SANDIA REPORT

SAND2014-18941
Unlimited Release
Printed October 2014

Ratio Distributions for Use in Spectral Radiation Detection Algorithms

Isaac R. Shokair

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@adonis.osti.gov
Online ordering:	http://www.osti.gov/bridge

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.fedworld.gov
Online order:	http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2014-18941
Unlimited Release
Printed October 2014

Ratio Distributions for Use in Spectral Radiation Detection Algorithms

Isaac R. Shokair

Sandia National Laboratories
Mail Stop 9408
P. O. Box 0969
Livermore, CA 94551-0969

Abstract

This report considers variance and confidence interval estimates of ratios of Poisson random variables using an approximation to the ratio distribution of these variables. One situation where this ratio arises is when normalized counts are considered for polyvinyl toluene (PVT) detectors with a small number of energy bins. These are the counts in an energy bin normalized to the total counts and represent the spectral shape for radiation portal monitor (RPM) measurements. Results of the variance estimate are shown for sample RPM data. It is also shown that a simple approximation of variance that does not make use of ratio distributions is accurate even at low counts. The ratio distribution is still required for evaluation of other statistical quantities such as confidence intervals.

Acknowledgements

Support of this work by the Department of Homeland Security's Domestic Nuclear Detection Office through its Technical Reachback Program is gratefully acknowledged.

Contents

1. Introduction

Radiation Portal Monitors (RPMs) have been deployed at border locations across the United States for detection of illicit radioactive materials with emphasis on special nuclear materials (SNM). These radiation portals measure gamma and neutron radiation time histories as a vehicle passes through the portal producing a series of gamma spectral and neutron count measurements for the vehicle.

Many primary inspection RPMs use polyvinyl toluene (PVT) gamma detectors, which is a material with poor energy resolution where the measured spectra aren't adequate for full spectrum analyses. Thus the measured counts are usually stored in a small number of energy bins or windows which span the energy range typically 0 – 3000 keV. The spectral information contained in measured counts in these energy windows is still useful for some differentiation between benign (mostly naturally occurring radioactive materials or NORM) and threat sources. Many analysis algorithms and methods use the spectral information content in the energy windows for detection of threat sources or for spectral characterization of radiation sources[1-4].

For anomaly detection and spectral characterization analyses discussed in references (2-4) the spatial profile of the measured counts is used to determine the spatial region(s) of interest (ROI) or subset of the measurement spatial extent that has the source signal above background. The counts for the time series of measurements in this region are added to obtain a single spectrum. In some cases the background is removed using an estimate of the suppressed background counts in the ROI. The counts for window i, $i = 1, \ldots, N$, which are assumed to follow Poisson statistics, are then normalized to the total counts in order to consider only the spectral shape of the source independent of the amplitude. Since the total counts contain the window i counts, the normalized counts for this window is a ratio of Poisson random variables that are partially correlated. Similar ratios also arise for multi-energy radiography systems where ratios of attenuated beam intensities at different x-ray energies are used to infer properties and distributions of the intervening materials[5].

It is anticipated that benign and threat radiation sources will have different spectral shapes but the differences are not expected to be large due to the response of PVT. Given measurements for relatively weak threat sources, it is extremely important to properly account for the expected variance due to Poisson noise in order to properly classify such sources and not confuse them with other naturally occurring radioactive materials (NORM). This is the subject of this report where we consider approximations to the ratio distribution of two Poisson random variables for the purpose of estimation of variance and confidence intervals of the normalized counts. The results of this analysis are then applied to sample RPM data with 8 energy windows.

2. Approximation of distribution of Ratios of Poisson Counts

For the case of a small number of spectral or energy bins (such as usually used for a PVT RPM), it is often of interest to use the normalized counts in order to consider only the spectral shape and remove dependence on the source strength[2-4]. For N spectral bins the normalized counts for a bin is given by:

$$\hat{n}_i = \frac{n_i}{n_t} \tag{1}$$

where n_i is the counts in the i-th bin and n_T is the total counts, that is, $n_T = \sum_{j=1}^{N} n_j$. We are interested in finding the distribution of $\hat{n}_i$ when the counts in all bins follow Poisson statistics keeping in mind that n_i and n_T are partially correlated. The total counts n_T can be partitioned into the sum of two uncorrelated quantities as: $n_T = n_i + \sum_{j \neq i} n_j = n_i + \nu$ and we obtain:

$$\hat{n}_i = \frac{n_i}{\nu + n_i} \tag{2}$$

For sufficiently large counts, the Poisson distribution can be approximated by a normal distribution (in an integral sense) and thus n_i and ν can be approximated by uncorrelated continuous random normal variables X and Y respectively. Let the mean values for X and Y be given by μ_x and μ_y respectively. Also, let the standard deviations be given by σ_x and σ_y. We consider the more general case of the ratio $F = \frac{X}{Y + \alpha X}$ where α is a constant. For normal distributions for X and Y, the distribution of F is derived in the appendix and is given by:

$$G(F) = \frac{1}{2\pi\, \sigma_x\, \sigma_y} \exp\left[-\left(\frac{\mu_x^2}{2\sigma_x^2} + \frac{\mu_y^2}{2\sigma_y^2} \right) \right] \left\{ \frac{1}{a^2} + \left(\frac{b\sqrt{\pi}}{2\, a^3} \right) \left[\exp\left(\frac{b^2}{4a^2} \right) \right] \left[2\, \Phi\left(\frac{b}{\sqrt{2}\, a} \right) - 1 \right] \right\} \tag{3}$$

$$a^2 = \left(\frac{F^2}{2\, \sigma_x^2} + \frac{(1 - \alpha F)^2}{2\, \sigma_y^2} \right)$$

$$b = \pm \left(\frac{\mu_x\, F}{\sigma_x^2} + \frac{\mu_y (1 - \alpha F)}{\sigma_y^2} \right)$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} dt\, e^{-t^2/2} = \frac{1}{2}\left[1 + \operatorname{erf}\left(x/\sqrt{2}\right)\right]$$

where $\operatorname{erf}(\,)$ denotes the well-known error function. The above expression for $\Phi(x)$ is the cumulative distribution function (CDF) of a standardized normal random variable. For the factor (*b*) the positive sign is for the range $-\infty < F \leq 1/\alpha$ and the negative sign is for $1/\alpha < F < \infty$. In the limit $\alpha \to 0$ Eq. (3) reduces to published results[6]. It should be noted that given a ratio distribution $D(R)$ with *R* being the ratio of uncorrelated continuous variables $R = \frac{X}{Y}$ (such as given in reference 6), the ratio distribution $G(F)$ with $F = \frac{X}{Y + \alpha X}$ can be obtained by a simple transformation. Since for any (*X*,*Y*) there is one-to-one correspondence between *R* and *F*, based on conservation of probability, we can write the differential form:

$$G(F)\, dF = D(R)\, dR$$

or

$$G(F) = D(R)\, \frac{dR}{dF} \tag{4}$$

We also have:

$$F = \frac{R}{(1 + \alpha\, R)}, \quad R = \frac{F}{(1 - \alpha\, F)}, \text{ and } \quad \frac{dR}{dF} = \frac{1}{(1 - \alpha F)^2}$$

The result for $G(F)$ is:

$$G(F) = \frac{1}{(1 - \alpha\, F)^2}\, D\left(\frac{F}{1 - \alpha\, F}\right) \tag{5}$$

Equation (5) can be shown to be equivalent to Eq. (3) with $D(R)$ as given in reference (6). This can also be verified numerically.

