*Journal of Radioanalytical and Nuclear Chemistry, Vol. 276, No. 2 (2008) 285-297* -- [PREPRINT]

# On the Detection of Rare, and Moderately Rare, Nuclear Events[a]

L. A. Currie

National Institute of Standards and Technology, Gaithersburg, MD 20899

Some of the more important developments in science and practical demands in commerce have been linked to attempts to detect rare events and rare contaminants, ranging from the early "counting" of solar neutrinos to the occurrence of dodder seeds in clover. For moderately rare events (≈5 to 50 counts) we consider limitations of the Poisson-normal approximation, together with the apparent problem of excessive false positives when a common expression is (mis-)used for detection decisions. For very rare events, rigorous approaches published more than half a century ago are applicable to such current problems as trace actinide contamination and nuclear treaty monitoring.

---

Key words: Rare event detection; international defining expressions; low-level detection capabilities: a consistent approach; $S_C$ as a random variable; Poisson-normal approximation; exact Poisson treatment: spurious false positives; nuclear treaty monitoring; historical perspective

---

## 1. *Introduction*

A principal objective of this article is to present a consistent approach, stemming directly from internationally recommended defining expressions, for the assessment of critical (decision) levels ($S_C$) and detection limits (minimum detectable signals) ($S_D$) in: a) the moderately rare event range, where the Poisson-normal approximation is adequate (nominally 5 counts to 50 counts) [Sect. 4], and b) the rare event range, where the exact Poisson distribution must be used [Sect. 5]. Since the groundwork was laid some 50 years ago (Poisson-normal)[1] to 70 years ago (exact Poisson)[2,3], the article contains also a substantial historical thread.[b]

A key element of the approach, for the moderately rare event range, is explicit consideration of the uncertainty of the (Poisson) variance estimate ($s_B^2$) of the background (B), given the B-count-based degrees of freedom ($\nu$). In this case both $S_C$ and $S_D$ depend on $\nu$, and the latter depends also on the (Poisson) variance function. A related issue is the mysterious occurrence of excessive false positives, said to occur when a "popular expression" for $S_C$ is used in the moderately rare count range -- extending to backgrounds of 100 counts or more! The false positive excess is found to vanish, however, when $S_C$ is treated as a random variable, with significance (detection) tests based on independent (Poisson) variance estimates and the use of Student's-t.

We begin with a cautionary note concerning the validity of the Poisson hypothesis and distributions of blanks in real measurement systems [Sect. 2], followed by a concise review of detection concepts and relevant international standards [Sect. 3].

## 2. IMPACT OF THE BLANK: a cautionary note

While the focus of this article relates to detection capabilities for (moderately) rare events that are taken to be strictly Poissonian, it must be recognized that for real measurement processes such a state is but an ideal asymptote. For low-level (decay, atom) counting, for example, there exists a substantial literature on the topic of "error multipliers" and incremental non-Poisson variance components -- generally manifest by an inflated Index of Dispersion (variance/mean)[5,6]. Fortunately, in most such cases if counts are not too few, the total variance can be estimated reliably by replication ($s^2$), and Student's-t applied for (detection) significance testing (*vide infra*).

More serious, however, is the situation where B-events (background, baseline, blank) lack independence, or where B is non-stationary[7]. A universal example in low-level counting (LLC) is the residual mu-meson ($\mu^-$) background component, which varies with cosmic ray intensity and barometric pressure. Unless "muon leakage" (direct, indirect) is excluded, such variations must be taken into account[7]. For extreme LLC, the ultimate solution to the muon problem is to go deep underground[8,9].

---

[b] Two Poisson-normal expressions for detection decisions ("equality of means"), introduced by Hald (1952),[1] are based on the signal/noise ratio and the Poisson (square-root) variance-stabilizing transformation, both including the half-integer continuity correction for small numbers of counts. This appears to be the earliest use of the techniques later adopted by Altshuler and Pasternack (subsequently by ISO-11929), and Stapleton, respectively. (See MARLAP[4] for discussion of the latter methods.)

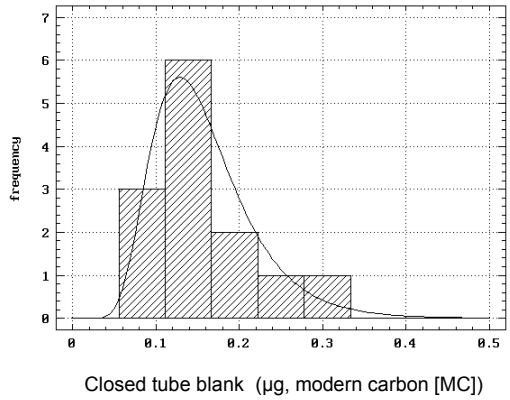



Fig. 1 Empirical distribution of 13 measurements of the NIST $^{14}C$ (modern carbon) closed tube combustion blank (CTB). Estimated parameters of the fitted lognormal distribution are: $\bar{x}$ = 0.158 µg, s = 0.061 µg (modern carbon).

When B derives from sources of contamination, its distribution is likely to be positively skewed. An example is taken from National Institute of Standards and Technology (NIST) studies of $^{14}C$ in remote atmospheric particles, using sub-micromole accelerator mass spectrometry (AMS). In Fig. 1 thirteen observations of the sample preparation (combustion) blank are fitted with a lognormal distribution, but with so few observations the positive tail of the distribution is quite uncertain, as is the form of the distribution. Herein lies the problem: In the absence of a rigorous theoretical basis, or a very large number of observations, uncertainties in the form and quantiles of the blank distribution may be too great for meaningful detection decisions, especially if the distribution is positively skewed. *A recommendation to ameliorate the problem:* Use replicate, paired observations to estimate the net signal ($\hat{S}$) whenever possible; that way approximate normality may be more readily achieved for the null case (when $\mu_S = 0$) [Ref. 7, Appendix].

## 3. DETECTION CAPABILITIES: the Poisson-normal approximation

### *3.1 Defining relations and international standards*

Building on foundations established in the literature of the 1960s, the International Union of Pure and Applied Chemistry (IUPAC)[10] and the International Organization for Standardization (ISO)[11,12] have prepared recently a series of international standards and recommendations on concepts and expressions for detection and quantification capabilities. Relevant national documents have been produced by the American Society for Testing and Materials (ASTM)[13,14] and MARLAP[4]. A recent International Committee for Radionuclide Metrology (ICRM) publication[8] covers issues specific to low-background counting. A very brief summary, using the notation of Ref. 10, follows.

