

# Revisiting Currie's Minimum Detectable Activity for Non-Destructive Assay By Gamma Detection Using Tolerance Intervals

E. Agboraw<sup>1</sup>, E. Bonner<sup>1</sup>, T. Burr<sup>1</sup>, S. Croft<sup>2</sup>, J.M. Kirkpatrick<sup>3</sup>, T. Krieger<sup>4</sup>, C. Norman<sup>1</sup>, P. Santi<sup>1</sup>, S. Walsh<sup>1</sup>

<sup>1</sup> International Atomic Energy Agency, Vienna Austria;

<sup>2</sup> Oak Ridge National Laboratory, USA ;

<sup>3</sup> Mirion Technologies (Canberra), Inc., USA

<sup>4</sup> Forschungszentrum Jülich GmbH, Inst. Energy and Climate Research, Jülich, Germany

## Abstract

Currie's paper [1] on estimating the minimum detectable activity (MDA) applied a Gaussian approximation to either Gaussian or Poisson data and remains the standard method to estimate radiological detection limits. This paper revisits the Currie method with attention to the false alarm probability (FAP) in Poisson and Gaussian data in non-destructive assay (NDA) by gamma (denoted as  $\gamma$ ) detection. The Currie detection limit  $L_D$  is an estimate of the smallest net signal count rate  $\lambda_N$  that can be detected with high probability and low FAP in the presence of non-zero background count rate  $\lambda_B$  that has been previously estimated. The MDA is the sample activity or mass corresponding to  $\lambda_N$ , defined as  $MDA = \frac{L_D}{\nu}$ , where in the case of  $\gamma$ -based NDA, the calibration factor  $\nu$  (a product of  $\gamma$ -ray yield, detector and geometric efficiency, counting time, and other factors) has measurement error that introduces systematic error in the estimate of the MDA. Kirkpatrick et al. [2] showed how to account for systematic uncertainties in the estimate of  $MDA = \frac{L_D}{\nu}$  using a modified version of Currie estimation [2,3]. The present paper combines the approach in [2] with a tolerance interval approach. It is shown that the FAP in signal detection can be significantly different from the nominal FAP if the nominal FAP is not based on a tolerance interval, and if the nominal FAP is based on a tolerance interval, then the MDA will be larger than Currie's estimated MDA.

## 1. Introduction

The Currie detection limit  $L_D$  is an estimate of the smallest net signal count rate  $\lambda_N$  that can be reliably detected with low FAP in the presence of non-zero background count rate  $\lambda_B$  [1]. The MDA is the sample activity (or mass through a conversion) corresponding to  $\lambda_N$ , defined as  $MDA = \frac{L_D}{\nu}$ , where the calibration factor  $\nu$  (a product of  $\gamma$ -ray yield, detector and geometric efficiency, counting time, and other factors) has measurement error that can introduce systematic error in the estimate of the MDA. Kirkpatrick et al. [2] showed how to account for such systematic uncertainties in the estimate of  $MDA = \frac{L_D}{\nu}$  using a modified version of the Currie estimation [2,3]. The MDA can be used prior to data

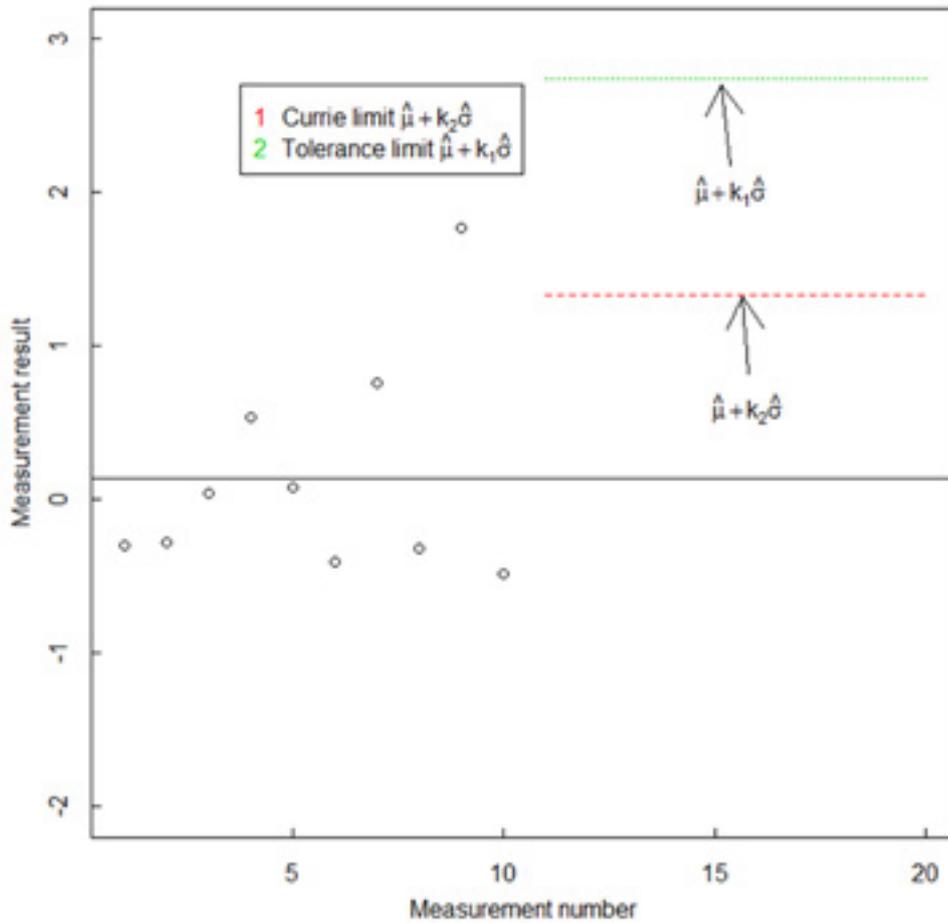
collection to compare different instruments and measurement scenarios, and can also be used as a quantitative measure on an item-specific basis after data collection. In  $\gamma$ -ray spectroscopy, the background is often estimated from the continuum beneath the peak(s) of interest, so the MDA is specific to the measurement conditions (including what other nuclides are present).

This paper revisits  $L_D$  with attention to the FAP (denoted  $\alpha$ ) in Poisson and Gaussian data, by using a tolerance interval approach [4,5]. Section 2 provides background, motivation, and example tolerance intervals. Section 3 provides a simulation approach and results for both Gaussian and Poisson data. Section 4 uses results from Section 3 to estimate the MDA while allowing for random and systematic errors in the calibration factor  $\nu$  in  $MDA = \frac{L_D}{\nu}$ . Section 5 is a discussion. Section 6 is a summary.