Problem with the Distribution G(F):

The ratio distribution $G(F)$ given by Eq. (3) is not well-behaved in the sense that the first and second moments don't exist. These moments are needed for evaluation of the mean and variance of the distribution. This problem can be seen by considering the limits μ_x and $\mu_y \to 0$ and σ_x and $\sigma_y \to 1$. In these limits the factor $b \to 0$ and $G(F) \to \frac{1}{\pi} \frac{1}{\left[F^2 + (1-\alpha F)^2\right]}$ which reduces to the Cauchy or Lorentz distribution in the limit of $\alpha \to 0$. Note that the quantity $F^2\, G(F)$ needed for evaluation of the variance is not integrable when the upper or lower limit is infinity. This problem arises from the singular point in the ratio F when $Y = -\alpha X$ since $(X,Y)\, \varepsilon \, [-\infty, \infty]$ (see Figure (A.1) in the appendix for an illustration of the singular points).

Approximation of the Poisson Distribution by a Truncated Normal Distribution

The above problem which is caused by negative values of X or Y can be resolved if instead of the full normal distribution, a truncated distribution is used where only positive *X* and *Y* values are allowed. This should provide a more physically acceptable approximation to the Poisson distribution especially for low counts. The form of the truncated distribution we use is:

$$f(X) = \mathrm{A}_x \exp\left[-\frac{(X-\mu_x)^2}{2\sigma_x^2}\right] \qquad X \ge 0$$
$$= 0 \qquad X < 0 \tag{6}$$

In this case the normalization factor is found from the requirement $\int_0^\infty dX \;\, f(X) = 1$ and is:

$$A_x = \frac{1}{\sqrt{2\pi}\,\sigma_x} \frac{2}{\left\{1 + erf\left(\frac{\mu_x}{\sqrt{2}\,\sigma_x}\right)\right\}},$$

where $erf(\,)$ denotes the error function. The equivalent distribution is used for the *Y* variable with a similar normalization factor A_y. Using these distributions the ratio distribution (see appendix for details) is given by:

$$G(F) = A_x\, A_y \;\exp\left[-\left(\frac{\mu_x^2}{2\sigma_x^2} + \frac{\mu_y^2}{2\sigma_y^2}\right)\right]\left\{\frac{1}{2a^2} + \left(\frac{b\sqrt{\pi}}{2\,a^3}\right)\, \exp\left(\frac{b^2}{4a^2}\right)\, \Phi\left(\frac{b}{\sqrt{2}\,a}\right)\right\} \tag{7}$$

$$a^2 = \left(\frac{F^2}{2\,\sigma_x^2} + \frac{(1-\alpha F)^2}{2\,\sigma_y^2} \right)$$

$$b = \left(\frac{\mu_x F}{\sigma_x^2} + \frac{\mu_y (1-\alpha F)}{\sigma_y^2} \right)$$

where $F \,\varepsilon\, [0, 1/\alpha]$ and the function $\Phi(x)$ is as defined previously. The mean, variance, and confidence intervals of the distribution given by Eq. (7) are valid since all moments of $G(F)$ are integrable. However these quantities are not easily found analytically and therefore numerical integration will be used for evaluation of any integral of this distribution.

3. Numerical Evaluation of Variance and Confidence Intervals

The mean and variance of the distribution given by Eq. (7) are not easily evaluated analytically and therefore numerical integration is used to evaluate these and any other quantities that involve integrals of this distribution. Since this function is differentiable and has no singular points, any simple numerical integration method can be used. For the results shown in this section we used Romberg numerical integration[7] for evaluation of all the integrals. For this evaluation the mean counts in Eq. (2) are specified by: $n_i \rightarrow \mu_x$ and $\nu \rightarrow \mu_y$. We also use $\sigma_x^2 = \mu_x$ and $\sigma_y^2 = \mu_y$ since we're dealing with approximations of Poisson distributions. The limit $\alpha \rightarrow 1$ is used. Figure (1) below shows examples of the ratio distribution given by Eq. (7) for the case $n_T = n_i + \nu = 40$.

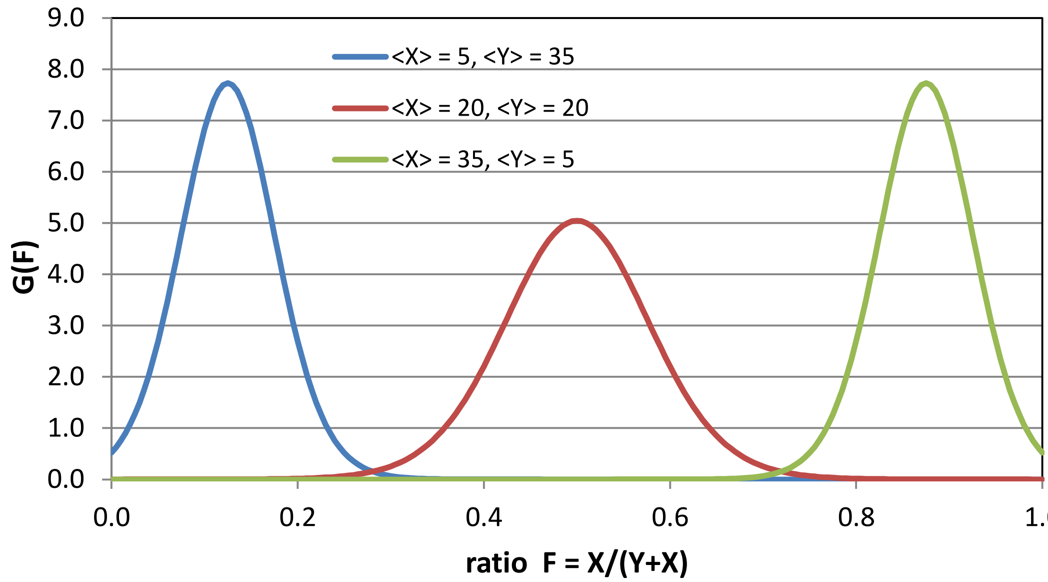

Figure 1. Examples of the ratio distribution given by Eq. (7) for three pairs of mean counts for total mean counts n_T = 40. The variances are $\sigma_x^2 = \mu_x$ = <X> and $\sigma_y^2 = \mu_y$ = <Y>.

Variance Estimates:

Two of the quantities of interest are the mean and the variance about the mean of the ratio distribution. We evaluate these quantities for a specified value of n_T and vary n_i from zero to n_T, that is, the ratio varies between 0 (zero counts in i-th window) and 1 (all the counts are in the i-th window). Each pair $(n_i, n_T - n_i)$ has an associated ratio distribution given by Eq. (7) and we evaluate the mean and variance for this distribution using numerical integration. Figures (2-3) show sample results. Along with these results we also show results of simulations using a large number of Poisson trials for each pair. Note that Poisson counts are generated independently for the mean values n_i and $(n_T - n_i)$ for each trial and then the ratio is calculated for each of the trials. The mean and variance of these simulated ratios are then calculated. Even for low total counts of $n_T = 10$ the approximation using the ratio distribution given by Eq. (7) is reasonable for the variance estimate. For large values of n_T the ratio distribution approximation is very good. Figure 4 shows the dependence of the variance on the total counts for a fixed ratio of the means. At total counts of about 25 the ratio distribution estimate of variance shows the largest departure (of about 10%) from the value obtained by simulations.