Defining relations:

Critical Value (Level) ($L_C$): $\Pr(\hat{L} > L_C \mid L=0) \leq \alpha$ (1)

Detection Limit ($L_D$): $\Pr(\hat{L} \leq L_C \mid L=L_D) = \beta$ (2)

Quantification Limit ($L_Q$): $L_Q = k_Q \sigma_Q$ where $k_Q = 1/\mathrm{RSD}_Q$ (3)

where α and β represent the hypothesis testing errors of the first and second kinds (false positives and false negatives) -- default value: 0.05 each; and $\mathrm{RSD}_Q$ represents the relative standard deviation at the Quantification Limit -- default value: 0.10. The generic

symbol ($L$) refers to net signals ($S$) or concentrations ($x$) above the background or baseline ($B$). The inequality (Eq. (1)) is required for discrete distributions, since only selected values of α exist in such cases. For this reason we introduce the symbol α' to represent the actual false positive risk, in contrast to the target value α (0.05 or 0.01 in the present text).[c]

*Notation:* In the remaining text we confine our attention to the Poisson signal domain, where the variate $S$ represents net counts (or events), with $Y$ and $B$ representing gross counts and background (or baseline) counts, respectively. In keeping with ISO-3534[15], we use y, b for observed values of the variates $Y$ and $B$, and $\mu_Y$, $\mu_B$ for their respective mean values (expectations).[d] For paired counting the estimate $\hat{S}$ equals ($y$-$b$); for the other asymptotic case (well-known blank), $\hat{S}$ equals ($y$-$\mu_B$). The variance of $S$ for the null state (when $\mu_S$=0) is represented by $\sigma_o^2$, where $\sigma_o^2 = \sigma_B^2$ for the well-known blank, or $2\sigma_B^2$ for paired counting. More generally, $\sigma_o^2 = \eta\sigma_B^2$, where $\eta = (r+1)/r$, with r being the number of background replicates (or ratio of background to gross sample counting times), when an average background correction is applied. Estimated variance is indicated by $s_o^2$, with ν degrees of freedom (ν=n−1, for simple replication). The notation (±) is used to denote standard uncertainties, with the exception of section 4.3.1.

Note that the treatment of "simple counting" considered here applies equally well to the net count resulting from subtraction of the background from the gross count, and to the estimation of the net peak area resulting from subtraction of the estimated baseline (spectroscopic, environmental) from the integrated gross peak.

### *3.2 Simplest realization: the normal distribution*

In the normal, homoscedastic (constant variance) case the defining equations yield the following simplified expressions for $S_C$ and $S_D$, where the expressions to the right of the arrows correspond to α = β = 0.05. The "b" equations apply to the case where variance is estimated by replication (as $s^2$) with ν degrees of freedom (df). Note that in Eq. (5b): (1) uncertainty in $\sigma_o$ renders the signal detection limit uncertain; (2) the approximate expression for the non-centrality parameter ($\delta$), in terms of ν and t, is accurate to within 1 % for α, β = 0.05 and ν ≥ 5.[10] (Here, the actual value of δ is 3.870.)

$$S_C = z_{1-\alpha}\,\sigma_o \rightarrow 1.645\,\sigma_o = 1.645\,\sigma_B\sqrt{\eta} \quad (4a)$$

$$S_C = t_{1-\alpha,\nu}\,s_o \rightarrow 2.015\,s_o = 2.015\,s_B\sqrt{\eta} \quad [5\text{ df}] \quad (4b)$$

$$S_D = S_C + z_\beta\,\sigma_D \rightarrow 2S_C = 3.29\,\sigma_o \quad (5a)$$

$$S_D = \delta_{\alpha,\beta,\nu}\,\sigma_o \approx 2t_{1-\alpha,\nu}\,(4\nu/(4\nu+1))\,\sigma_o \rightarrow 3.84\,\sigma_o \quad [5\text{ df}] \quad (5b)$$

$$S_Q = k_Q\sigma_o = 10\,\sigma_o \quad (6)$$

---

[c] If the assumptions are correct, and the significance (detection) test is performed properly, α' should equal α for continuous variables, and α' ≤ α for discrete variables. (See section 5.) Incorrect assumptions or testing can lead to excessive false positives (α' > α ), as illustrated in sections 4.1 and 4.2.

[d] Although observed values (y, b, ...) are necessarily integers, the expectations are real numbers having the same dimensionless units (counts).

### *3.3 The Poisson-normal approximation*

The asymptotic (large-count) Poisson-normal distribution is taken as a continuous, normal distribution with the special parametric characteristic of the Poisson distribution: equality of the mean ($\mu$) and variance ($\sigma^2$)[e]. Three consequences are: (1) the variance function (heteroscedasticity) must be taken into account in deriving $S_D$ and $S_Q$; (2) $\sigma_B^2$, which equals $\mu_B$, can be estimated as $s_B^2 = b$ (observed background counts); (3) the confidence interval for $\mu_B$, given $b$, is now asymmetric (Section 4.3.1). For $\sigma$-known, $\alpha=\beta=0.05$, $\eta=2$ (paired counting), and $k_Q=10$, the resulting *asymptotic* expressions are

$$S_C = z_{1-\alpha}\,\sigma_o = z_{0.95}\sigma_B\sqrt{\eta} \rightarrow 1.645\,\sqrt{(2\mu_B)} = 2.326\,\sqrt{\mu_B} \qquad (7a)$$

$$S_D = S_C + z_{1-\beta}\,\sigma_D \rightarrow z_{0.95}^2 + 2\,S_C \rightarrow 2.71 + 4.65\,\sqrt{\mu_B} \qquad (8a)$$

$$S_Q = (k_Q^2/2)\,(1 + [1 + 4\sigma_o^2/k_Q^2]^{1/2}) \rightarrow 50\,(1 + [1 + 2\mu_B/25]^{1/2}) \qquad (9)$$

Asymptotic expressions that ignore the effects of heteroscedasticity can be given for $S_D$ ($3.29\,\sigma_o = 3.29\,\sigma_B\,\sqrt{2} = 4.65\,\sigma_B$) and $S_Q$ ($10\,\sigma_o = 10\,\sigma_B\,\sqrt{2} = 14.1\,\sigma_B$). These expressions are correct to within 10 % of the complete expressions (Eqs. (8a), (9)) for $\mu_B > 33.8$ counts (67.6 counts for $\alpha$, $\beta = 0.01$), and $\mu_B > 1371.9$ counts, respectively.[f]

