainty $u(y)$ associated with y is calculated according to the GUM [1] as the positive square root of the variance

$$u^2(y) = \sum_{i=1}^{n} \left(\frac{\partial G}{\partial x_i} \right)^2 \cdot u^2(x_i) \text{ with the sensitivity coefficients } c_i \equiv \left. \frac{\partial G}{\partial x_1} \right|_{X_1 = x_1, ..., X_n = x_n} . \tag{3}$$

If the input quantities X_i are correlated, the standard uncertainty $u(y)$ has to be calculated using covariances $u(x_i,x_j)$; see the GUM [1] for details:

$$u^2(y) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_i \cdot c_j \cdot u(x_i, x_j) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_i \cdot c_j \cdot u(x_i) \cdot u(x_j) \cdot r(x_i, x_j) \tag{4}$$

with $r(x_i, x_i)$ being the correlation coefficients between x_i and x_j . It holds $u(x_i, x_i) = u^2(x_i)$. Given m repeated measurements of the input quantities X_i and X_j with the measured values $x_{i,k}$ and $x_{j,k}$ (k = 1,…,m) and the respective arithmetic means $\bar{x}_{i,k}$ and $\bar{x}_{j,k}$, then the estimates $x_i = \bar{x}_{i,k}$ and $x_j = \bar{x}_{j,k}$ follow and the covariances $u(x_i,x_j)$ are calculated by:

$$u(x_i, x_j) = \frac{1}{m-1} \sum_{k=1}^{m} (x_{i,k} - x_i)(x_{j,k} - x_j) . \tag{5}$$

Since, in general, the covariances can directly be calculated from the standard uncertainties and the correlation coefficient via $u(x_i, x_j) = u(x_i) \cdot u(x_j) \cdot r(x_i, x_j)$, their determination does not cause calculational problems.

If the partial derivatives are not explicitly available, they can be numerically sufficiently well approximated by using the standard uncertainty $u(x_i)$ as an increment of x_i:

$$\frac{\partial G}{\partial x_i} = \frac{1}{u(x_i)} \cdot \left(G(x_1, ..., x_i + u(x_i)/2, ..., x_n) - G(x_1, ..., x_i - u(x_i)/2, ..., x_n) \right) \quad . \qquad (6)$$

The standard uncertainties have to be evaluated according to the GUM [1]. In the GUM, uncertainties are evaluated either by “statistical methods” (type A) or by “other means” (type B). Type A uncertainties can be evaluated from repeated or counting measurements, while Type B uncertainties cannot. They are, for instance, uncertainties given in certificates of standard reference materials or of calibration radiation sources which are used in the evaluation of a measurement.

It is the distinction between the two ways (Type A and Type B), by which uncertainties are evaluated, which causes the problem with Bayesian and conventional statistics. Conventional statistics can only handle Type A uncertainties, but not Type B ones. Only by Bayesian statistics, uncertainties of both types can be consistently determined. They both express quantitatively the actual state of incomplete knowledge about the quantities involved.

Though many results of the conventional and the Bayesian approaches are numerically practically equal, they must not be confused with each other because the understanding of the term “*probability*” is completely different in both statistics. The conventional or frequentist view is “*Probability is the stochastic limit of relative frequencies”* while the Bayesian view is “*Probability is a measure of the degree of belief an individual has in an uncertain proposition”*. But, there are frequencies which do not represent probabilities and there are probabilities which cannot be expressed as frequencies. Bayesian statistics provides a more intuitive assessment methodology than conventional statistics, closer to the scientific thinking than conventional ones. For more details of these questions see e.g. [19 - 27].

Another complication when applying the GUM, which is not related to the type statistics used, arises from the fact that the GUM uncertainties are the result of a first order Taylor expansion and only applies to linear models or those which can be reasonably approximated by such a model – at least in the proximity of the actual data. This is frequently overlooked and the GUM is applied without checking first whether the model in question fulfils this requirement. However, the GUM Supplement 1 [13] describes a generally applicable methodology based on Monte Carlo techniques which resolves this issue. Methods for the determination of characteristic limits in this case are also available [28].

3 Bayesian Statistics in Measurement

As mentioned above, the basic difference between conventional and Bayesian statistics lies in the different use of the term probability. Considering measurements, conventional statistics describes only the probability distribution[♠] $f(y|\tilde{y})$, i.e. the conditional distribution of estimates y given the true value $\tilde{y}$ of the measurand Y. However, since the true value $\tilde{y}$ of a measurand is principally unknown, it is the basic task of an experiment to make statements about the probability of $\tilde{y}$. Bayesian statistics allows the calculation of the PDF $f(\tilde{y}|y)$ of the true value $\tilde{y}$ of a measurand Y given the measured estimate y as well as of $f(y|\tilde{y})$. The (standard) measurement uncertainty and the characteristic limits are based on both distributions $f(y|\tilde{y})$ and $f(\tilde{y}|y)$. Characteristic limits implicitly depend on further conditions and information such as the model, measurement data and associated uncertainties.

In order to establish the posterior PDF $f(\tilde{y}|y)$, one uses an approach which separates the information about the measurand obtained from the actual experiment (data prior or likelihood) from other information available about the measurand (model prior) by

$$f(\tilde{y}|y) = C \cdot f_0(\tilde{y}|y) \cdot f(\tilde{y}) \tag{7}$$

♠ Frequently, the term probability density function (PDF) is used instead of probability distribution. This terminology is also used here in accordance with its use in the GUM Supplement 1 [13].

The likelihood $f_0(\tilde{y}|y)$ is the PDF that the measurand Y has the true value $\tilde{y}$ if only the measured value y and the associated standard uncertainty $u(y)$ are given. It only accounts for the measured values and neglects any other information about the measurand. The model prior $f(\tilde{y})$ represents all the information about the measurand available before the experiment is performed. Therefore, it does not depend on y. C is a normalization constant.