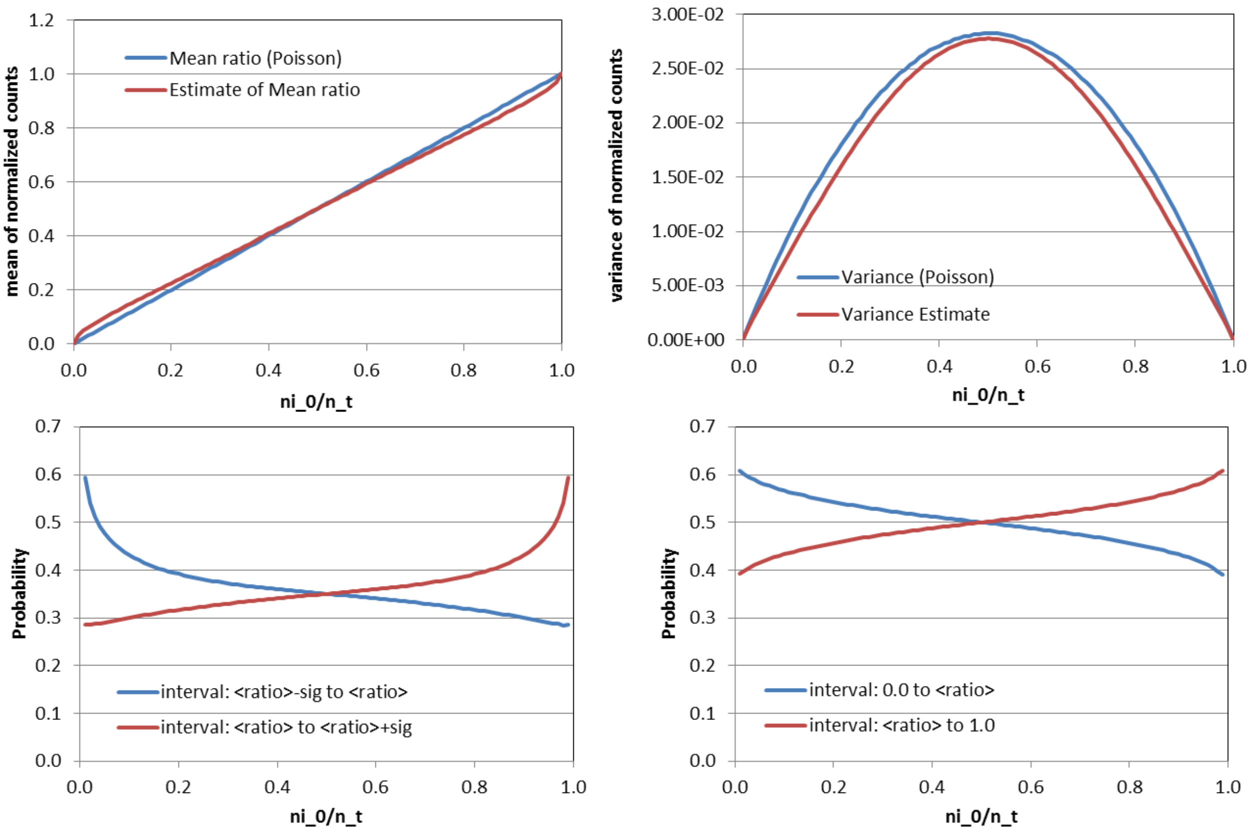

Figure 2. The top two plots compare the mean ratio and variance based on the ratio distribution given by Eq. (7) with results of simulations using Poisson random variables (blue curves) for mean total counts of n_T = n_t = 10. Each point along the x-axis represents a different value of n_i = ni_0 (the mean counts in a specific energy window) and ranges from zero (no counts in the window) to 1.0 (all counts are in the window). For the simulations, for each pair of mean values $(n_i, n_T - n_i)$, one million trials were used to estimate the mean ratio and variance. Left bottom plot: cumulative probability of the ratio distribution over two intervals: (1) mean ratio – sigma to mean ratio and (2) mean ratio to mean ratio + sigma where sigma is the square root of the calculated variance. Except near the boundaries, the sum of these two probabilities is close to 70% which close to the 68% for a normal distribution. Bottom right plot: cumulative probability over two intervals: (1) 0.0 to mean ratio and (2) mean ratio to 1.0 to show asymmetry of the distribution. The sum of these two probabilities is equal to unity.

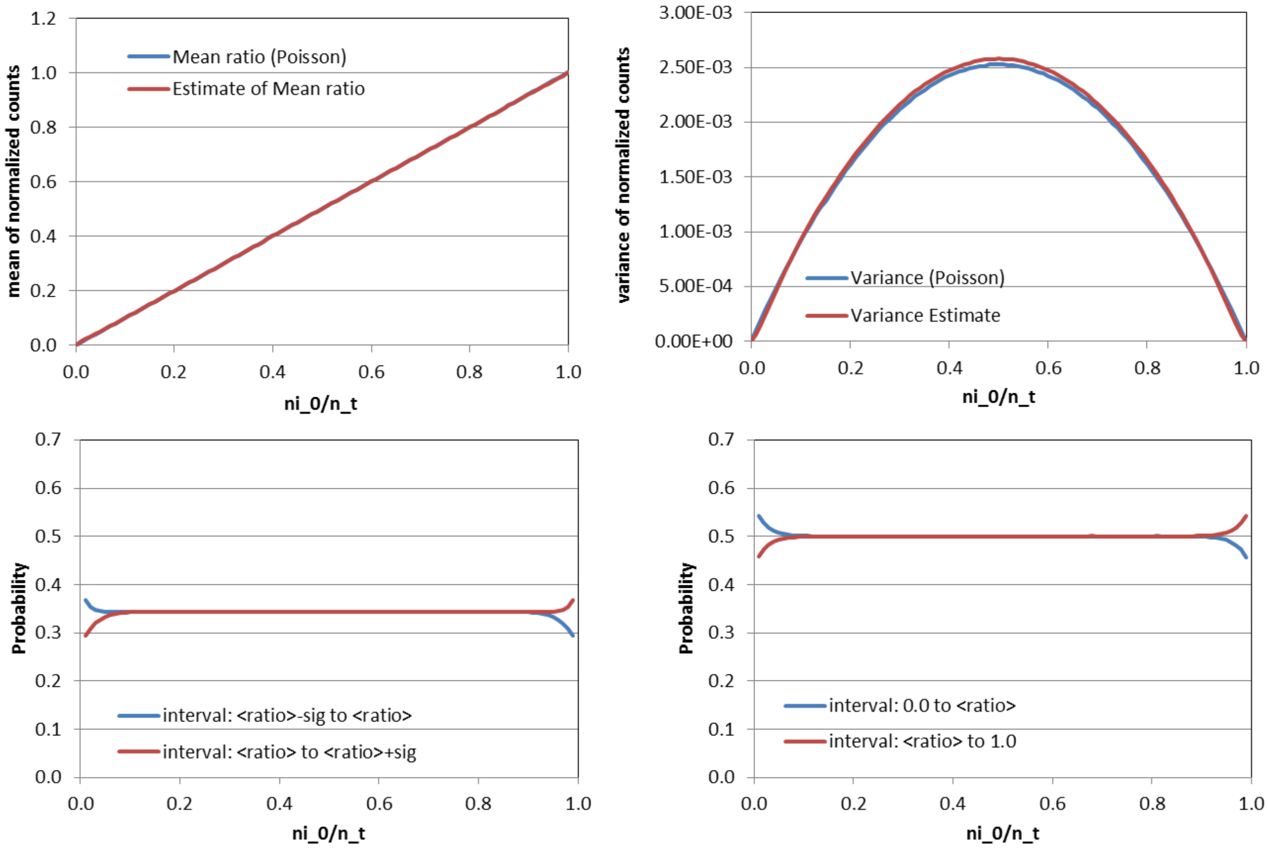

Figure 3. Same as Figure 2 with n_T = 100.