### *3.4 "Systematic" uncertainties in $S_C$, $S_D$, and $\alpha$*[g]

When a *fixed* estimate $b$ is substituted for $\mu_B$ in Eqs. (7a) and (8a), the relative uncertainties of $S_C$ and the asymptotic term ($2S_C$) of $S_D$ are equal to the relative uncertainty ($u_r$) of $\sqrt{b}$ -- i.e., $(u(\sqrt{b})/\sqrt{\mu_B})$. The numerator reflects the well known *variance stabilizing transformation* of Poisson variates, corresponding to a constant standard uncertainty of 0.5. Thus, to achieve $u_r \leq 10$ %, requires $\mu_B \geq 25.0$ counts. To treat the critical level as a fixed discriminator in this case, however, it is interesting to project the uncertainty onto $z_{1-\alpha}$ and ultimately, $\alpha$. The requisite value of $u_r(\sqrt{B})$ is then redefined in terms of the acceptable uncertainty in $\alpha$, as manifest by the relative uncertainty of $z_{1-\alpha}$. Constraining $\alpha$ to a standard uncertainty interval of 0.04 to 0.06, for example, corresponds to $z_{1-\alpha}$: 1.555 to 1.751 -- i.e., $u_r(z) = u_r(\sqrt{B}) \approx 6$ %. To achieve such a limit for the $\alpha$-uncertainty requires $\mu_B \geq (2(0.06))^{-2}$, or 69.4 counts -- consistent with the author's recommendation of some 35 years ago [Sect. 5, Ref. 21].

## 4. MODERATELY RARE (≈ 5 TO 50) BACKGROUND EVENTS[h]; $S_C$ as a Random Variable

*Critical value ($S_C$).* When replication variance is estimated as $s^2$ with (n-1) df, as in Eq. (4b), reliable detection decisions can be made for moderate numbers of counts

---

[e] An improved approximation, for small numbers of counts (n), includes a half-integer "continuity correction" $n \rightarrow (n+½)$ [Ref. 6, section 2.2].

[f] The minimum $\mu_B$ value for the $S_Q$ asymptote represents a small correction to the value (1250 counts) given in footnote-c of Table II in Currie [1968][16]. The corrected value for the well-known blank case ($\eta$=1) is 2743.8 counts. Non-significant digits are included here to illustrate the calculations and to emphasize the fact that the $\mu$'s are real numbers.

[g] The Poisson quantification limit is not considered further, since $S_Q$ requires a *minimum* of 100 counts (events) [Eq. (9)]., which is generally beyond the scope of "rare and moderately rare" events.

[h] For perspective on the nominal range, see Section 4.3.

using $t_{1-\alpha,\ n-1}$. Alternatively, for counting data, an independent estimate of the variance (counting variance) is given by the number of counts -- i.e., for the background $s_B^2=b$, with $\nu=2s_B^2$ degrees of freedom. In this case (also) $S_C(b)$ is a random variable, as shown in Eq. (7b).

$$S_C = t_{1-\alpha,\nu}\, s_o = t_{1-\alpha,\nu}\, s_B\sqrt{\eta} = t_{1-\alpha,\nu}\sqrt{(\eta b)} \rightarrow t_{0.95,\,2b}\sqrt{(\eta b)} \qquad (7b)$$

For $\eta=2$ (paired comparison) and b = 10 counts, for example, $S_C = t_{0.95,20}\sqrt{(2b)} = 2.44\sqrt{b} =$ 7.71 counts. More generally, especially for very few counts, the background variance may be estimated as $(b+\varepsilon)$ where $\varepsilon$ is a small correction term.[i]

*Detection Limit ($S_D$).* When $\sigma_o^2$ is estimated as $s_o^2$, but $\sigma$ varies with signal magnitude, then the simple expression for $S_D$ (Eq. (5b)) does not apply. In lieu of the rigorous treatment of the heteroscedastic non-central-t problem[11] [also, Ref. 4, Sect. 20A.3.1], when $\alpha = \beta$ one may use the following approximation[17], which is conservative but accurate to within a few percent when $\alpha, \beta = 0.05$ and $\nu \geq 5$.

$$S_D \approx (\delta/2)\,\sigma_o + (\delta/2)\,\sigma_D = \delta\,\bar{\sigma} \qquad (8b)$$

In the general case, an iterative solution involving the variance function is required, since $\sigma_D$ is a function of $S_D$. For a pure Poisson process, where $\sigma_D^2 = S_D + \sigma_o^2$, an algebraic solution obtains,

$$S_D \approx (\delta/2)^2 + \delta\,\sigma_o \qquad (8c)$$

Although the approximation (Eq. 8b) is reasonably accurate, $S_D$ can have substantial uncertainty when $\sigma_o$ is estimated as $s_o$; a convenient approximation for the relative standard uncertainty of the latter is $1/\sqrt{(2\nu)}$, or about 20 % for 12 degrees of freedom.

### *4.1 The false positive dilemma*

When $\mu_B$ is large, the Poisson-normal expression for $S_C$ (Eq. 7a) is commonly replaced by the approximate relation

$$S_C \approx z_{0.95}s_o = z_{0.95}s_B\sqrt{\eta} = 1.645\sqrt{(2b)} = 2.326\sqrt{b} \qquad (7c)$$

Eq. (7c) is an excellent approximation to the correct Poisson-normal expression (7b) for large b; and it might be considered also for b as small as 15 counts (30 degrees of freedom), considering the convergence between $t_{0.95,\nu}$ and $z_{0.95}$. (For few counts observed, use of $b+\varepsilon$ is recommended.) Surprisingly, however, use of expression (7c), said to have been popularized by Currie (1968),[16] has been reported in the literature to result in unacceptable levels of false positives, even for $\mu_B$ as large as 100 counts [Ref. 4, Sect. 20.4, 20.A.2, and references therein.].

---

[i] Degrees of freedom ($2s_B^2$) for the t-test, and recommended values for $\varepsilon$ (ranging from 0 to 1), are based in part on Ref. 4 recommendations [Attachment 19D], and expressions for the confidence interval of the Poisson parameter (mean, variance)[18]. For the purpose of this article, we take $\varepsilon = 0.5$, which is equal to the asymmetry (heteroscedasticity) correction for the central ("1σ") confidence interval of the Poisson parameter, derived from the Poisson-normal approximation. (See section 4.3.1.)