## 2. Background

Currie [1] provided approximate MDA calculations for the desired FAP based on the assumption that the measurement data has a Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ , denoted  $X \sim N(\mu, \sigma^2)$ . In  $\gamma$ -ray spectroscopy, the measurement data are  $\gamma$ -ray counts at certain energies, which are often well modeled with a Poisson distribution, which for a large mean count rate is well approximated by a Gaussian distribution. Because  $\mu$  and  $\sigma$  must be estimated, the well-known frequentist approach to a confidence interval for  $\mu$  from  $n$  measurements  $x_1, x_2, \dots, x_n$  is  $\bar{x} \pm t_{1-\alpha/(n-1)} s / \sqrt{n} = \hat{\mu} \pm t_{1-\alpha/(n-1)} \hat{\sigma} / \sqrt{n}$ , where  $t_{1-\alpha/(n-1)}$  denotes the  $(1-\alpha)$  quantile of the  $t$  distribution with  $n-1$  degrees of freedom,  $\bar{x} = \sum_{i=1}^n x_i / n$ , and  $s^2 = \hat{\sigma}^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)$  [1,4,5].

In nuclear safeguards (Sections 3 and 4), background measurements are often used to estimate an alarm threshold that has a small nominal  $\alpha$ , such as  $\alpha = 0.05$ . So, instead of requiring a confidence interval for  $\mu$ , the need is to estimate a threshold (the 0.95 quantile of the distribution of  $X$ ), denoted  $T_{0.95}$ , that corresponds to  $\alpha = 0.05$ . The threshold  $T_{0.95}$  is the upper limit of a one-sided interval of the distribution of  $X$  if doing one-sided testing for a positive mean shift. In contrast to a confidence interval, a tolerance interval is an interval that bounds a fraction of a probability



**Fig. 1:** Illustration of the tolerance limit  $k_1 = 3.7$  compared to the Currie limit  $k_2 = 1.7$  for future data in the case of using 10 Gaussian observations to estimate  $\mu$  and  $\sigma$  and the corresponding Gaussian quantile,  $\hat{T}_{0.95} = \hat{\mu} + k\hat{\sigma}$ .

distribution with a specified confidence (frequentist) or probability (Bayesian approach) [4,5]. Both frequentist and Bayesian tolerance interval approaches will be presented in this paper. The frequentist tolerance interval estimators presented have the form  $\hat{T}_{0.95} = \hat{\mu} + k\hat{\sigma}$ , where  $k$  is the coverage factor that depends on  $n$ . The goal in both the frequentist and Bayesian approaches is to achieve  $P(\hat{T}_{0.95} \geq T_{0.95}) = p$ , where  $p$  is a user-specified probability (the frequentist confidence level), such as  $p = 0.99$  [4,5]. In the Bayesian approach,  $\mu$  and  $\sigma$  are random unknown parameters so  $P(\hat{T}_{0.95} \geq T_{0.95}) = P_{\mu,\sigma}(\hat{T}_{0.95} \geq T_{0.95})$  is computed with respect to  $\mu$  and  $\sigma$ . In the frequentist approach,  $\hat{\mu}$  and  $\hat{\sigma}$  are random while  $\mu$  and  $\sigma$  are fixed unknowns so  $P(\hat{T}_{0.95} \geq T_{0.95}) = P_{X_1, X_2, \dots, X_n}(\hat{T}_{0.95} \geq T_{0.95})$  is computed with respect to random samples of size  $n$ .

In any frequentist approach, probabilities such as  $\alpha$  are calculated with respect to the distribution of  $X$  for fixed  $\mu$  and  $\sigma$ . A frequentist tolerance interval has an associated confidence, which is the long-run relative frequency (probability) that an interval such as  $(0, \hat{T}_{0.95} = \hat{\mu} + k\hat{\sigma})$  will include a future observation  $X$  from the same distribution as the training data used to estimate  $\hat{\mu}$  and  $\hat{\sigma}$ . In any Bayesian approach, probabilities are calculated with respect to the joint posterior distribution  $f_{posterior}(\mu, \sigma)$  for fixed  $X$  [5].

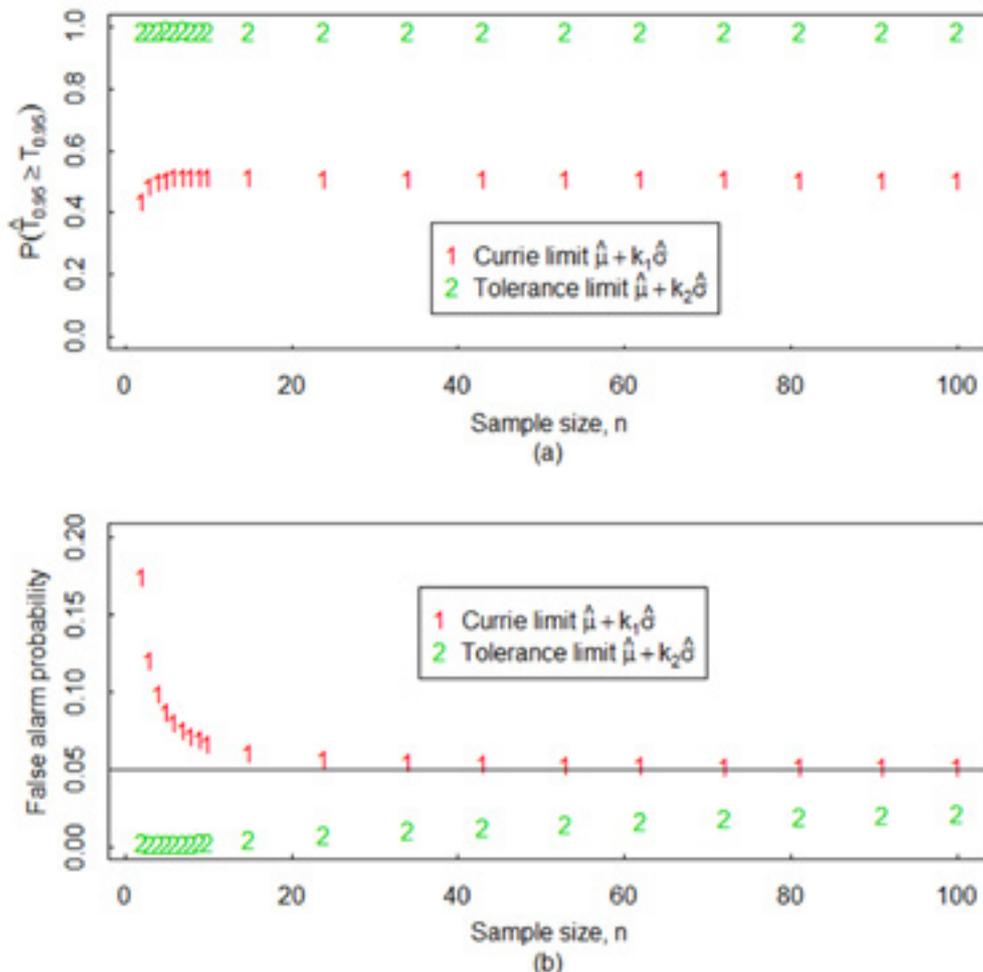
To illustrate the frequentist approach, assume that  $n = 10$  measurements are used to construct an upper limit that bounds at least  $p = 0.95$  ( $\alpha \leq 0.05$ ) of future data with probability  $p = 0.99$ . Fig. 1 plots a single realization of the  $n = 10$  measurements and compares the Currie limit to the tolerance interval limit. To achieve a user-specified  $\alpha$  for future measurements aimed to detect whether any signal is present in a background measurement, Currie [1] used the detection threshold  $\hat{T} = \hat{\mu}_B + k_{1-\alpha}\hat{\sigma}_B$  where  $k_{1-\alpha}$  is the  $(1-\alpha)$  quantile of the Gaussian distribution, and  $\hat{\sigma}_B = \sqrt{\hat{\sigma}^2 + \hat{\sigma}^2/n}$ , and the term  $\hat{\sigma}^2/n$  is the estimated variance of the estimate of the unknown mean  $\mu_B$ . Regarding notation, in this paper, the subscript B denotes background, and the subscript N denotes net, and both the B and N subscripts will sometimes be omitted, depending on the context, to avoid cluttering the notation. Currie regarded this value of  $\hat{T}$  as an approximate value if the underlying data is non-Gaussian (such as Poisson; see Section 3). If one uses  $k_{1-\alpha} = t_{n-1}(\delta)/\sqrt{n}$  instead of the  $(1-\alpha)$  quantile of the Gaussian distribution in Currie's calculation  $\hat{T} = \hat{\mu}_B + k_{1-\alpha}\hat{\sigma}_B$ , with noncentrality parameter  $\delta = z_p\sqrt{n}$  where  $z_p$  is the  $1-p$  quantile of the standard Gaussian, then the calculation is exact if the underlying data has a Gaussian distribution [4].