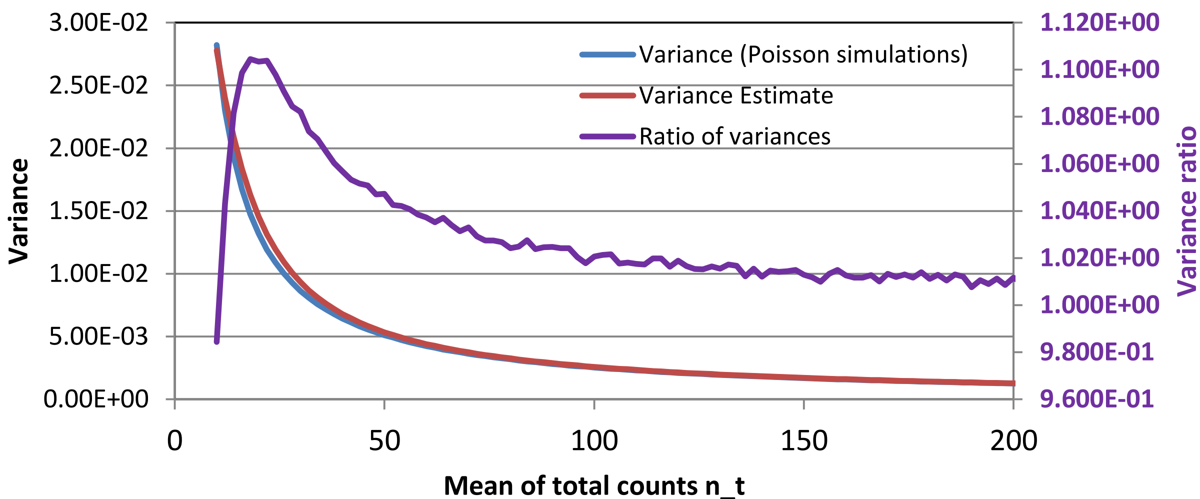

Figure 4. Comparison of variance of ratio of counts n_i / n_T using the ratio distribution and exact values (based on simulations) varying the total counts n_T for a fixed ratio of means $n_i / n_T = 0.5$.

A Simpler Method for Approximation of Mean and Variance:

When the total counts are sufficiently large an alternate and much simpler method of mean and variance estimation of the normalized counts can be used. Going back to Eq. (2), if we consider a single measurement, we can express the counts as the sum of a mean and a deviation from the mean, that is:

$$n_i = n_{io} + \delta n_i$$

$$\nu = \nu_o + \delta \nu \tag{8}$$

$$n_t = n_o + \delta n \qquad \delta n = \delta n_i + \delta \nu$$

and thus:

$$\hat{n}_i = \frac{n_i}{n_t} = \frac{(n_{io} + \delta n_i)}{(n_o + \delta n)} \tag{9}$$

For reasonably large total counts n_t a Taylor expansion to second order in deviation results in:

$$\hat{n}_i \approx \frac{(n_{io} + \delta n_i)}{n_o}\left(1 - \frac{\delta n}{n_o} + \left[\frac{\delta n}{n_o}\right]^2 + \cdots\right)$$

This expression can be expanded to obtain:

$$\hat{n}_i \approx \frac{n_{io}}{n_o} + \frac{\delta n_i}{n_o} - n_{io}\frac{\delta n}{n_o^2} + n_{io}\frac{(\delta n)^2}{n_0^3} - \frac{\delta n_i \; \delta n}{n_o^2} + \frac{\delta n_i \; (\delta n)^2}{n_o^3} + \cdots \tag{10}$$

Using $\delta n = \delta n_i + \delta \nu$ and taking the mean of Eq. (10) over many measurements we obtain:

$$\langle \hat{n}_i \rangle \approx \frac{n_{io}}{n_o} + n_{io}\frac{\langle (\delta n_i)^2 \rangle + \langle (\delta \nu)^2 \rangle}{n_0^3} - \frac{\langle (\delta n_i)^2 \rangle}{n_o^2} + \frac{\langle (\delta n_i)^3 \rangle}{n_o^3} + \cdots$$

where $\langle \; \rangle$ denotes the mean of a quantity over the distribution of values. The values $\langle \delta n_i \rangle$ and $\langle \delta \nu \rangle$ are zero by definition of the mean and the means of products δn_i and $\delta \nu$ or their powers are separable because of the independence property, that is, these quantities are not correlated. So far this applies to any distribution. If the counts follow Poisson statistics: $\langle (\delta n_i)^2 \rangle = n_{io}$, $\langle (\delta \nu)^2 \rangle = \nu_o$, and the above expression reduces to:

$$\langle \hat{n}_i \rangle \approx \frac{n_{io}}{n_o} + \cdots \tag{11}$$

The variance of the normalized counts is given by:

$$\sigma_i^2 = \left\langle \left(\hat{n}_i - \langle \hat{n}_i \rangle\right)^2 \right\rangle$$
$$= \left\langle \hat{n}_i^{\,2} \right\rangle - \langle \hat{n}_i \rangle^2$$

Using the Taylor expansion expression from Eq. (10) and keeping up to quadratic terms in deviation we obtain:

$$\begin{aligned}
\sigma_i^2 \approx{}& \left(\frac{n_{io}}{n_o}\right)^2 - 2\frac{n_{io}}{n_o^{\,3}}\left\langle (\delta n_i)^2 \right\rangle + 2\frac{n_{io}^{\,2}}{n_o^{\,4}}\left\langle (\delta n_i)^2 + (\delta \nu)^2 \right\rangle + \frac{\left\langle (\delta n_i)^2 \right\rangle}{n_o^2} - 2\frac{n_{io}}{n_o^{\,3}}\left\langle (\delta n_i)^2 \right\rangle + \\
& + \frac{n_{io}^{\,2}}{n_o^{\,4}}\left\langle (\delta n_i)^2 + (\delta \nu)^2 \right\rangle + \cdots \\
& - \left\{ \left(\frac{n_{io}}{n_o}\right)^2 - 2\frac{n_{io}}{n_o^{\,3}}\left\langle (\delta n_i)^2 \right\rangle + 2\frac{n_{io}^{\,2}}{n_o^{\,4}}\left\langle (\delta n_i)^2 + (\delta \nu)^2 \right\rangle \right\} + \cdots \\
\approx{}& \frac{\left\langle (\delta n_i)^2 \right\rangle}{n_o^{\,2}} - 2\frac{n_{io}}{n_o^{\,3}}\left\langle (\delta n_i)^2 \right\rangle + \frac{n_{io}^{\,2}}{n_o^{\,4}}\left\langle (\delta n_i)^2 + (\delta \nu)^2 \right\rangle + \cdots \\
\approx{}& \frac{n_{io}}{n_o^{\,2}} - 2\frac{n_{io}^{\,2}}{n_o^{\,3}} + \frac{n_{io}^{\,2}}{n_o^{\,4}}\left(n_{io} + \nu_o\right) + \cdots \\
\approx{}& \frac{n_{io}}{n_o^{\,2}} - \frac{n_{io}^{\,2}}{n_o^{\,3}} + \cdots
\end{aligned} \qquad (12)$$

In the above expression properties of Poisson distributions were assumed, namely $\left\langle (\delta n_i)^2 \right\rangle = n_{io}$ and $\left\langle (\delta \nu)^2 \right\rangle = \nu_o$. It turns out that the result given by Eq. (12) is identical to what would be obtained using the linear deviation approximation for a function of two variables:

$$\delta f = \frac{\partial f(x,y)}{\partial x}\delta x + \frac{\partial f(x,y)}{\partial y}\delta y ,$$

where the derivatives are evaluated at the mean values of the variables. For uncorrelated variables the overall variance is given by[8]:

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2$$

This equation is independent of the distributions involved. For Poisson distributions the result is identical to Eq. (12). For large counts the expressions given by Equations (11-12) should approach the exact values. This was verified by numerical tests with large numbers of Poisson trials of n_i and ν ($1x10^6$ trials were used) and evaluation of $\langle \hat{n}_i \rangle$ and σ_i^2 using the mean and variance over all these trials as a good estimate of the exact values. The comparison is shown Figures (5-6) below and shows the approximation is good for total counts as low as 10. This result is surprisingly accurate for such low counts and was not expected before considering the ratio distributions. Based on this result the ratio distribution is only needed when other statistical quantities such as confidence intervals are required.