To explore the nature of this apparent problem, we first reformulated expressions (7b) and (7c) to standard form, to give the respective expressions (10b) and (10c) for the critical values of the detection test statistics ($t_{P,\nu}$, $z_P$, respectively), *as applied*.

$$(\hat{S}/s_o)_C = [(y-b)/(s_B\sqrt{\eta})]_C = [(y-b)/\sqrt{2(b'+\varepsilon)}]_C = t_{0.95,\,2(b'+\varepsilon)} \quad (10b)$$
$$(\hat{S}/s_o)_C \approx [(y-b)/(s_B\sqrt{\eta})]_C = [(y-b)/\sqrt{(2b)}]_C = z_{0.95} \quad (10c)$$

where b' (or, b'+ε) represents an *independent estimate* of the variance of the blank.

There are three notable differences. (1) The commonly-applied expression (10c) carries the assumption that the test ratio ($\hat{S}/s_o$) is normally distributed, with its upper critical value given as the 95$^{th}$ quantile of the normal variate (z) rather than that of the t-variate. (2) Expression (10b) provides improved estimates for the very small count case, when used with a non-zero value for ε. This averts the "zero catastrophe" (unbounded or indeterminate values of the test ratio and zero df for t), which, however, can be minimized by restricting the Poisson-normal detection test to $\mu_B$'s such that the probability of zero counts -- $\exp(-\mu_B)$ -- is negligible; for example, $\mu_B > \ln 1000 = 6.91$ counts, or $\mu_B > 3.45$ counts if 2b is replaced by the sum of two independent b-values (b'+b''). (3) Most serious is the subtle application of the *same observed value* of the background counts (b) in the numerator and denominator of the test ratio in (10c).[j] Unless independent b-observations are used for the estimate of the null signal (numerator), and the estimate of its variance (denominator), *the assumption of normality is invalid*, and neither $z_P$ nor $t_{P,\nu}$ can be expected to give correct false positive rates. (The problem is greatly reduced for large counts; and it does not arise in the conventional application of the t-test (Eq. 4b), because the replication variance estimate ($s^2$) is necessarily independent of the estimated mean.)

The last point is crucial: The failure of the approximate expression (10c) for low to moderate numbers of background counts is *not* primarily a result of a faulty expression for $S_C$ (Eq. 7c), but rather the manner in which it is commonly applied, with *b* redundancy in numerator and denominator (Eq. 10c).

### *4.2 Empirical ($\hat{S}/s_o$) distributions and false positive functions*

*Empirical distributions* of the correct (independent-b, $t_{P,\nu}$) (Eq. 10b) and the approximate (dependent-b, $z_P$) (Eq. 10c) test ratios ($\hat{S}/s_o$) were generated by Monte Carlo sampling of Poisson variates as a function of $\mu_B$ -- noting that for the independent t-test for a given $\mu_B$, the critical value $t_{1-\alpha,\nu}$ had to be adjusted for each "observed" background count b', according to the relation $\nu = 2s_B^2 = 2(b'+\varepsilon)$. Fig. 2 shows a 4-way comparison of empirical distributions (N = 2000 each) for (continuous) Poisson-normal variates and discrete Poisson variates, for both independent (Eq. 10b) and dependent (Eq. 10c) cases, for a moderately small value for $\mu_B$ (8.52) -- selected to be a non-integer, and small, yet large enough that the probability that b=0 would be negligible.

---

[j] According to Ref. 4 [Sect, 20.A.2], use of the *same observed value* b for both the estimated net signal and its estimated variance is a common practice, even for low-level counting, despite the correlation introduced.

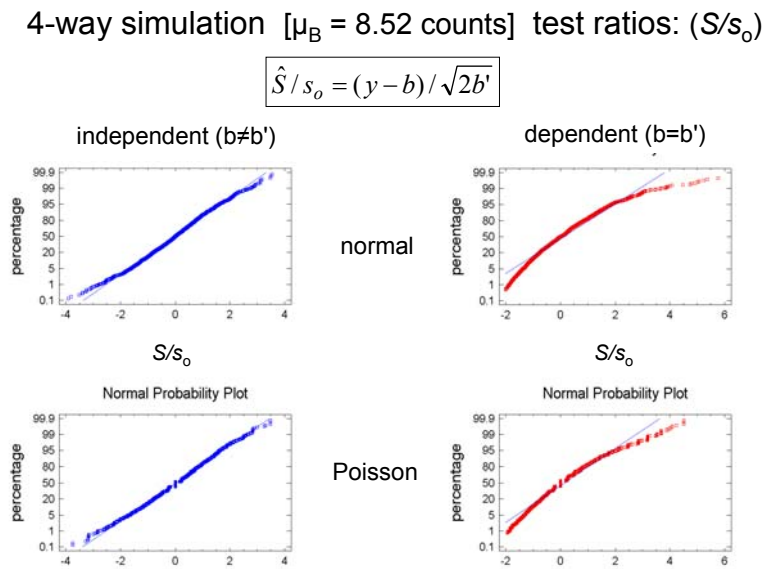



Fig. 2 Normal probability plots of empirical distributions of the test ratio $\hat{S}/s_o = (y-b)/\sqrt{2b'}$ for the null case, for Poisson-normal data (top pair) and Poisson data (bottom pair) . Numerator and denominator are independent for the plots at the left (b ≠ b'); they are not, for the plots at the right (b = b'). Each data vector (y, b, b') comprised N = 2000 computer-generated random samples from probability distributions with expectation ($\mu_B$) 8.52 counts.

It is evident from the plots that: the independent-b ratios are generally similar to t-distributions, while the dependent-b ratios show a pronounced positive skew, consistent with excessive false positives, and certainly not normal as had been assumed in the application of $z_{0.95}$ as the test statistic in Eq. (10c). Also, the shapes of the (Poisson-normal, and discrete Poisson) data distributions are generally the same, with the exception of the pronounced granularity in the Poisson case.

The conclusion is clear: Whether the distribution is discrete (Poisson) or continuous (Poisson-normal), at the moderate count level of Fig. 2 (8.52 counts) the use of independent background counts in the t-test (α=0.05) gives reasonable false positive rates (α' ≈ 0.049±0.002), whereas use of dependent $\hat{S}/s_o$ ratios yields nearly doubled false positive rates (α' ≈ 0.085±0.003). Increased background counts, however, result in an asymptotic approach to normality (Fig. 3), with a reduction (1) in granularity and excessive false positives, and (2) in the difference between the actual critical value ($k_{0.95}$) and that of the normal variate ($z_{0.95}$). Although the latter difference is reduced to about 10 % for $\mu_B$ = 50.0 counts, actual false positives (α') are still needlessly inflated at 0.062 (for α=0.05) and 0.019 (for α=0.01).