Perhaps surprisingly, an exact expression for a tolerance interval is only available in the one-sided Gaussian case just described [4-7]. However, good approximate expressions for many other cases are available [5-7]. Alternatively, and in the approach taken in this paper, tolerance intervals can be well estimated using simulation to approximate an alarm threshold that is designed to contain at least  $1 - \alpha$  percent of future observations with a specified coverage probability  $p$ . Currie did not consider the probability  $p$  and note from Fig. 1, that for  $p = 0.99$ , the decision limit is much larger than Currie's limit, with  $k_1 = 3.7$  (tolerance) versus  $k_2 = 1.7$  (Currie). As shown in Section 3, using the value  $k_1 = 3.7$  corresponds to  $p = 0.99 = P(\hat{T}_{0.95} \geq T_{0.95})$ , while using  $k_2 = 1.7$  gives  $p = 0.52$ .

Fig. 2 plots  $P(\hat{T}_{0.95} \geq T_{0.95})$  (Fig. 2a) and the true average FAP (Fig. 2b) for a range of sample sizes  $n$  if the data is Gaussian for both the tolerance method (using  $p = 0.99$ ) and the Currie method. The tolerance value for  $k_1$  (which depends on  $n$ ),  $P(\hat{T}_{0.95} \geq T_{0.95})$ , and the true FAP are

easily calculated using simulation in R [8] as shown in Section 3.

A Bayesian analysis specifies a prior probability for parameter(s)  $\theta$ , a likelihood (such as Gaussian or Poisson in this paper)  $P(X | q)$ , and then finds the posterior distribution of  $\theta$ ,  $f_{posterior}(\theta)$ . Bayesian tolerance interval construction then finds an estimate  $\hat{T}_{1-\alpha}$  such that  $P(X < \hat{T}_{1-\alpha} | \theta) = 1 - \alpha$  with specified coverage probability  $p$ . In the Gaussian case with unknown  $\mu$  and  $\sigma$ ,  $\theta = (\mu, \sigma)$ . In the Poisson case,  $\theta = (\lambda_G, \lambda_B)$  if both a gross count rate and background count rate are required, and  $\theta = (\lambda)$  if the count rate at a single region of interest is required. For Gaussian and Poisson data, conjugate prior pdfs are available, which have the convenient property that the posterior pdf is in the same family as the prior, but with updated parameters. For example, the conjugate prior for the Gaussian with unknown  $\mu$  and  $\sigma$  is the Gaussian-inverse-Gamma and the conjugate prior for the Poisson is the Gamma distribution [5].



**Fig. 2:** The true value of  $P(\hat{T}_{0.95} \geq T_{0.95})$  in (a) and true FAP in (b) if data is Gaussian. The tolerance interval method is conservative, so has a FAP that is smaller than 0.05 by construction. The Currie method has FAP much larger than 0.05 for small sample sizes.

### 3. Simulation to estimate $\hat{T}_{0.95}$

The user seeks  $\hat{T}_{1-\alpha}$  such that  $P(X < \hat{T}_{1-\alpha} | \theta) = 1 - \alpha$  with specified coverage probability  $p$ .

#### 3.1 A simulation-based trial-and-error frequentist approach for Gaussian data

The simulation-based trial-and-error frequentist approach to estimate  $k$  is as follows.

1. Specify  $n, \mu, \sigma$ .
2. For each of many (typically  $10^5$  or more) simulations, generate  $X_i \sim N(\mu, \sigma^2)$ , for  $i = 1, 2, \dots, n$
3. Compute  $\hat{\mu} + k\hat{\sigma}$  for a grid of trial values for  $k$  using 
$$\hat{\mu} = \bar{x} = \sum_{i=1}^n x_i / n, \quad \hat{\sigma} = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)}.$$
4. Select the trial value of  $k$  that includes at least 95% of the population (of future  $X$  values) with probability  $p = 0.99$ ; that is,  $P(\hat{T}_{0.95} \geq T_{0.95}) = 0.99$ , where  $\hat{T}_{0.95} = \hat{\mu} + k\hat{\sigma}$ .

For example, with  $n = 10$  and for any values of  $\mu$  and  $\sigma$ , the exact result is  $k = 3.738$ , and the simulation-based result in R [8] is  $k = 3.74$ , which is within the small simulation error in a large but finite number ( $10^5$ ) or simulations. Similarly, simulation can also estimate the probability that the Currie-based  $k$  value bounds at least 95% of the probability density function (pdf) of  $X$  (so the FAP is 0.05 or less), and in this example with  $n = 10$ , there is a probability of approximately 0.52 that the Currie-based value of  $k$  has a FAP of 0.05 or less.