If, for instance, an activity of a radiation source or a concentration of an element is the measurand, there exists the meaningful information that the measurand is non-negative $(\tilde{y} \geq 0)$ before the measurement is carried out. This yields for the model prior $f(\tilde{y})$:

$$f(\tilde{y}) = \begin{cases} \text{const} & (\tilde{y} \geq 0) \\ 0 & (\tilde{y} < 0) \end{cases} \tag{8}$$

Note, that the actual result y of a measurement, for instance a net count rate, can be negative. But the experimentalist knows *a priori* without performing an experiment that the true value $\tilde{y}$ is non-negative. All non-negative values of the measurand have the same *a priori* probability if there is no other information about the true value of the measurand before the measurement has been performed.

Since the likelihood $f_0(\tilde{y}|y)$ in essence considers the experimental information, the expectation $\mathrm{E}_0(\tilde{y}) = y$ and the variance $\mathrm{Var}_0(\tilde{y}) = u^2(y)$ should hold true for the probability density distribution $f_0(\tilde{y}|y)$.

According to Weise and Wöger [15, 16], the posterior PDF $f(\tilde{y}|y)$ can be determined by applying the principle of maximum (information) entropy [27] S:

$$S = -\int f_0(\tilde{y}|y) \cdot \ln\left(f_0(\tilde{y}|y)\right) \mathrm{d}\,\tilde{y} = \max \tag{9}$$

It is to emphasize that the PME is also used in the GUM Supplement 1 [13] to derive PDFs for a variety of differing information scenarios.

Equation 9 can be solved with the constraints $\mathrm{E}_0(\tilde{y}) = y$ and $\mathrm{Var}_0(\tilde{y}) = u^2(y)$ of $f_0(\tilde{y}|y)$ by the method of Lagrangian multiplicators and one obtains the result

$$f(\tilde{y}|y) = C \cdot f(\tilde{y}) \cdot \exp\left(-(\tilde{y}-y)^2 /(2 \cdot u^2(y))\right) \tag{10}$$

Accordingly, the distribution $f(\tilde{y}|y)$ is a product of the model prior $f(\tilde{y})$ and a Gaussian N(y,u(y)), i.e. a truncated Gaussian (Fig. 1). Note, that the Gaussian in equation 10 is not an approximation as in conventional statistics or a distribution of measured values from repeated or counting measurements. It is instead the explicit result of maximizing the information entropy and expresses the state of knowledge about the measurand *Y*.

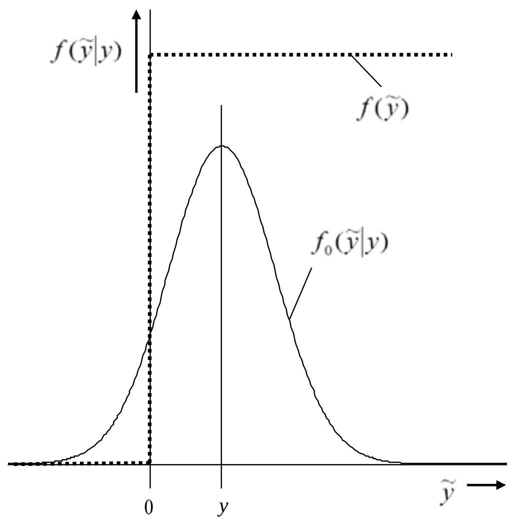

Fig. 1. Illustration of the likelihood $f_0(\tilde{y}|y)$, the model prior $f(\tilde{y})$, and the posterior $f(\tilde{y}|y)$ given in equation 10 for a non-negative measurand *Y*.

After $f(\tilde{y}|y)$ is obtained, the Bayes theorem (equation 11) also allows the calculation of the PDF $f(y|\tilde{y})$ of an estimate y given the true value $\tilde{y}$ of the measurand *Y*:

$$f(y|\tilde{y}) \cdot f(\tilde{y}) = f(\tilde{y}|y) \cdot f(y)\,. \tag{11}$$

The PDF $f(y)$ is uniform for all possible measurement results y and $f(\tilde{y})$ is uniform for all $\tilde{y} \geq 0$ according to equation 8. Thus, $f(y|\tilde{y})$ is obtained from equations 10 and 11 by approximating the now not available $u(y)$ by a function $\tilde{u}(\tilde{y})$

$$f(y|\tilde{y}) = C \cdot \exp\left(-(y-\tilde{y})^2 /(2 \cdot \tilde{u}^2(\tilde{y}))\right) \text{ with } (\tilde{y} \geq 0) \tag{12}$$

The PDF $f(y|\tilde{y})$ is a Gaussian for a given true value $\tilde{y}$ of the measurand with the standard uncertainty $\tilde{u}(\tilde{y})$. Note, that the true value $\tilde{y}$ of the measurand Y is now a parameter in equation 12 and that the variance $u^2(y)$ of the PDF $f(\tilde{y}|y)$ equals the variance $\tilde{u}^2(\tilde{y})$ of the PDF $f(y|\tilde{y})$, i.e. $u^2(y) = \tilde{u}^2(\tilde{y})$.

4 Calculation of the Standard Uncertainty as a Function of the True Value of the Measurand

For the provision and numerical calculation of the decision threshold and detection limit, the standard uncertainty of the measurand is needed as a function $\tilde{u}(\tilde{y})$ of the true value $\tilde{y} \geq 0$ of the measurand. This function has to be determined in a way similar to $u(y)$ within the framework of the evaluation of the measurements by application of GUM [1]. In most cases, $\tilde{u}(\tilde{y})$ has to be formed as a positive square root of a variance function $\tilde{u}^2(\tilde{y})$ calculated first. This function must be defined, unique and continuous for all $\tilde{y} \geq 0$ and must not assume negative values.