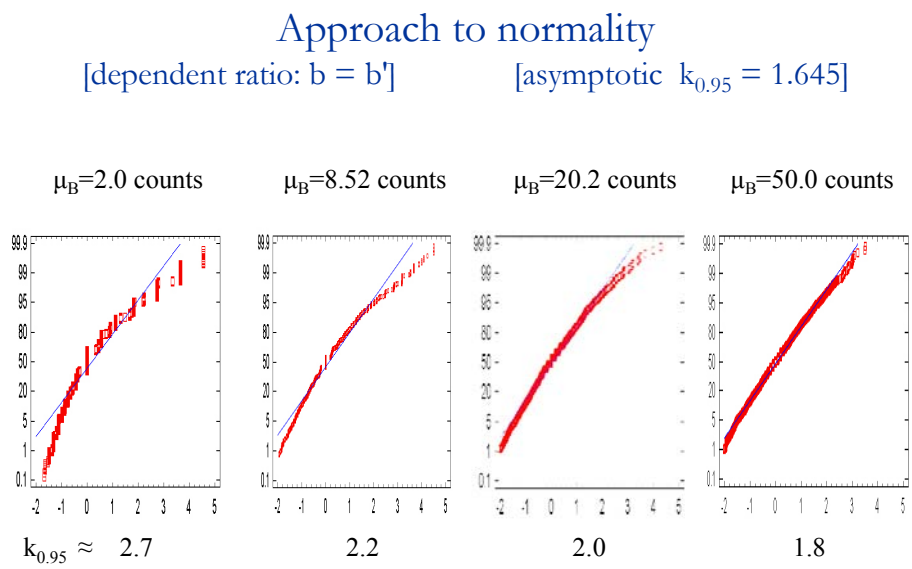



Fig. 3 Empirical (N = 2000) dependent, Poisson test ratios and 95th percentile estimates ($k_{0.95}$) vs $\mu_B$. As $\mu_B$ increases, skewness and granularity decrease, and $k_{0.95}$ approaches the normal asymptote ($z_{0.95}$). (Abscissa is $\hat{S}/s_o$ ; ordinate is the cumulative percentage.)

*False positive (α') functions* for the moderate range of background events are displayed in Fig. 4 for the target α equal to 0.05 and 0.01, respectively -- covering the range: $\mu_B$ = 2.0 counts to $\mu_B$ = 50.0 counts. The three curves in each figure correspond to the "correct" independent t-test (**o:**curves), the independent z-test (+:curves), and the common, but incorrect, dependent z-test (**x**:curves). It is clear that the independent t-test gives acceptable results over the entire moderate count range, with the independent z-test showing relatively small bias (≈10%) for $\mu_B$>15 counts (α=0.05) or 50 counts (α=0.01). (These values derive from the relation: $t_{1-\alpha',\nu} = z_{1-\alpha}$ where α'/α is the bias ratio.) The common, but faulty (dependent) z-test (**x**:curves), however, produces excessive false positives over the entire range. For the best performance of the independent-t test for the smallest background counts ($\mu_B$<10 counts), $\sigma_o^2$ was estimated as (b'+b"+ε) rather than 2(b'+ε), as discussed in the following section.

Consistency of the observed α' with the target value (0.05, 0.01) for the t-curves (**o:**curves) is seen for all points in Fig. 4 for $\mu_B$≥5.0 -- i.e., $\mu_B$ = 5.0, 8.52, 20.2, 35.4, 50.0. In fact, the observed α'-means for the two sets of five points are: 0.0498±0.0010 (α=0.05) and 0.0097±0.0004 (α=0.01).

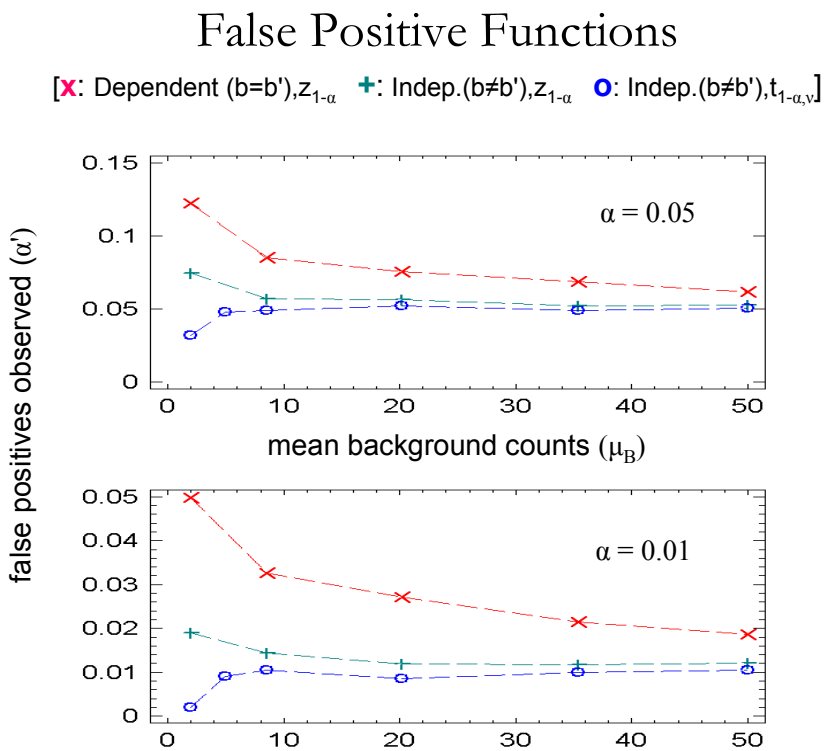



Fig. 4 False positive functions (with standard uncertainties ≈ symbol size): α' vs $\mu_B$, for the target α equal to 0.05 (or 0.01) and three ways of performing the detection test, as discussed in Section 4.2 ($N \geq 10^4$ per point.). The upper, red (**x**) curves represent spurious false positives because of the lack of independence between the observed differences $\hat{S}$ and the variance estimates $s_o^2$ -- both being based on the same background estimates (b = b'). The other curves are based on independent test ratios (b ≠ b') and $z_{1-\alpha}$ (green, + curves) or $t_{1-\alpha,\nu}$ (blue, **o** curves). While $z_{1-\alpha}$ is asymptotically correct, $t_{1-\alpha,\nu}$ must be used for small numbers of counts. For the correct, t-curves (**o**) all points between $\mu_B$=5.0 and $\mu_B$=50.0 are consistent with the target α (0.05, 0.01).