One nuclear safeguards application for tolerance intervals for Gaussian data is inspector (i) measurements of operator (o) declarations of  $n$  items sampled for verification. In each of  $n$  values of the operator-inspector difference statistic  $d_j = (o_j - i_j)/o_j$ , if  $|d_j| > k\delta$  (in two-sided testing), then the  $j$ -th item selected for verification leads to an alarm, where  $\delta_T = \sqrt{\delta_R^2 + \delta_S^2}$  (with  $\delta_T$  the total RSD,  $\delta_S$  the between-period short-term systematic error RSD, and  $\delta_R$  the within-period reproducibility) and  $k = 3$  is a common choice that corresponds to a small  $\alpha$  of approximately 0.001. The null hypothesis is  $\mu = 0$ , and  $\delta_T$  can be estimated by applying analysis of variance (ANOVA) [9-12]. If one assumes  $\hat{\delta}_T = \delta_T$  then choosing  $k = 1.65$  corresponds to  $\alpha = 0.05$  (Gaussian approximation); however, as an example, if  $n = 10$  paired measurements in each of 3 prior inspection periods are available, and  $\delta_S = \delta_R = 0.03$  [9 12], then choosing  $k = 1.65$  leads to an actual FAP of 0.05 or less with probability 0.38. If one desires a high probability  $p = 0.99$  that the actual FAP is as small as the nominal FAP, then simulation [9,12] indicates that instead of  $k = 1.65$ , one must choose, for example,  $k = 2.58$  for 5 groups of 10 measurements,  $k = 2.94$  for 3 groups of 10 measurements and  $k = 4.35$  for 2 groups of 5 measurements. Unlike the single-component Gaussian case, these values of  $k$  depend on the values of the ratio  $\delta_S/\delta_R$ , which is unknown, so approximate frequentist or Bayesian methods

are needed. Note that any Bayesian method can be regarded as approximate because one almost never knows the exact prior probability distribution. The accuracy of these approximate methods can be assessed using simulation and/or by analysis of historical data.

#### 3.2 Poisson data

Fig. 3 shows that the true FAP of Currie's method can be quite different from the nominal FAP, so tolerance interval construction should be considered. In Fig. 3, the simulated data is  $n = 1$  observation of  $X \sim \text{Poisson}(\lambda)$ , with  $\lambda = 1, 10$ , or 100. A count time of  $t = 1$  second is used to estimate the background and to test whether a subsequent measurement corresponds to the same background rate  $\lambda$  (See Section 5.1). For comparison,  $P(X_{test}/t = \hat{\lambda}_{test} > T)$  of the corresponding Gaussian distribution is shown, where  $\hat{T}_{0.95} = \hat{\lambda} + k\sqrt{\hat{\lambda}(1+1/n)/t}$ , which is Currie's [1] approach to estimate  $T$  by using the Gaussian approximation for both Gaussian and Poisson data, and using the factor  $\sqrt{1+1/n}$  to quantify the impact of uncertainty in the estimated mean on the estimated background standard deviation. Note (Fig. 3b) that for large values of  $\lambda$  (and/or large count times) such as  $\lambda \geq 100$ , then the Gaussian approximation (with the factor  $\sqrt{1+1/n}$  but without the notion of a tolerance interval) to the Poisson is adequate. The reason for this good accuracy is that the variance of the Poisson distribution is equal to its mean  $\lambda$ , so the Poisson standard deviation can be estimated with less uncertainty than that of the Gaussian.

Recall from Example 3.1 that estimating the standard deviation of the Gaussian requires  $n > 1$ , and that the notion of tolerance intervals is needed; the estimated threshold  $T$  is much too small if uncertainties in  $\hat{\mu}, \hat{\sigma}$  are not accounted for properly. Without using tolerance intervals, references [2, 13-14] extended Currie's treatment of Poisson data [1] by using the Poisson distribution rather than an approximating Gaussian. Particularly when count rates and/or count times are small, it is prudent to use the Poisson distribution rather than the approximating Gaussian. As an example (also used in Section 4), let  $n = 5, \mu = 10$ , and  $x_1, x_2, \dots, x_5$  are 10,12,10,10,8, so  $\bar{x} = 10$  and Currie's  $\hat{T}_{0.95} = \hat{\mu}_B + k_{1-\alpha}\hat{\sigma}_B = 12.5$ , which is rounded up to 13 (and in 93% of  $10^5$  simulations, test measurements exceed the 13 limit, so the FAP can be much larger than 0.05). In the same example, a one-sided tolerance interval using the R code in Section 3.3 below leads to  $\hat{T} = 21.5$ , rounded up to 22 for 99% confidence that the FAP is 0.05 or smaller. Also for the same example, a Bayesian tolerance interval approach is illustrated in Section 3.3 using a prior probability density  $f_{prior}(\lambda) = \text{Gamma}(\alpha_{prior} = 1, \beta_{prior} = 0.75)$  (the conjugate prior for the Poisson, and this particular prior has mean  $\alpha_{prior}/\beta_{prior} = 1/0.75 = 13.3$  and standard deviation  $\sqrt{\alpha/\beta^2} = 1/0.75 = 13.3$ ) has  $f_{posterior}(\lambda) = \text{Gamma}(\alpha_{prior} + \sum_{i=1}^n x_i, \beta_{prior} + n) = \text{Gamma}(1+50, 0.75+5)$ , which has