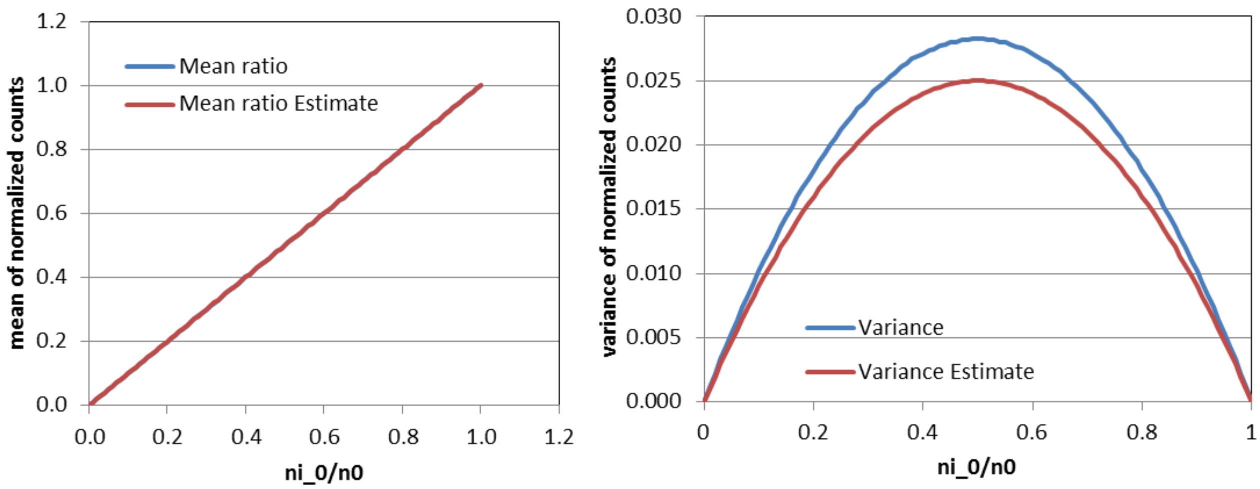

Figure 5. Mean (left) and variance (right) of the normalized counts varying the ratio of mean counts in the i^{th} window to the mean total counts. The mean total counts is n_o = 10. The blue curves are obtained using 1 million statistical trials to evaluate the mean and variance of the normalized counts. The red curves are obtained using the Taylor expansion approximations given by Equations (11-12).

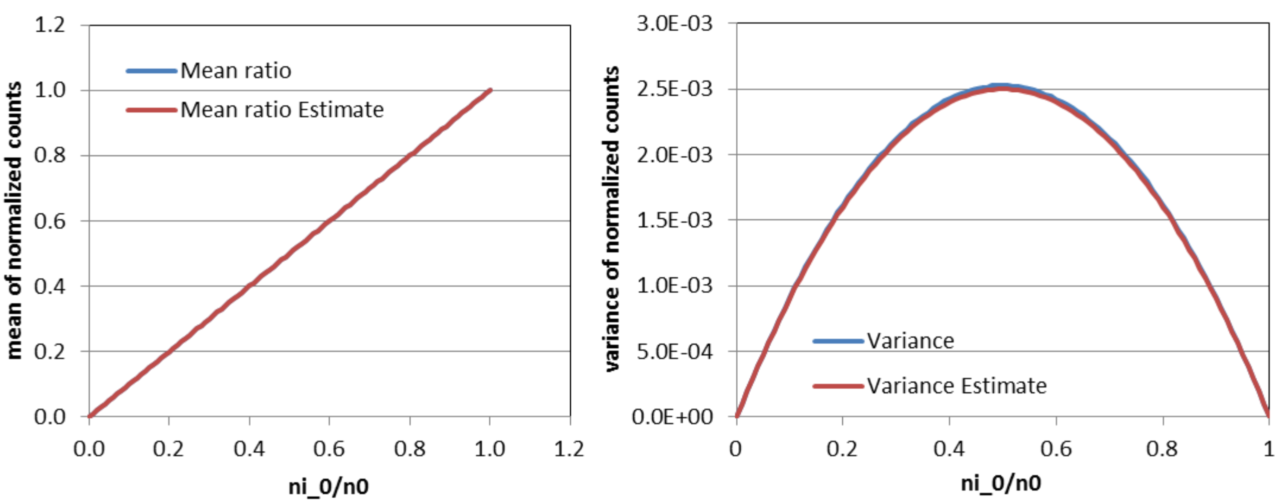

Figure 6. Same as Figure (5) with n_o = 100.

Estimates of Confidence Intervals:

In addition to variance, when only one measurement of a specific quantity is available it is often desirable to estimate the interval for this quantity (around the measured value) that is likely to contain a specified percentage of other equivalent measurements. This is known as the confidence interval. For a continuous variable with a well-behaved distribution, there is an infinite number of intervals that would result in the same probability. Thus the definition of the confidence interval has to be narrowed down to be the smallest possible interval for the specified probability. This means that the distribution function (probability density function or PDF) for any value outside the confidence interval has to be smaller than the minimum PDF inside the interval. Physically, this means that the confidence interval is that which most likely will contain the specified fraction of measurements of the quantity in question. The previous definition of course holds only for distributions with a single maximum. Distributions with multiple maxima are not of interest for this analysis.

For continuous and symmetric probability distributions the confidence interval is centered about the mean value (point of maximum probability) and extends by equal amounts on either side of the mean. For example, for a normal distribution, the 68% confidence interval extends from $-\sigma$ to σ around the mean of the distribution.

For non-symmetric distributions such as the ratio distribution considered here, the mean is not necessarily the point of maximum probability. For this case we estimate the confidence interval by first finding the ratio at maximum probability (R_{max}) and then using equal contributions from either side of R_{max} to obtain the total specified probability or confidence value. For cases where the integrated probability from the R = 0 to R_{max} is less than half the confidence value, R = 0 is used as the start of the confidence interval, and the right bound of the interval is adjusted to get the total confidence value. The same is done for cases where the integrated probability from R_{max} to R = 1 is less than half the confidence value. Simple bisection algorithms are used to find R_{max} and the confidence intervals. All integrals of the ratio distribution are evaluated numerically. We should note that this estimate of confidence intervals is not strictly consistent with the above definition (i.e. PDF outside the interval has to be smaller than inside) and is only used an illustrative estimate for this report.

The two figures below show results of the estimated confidence intervals as the quantity n_i/n_t is varied for the case of 80% confidence value. Along with these plots the "exact" results using simulations with a large number of Poisson trials are also shown. The agreement is good for the two cases shown. The discontinuous results for the simulations is due to the discrete nature of the two Poisson distributions (and therefore the resulting ratio distribution) that becomes very clear for low counts. For low counts the probable values of the ratio (those ratios with significant probability of occurrence) are relatively few and this leads to the large discontinuities in the confidence intervals seen in the figures. Discontinuities in integrated quantities will always be present for Poisson-related ratio distributions, but the amplitudes of these discontinuities become smaller as the total counts and therefore the number of probable ratios gets larger.

When only a single measurement of the counts is available, absent any other information, the best estimate for the mean counts are these measured counts and they can be used to estimate the confidence intervals for the ratio (normalized counts) for an energy window using the method described above.