### *4.3 At what point must we abandon the Poisson-normal approximation?*

There is no unique answer to this question, but a number of criteria lead to similar results, generally consistent with the notion that the approximation " ... is quite accurate enough even below μ = 10 counts." [Ref. 6, p. 22]. Apart from the discrete-continuous dichotomy, perhaps the most serious difference lies in the skewness of the Poisson distribution vs the symmetry (and negative tail) of the Poisson-normal distribution for very small $\mu_B$, setting limits to the validity of the t-test. Considering the expression (Eq. 7b) for $S_C$ as $t_{1-\alpha,\nu}s_o$ or $[t_{1-\alpha,\nu}s_B\sqrt{2}]$ with variance estimates $s_o^2 = (b'+b''+\varepsilon)$ or $[s_B^2 = (b'+\varepsilon)]$, we would require that the $\Pr(b'+b''\leq 0)$ or $[\Pr(b'\leq 0)]$, given $\mu_B$, be negligible for both distributions (equality for Poisson, inequality for Poisson-normal). If "negligible" is taken as $10^{-3}$, then the replicate background (b'+b'') with expectation $2\mu_B$ corresponds to minimum $\mu_B$ values of (ln 1000)/2 = 3.45 counts (Poisson) and $z_{0.001}^2/2 = (-3.09)^2/2$ = 4.77 counts (Poisson-normal). For the single background count (b' with expectation $\mu_B$),

twice as many counts would be required: $\mu_B$ = 6.91 and 9.55 counts, respectively. The empirical (independent-t) false positive functions in Fig. 4, which use (b'+b") for $\mu_B < 10$ counts, are consistent with these bounds, showing reasonable agreement with the target α's (0.05, 0.01) for mean backgrounds of 5.0 counts and above. Divergence below that point (e.g., at $\mu_B$ = 2.0 counts) yields conservative results (too few false positives) -- an effect that becomes more pronounced for smaller target α's.

### *4.3.1 Confidence intervals (CI) for μ, given b*

Comparison of Poisson-normal vs exact Poisson confidence intervals (1) serves as an objective guide to the effect of α on the transition point, and (2) provides the basis for the rigorous Poisson treatment of detection capabilities in the next section.

*The Poisson-normal confidence interval* is the solution to the two-sided equality, $(\mu_- + z_P\sqrt{\mu_-}) = b = (\mu_+ - z_P\sqrt{\mu_+})$, having a (2P-1) two-sided confidence level with limits $\mu_-$, $\mu_+$. The commonly-used first order solution is $\mu_\pm = b \pm z_P\sqrt{b}$, . The second order solution corrects for heteroscedasticity bias with the addition of the term $z_P^2/2$ -- the basis for our having selected ε = 0.5 for the midpoint (b+ε) of the "central" (1σ) confidence interval for $\sigma_B^2$. The full solution, given in Eq. (11), can be further improved with a continuity correction [b→(b−0.5) for $\mu_-$, and b→(b+0.5) for $\mu_+$ ].

$$\mu_\pm = b + z_P^2/2 \pm z_P\sqrt{(b + z_P^2/4)} \qquad (11)$$

*The exact Poisson confidence interval* was derived by Garwood[18] [1936], who included a table of 2-sided limits for α = 0.05 and 0.01, and b = 0 to 50. A preferred formulation of Garwood's solution is

$$\mu_\pm = (s_B^2(\chi^2/\nu)_{P,\nu})_\pm \qquad (12)$$

where $\nu = 2s_B^2$ with lower and upper limits based on $[s_B^2=b;\ P=\alpha]_-$, $[s_B^2=(b+1);\ P=1-\alpha\ ]_+$, respectively.

To illustrate, taking b=9 counts and α=0.05, we find ($\mu_-$, $\mu_+$) limits for 90 % confidence intervals to be (5.24, 15.47) [Eq. (11)], and (4.70, 15.71) [Eq. (12)]. The two types of CIs differ primarily in their lower limits, which agree to within about 10 % for b≥10 (α=0.05) and b≥16 (α=0.01) -- corresponding to Poisson $\mu_-$'s of 5.43 and 8.18. Including the continuity correction reduces the "10 % b-values" to 4 and 10, respectively.[k]

## 5. DETECTION OF RARE EVENTS: some ancient (and modern) history

The final segment of this article treats the "rare" count region, within which the Poisson-normal approximation is no longer adequate, and where the granularity (discreteness) is extreme. As a result of the broad importance of rare event detection in

[k] The first few μ's from Eq. (12) for b = 0, 1, 2, 3 are: [α = 0.05] $\mu_-$ = 0, 0.0513, 0.355, 0.818; $\mu_+$ = 2.996, 4.74, 6.30, 7.75; [α = 0.01] $\mu_-$ = 0, 0.0101, 0.149, 0.436; $\mu_+$ = 4.61, 6.64, 8.41, 10.05. Such solutions to Eq (12) are central to the construction of critical values and detection limits for the exact Poisson treatment in section 5.1 (and Fig. 5.)

numerous disciplines and eras, formulation of exact Poisson approaches appeared many years ago. As early as 1937 research was published on the use of a modified-Bessel function to describe differences between two Poisson variates[19]; such work continues today[20]. The accounts of rare event detection (decisions and capabilities) discussed here, representing the two limiting cases: (well-known background) and (paired counts), are based on publications from 1972[21] and 1939[3], respectively. The stimuli for these early works, indicated in Sections 5.1 and 5.2, were both important and diverse. Current applications are no less so, including trace radionuclide monitoring in connection with the Comprehensive Nuclear Test Ban Treaty[22], and the search for extremely rare events in particle physics[9]. (See Section 6.)