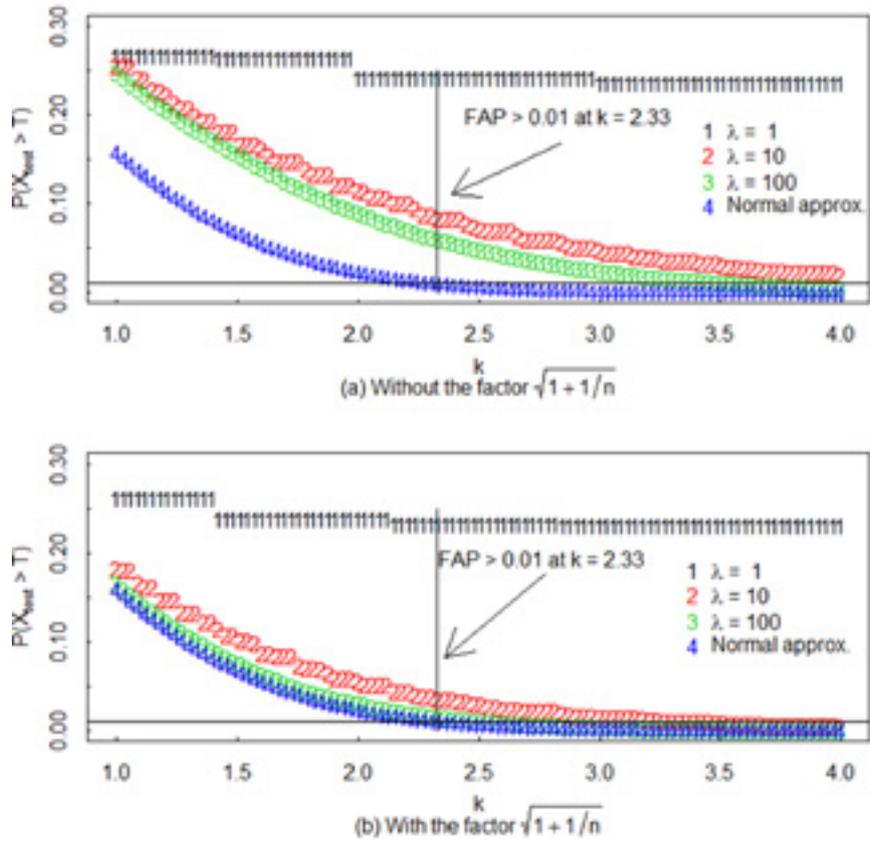


Fig. 3: The probability  $P(X_{test}/t = \hat{\lambda}_{test} > T)$  versus  $k$  for  $\lambda = 1, 10,$  and  $100$ . The normal approximation is also plotted. Currie's factor  $\sqrt{1+1/n} = \sqrt{2}$  is ignored in (a), included in (b).

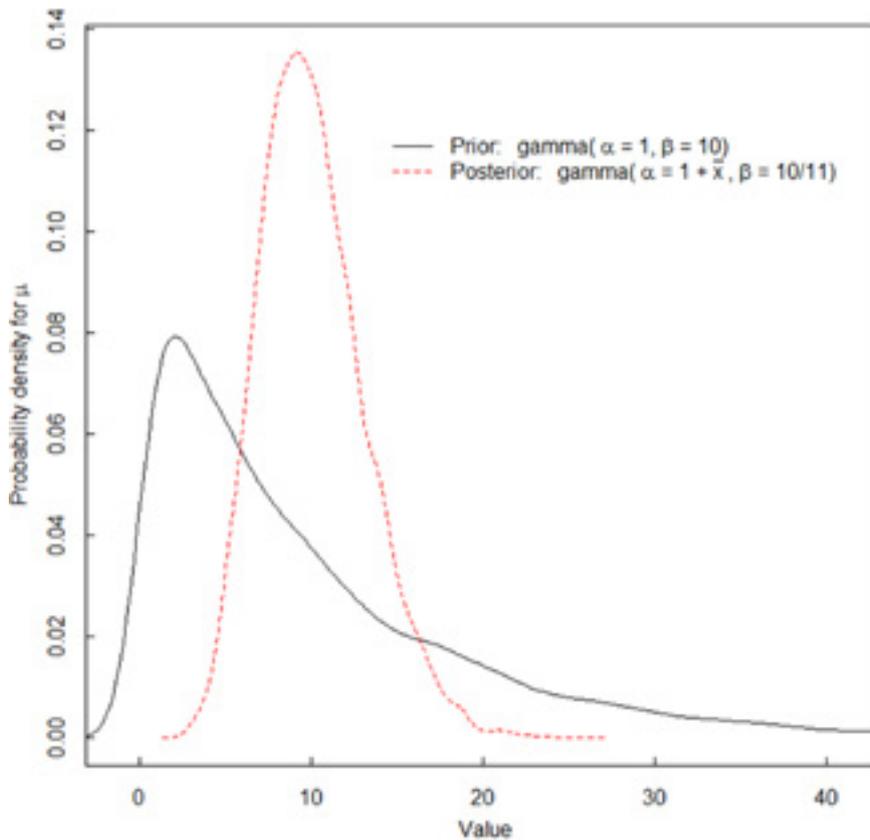


Fig. 4: The prior and posterior distribution for  $\lambda$  for  $n = 5, \mu = 10,$  and  $\bar{x} = 10$ .

mean 10.05 and standard deviation 1.41; see Fig. 4. Note that the Gamma parameters, conventionally denoted as  $\alpha_{prior}$  and  $\beta_{prior}$  are not related to the FAP  $\alpha$  or the nondetection probability  $\beta$ .

### 3.3 Simulation for Poisson data for frequentist and Bayesian approaches

#### 3.3.1 Frequentist approach

1. Specify  $\lambda$  and  $n$ .
2. For each of many ( $10^5$  or more) simulations, generate  $X_i \sim Poisson(\lambda)$ ,  $i = 1, 2, \dots, n$ .
3. Compute  $\hat{\lambda} + k\sqrt{\hat{\lambda}/n}$  for a grid of trial values for  $k$  using  $\hat{\lambda} = x$ .
4. Select the trial value of  $k$  that includes at least 95% of the population (of future  $X$  values) with probability  $\gamma = 0.99$ ; that is,  $P(\hat{T}_{0.95} \geq T_{0.95}) = 0.99$ , where  $\hat{T}_{0.95} = \hat{\lambda} + k\sqrt{\hat{\lambda}/n}$ .

With  $n = 5$ ,  $\lambda = 10$ , the Currie approximation is  $\hat{T}_{0.95} = 12.5$  and the exact value using simulation (to within negligible simulation error) is  $\hat{T}_{0.95} = 21.5$ . The probability that the FAP is 0.05 or less is only 0.07 with the Currie value and is, by design, 0.99 with the simulation approach. Unlike with Gaussian data, for Poisson data, the value of  $k$  depends on  $\lambda$ , so  $\lambda$  must be replaced with  $\hat{\lambda}$ .

#### 3.3.2 Bayesian approach

1. Specify  $n$  and the parameters of the Gamma prior  $\alpha_{prior}$  and  $\beta_{prior}$ . In this example  $\alpha_{prior} = 1$  and  $\beta_{prior} = 0.075$  (a very wide prior with mean and standard deviation of 13.3).
2. For each of many (typically  $10^5$  or more) simulations, generate  $\lambda \sim Gamma(\alpha_{prior}, \beta_{prior})$  and  $X_i \sim Poisson(\lambda)$ ,  $i = 1, 2, \dots, n$ .
3. Compute  $\alpha_{post} = \sum_{i=1}^n X_i + \alpha_{prior}$  and  $\beta_{post} = n + \beta_{prior}$ .
4. Choose the quantile of the posterior  $\lambda_{post} \sim Gamma(\alpha_{post}, \beta_{post})$  such that  $P(\hat{T}_{0.95} \geq T_{0.95}) = \rho = 0.99$ . This is the count value that is greater than 95% of the distribution of  $X$  for 99% of the  $\lambda$  values generated in the simulations.