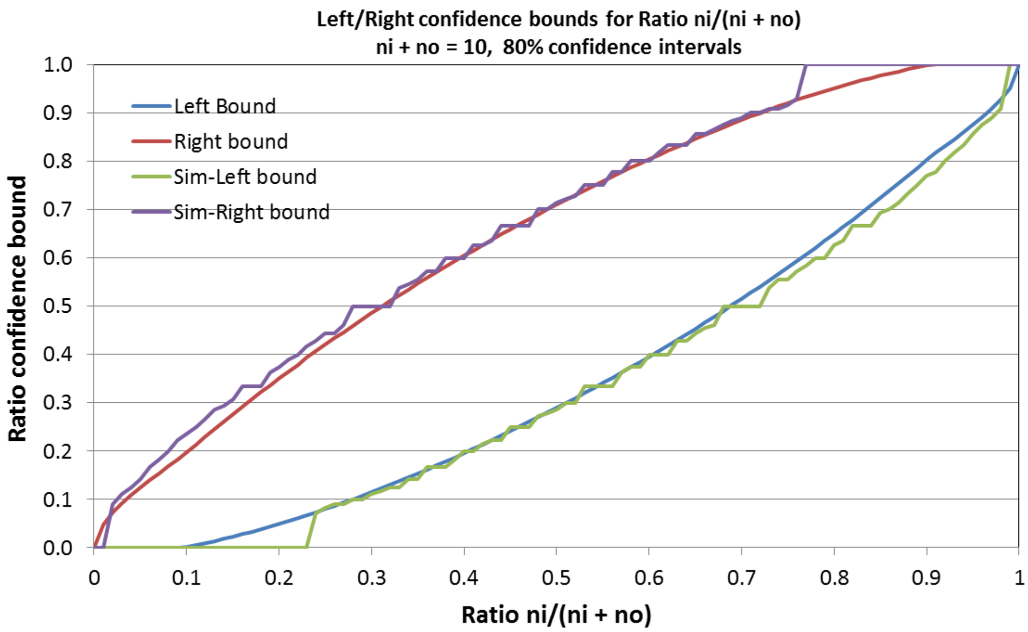

Figure 7. Confidence intervals for the ratio vs. the mean ratio n_i/n_t for 80% confidence value for the case with n_t = 10 counts. The smooth curves were obtained using the ratio distribution. The discontinuous curves are the results of simulations using $1x10^6$ Poisson trials for each ni and nt value.

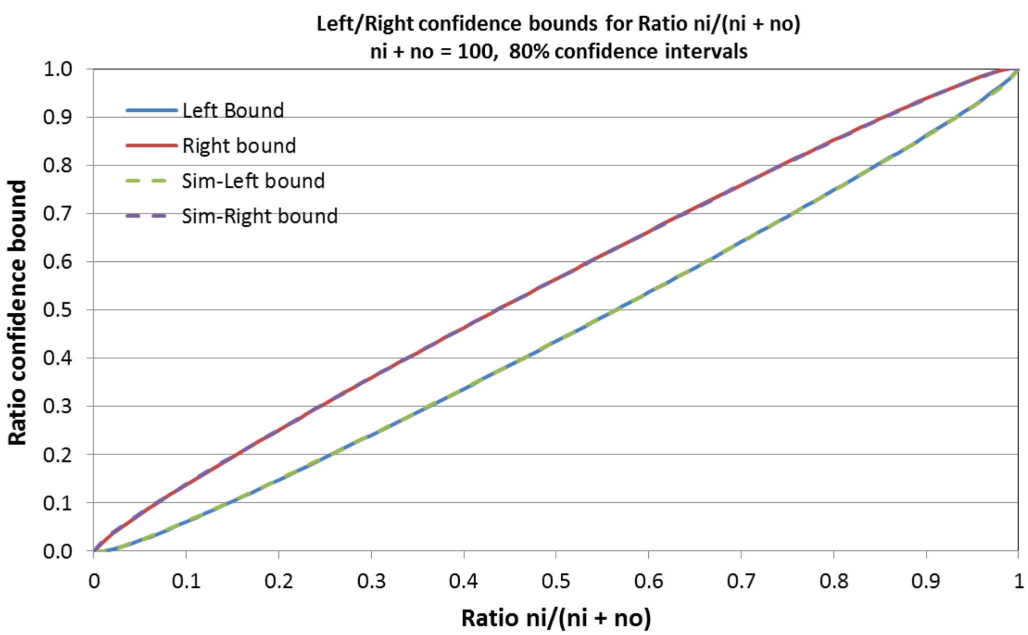

Figure 8. Same as Figure 7 for the case with n_t= 100.

4. Sample Results for RPM Data

The estimated variance of normalized counts as obtained based on the ratio distribution discussed in the previous section can be used in spectral characterization analysis and anomaly detection. In these types of analyses a distance metric is usually used and this distance can be weighted by the estimated variance to account for spread in measured spectra due to noise effects. The Mahalanobis is such a distance metric.

In this section we will only show a very simple illustration of the effect of variance for measured PVT RPM spectra. The data used is for typical alarm occupancies encountered in the stream-of-commerce at ports of entry. The analyzed spectra are made up of counts in eight spectral bands (also referred to as energy windows) which are summed over the spatial region that is determined to contain the radiation source. This region is often referred to as the spatial region of interest (ROI) and is determined from the time profile of measured gross counts as a vehicle traverses the RPM. Note that the background is not subtracted in this analysis. The variance of the normalized counts for all windows is estimated using the approximate ratio distributions of the quantities:

$$\hat{n}_i = \frac{n_i}{\sum_{k=1}^{N} n_i}, \quad i = 1, \ldots, N \text{ , with } N = 8.$$

Figure 9 shows the normalized counts with the variance estimates in the form of a rectangle for each pair of energy windows. Each rectangle is centered around the normalized counts and has extent of 2σ in each dimension where σ is the square root of the variance estimate for the window in question.

The variance estimate is different for each measurement and depends on the source signal amplitude and the number of time samples in the ROI. The results show several spectral clusters when all energy windows are considered. It is presumed that this data is for sources that contain naturally occurring radioactive materials (NORM) or are due to medical treatments. Cluster analysis of this data for spectral characterization purposes along with analysis of threat source spectra should help assess the capability of PVT RPMs to discriminate between benign NORM and threat sources. Anomaly detection is one method that is used for this purpose.