### *5.1 Exact Poisson Case-I: well-known background*[21]

The stimulus for the method presented here was the need to assess detection capabilities for very low levels of noble gas radionuclides, including $^{37}$Ar produced by solar neutrinos[23]. Knowledge of the background makes the development quite straightforward and amenable to convenient graphical and tabular solutions, and broadly applicable to extreme low-level counting systems having stable backgrounds or baselines. Given the expectation for the background counts ($\mu_B$), one can utilize the cumulative Poisson distribution to calculate the critical number of (gross) counts $y_C$, considering the error of the first kind ($\alpha$), and then the gross count detection limit $y_D$, considering the error of the second kind ($\beta$). These values follow from the defining equations (1) and (2), adapted to this special case.[1]

Critical value: $\Pr(y>y_C \mid \mu_Y=\mu_B) \leq \alpha$ (13a)

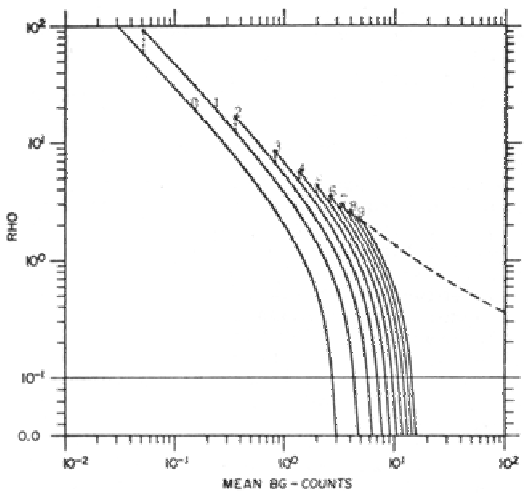
Detection limit: $\Pr(y\leq y_C \mid \mu_Y=y_D) = \beta$ (13b)

Derived relations for the net signal are then: $S_C = y_C - \mu_B$, and $S_D = y_D - \mu_B$.

As a convenient overview, a combined graphical-tabular representation of the results is presented in Fig. 5. For the table, the row in which $\mu_B$ falls (col. 2) is used to find $y_C$ (col. 3) and $y_D$ (col. 4). For the graph, the envelope of the saw-tooth curve represents the relation between $\mu_B$ (abscissa) and the background equivalent activity ratio, RHO = $S_D/\mu_B$ (ordinate). Corresponding $y_C$ values are given by the integers above the curve.

[1] Alternatively, $y_C$ and $y_D$ can be derived from Eq. (12), taking advantage of the duality between confidence intervals and significance tests.[25]

Exact Poisson-I (well-known blank)
graphical & tabular critical levels ($y_C$) and detection limits ($y_D$)




| $\alpha$ (minimum)[a] | $\mu_B$ (range) | $y_C = S_C + \mu_B$ ($\alpha \leq 0.05$) | $y_D = S_D + \mu_B$ ($\beta = 0.05$) |
|---|---|---|---|
| -- | 0 - 0.051 | 0 | 3.00 |
| 0.0013 | 0.052 - 0.35 | 1 | 4.74 |
| 0.0060 | 0.36 - 0.81 | 2 | 6.30 |
| 0.0089 | 0.82 - 1.36 | 3 | 7.75 |
| 0.013 | 1.37 - 1.96 | 4 | 9.15 |
| 0.016 | 1.97 - 2.60 | 5 | 10.51 |
| 0.018 | 2.61 - 3.28 | 6 | 11.84 |
| 0.020 | 3.29 - 3.97 | 7 | 13.15 |
| 0.021 | 3.98 - 4.69 | 8 | 14.44 |
| 0.022 | 4.70 - 5.42 | 9 | 15.70 |

a. For each $\mu_B$ range, $\alpha$ varies monotonically from $\alpha_{min}$ to 0.05

Fig. 5 Exact Poisson-I (well-known blank) critical levels and detection limits[21] ($\alpha$, $\beta$ = 0.05). Integers in the graphical and tabular representations are critical values of gross counts ($y_C$). Abscissa and column-2 are background counts ($\mu_B$). The gross count detection limit ($y_D$) is shown in column-4. The saw-tooth curve links $\mu_B$, $y_C$ (integers above the curve), and the background equivalent net count detection limit (RHO = $S_D/\mu_B$).

### *5.1.1 Example: application to low-level detection of $^{85}$Kr in the atmosphere*[24].

To illustrate, we consider the ($\alpha$, $\beta$ = 0.05) detection capabilities of the NIST low-level 5 mL gas GM/proportional counting system in an hypothetical 85 min screening experiment for small samples of atmospheric $^{85}$Kr. The pertinent characteristics of the NIST system are the "well-known" background rate (1.2 $h^{-1}$) and the $^{85}$Kr counting efficiency (Eff = 0.65). As a result, $\mu_B$ = 1.7 counts. Referring to Fig. 5, we find: that the corresponding value for $y_C$ is 4 counts (integer), that the false positive risk ($\alpha'$) falls within the range 0.013 to 0.050[m], and that the gross count detection limit $y_D$ is 9.15 counts. Thus, $S_D$ = (9.15-1.70) = 7.45 counts; and the corresponding detection limit for $^{85}$Kr equals $S_D/(0.65 \cdot 85)$ disintegrations per minute, or 2.25 mBq. If y=1 count were observed in an 85 minute screening experiment, we would conclude "n.d." (not detected), with a 90 % (gross count) confidence interval of (0.051, 4.74) counts -- corresponding to an upper limit of 0.92 mBq $^{85}$Kr. Note that the alternate graphical solution (Fig. 5) permits one to see at a glance that $\mu_B$ = 1.7 counts intersects the curve at $y_C$ = 4 counts; and that a horizontal line extended from that point meets the ordinate at RHO = $S_D/\mu_B$ = 4.4. Thus, the minimum detectable activity is 4.4 times the background equivalent

[m] The actual value ($\alpha'$ = 0. 030) equals Pr(y>4 counts| $\mu_Y$=1.70 counts).

activity -- i.e., 4.4 (0.51 mBq) = 2.24 mBq. If more stringent criteria were required (e.g., $\alpha$, $\beta$ = 0.01; and 98 % CI), application of Eqs. (12) and (13) give (for $\mu_B$ = 1.70 counts): $y_C$ = 5 counts and $y_D$ = 13.11 counts. The latter corresponds to a minimum detectable net signal of 11.41 counts, which is equivalent to a minimum detectable $^{85}$Kr activity of 3.44 mBq. If y=1 count were observed (in 85 min), the result remains "n.d.", but the more conservative confidence limits (0.01, 0.99) correspond to 0.010, 6.64 gross counts, such that the upper limit for $^{85}$Kr activity would equal 1.49 mBq.