In some cases, $\tilde{u}(\tilde{y})$ can be explicitly specified, provided that $u(x_1)$ is given as a function $h_1(x_1)$ of x_1. In such cases, y has to be replaced by $\tilde{y}$ and equation 2 must be solved for the estimate x_1 of the input quantity X_1 which in the following is always taken as the gross effect quantity. With a specified $\tilde{y}$, the value x_1 can also be calculated numerically from equation 2, for instance, by means of an iteration procedure, which results in x_1 as a function of $\tilde{y}$ and $x_2, ..., x_m$. This function has to replace x_1 in equation 3 and in $u(x_1) = h_1(x_1)$, which finally yields $\tilde{u}(\tilde{y})$ instead of $u(y)$. In most cases of the models dealt with in this paper one has to proceed in this way. Otherwise, $\tilde{u}(\tilde{y})$ can be obtained as an approximation by interpolation from the data y_j and $u(y_j)$ of several measurements. If only $\tilde{u}(0)$, y, and $u(y)$ are known, the approximation by linear interpolation according to equation 13 is often sufficient for $y > 0$:

$$\tilde{u}^2(\tilde{y}) = \tilde{u}^2(0) \cdot (1 - \tilde{y}/y) + u^2(y) \cdot \tilde{y}/y \qquad (13)$$

5 Decision Threshold and Detection Limit

Without a detailed mathematical foundation of Bayesian characteristic limits, which may be found elsewhere [17, 18, 29, 30], we can now define the characteristic limits for a non-negative measurand Y which is, for instance, a concentration of an element or an activity of a radionuclide in a sample. The true value $\tilde{y}$ is zero if the element or the radionuclide is not present. The decision threshold and the detection limit are defined [10, 17, 18] on the basis of decision about the acceptance of the null hypothesis H_0: $\tilde{y} = 0$ against the acceptance of the alternative hypothesis H_1: $\tilde{y} > 0$.

To solve this decision problem in Bayesian statistics [22], it is necessary to specify an appropriate loss function, measuring the consequences of accepting or rejecting H_0 as a function of the actual measurement result y. In ISO 11929 a quadratic loss function is used as the simplest function since there is no further information about the consequences of accepting or rejecting H_0 and both actions are equally weighted.

The entire procedure then works as follows. As result of the measurement, y and the associated standard uncertainty $u(y)$ are derived according to the GUM [1] as a complete result of the measurement. y and $u(y)$ have to be derived by evaluation of measured quantities and of other information by way of the mathematical model which takes into account all relevant quantities. Generally, it will not be explicitly made use of the fact that the measurand is non-negative. Therefore, y may become negative, in particular, if the true value of the measurand is close to zero.

For the determination of the decision threshold and the detection limit, the standard uncertainty of the decision quantity has to be calculated, if possible, as a function $\tilde{u}(\tilde{y})$ of the true value $\tilde{y}$ of the measurand. In the case that this is not possible, approximate solutions are described below.

Then, the decision threshold y^* (Fig. 2) is a characteristic limit which when exceeded by a result y of a measurement one decides that the element or radionuclide is present in the sample. If $y \leq y^*$, one decides to accept the null hypothesis, H_0: $\tilde{y} = 0$, and concludes that the element or radionuclide is not found in this sample. If this decision rule

$$P(y > y^* | \tilde{y} = 0) = \alpha \tag{14}$$

is observed, a wrong acceptance of the alternative hypothesis, H_1: $\tilde{y} > 0$, occurs with the probability α which is the probability of the error of the first kind of the decision made.

The decision threshold is given by

$$y^* = k_{1-\alpha} \cdot \tilde{u}(0) \tag{15}$$

with $k_{1-\alpha}$ being the (1-α)-quantile of the standardized normal distribution. $\tilde{u}(0)$ is the uncertainty of the measurand if its true value equals zero. If the approximation $\tilde{u}(\tilde{y} = 0) = u(y)$ is sufficient, one obtains

$$y^* = k_{1-\alpha} \cdot u(y) . \tag{16}$$

The detection limit $y^{\#}$ (Fig. 2) is the smallest true value of the measurand detectable with the measuring method. It is defined by

$$P(y < y^* | \tilde{y} = y^{\#}) = \beta . \tag{17}$$

The detection limit $y^{\#}$ is sufficiently larger than the decision threshold y^* such that the probability of $y < y^*$ equals the probability β of the error of the second kind in the case of $\tilde{y} = y^{\#}$.

The detection limit is calculated by

$$y^{\#} = y^{*} + k_{1-\beta} \cdot \tilde{u}(y^{\#}) \qquad (18)$$

with $k_{1-\beta}$ being the (1-β)-quantile of the standardized normal distribution.

Equation 18 is an implicit one. The detection limit can be calculated from it by iteration using for example the starting approximation $y^{\#} = 2 \cdot y^{*}$.

For the numerical calculation of the decision threshold and the detection limit the function $\tilde{u}(\tilde{y})$ is needed which gives the standard uncertainty of the decision quantity as function of the true value $\tilde{y}$ of the measurand. This function generally has to be determined in the course of the evaluation of the measurement according to the GUM [1]. Often this function is only slowly increasing. Therefore it is justified in many cases to use the approximation $\tilde{u}(\tilde{y}) = u(y)$. If the approximation $\tilde{u}(\tilde{y}) = u(y)$ is sufficient for all true values $\tilde{y}$, then

$y^{\#} = (k_{1-\alpha} + k_{1-\beta}) \cdot u(y)$ is valid. (19)

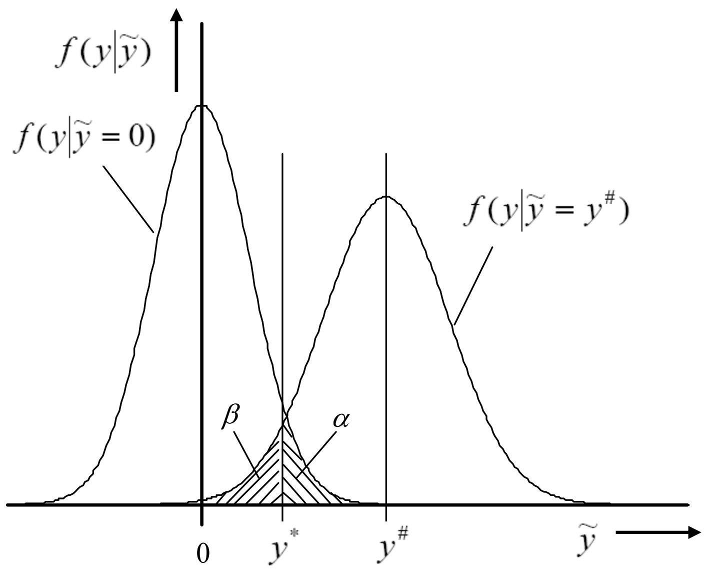

Fig. 2 Illustration of the decision threshold y^{*} and the detection limit $y^{\#}$.