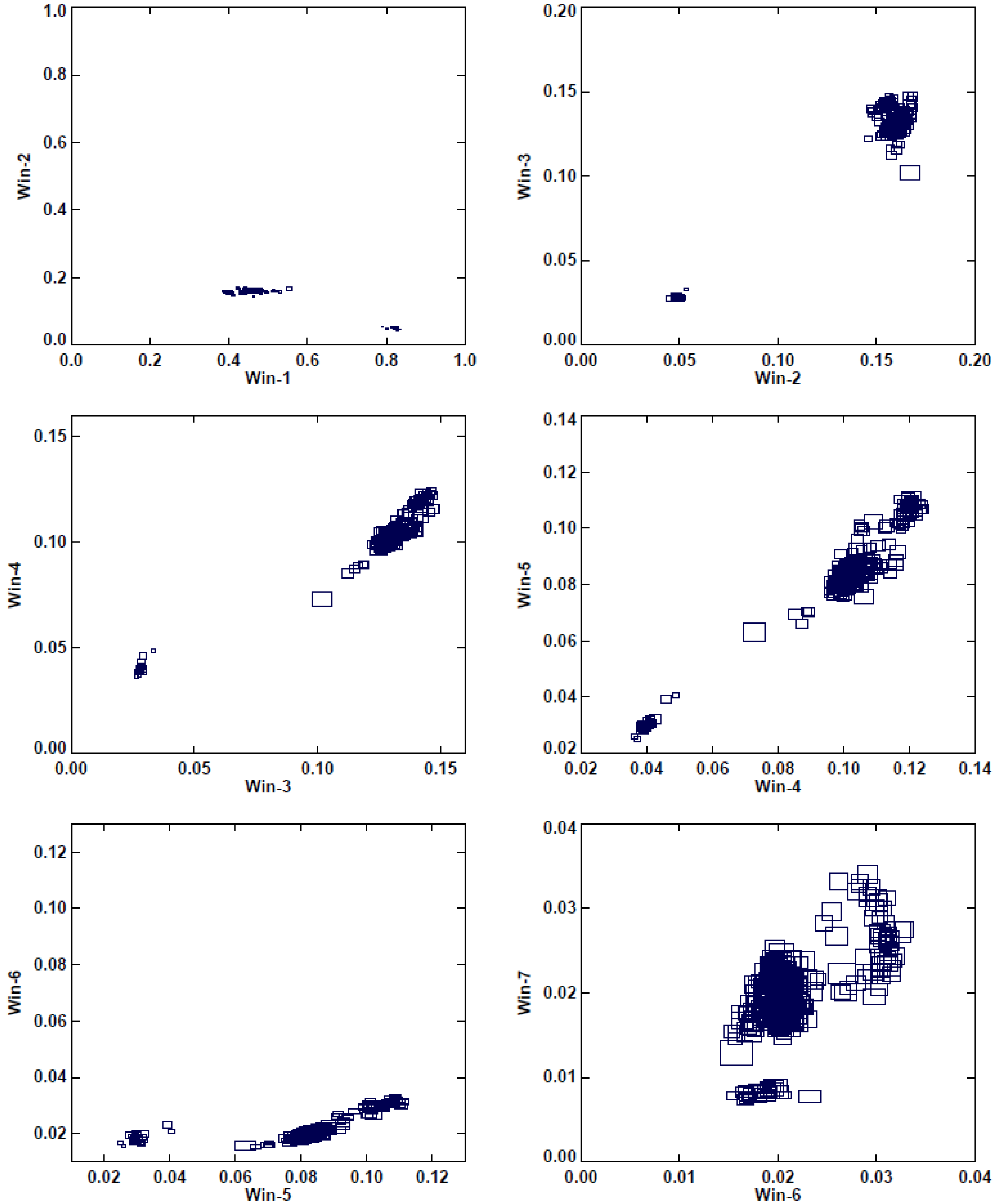

Figure 9. Normalized spectra for RPM measurements showing the magnitude of variance due to measurement noise. Each rectangle shown represents the normalized counts for two energy windows (center of rectangle). The extent of each rectangle in each dimension is twice the square root of the estimated variance based on the ratio distribution. The background was not subtracted from the ROI spectra in this analysis.

6. Summary

This report considered estimation of variance and confidence intervals of normalized counts using analysis based on ratio distributions. The normalized counts for an energy window is the counts divided by the total counts and can be represented as a ratio that is a function of two uncorrelated random variables, the counts in the window and the counts summed over all the other windows. Truncated normal distributions were used to approximate the Poisson distribution for the counts where a normal-like distribution is used, but the counts are constrained to positive values. This turns out to be a good approximation and avoids problems that result from singular values in the ratio when all real values are allowed. The variance of the ratio distribution was estimated numerically. Using simulations it was shown that this variance estimate is reasonably accurate even for low total counts. This variance estimate was applied to sample alarm RPM data for 8 energy windows to show the spread in the measured normalized spectra.

It turns out that using a simple approximation for variance is equally good for low counts without the need for detailed analysis involving ratio distributions. However the ratio distribution is still required for obtaining other statistical quantities such as confidence intervals.

Appendix

In this section the derivation of the ratio distribution of the quantity $X/(Y+\alpha X)$ is provided where *X* and *Y* are uncorrelated random normal variables and α is a constant.

We start by considering the general case of finding the distribution of the general function $F = F(X,Y)$ where X and Y are random uncorrelated variables with distribution functions $f_1(X)$ and $f_2(Y)$ respectively. We need to find the distribution $G(F)$ so that $G(F)dF$ is the probability that F is in the range F to $F+dF$. In the X,Y plane contours of constant F need to be determined. Let S denote a variable along such a contour and let U denote the variable orthogonal to the contour. The distribution $G(F)$ is then found from:

$$G(F)\,dF = \left(\int_{\text{Contour of constant}\,F} dS\, J(X,Y)\, f_1(X)\, f_2(Y) \right) dU \qquad \text{(A.1)}$$

Where *J* is the the Jacobian of the transformation from (X,Y) to (S,U). Thus given F we need to find (S,U) in terms of (X,Y). The above equation must also obey the normalization:

$$\int_{\text{All F}} G(F)\, dF = \int_{\text{All S}} dS \int_{\text{All U}} dU\ J\ f_1\ f_2 = \int_{\text{All X}} dX \int_{\text{All Y}} dY\ f_1\ f_2 \qquad \text{(A.2)}$$

(A.3)

Normal distributions of the following forms are assumed to start with:

$$f_1(X) = \frac{1}{\sqrt{2\pi}\,\sigma_x} \exp\left[\frac{-(X-\mu_x)^2}{2\,\sigma_x^2} \right] \qquad \text{(A.4)}$$

$$f_2(Y) = \frac{1}{\sqrt{2\pi}\,\sigma_y} \exp\left[\frac{-(Y-\mu_y)^2}{2\,\sigma_y^2} \right] \qquad \text{(A.5)}$$

For the case under consideration in this report we have:

$$F = X/(Y+\alpha X) \text{ or } Y = \left(\frac{1-\alpha F}{F} \right) X$$

where α is a positive value. Thus the contours of constant F are lines that pass through the origin as shown in the figure below. These contours can be described simply by:

$$X = r\cos(\theta) \text{ and } Y = r\sin(\theta)$$

The resulting expression for F is:

$$F = \frac{\cos(\theta)}{\sin(\theta) + \alpha \cos(\theta)} \tag{A.6}$$

On the dashed line in the figure $\sin(\theta_0) + \alpha \cos(\theta_0) = 0$, thus $\theta_0 = \arctan(-\alpha)$. For $\alpha \to 0,\ \theta_0 \to \pi$ and for $\alpha \to 1,\ \theta_0 \to 3\pi/4$.

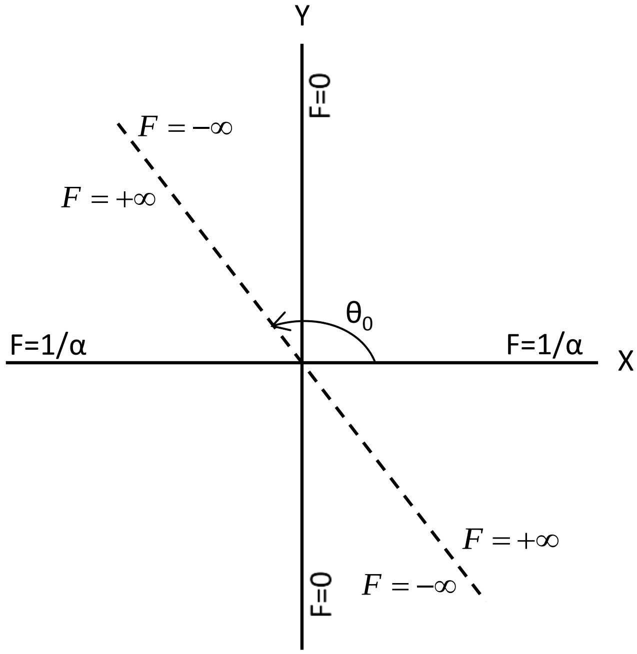

Figure (A.1) Contours of constant F values.