The Bayesian result is  $\hat{T}_{0.95} = 23.4$  for the same Poisson example. Recall that Currie's value of  $\hat{T}_{0.95}$  is 13, the frequentist  $\hat{T}_{0.95}$  given above is 21.5, and all values of  $\hat{T}_{0.95}$  are approximations. The Bayesian estimate  $\hat{T}_{0.95}$  is approximate because there is always mismatch between the true and assumed prior. The frequentist estimate  $\hat{T}_{0.95}$  is approximate because it depends on the true value of  $\lambda$  so in practice, one uses  $\lambda = \hat{\lambda}$ . Currie's  $\hat{T}_{0.95}$  is approximate for the reasons given. Recall that the accuracy of these

approximate methods can be assessed using simulation and/or by analysis of historical data.

### 3.4 Example with two Poisson counts in each assay

Detection of  $g$  counts often requires measurement of both the nearby-in-energy "background" counts and the peak region "gross" counts (Section 5.1). The gross mean count rate is  $\lambda_G = \lambda_B + \lambda_N$  [2,13,14]. The Bayesian approach is effective in this context for two main reasons: a conjugate prior (Gamma) can be specified for  $\lambda_G$  and  $\lambda_B$ , so the measured G and B counts each lead to  $f_{posterior}(\lambda) = Gamma(\alpha_{prior} + \sum_{i=1}^n X_i, \beta_{prior} + n)$ , and it is simple to enforce  $\lambda_N \geq 0$ . Although the choice of prior parameters  $\alpha$  and  $\beta$  for both  $\lambda_G$  and  $\lambda_B$  is subjective, the user often can bound the range for both  $\lambda_G$  and  $\lambda_B$  from prior data, so  $\alpha_{prior}$  and  $\beta_{prior}$  can each be within some modest range. If the Bayesian approach is applied repeatedly, its long-run behavior can be evaluated to check, for example, whether the nominal FAP is close to the actual FAP.

To illustrate, choose  $\alpha_{prior} = 1$  and  $\beta_{prior} = 0.075$  for  $\lambda_G$  and  $\lambda_B$  as in the previous Bayesian example for Poisson data. Generate  $G \sim Poisson(\lambda_G)$  and  $B \sim Poisson(\lambda_B)$ . For each of many ( $10^5$  or more) simulations, generate  $\lambda_G$  from its posterior  $Gamma(1+G, 0.075+1)$  and generate  $\lambda_B$  from its posterior  $Gamma(1+B, 0.075+1)$  and for those simulations for which  $\lambda_G \geq \lambda_B$  (because  $\lambda_N \geq 0$ ), compute  $G - B$ . Determine the threshold  $T$  for  $G - B$  such that with probability at least  $\rho = 0.99$ ,  $P(G - B \geq T) \leq 0.05$ . The result for  $G = 30$  and  $B = 10$  is  $T = 34$  (Currie) and  $T = 45$  (Bayesian tolerance, using the Skellam distribution, which is the distribution of the difference in two Poisson random variables). Then  $L_D$  is an estimate of the smallest net signal count rate  $\lambda_N$  that can be detected with high probability and low FAP in the presence of nonzero background count rate  $\lambda_B$  that has been previously estimated. Ignoring errors in the calibration factor  $\nu$  (assuming  $\nu = \nu_{True}$  and for simplicity here also assuming  $\nu_{True} = 1$ ), the Currie-based MDA is 39 and the tolerance interval-based MDA is 77. Allowing for 5% RSD in the total error as in the previous example and assuming that  $\nu_{Meas} = \nu_{True}(1+S+R)$  has a Gaussian distribution (any distribution is simple to accommodate here), then the Currie-based MDA, which corresponds to the net count rate assuming zero external background (see Section 5.1), increases from 39 to 45 and the tolerance interval-based MDA increases from 77 to 87.

### 4. Implications for the MDA

Recall the Poisson example in Section 3.3 for which Currie's  $\hat{T}_{0.95} = \hat{\mu}_B + k_{1-\alpha} \hat{\sigma}_B = 12.5$  (which is rounded up to 13), and the one-sided tolerance limit is  $\hat{T}_{0.95} = 22$  for 99% confidence that the FAP is 0.05 or smaller. Therefore, the

estimated MDA based on the tolerance interval limit will be larger than the estimated MDA based on the Currie limit. Specifically, if the mean count rate shifts from  $\mu = 10$  to  $\mu = 19.5$  any future observation  $X \sim \text{Poisson}(\lambda = 19.5)$  satisfies  $P(X \geq 13) \geq 0.95$  for the Currie mean shift and  $X \sim \text{Poisson}(\lambda = 34.4)$  satisfies  $P(X \geq 22) \geq 0.95$  for the tolerance interval mean shift. The mean shift values  $\lambda = 19.5$  and  $\lambda = 34.4$  are easily computed by numerical search. The MDA is then calculated by converting the mean shift to an activity, which requires calibration.

Recall that the MDA is defined as  $MDA = \frac{L_D}{\nu}$ , where in this example  $L_D = 19.5$  (Currie) or  $L_D = 34.4$  (tolerance) and the calibration factor  $\nu$  (a product of  $\gamma$ -ray yield, detector and geometric efficiency, counting time, and other factors) has measurement error that can introduce systematic error in the estimate of the MDA. References [2,13,14] account for systematic uncertainties in the estimate of the MDA using a modified version of the Currie estimator [2,3].

To allow for random and/or systematic errors in  $\nu$ ,  $\nu_{Meas} = \nu_{True}(1 + S + R)$ , implies that the mean shift when the signal is present has uncertainty. To illustrate, assume that it is desired to have at least 99% confidence that the mean shift is above some limit. Assuming Gaussian calibration errors, then, for example, assuming 5% relative standard deviation (which is assumed here to include both random and systematic components) in converting the mean shift to activity using  $MDA = \frac{L_D}{\nu}$  increases the estimated mean shift that can be detected with high probability from 19.5 to 22.1 (Currie approximation) and from 34.4 to 38. (tolerance interval approximation).

## 5. Discussion

This section describes three additional topics related to MDA calculations.

### 5.1 Definition of the background

In some  $\gamma$ -based NDA applications, the challenge to define and measure the relevant background is important. For example, in attribute measurements of fresh fuel assemblies, one task is to assess whether a given assembly is a dummy (not containing  $^{235}\text{U}$ ). In this case, the background is defined as the response of the detector to  $\gamma$  emissions from neighboring assemblies if the assembly being measured were a dummy. That is, measurement behavior needs to be characterized if  $\gamma$  emissions could be measured from only the neighboring assemblies at the location of the assembly being measured. The measurement seeks to provide evidence that a signature from the item was detected (thereby verifying presence of  $^{235}\text{U}$ ) and that the measured signature originated from the item, not from radiation outside the item.