Frequently, the value of y is calculated as the difference (net effect) of two quantity values of approximately equal size with x_1 being the gross effect and x_0 being the background or blank effect, both obtained from independent measurements. In this case of $y = x_1 - x_0$ one gets

$u^2(y) = u^2(x_1) + u^2(x_0)$ with the standard uncertainties $u(x_1)$ and $u(x_0)$ associated with x_1 and x_0, respectively. From this, one obtains $\tilde{u}^2(\tilde{y} = 0) = 2 \cdot u^2(x_0)$, since for $\tilde{y} = 0$ one expects $x_1 = x_0$.

If only $\tilde{u}^2(0)$, y and $u(y)$ are known, the approximation by linear interpolation according to Equation 18 is often sufficient for $y > 0$:

$$\tilde{u}^2(\tilde{y}) = \tilde{u}^2(0) \cdot (1 - \tilde{y}/y) + u^2(y) \cdot \tilde{y}/y\,. \qquad (20)$$

With this interpolation formula one gets the approximation for the detection limit

$$y^\# = a + \sqrt{a^2 + (k_{1-\beta}^2 - k_{1-\alpha}^2) \cdot \tilde{u}^2(0)} \qquad (21)$$

with

$$a = k_{1-\alpha} \cdot \tilde{u}(0) + \frac{1}{2}(k_{1-\beta}^2 / y) \cdot (u^2(y) - \tilde{u}^2(0)) \qquad (22)$$

For $\alpha = \beta$, one receives $y^\# = 2 \cdot a$.

6 Limits of the Confidence or Coverage Interval

The confidence or coverage interval (Fig. 3) includes for a result y of a measurement, which exceeds the decision threshold y^*, the true value of the measurand with a probability 1 - γ. It is enclosed by the lower and upper limit of the confidence or coverage interval, respectively $y^\triangleleft$ and $y^\triangleright$, which are defined by

$$\int_0^{y^\triangleleft} f(\tilde{y}|y)\,\mathrm{d}y = \int_{y^\triangleright}^{\infty} f(\tilde{y}|y)\,\mathrm{d}y = \gamma/2 \qquad (23)$$

so that $\int_{y^{\triangleleft}}^{y^{\triangleright}} f(\tilde{y}|y)\,\mathrm{d}y = 1-\gamma$ holds.

Explicitly, the lower and upper limit of the confidence or coverage interval, $y^{\triangleleft}$ and $y^{\triangleright}$ are calculated via

$$y^{\triangleleft} = y - k_p \cdot u(y) \text{ with } p = \omega \cdot (1-\gamma/2) \tag{24}$$

$$y^{\triangleright} = y + k_q \cdot u(y) \text{ with } q = 1-\omega \cdot \gamma/2 \tag{25}$$

The parameter ω is given by

$$\omega = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y/u(y)} \exp(-z^2/2)\,\mathrm{dz} = \Phi(y/u(y)) \tag{26}$$

Values of the function $\Phi(t)$, which is the distribution function of the standardized normal distribution, and the quantiles k_p of the standardized normal distribution are tabulated, e.g. [31], but also available in most spread-sheet applications.

The limits of the confidence or coverage interval are not symmetrical around the expectation $\hat{y} = \mathrm{E}(f(\tilde{y}|y))$. The probabilities of $\tilde{y} < y^{\triangleleft}$ and $\tilde{y} > y^{\triangleright}$, however, both are equal to $\gamma/2$ and the relationship $0 < y^{\triangleleft} < y^{\triangleright}$ is valid. If y and $u(y)$ are of similar size, this asymmetry of the confidence or coverage interval is clearly visible. But for $y >> u(y)$, the well known formula

$$y^{\triangleleft,\triangleright} = y \pm k_{1-\gamma/2} \cdot u(y) \tag{27}$$

is valid as an approximation. Equation 27 is applicable if $y >\approx 2 \cdot k_{1-\gamma/2} \cdot u(y)$.

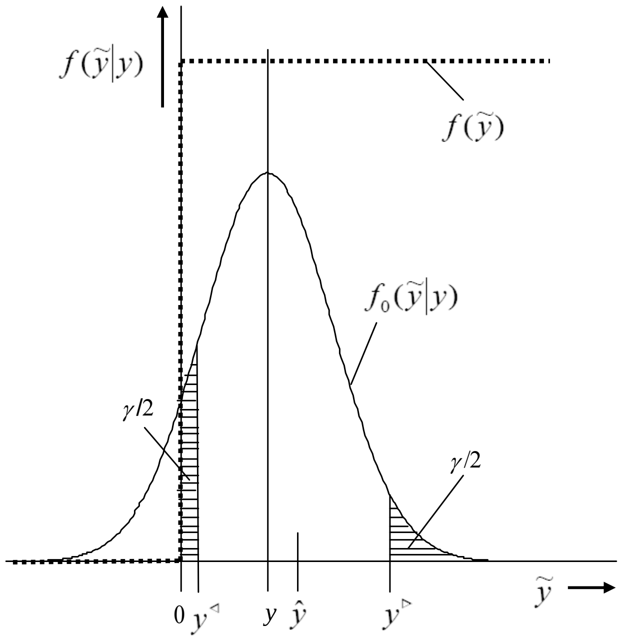

Fig. 3. Illustration of the limits $y^{\triangleleft}$ and $y^{\triangleright}$ of the confidence or coverage interval and of the best estimate $\hat{y}$ of the true value $\tilde{y}$ of a non-negative measurand *Y*.