Using Eq. (A.1) we can write:

$$G(F)dF = \int_0^\infty dr\,(r\,d\theta)\{\ f_1(r\cos\theta)\ f_2(r\sin\theta)\ +\ f_1(r\cos(\theta+\pi))\ f_2(r\sin(\theta+\pi))\ \} \tag{A.7}$$

Using the explicit forms for f_1 and f_2 we obtain:

$$G(F) = \frac{1}{2\pi\,\sigma_x\,\sigma_y}\left(\frac{d\theta}{dF}\right)\int_0^\infty dr\,r\left\{\exp\left[-\frac{(r\cos\theta-\mu_x)^2}{2\sigma_x^2} - \frac{(r\sin\theta-\mu_y)^2}{2\sigma_y^2}\right] + \text{same for } (\theta+\pi)\right\} \tag{A.8}$$

Using Eq. (A.6) we find:

$$\frac{d\theta}{dF} = -\frac{1}{\left[F^2 + (1-\alpha F)^2\right]} \tag{A.9}$$

$$\cos(\theta) = \pm \frac{F}{\left[F^2 + (1-\alpha F)^2\right]^{1/2}} \qquad (+) \text{ for } -\infty < F \le 1/\alpha \quad , \quad 0 < \theta < \theta_0$$

$$\sin(\theta) = \pm \frac{(1-\alpha F)}{\left[F^2 + (1-\alpha F)^2\right]^{1/2}} \qquad (-) \text{ for } 1/\alpha < F \le \infty \quad , \quad \theta_0 < \theta < \pi \tag{A.10}$$

After some straightforward algebra Eq. (A.8) results in:

$$G(F) = \frac{1}{2\pi\,\sigma_x\,\sigma_y} \exp\left[-\left(\frac{\mu_x^2}{2\sigma_x^2} + \frac{\mu_y^2}{2\sigma_y^2}\right)\right]\left\{\frac{1}{a^2} + \frac{b\sqrt{\pi}}{2a^3} e^{b^2/4a^2}\left[2\Phi\left(\frac{b}{\sqrt{2}\,a}\right) - 1\right]\right\} \tag{A.11}$$

$$a^2 = \left(\frac{F^2}{2\sigma_x^2} + \frac{(1-\alpha F)^2}{2\sigma_y^2}\right)$$

$$b = \pm\left(\frac{\mu_x F}{\sigma_x^2} + \frac{\mu_y (1-\alpha F)}{\sigma_y^2}\right) \qquad \begin{array}{l} (+) \text{ for } -\infty < F \le 1/\alpha \\ (-) \text{ for } 1/\alpha < F < \infty \end{array}$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} dt\; e^{-t^2/2} = \frac{1}{2}\left[1 + erf\,(\frac{x}{\sqrt{2}})\right] = \text{CDF for standardized normal distribution.}$$

In the limit $\alpha \to 0$ Eq. (A.11) reduces to published results[(6)]. The distribution $G(F)$ given by Eq. (A.11) does not possess integrable moments and thus presents a problem when one is interested in evaluation of statistical quantities such as variance. This is a result of the original random variables X and Y allowed to have negative values. To overcome this problem we consider truncated distributions that are only allowed to have positive X and Y values.

<u>Truncated Normal Distributions</u>:

For modified normal distributions that are only allowed to have positive values the problem with singular values of the ratio $F = X/(Y + \alpha X)$ are eliminated. In this case the ratio is limited to the range $0 \le F \le 1/\alpha$, or the top-right quadrant of Figure (A.1).

The truncated distribution $f_1(X)$ considered here is given by $f_1(X) = A_x \exp\left[-\frac{(X-\mu_x)^2}{2\sigma_x^2}\right]$,

where *X* is limited to positive values and A_x is found from the normalization $1 = \int_0^{\infty} dX\; f_1(X)$, or:

$$A_x = \frac{1}{\sqrt{2\pi}\,\sigma_x} \frac{2}{\left\{1 + erf\left(\mu_x/\sqrt{2}\sigma_x\right)\right\}}.$$

Similar expressions are used for $f_2(Y)$. With these modified distributions for X and Y, the distribution $G(F)$ is found in the same way as before using the normalizations and Eq. (A.8) without the $(\theta+\pi)$ term. The result is:

$$\begin{aligned} G(F) &= A_x\, A_y \left(\frac{d\theta}{dF}\right) \int_0^{\infty} dr\, r \left\{ \exp\left[-\frac{(r\cos\theta-\mu_x)^2}{2\sigma_x^2} - \frac{(r\sin\theta-\mu_y)^2}{2\sigma_y^2} \right] \right\} \\ &= A_x\, A_y\, \exp\left[-\left(\frac{\mu_x^2}{2\sigma_x^2} + \frac{\mu_y^2}{2\sigma_y^2} \right) \right] \left\{ \frac{1}{2a^2} + \frac{b\sqrt{\pi}}{2a^3} e^{b^2/4a^2}\, \Phi\left(\frac{b}{\sqrt{2}\, a} \right) \right\} \end{aligned} \tag{A.12}$$

where $a, b,$ and Φ are as defined previously with only the (+) option for b required. One method of verification of Eq. (A.12) is to numerically test the normalization condition given by:

$$\int_0^{1/\alpha} G(F)\, dF = 1$$

This was found to be correct to within the expected numerical accuracy.

References

1. Ely, J., Kouzes, R., Schweppe, J., Siciliano, E., Strachan, D., Weier, D.,"The Use of Energy Windowing to Discriminate SNM from NORM in Radiation Portal Monitors," Pacific Northwest National Laboratory SA 44680, *Nuclear Instruments and Methods in Physics Research A*, 560: 373-387, 2006.

2. Shokair, I., Cohen, D., Mengesha, W., and Walsh, E., "Anomaly Detection for PVT RPM Measurements", Sandia National Laboratories report SAND2012-2081, January 2012.

3. Shokair, I. and Cohen, D., "Anomaly Detection and Optimization of Energy Windows for PVT RPM Measurements", Sandia National Laboratories report SAND2012-2941, July 2012.

4. Secondary Reachback Data Mining Group, "Anthony New Mexico Analyses", Sandia National Laboratories report SAND2012-6622, August 2012.

5. Ogorodnikov, S. and Petrunin, V., "Processing of interlaced images in 4-10 MeV dual energy customs system for material recognition", Physical Review Special Topics – Accelerators and Beams, vol. 5, 104701 (2002).

6. Hinkley, D. V., "On the ratio of two correlated normal random variables", Biometrika, vol. 56, 3, p. 635, 1969.

7. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, Numerical Recipes in C, Cambridge University Press, 1992.

8. Any text on data analysis or statistics, for example, Beers, Yardley, Introduction to the Theory of Error, Addison-Wesley, 1962.

Distribution

1 Department of Homeland Security
Domestic Nuclear Detection Office
Attn: Brian Savage (electronic)

1 Department of Homeland Security
Domestic Nuclear Detection Office
Attn: Guy Hartsough (electronic)

1	MS 9004	James C. Lund	8130	(electronic copy)
1	MS 9159	Noel M. Nachtigal	8958	(electronic copy)
1	MS 9402	Alfredo M. Morales	8126	(electronic copy)
1	MS 9402	Kristin L. Hertz	8126	(electronic copy)
1	MS 9402	William C. Johnson	8126	(electronic copy)
1	MS 9402	Christopher L. Kunz	8126	(electronic copy)
1	MS 9406	Craig R. Tewell	8127	(electronic copy)
1	MS 9406	David Reyna	8127	(electronic copy)
1	MS 9407	Lorraine E. Sadler	8118	(electronic copy)
1	MS 9408	Willard R. Bolton	8137	(electronic copy)
1	MS 9408	Isaac R. Shokair	8137	
1	MS0899	RIM-Reports Management	9532	(electronic copy)

Sandia National Laboratories