The minimum detectable quantity is not usually defined for attribute testing; however, it is sometimes desired (beyond the scope of this example) to estimate the probability that the test alarms for large mean shifts, such as a mean shift associated with 50% or more nuclear material missing.

Gamma-ray detectors detect distinct  $\gamma$ -rays energies. Sodium Iodide (NaI) and Cadmium-Zinc-Telluride (CZT) are common detector types. The presence of  $^{235}\text{U}$  is verified in fresh fuel by estimating the area in the peak region of interest (ROI) associated with the 185.7 keV  $\gamma$ -ray. If the estimated peak area exceeds 3 times its estimated standard deviation, then based on the acceptance criteria established by the IAEA corresponding to a 99.7% confidence level (assuming no estimation error in the estimated standard deviation, so tolerance interval concepts are not being used; see the final paragraph in section 5.1) in the presence of the peak, the peak is considered to be present in the spectrum and the presence of  $^{235}\text{U}$  is verified within the fuel. Because  $\gamma$ -rays at such energies interact with materials primarily through both the photoelectric effect (in which the  $\gamma$ -ray transfers all its energy to the detector medium) and Compton scattering (in which the  $\gamma$ -ray scatters off an electron in the medium or surrounding mediums causing a partial transfer of its energy to the detector medium), each measured peak in a  $\gamma$ -ray spectrum lies on top of a background caused by higher energy  $\gamma$ -rays that underwent Compton scattering within the detector. An example of this can be seen in Fig. 5 in which a  $\gamma$ -ray spectrum of a fresh fuel assembly as measured by a CZT detector is shown for  $\gamma$ -ray energies ranging from 20 keV to 305 keV. Because a fresh fuel assembly contains both  $^{235}\text{U}$  and  $^{238}\text{U}$ , and because the  $\gamma$ -rays that are associated with the decay of  $^{238}\text{U}$  exist at energies between 700 and 1001 keV, the 186 keV  $\gamma$ -ray photo peak from  $^{235}\text{U}$  will always be present on top of a Compton background associated with the scattering of  $\gamma$ -rays from  $^{238}\text{U}$  within the CZT detector.

In fresh fuel verification scenarios where shielding and collimation can be used to detect  $\gamma$ -rays only from the selected assembly, the peak area is estimated as the difference between the total counts in the ROI that includes the peak and the counts associated with the Compton background in that region (recall Example 3.3). To assist in determining whether the attribute test condition has been satisfied, most software programs automatically notify the inspector when the net area of the peak ROI above the Compton background is larger than 3 times its estimated standard deviation.

In cases where attribute measurements are performed near other items containing the same type of nuclear material, a background measurement is needed to estimate the peak's count rate as detected from the surrounding environment. The background-only measurement corresponding to the item-plus-background measurement in Fig. 5 had a similar spectrum shape as that in Fig. 5, but the peak ROI counts

were approximately 60% lower. For this background-only measurement, the same CZT detector that was used for item-plus-background was put in an empty slot of a storage rack containing fresh fuel assemblies. The background spectrum was measured during the same training exercise and for the same count time as the spectrum in Fig. 5, which was from an attribute test measurement of a fresh fuel assembly within that same storage rack. Assuming zero room background, applying the attribute test to the background spectrum would yield a positive (and incorrect) verification because the estimated peak area was approximately 13 times its estimated standard deviation. Therefore, to ensure proper verification of items using the attribute test, careful consideration must be given regarding how the room background is measured in order to reject the possibility that the measured spectrum was the result of room background and not from the item to be verified.

When the background spectrum shows evidence that the peak of interest is present from measuring the surrounding environment, the verification of an item using the attribute test is executed using

$$(\text{Measured rate}) - (\text{background rate}) > 3\sqrt{\left(\frac{\hat{\sigma}_B}{T_B}\right)^2 + \left(\frac{\hat{\sigma}_M}{T_M}\right)^2}$$

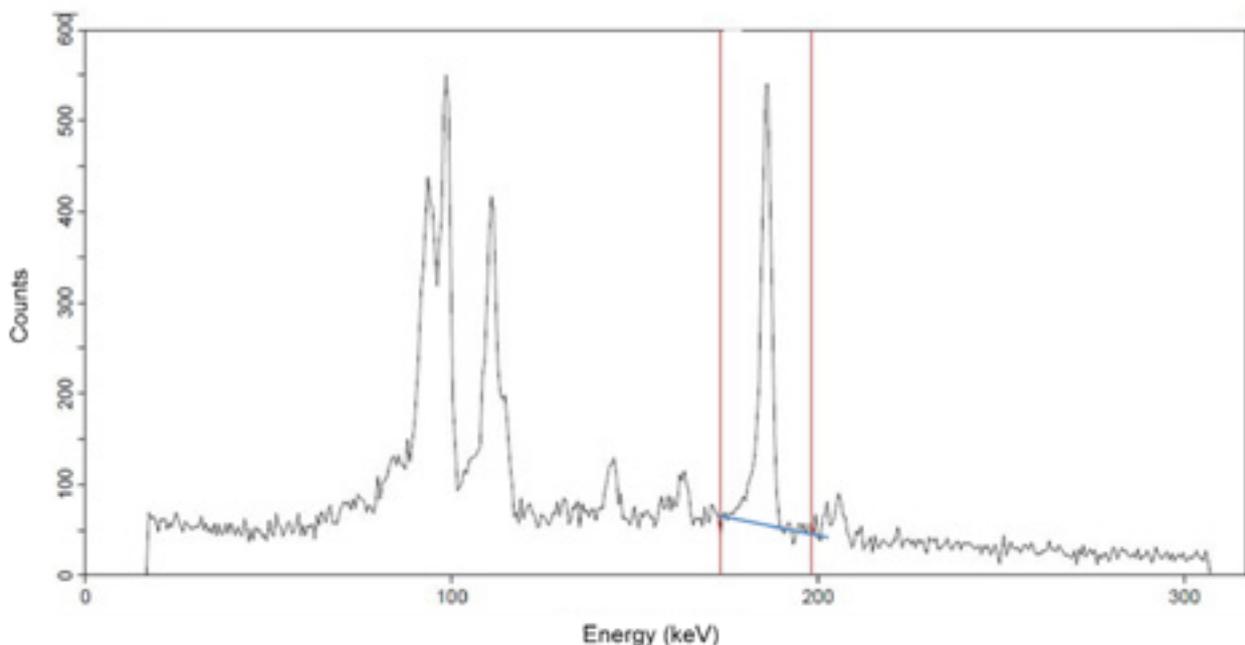
where  $\hat{\sigma}_B$  is the estimated standard deviation in the estimated peak area in the background measurement,  $\hat{\sigma}_M$  is the estimated standard deviation in the area of the peak in the measured spectrum from the item, and  $T_B$  and  $T_M$  are the count times corresponding to the background and item measurement, respectively. Both the measured and background rates are corrected for the background caused by Compton scattering by estimating the count rate of the net

peak, which involves a difference of two quantities as in Example 3.3. In cases where the attribute test software is unable to account for room background in automatically calculating whether the attribute test has been passed, the inspector performs the attribute test calculation for each item. The attribute test aims to answer the question ‘Does the item contain the material as declared?’, and an inspector’s time is quite limited, so inspectors sometimes apply more stringent statistical tests to help confirm the attribute test result. One example of such a stringent test is

$$\text{Measured rate} - 3\frac{\sigma_M}{T_M} > \text{Background rate} + 3\frac{\sigma_B}{T_B}.$$

Inspectors typically perform the background measurement before performing the verification measurements, and the quantity on the right side of the stringent inequality is a single calculated value, which makes the evaluation simple to do while performing verification measurements. Measurements that do not pass the stringent test can be tested against the more formal method.