7 Assessment of an Analytical Technique

Having performed a measurement and an evaluation of the measurement according to the GUM [1], the performance of the analytical technique can be assessed in the following way:

A measured result *y* has to be compared with the decision threshold y^* calculated by means of equation 15. If a result of the measurement *y* is larger than the decision threshold y^* one decides that a non-zero effect quantified by the measurand is observed and that the element or activity is present in the sample.

To check whether a measurement procedure is suitable for measuring the measurand, the calculated detection limit $y^{\#}$ has to be compared with a specified guideline value, e.g. according to specified requirements on the sensitivity of the measurement procedure from scientific, legal or other reasons. The detection limit has to be calculated by means of equation 18. If the detection limit thus determined is smaller than the guideline value, the procedure is suitable for the measurement, otherwise it is not.

If a non-zero effect is observed, i.e. $y > y^*$, the best estimate $\hat{y}$ of the measurand (Fig. 3) can be calculated as the expectation of the PDF $f(\tilde{y}|y)$ and the standard deviation of $f(\tilde{y}|y)$ is the standard uncertainty $u(\hat{y})$ associated with the best estimate $\hat{y}$ of the measurand *Y*.

$$u(\hat{y}) = \sqrt{\mathrm{Var}(f(\tilde{y}|y)} \tag{28}$$

Using ω from equation 26, the best estimate $\hat{y}$ is calculated by

$$\hat{y} = \mathrm{E}(f(\tilde{y}|\mathrm{y})) = y + \frac{u(y) \cdot \exp\left(-y^2/(2 \cdot u^2(y))\right)}{\omega \cdot \sqrt{2\pi}} \tag{29}$$

with the associated standard uncertainty $u(\hat{y})$

$$u(\hat{y}) = \sqrt{u^2(y) - (\hat{y} - y) \cdot \hat{y}}\ . \tag{30}$$

The following relationships hold: $\hat{y} \geq y$ and $\hat{y} \geq 0$ as well as $u(\hat{y}) \leq u(y)$. For $y >> u(y)$ the approximations $\hat{y} = y$ and $u(\hat{y}) = u(y)$ are valid. See Fig. 3 for an illustration of the confidence or coverage interval and the best estimate of the measurand.

8 Applications

8.1 Frequently Used Models

Many applications, also in other fields than measurements of ionising radiation, use models of evaluation of the general mathematical form:

$$Y = G(X_1,\ldots,X_m) = (X_1 - X_2 \cdot X_3 - X_4) \cdot \frac{X_6 \cdot X_8 \cdots}{X_5 \cdot X_7 \cdots} = (X_1 - X_2 \cdot X_3 - X_4) \cdot W \tag{31}$$

with $W = \frac{X_6 \cdot X_8 \cdots}{X_5 \cdot X_7 \cdots}$.

In measurements of ionizing radiation, $X_1 = R_\mathrm{g}$ and $X_2 = R_0$ frequently are the counting rates of a gross and a background measurement, respectively. X_3 can, for instance, be a shielding correction and X_4 an additional general background correction. X_5, X_6, … are calibration and correction factors. If X_3 or X_4 are not needed, $x_3 = 1$ and $u(x_3) = 0$ or $x_4 = 0$ and $u(x_4) = 0$ have to be set.

By replacing the quantities in equation 31 by their actual estimates x_i and w for W one obtains with $x_1 = r_\mathrm{g} = n_\mathrm{g} / t_\mathrm{g}$ and $x_2 = r_0 = n_0 / t_0$:

$$y = G(x_1,...,x_m) = (x_1 - x_2 \cdot x_3 - x_4) \cdot w = (r_\mathrm{g} - r_0 \cdot x_3 - x_4) \cdot w = \left(\frac{n_\mathrm{g}}{t_\mathrm{g}} - \frac{n_0}{t_0} x_3 - x_4 \right) \cdot w \qquad (32)$$

n_g and n_0 are the numbers of counted events in the gross and the background measurement of duration t_g and t_0, respectively.

With the partial derivatives $\frac{\partial G}{\partial X_1} = W$; $\frac{\partial G}{\partial X_2} = -X_3 \cdot W$; $\frac{\partial G}{\partial X_3} = -X_2 \cdot W$; $\frac{\partial G}{\partial X_4} = -W$; $\frac{\partial G}{\partial X_i} = \pm \frac{Y}{X_i}$ $(i \geq 5)$, and by substituting the estimates x_i, w and y, equation 3 yields the standard uncertainty $u(y)$ of the measurand Y associated with y:

$$\begin{aligned} u^2(y) &= w^2 \cdot (u^2(x_1) + x_3^2 \cdot u^2(x_2) + x_2^2 \cdot u^2(x_3) + u^2(x_4)) + y^2 \cdot u_\mathrm{rel}^2(w) \\ &= w^2 \cdot (r_\mathrm{g} / t_\mathrm{g} + x_3^2 r_0 / t_0 + r_0^2 \cdot u^2(x_3) + u^2(x_4)) + y^2 \cdot u_\mathrm{rel}^2(w) \end{aligned} \qquad (33)$$

where $u_\mathrm{rel}^2(w) = \sum_{i=5}^{m} \frac{u^2(x_i)}{x_i^2}$ is the sum of the squared relative standard uncertainties of the quantities X_5 to X_m. For $m < 5$, the values $w = 1$ and $u_\mathrm{rel}^2(w) = 0$ apply.

For a true value $\tilde{y}$ one expects

$$n_g = \tilde{y}/w + r_0 \cdot x_3 + x_4 \quad (34)$$

and with equation 33 one obtains

$$\tilde{u}^2(\tilde{y}) = w^2 \cdot [(\tilde{y}/w + r_0 \cdot x_3 + x_4)/t_{\mathrm{g}} + x_3^2 \cdot r_0/t_0 + r_0^2 \cdot u^2(x_3) + u^2(x_4)] + \tilde{y}^2 \cdot u_{\mathrm{rel}}^2(w) . \quad (35)$$

This yields for $\tilde{y} = 0$ the decision threshold

$$y^* = k_{1-\alpha} \cdot \tilde{u}(0) = k_{1-\alpha} \cdot w \cdot \sqrt{(r_0 \cdot x_3 + x_4)/t_{\mathrm{g}} + x_3^2 \cdot r_0/t_0 + r_0^2 \cdot u^2(x_3) + u^2(x_4)} \quad (36)$$

Note that in this class of model functions according to equation 31 the decision threshold does not depend on the uncertainty of *w*.

For the detection limit one obtains

$$y^{\#} = y^* + k_{1-\beta} \cdot \tilde{u}(y^{\#})$$
$$= y^* + k_{1-\beta} \cdot \sqrt{w^2 \cdot [(y^{\#}/w + r_0 \cdot x_3 + x_4)/t_{\mathrm{g}} + x_3^2 \cdot r_0/t_0 + r_0^2 \cdot u^2(x_3) + u^2(x_4)] + y^{\#2} \cdot u_{\mathrm{rel}}^2(w)} \quad (37)$$

which has to be solved for $y^{\#}$. Equation 37 has a solution if $k_{1-\beta}^2 \cdot u_{\mathrm{rel}}^2(w) < 1$.

8.2 Determination of an Activity

In ionizing-radiation measurements, often the activity is determined from a measurement of a net count rate value $r_{\mathrm{n}} = (r_{\mathrm{g}} - r_0)$ as the difference of a gross count rate value $r_{\mathrm{g}} = n_{\mathrm{g}}/t_{\mathrm{g}}$ and a background count rate value $r_0 = n_0/t_0$ with time preselection multiplied by a calibration factor *w* with the standard uncertainties $u(r_{\mathrm{g}}) = r_{\mathrm{g}}/t_{\mathrm{g}}$, $u(r_0) = r_0/t_0$, and $u(w)$, respectively. This yields the simple model

$$y = (r_{\mathrm{g}} - r_0) \cdot w = (n_g / t_g - n_0 / t_0) \cdot w \quad (38)$$

which just is a special case of equation 32 with $x_1 = r_g$, $x_2 = r_0$, $x_3 = 1$, $u(x_3) = 0$, $x_4 = 0$ and $u(x_4) = 0$. Since, however, it is used so frequently, it shall be explicitly dealt with here.

Equation 3 yields the standard uncertainty $u(y)$ of the measurand Y associated with y:

$$u^2(y) = w^2 \cdot (r_\mathrm{g} / t_\mathrm{g} + r_0 / t_0) + y^2 \cdot u_\mathrm{rel}^2(w) \text{ with } u_\mathrm{rel}(w) = u(w)/w. \tag{39}$$

With this information, $\tilde{u}(\tilde{y})$ can be explicitly calculated since one expects for a true value $\tilde{y}$ of the measurand a number of counts n_g in the gross measurement

$$n_g = \tilde{y} / w + r_0 \cdot t_g \tag{40}$$

Since $u^2(n_\mathrm{g}) = n_\mathrm{g}$ is valid for a Poisson process, one can calculate $\tilde{u}(\tilde{y})$ using equation 39 as

$$\tilde{u}^2(\tilde{y}) = w^2 \cdot ((\tilde{y} / w + r_0) / t_g + r_0 / t_0) + \tilde{y}^2 \cdot u_\mathrm{rel}^2(w) \tag{41}$$

and obtains for $\tilde{y} = 0$ the decision threshold

$$y^* = k_{1-\alpha} \cdot \tilde{u}(0) = k_{1-\alpha} \cdot w \cdot \sqrt{r_0 \cdot \left(\frac{1}{t_\mathrm{g}} + \frac{1}{t_0} \right)} \tag{42}$$

and for the detection limit

$$y^\# = y^* + k_{1-\beta} \cdot \tilde{u}(y^\#) = y^* + k_{1-\beta} \cdot \sqrt{w^2 \cdot ((y^\# / w + r_0) / t_\mathrm{g} + r_0 / t_0) + {y^\#}^2 \cdot u_\mathrm{rel}^2(w)} \tag{43}$$

which can conveniently be solved by iteration with the starting value $y^\# = 2y^*$. For $\alpha = \beta$, i.e. $k_{1-\alpha} = k_{1-\beta}$, equation 43 has the simple explicit solution:

$$y^{\#} = \frac{2 \cdot y^{*} + (k_{1-\alpha}^{2} \cdot w)/t_{g}}{1 - k_{1-\alpha}^{2} \cdot u_{\mathrm{rel}}^{2}(w)} \quad . \tag{44}$$

8.3 General Measurement of a Net Quantity with Calibration

The simple model of equation 38 can also be used to demonstrate that the approach described here is not limited to measurements of ionizing radiation. A model in the form of equation 7 describes an evaluation of any measurand which is derived from a gross or sample measurement and a background or blank measurement. The value y of the measurand Y is the difference of the gross signal x_g and the blank signal x_0 multiplied by a calibration factor w with their respective standard uncertainties $u(x_g)$, $u(x_0)$, and $u(w)$.

$$y = (x_g - x_0) \cdot w \tag{45}$$

Then the standard uncertainty $u(y)$ associated with y is given by:

$$u^2(y) = w^2 \cdot (u^2(x_g) + u^2(x_0)) + y^2 \cdot u_{\mathrm{rel}}^2(w) \text{ with } u_{\mathrm{rel}}(w) = u(w)/w \, . \tag{46}$$

The minimum information requirement to allow for the calculation of the decision threshold and the detection limit is that the experiment was successfully performed at least one times each for the gross and the background measurements. This means that x_g, $u(x_g)$, x_0, $u(x_0)$, w, and $u(w)$ are available. In particular, it is not needed for the following that x_g and x_0 result from a Poisson process.