Recall that the Gaussian approximation to the Poisson is adequate for tolerance interval estimation if the Poisson mean  $\lambda \geq 100$  (Fig. 3), so the factor of 3 used above is justified because for the data in Fig. 5, the quantiles of the Gaussian provide an adequate approximation to the true FAP, assuming  $\hat{\lambda} = \lambda$  (but one should be aware of the need for the factor  $\sqrt{1+1/n} = \sqrt{2} = 1.41$  as in Fig 3a versus Fig 3b). A more complicated method than a sum of Poisson counts below and above the peak ROI is often used to estimate the background under the blue line in Fig. 5; so  $\hat{\sigma}_B$  and  $\hat{\sigma}_M$  can involve more than the Poisson distribution (beyond this paper’s scope).



**Fig. 5:**  $\gamma$ -ray spectrum of a fresh fuel assembly from a CZT detector. The red lines indicate the ROI used to estimate 186 keV peak area while the blue line is an estimate of the Compton background beneath the 186 keV peak based on a linear interpolation of the background at  $\gamma$ -ray energies that are just above and below the ROI.

## 5.2 Tolerance interval versus prediction interval

A prediction interval is another approach to the MDA that leads to larger MDA values than the Currie-based MDAs and to smaller MDAs than the tolerance interval-based MDAs. The prediction interval approach averages over the parameter(s)  $\theta$  so there is no probability statement regarding confidence in coverage [5].

## 5.3 Impact of analyzing predicted counts rather than estimated activity

Zykov [15] describes a pass-fail criterion for verification measurements (operator declarations compared to inspector measurements, as in Section 3.1) regarding the minimum detectable defect size (the minimum amount of missing radioactive material) if analysis of the inspector measurements is based on measurements that are predicted using modeling and the operator declarations. Such an approach would avoid explicit inversion of measurements to activity or nuclear material mass, and simulations to be presented elsewhere suggest that the minimum detectable defect size would be smaller. The minimum detectable defect size would be based on a tolerance interval approach, because that is more conservative than the Currie approach, as this paper has shown. This methodology could lead to more efficient verification sampling plans. Regarding testing for patterns, recall that the overall test for a pattern is based on the average difference statistic,  $D = \frac{1}{n} \sum_{j=1}^n \frac{o_j - i_j}{o_j}$  [14], which could be defined on the basis of measured masses or on the basis of predicted and observed measurements. However, whenever an estimate of the D statistic (at the stratum, material balance component, or material balance area level) is needed, e.g. for the detection of diversion into D through material balance evaluation, it would be necessary for the inspector to estimate item mass, so explicit inversion of inspector measurements to item mass would be required.

## 6. Summary

This paper revisited Currie's MDA with attention to the FAP in Poisson and Gaussian data in NDA by  $\gamma$ -ray detection. It was shown that the actual FAP can be significantly larger than the nominal FAP if the nominal FAP is not calculated based on a tolerance interval; and, if the nominal FAP is calculated based on a tolerance interval, then the MDA is increased compared to the Currie approximation. Implications for safeguards have not yet been evaluated. A simple way to accommodate random and/or systematic errors in converting from a mean shift to an activity shift was illustrated.

## 7. References

[1] Currie, L., Limits for qualitative detection and quantitative determination: application to radiochemistry, *Analytical Chemistry* 40, 58-5936, 1968.

- [2] Kirkpatrick, J., Venkataraman, R., Young, B., Minimum detectable activity, systematic uncertainties, and the ISO 11929 standard, *Journal of Radioanalytical and Nuclear Chemistry* 296, 1005-1010, 2013.
- [3] International Organization for Standardization ISO 11929 determination of the characteristic limits (decision threshold, detection limit, and limits of the confidence interval) for measurements of ionizing radiation-fundamentals and application, Geneva, 2010.
- [4] Young, D., Tolerance: an R package for estimating tolerance intervals, *Journal of Statistical Software* 36(5),1-39 2010.
- [5] Hamada, M., Johnson, V., Moore, L., Wendelberger, J. Bayesian prediction intervals and their relation to tolerance intervals, *Technometrics* 46(4), 452-459, 2004.
- [6] Janiga, I., Garaj, I., Two-sided tolerance limits of normal distributions with unknown means and unknown common variability, *Measurement Science Review* (3) (1), 75-78, 2003.
- [7] Eberhardt, K., Mee, R., Reeve, C., Computing factors for exact two-sided tolerance limits for a normal distribution, *Communication in Statistics- Simulation and Computation*, 18(1), 397-413,1989
- [8] *R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org/>, 2012.*
- [9] Bonner, E., Burr, T., Martin, K., Norman, C., Santi, P., Uncertainty quantification as presented in training courses for safeguards inspectors, ESARDA 2017.
- [10] Miller, R., *Beyond ANOVA: Basics of Applied Statistics*, Chapman & Hall, 1998.
- [11] Liao, C., Lin, T., Iyer, H., One- and two-sided tolerance intervals for general balanced mixed models and unbalanced one-way random models, *Technometrics* 47(3), 323-335, 2005.
- [12] Burr, T., Krieger, T., Krzysztozek, K., Norman, C., Tolerance intervals for measurements with systematic and random errors, IAEA report 2017.
- [13] Kirkpatrick, J., Young, B., Poisson statistical methods for the analysis of low-count gamma spectra, *IEEE Transactions on Nuclear Science* 56(3), 1278-1282, 2009.
- [14] Kirkpatrick, J., Venkataraman, R., Young, B., Calculation of the detection limit in radiation measurements with systematic uncertainties, *Nuclear Instruments and Methods in Physics Research A* 784, 306-310, 2015.
- [15] Zykov, S. Technical challenges and technological gaps in IAEA safeguards, *Proceedings of the Institute of Nuclear Materials Management*, 2015.