For $\tilde{y} = 0$, one expects $x_g = x_0$ and obtains with equation 46 $\tilde{u}(0) = 2 \cdot u(x_0)$ Then, the decision threshold is calculated by:

$$y^{*} = k_{1-\alpha} \cdot w \cdot \sqrt{2} \cdot u(x_0) \tag{47}$$

If no further information on the measurement procedure and on $\tilde{u}(\tilde{y})$ is given, the detection limit can only be calculated using the interpolation formula (equation 20) and obtains an explicit formula for the detection limit according to equations 21 and 22. Note, however, that this explicit formula for the detection limit is only an approximation which works best if $y \approx\geq 2y^{*}$.

8.4 Further and Future Applications: Beyond ISO 11929

The procedures described in this paper can also be applied to other measurements. Further applications of the approach presented in this paper are described elsewhere: measurements with ratemeters [17, 18, 32], albedo dosimeters [33], counting measurements on moving objects [17, 18], repeated counting measurements with random influences [17], counting measurements on filters during accumulation of radioactive materials [17], alpha spectrometry [16], spectrometric measurements [17, 18, 34], unfolding in spectrometric measurements [17, 34]. Numerical examples are given in refs. [17, 30].

The constraints $\mathrm{E}_0(\tilde{y}) = y$ and $\mathrm{Var}_0(\tilde{y}) = u^2(y)$ of $f_0(\tilde{y}|y)$ used in solving Equation 9 is scientifically meaningful if the measurement is unbiased and all components of the uncertainty $\tilde{y}$ results from the measurement process. If this is cannot be assumed or if the available information about y and $\tilde{y}$ does not justify these constraints, other types of PDFs have to be used which can be handled using Monte Carlo techniques. This methodology is described in the GUM Supplement 1 [13]. Also in such cases characteristic limits can be consistently derived as shortly described below and given in detail elsewhere [28].

Assume a model $Y = G(\mathbf{X})$ according to equation 1 with the vector **X** of input quantities and the output quantity *Y*. Let $\Im_{\mathrm{prior}}$ be any prior information, e.g. that the measurand is non-negative. According to the GUM Supplement [13] one has to establish the PDFs $f_{\mathbf{X}}(\mathbf{X})$ taking into account the measurement results **x** for **X** as well as all further information available on the behavior of **X**. By performing an experiment one obtains the vector of measurement results $\mathbf{x} = \mathrm{E}(\mathbf{X})$ together with its covariance matrix $\mathbf{U_x} = \mathrm{Cov}(\mathbf{X})$ of $f_{\mathbf{X}}(\mathbf{X})$ without the prior information. Then, the Markov formula

$$f_Y(Y|\mathbf{X}) = C \cdot \int f_{\mathbf{X}}(\mathbf{X}) \cdot \delta(Y - G(\mathbf{X}))\, \mathrm{d}\mathbf{X} \qquad (48)$$

allows calculating the likelihood PDF $f_Y(Y|\mathbf{X})$ and one obtains $y = \mathrm{E}(Y)$ and $u^2(y) = \mathrm{Var}(Y)$ of $f_Y(Y|\mathbf{X})$. From $f_Y(Y|\mathbf{X})$ the decision threshold is calculated according to its definition in equation 14.

The posterior PDF $f(Y|\mathbf{X},\Im_{\text{prior}})$ is given with equation 7 by

$$f(Y|\mathbf{X},\Im_{\text{prior}}) = C \cdot f(Y|\Im_{\text{prior}}) \cdot f(Y|\mathbf{X}) \tag{49}$$

Using the already obtained decision threshold and the definition of the detection limit according to equation 17, the detection limit is derived from further Monte Carlo simulations, e.g. by a numerical evaluation of the probability β of the error of 2nd kind as a function of the true value $\tilde{y} = y^{\#}$. The limits of the coverage interval $y^{\triangleleft}$ and $y^{\triangleright}$ are defined consistently with equation 23 and can be calculated from the posterior PDF $f(Y|\mathbf{X},\Im_{\text{prior}})$. Finally, the best estimate $\hat{y}$ and its associated standard uncertainty $u^2(\hat{y})$ are calculated as $\hat{y} = \mathrm{E}(Y)$ and $u^2(\hat{y}) = \mathrm{Var}(Y)$ of the posterior PDF $f(Y|\mathbf{X},\Im_{\text{prior}})$. A detailed publication on these future applications can be found elsewhere [28].

9 Conclusions

- With the GUM there exists an internationally accepted, standardized procedure for the determination of measurement uncertainties.
- Bayesian statistics provides the methodological basis for the GUM which allows taking into account both Type A and Type B uncertainties in a consistent way.
- With standard uncertainties according to the GUM, characteristic limits can be calculated for any measurement procedure according to ISO 11929-7 using Bayesian statistics.
- The procedures described in this paper and in refs. [17, 18, 29, 30] provide a basis for the revision of ISO 11929 Parts 1 to 4. With the revised standard ISO 11929, a consistent standardization of the calculation of characteristic limits will be provided, covering an extremely wide range of models and applications.

- With the Supplement to the GUM, the range of applications of the GUM methodology will be widely broadened using Monte Carlo techniques. Characteristic limits can be consistently derived also for such future applications.

References

[1] ISO Guide to the Expression of Uncertainty in Measurement. (Geneva: ISO) (1993), corrected reprint (1995).

[2] EURACHEM/CITAC; Guide CG 4, Quantifying uncertainty in analytical measurement. 2nd edition. (2000), http://www.eurachem.ul.pt/guides/QUAM2000-1.pdf .

[3] IAEA; Quantifying uncertainty in nuclear analytical measurements, IAEA Tecdoc 1401, July 2004, IAEA Vienna.

[4] ISO 11929-1 Determination of the detection limit and decision threshold for ionizing radiation measurements – Part 1: Fundamentals and application to counting measurements without the influence of sample treatment. (Geneva, ISO) (2000).

[5] ISO 11929-2 Determination of the detection limit and decision threshold for ionizing radiation measurements – Part 2: Fundamentals and applications to counting measurements with the influence of sample treatment. (Geneva, ISO) (2000).

[6] ISO 11929-3 Determination of the detection limit and decision threshold for ionizing radiation measurements – Part 3: Fundamentals and application to counting measurements by high resolution gamma spectrometry without the influence of sample treatment. (Geneva, ISO) (2000).

[7] ISO 11929-4 Determination of the detection limit and decision threshold for ionizing radiation measurements – Part 4: Fundamentals and application to measurements by use of linear-scale analogue ratemeters, without the influence of sample treatment. (Geneva, ISO) (2001).

[8] ISO 11929-5 Determination of the detection limit and decision threshold for ionizing radiation measurements – Part 5: Fundamentals and applications to counting measurements on filters during accumulation of radioactive material. (Geneva, ISO) (2005).

[9] ISO 11929-6 Determination of the detection limit and decision threshold for ionizing radiation measurements – Part 6: Fundamentals and applications to measurements by use of transient mode. (Geneva, ISO) (2005).

[10] ISO 11929-7 Determination of the detection limit and decision threshold for ionizing radiation measurements – Part 7: Fundamentals and general applications. (Geneva, ISO) (2005).

[11] ISO 11929-8 Determination of the detection limit and decision threshold for ionizing radiation measurements – Part 8: Fundamentals and application to unfolding of spectrometric measurements without the influence of sample treatment. (Geneva, ISO) (2005).

[12] ISO 11843-1 Capability of detection - Part 1: Terms and definitions, (Geneva, ISO) (1997).

[13] ISO Guide to the expression of uncertainty in measurement (GUM) - Supplement 1: Propagation of distributions using a Monte Carlo method, ISO; Geneve, 2008.

[14] International Vocabulary of Basic and General Terms in Metrology (VIM) — Third Edition, ISO/IEC (2007).

[15] Weise, K. and Wöger, W.; A Bayesian theory of measurement uncertainty. Meas. Sci. Technol. 4, 1-11 (1993).

[16] Weise, K. and Wöger, W.; Meßunsicherheit und Meßdatenauswertung. (Berlin: Wiley-VCH) (1999).

[17] Weise, K., Hübel, K., Michel, R., Rose, E., Schläger, M., Schrammel, D., Täschner, M.; Determination of the detection limit and decision threshold for ionizing-radiation measurements: fundamentals and particular applications. (Köln: TÜV Verlag Rheinland) (2005) ISBN 3-8249-0945-6.

[18] Weise, K. Hübel, K., Rose, E., Schläger, M., Schrammel, D., Täschner, M., Michel R.; Bayesian decision threshold, detection limit and confidence limits in ionising-radiation measurement, Radiat. Prot. Dosimetry 121(1) (2006) 52 – 63; doi:10.1093/rpd/ncl095; RPD Advance Access published July 25, 2006.

[19] Berger, J. O.: 1985, Statistical Decision Theory and Bayesian Analysis. (New York, Springer)

[20] Lee, P.M.; Bayesian Statistics: An Introduction. (Oxford University Press, New York) (1989).

[21] Bernardo, J. M., Smith, A.F.M.; Bayesian Theory. (Chicester, UK: Wiley) (1994).

[22] Bernardo, J.M.; Bayesian Statistics, in: Probability and Statistics (R.Viertl, ed.) of the Encyclopedia of Life Support Systems (EOLSS). (Oxford, UK, UNESCO) (2003).

[23] Jaynes, E.T.; Probability Theory - The Logic of Science, edited by G.L. Bretthorst. (Cambridge University Press, Cambridge) (2003).

[24] Robert, Ch.P.; The Bayesian Choice (2nd. ed.). (New York: Springer) (2001).

[25] Gelman, A., Carlin, J. B., Stern, H. and Rubin, D.B.; Bayesian Data Analysis (2nd edition). (London, Chapman) (2003).

[26] Gregory, P.C.; Bayesian logical data analysis for the physical sciences. (Cambridge University Press, Cambridge) (2005).

[27] Jaynes, E.T.; Papers on probability, statistics and statistical physics. ed. R.D. Rosenkrantz, (Boston, New York, Dordrecht: Kluwe Publ.) (1989).

[28] Weise, K., Kanisch, G., Michel, R., Schläger, M., Schrammel, D., Täschner, M.; Monte Carlo determination of the characteristic limits in measurement of ionising radiation: Fundamentals and numerics, Radiation Protection Dosimetry 135 No. 3 (2009) 169 – 196; doi:10.1093/rpd/ncp105.

[29] Weise, K.; Bayesian-statistical decision threshold, detection limit, and confidence interval in nuclear radiation measurement. Kerntechnik 63, 214 – 224 (1997).

[30] Michel, R.; Quality assurance of nuclear analytical techniques based on Bayesian characteristic limits. J. Radioanal. Chem. 245, 137 – 144 (2000).

[31] Abramowitz, M., Stegun, I.; Handbook of Mathematical Functions. 5th ed., Chap. 26, Dover Publications (New York) (1968).

[32] Weise, K.; The Bayesian count rate probability distribution in measurement of ionizing radiation by use of a ratemeter. PTB Report Ra-44. (Braunschweig: Physikalisch-Technische Bundesanstalt) (2004).

[33] DIN 25482-11 Detection limit and decision threshold for ionizing radiation measurements – Part 11: Measurements using albedo dosimeters. (in German) (Berlin: Beuth-Verlag) (2003).

[34] Weise, K., Michel, R.; Erkennungsgrenze, Nachweisgrenze und Vertrauensbereich in der allgemeinen Kernstrahlungs-Spektrometrie, Kerntechnik 60 No. 4, 189 – 196 (1995).