### *5.2 Exact Poisson Case-II: paired sample, background counts*[3]

The stimulus for the development of a rigorous treatment for the second asymptotic case, involving the comparison of two Poisson variables, was totally different: the need to detect rarely occurring dodder seeds in large amounts of clover.[n] This was a problem of some practical importance, however, because the contaminating seeds belong to the class of twining herbs that are parasitic to plants. It is noteworthy that dodder seed research is still of considerable interest: An article by J.B. Runyon, et al. in the 29 September 2006 issue of the journal *Science* (v. 313, p. 1964) presents new insight into the role of "volatile chemical cues" in guiding a parasitic plant (dodder) to a host (tomato plant). Such directed growth is essential for survival; the rootless parasite must locate and attach to a host within a just a few cm and a few days, or it dies. In the context of the present article, the detection of the rarely occurring dodder seeds is the analog of the detection of rarely occurring nuclear particles or decays.

Unlike the previous section, where the challenge was to detect a significant signal above a well-known background, the challenge here is to detect a significant difference between two low-level Poisson variables. The sample space is now 2-dimensional, with a critical *boundary* replacing the critical *level* of the single Poisson variable. Przyborowski and Wilenski formulated the problem by first expressing the joint probability law for observations x, y as

$$\Pr(x,y \mid \mu_x,\mu_y) = (\mu_x^{\,x}\cdot\mu_y^{\,y}/x!y!)\cdot\exp(-(\mu_x+\mu_y)) \qquad (14)$$

where, in terms of low-level counting, x and y represent background counts (b) and gross counts, respectively; $\mu_x$ and $\mu_y$ are their expectations.

Eq. (14) can be transformed into a more interesting form (Eq. 15) using the following substitutions: $n = x + y$, $\mu = \mu_x+\mu_y$, $\rho = \mu_y/(\mu_x+\mu_y)$.

$$\Pr(x,y \mid \rho, \mu) = [(\mu^n/n!)\cdot\exp(-\mu)][(n!/(y!(n-y)!))\rho^y(1-\rho)^{n-y}] \qquad (15)$$

*Critical region.* The critical, or rejection, region (w) is defined on a 2-dimensional (integer) grid of the possible sample points E(x,y) lying beyond the critical boundary. For a given n, the partition into y and x = n - y is governed only by the second factor in Eq. (15), which is a term in the binomial expansion of $[(1-\rho)+\rho]^n$. For the null hypothesis, $\mu_y = \mu_x$, so $\rho$ = ½. Thus, for each n, taking $\alpha$ = 0.05, the critical value for y is

---

[n] The contamination level of concern (to the International Seed Testing Association, in the mid-1930s), was in the range of a few dodder seeds per 100 g sample of clover seeds (≈5 x $10^4$) -- i.e., a number concentration of ca. $10^{-4}$. Observed number concentrations of the contaminant seeds varied from 0 to 36 per kg clover[2]

given by $\Pr(y>y_C \mid n, \rho=0.5) \leq 0.05$, independent of $\mu$.[o] The $y_C$ are simply 1-sided critical values for proportions; the critical boundary is their union. If n = 12 counts, for example, the integer $y_C$ derives from the 95+ percentile of the binomial distribution F(12, 0.5), which equals 9 counts. (In this particular case, $\alpha' = 0.019$.) In Fig. 6, for $x \leq 10$ and $\alpha' \leq 0.05$, the critical region corresponds to the 2-dimensional integer space above the dashed curve, which links the $y_C$ points.

*Detection limit.* Evaluation of the detection limit requires that the full probability equation (15) be considered, with the power function (1-β, given α) given by

$$P\{E \in w \mid \rho, \mu\} = \sum_{n=0}^{\infty} (\mu^n/n!)\cdot\exp(-\mu) \sum_{w(n,\alpha)} (n!/(y!(n-y)!))\rho^y(1-\rho)^{n-y} \qquad (16)$$

where ($E \in w$) refers to all observable points E(x,y) that lie within the critical region w. In Ref. 3, on the basis of Eq. (16), contours of fixed power (given α) are presented as a function of the expectations $\mu_x$, $\mu_y$. From these data, taking α and β to be 0.05, we derived the minimum detectable gross count ($y_D$) function, shown as the solid curve in Fig. 6. For comparison, $y_D$ values, taken from MARLAP's "true values of $S_D$" tabulation [Ref. 4, Table 20.3 (final column)] are shown as points falling just below the $y_D$ line in Fig. 6. (MARLAP values are tabulated as $S_D$ [last column]; MARLAP $y_D$'s were calculated by adding the corresponding $\mu_B$'s [first column of Table 20.3].)

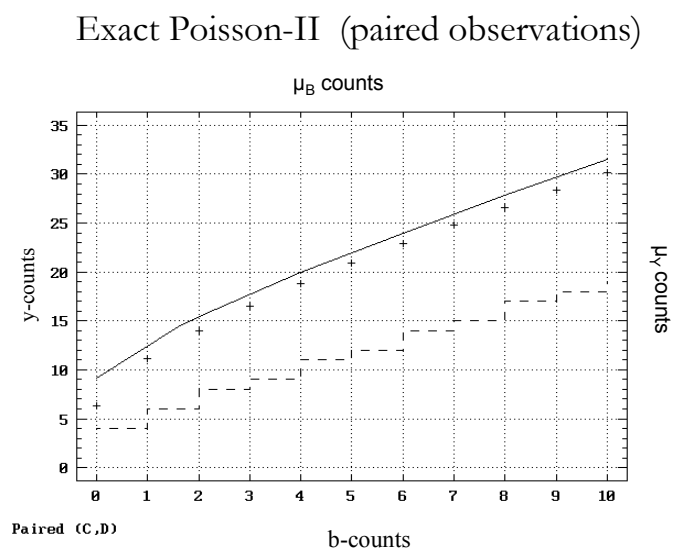



Fig. 6 Exact Poisson-II (paired observations) critical region and detection limits[3] (α, β = 0.05). The dashed curve is the critical boundary, linking critical value points to the discrete observables x (b-counts), and y (gross counts). (The critical, rejection region [*w*] lies above the dashed curve.) The solid curve represents the (gross count) detection limit ($y_D$) as a function of the continuous variable $\mu_B$. Selected $y_D$ values (+) from Ref. 4 [Table 20.3 (last column, $S_D$) plus (first column, $\mu_B$)] are given for comparison.

[o] There is a subtle difference in critical values in Ref. 3, where the significance test was equivalent to $\Pr(y \geq y_C \mid n, \rho=0.5) \leq 0.05$. The critical region w, however, is unaffected; and the $y_C$'s given here are consistent with the test as defined above ($y>y_C$). Also: ρ, as used here, is the complement of the ρ in Ref. 3.

### *5.2.1 Application*.

The most stringent application of Fig. 6 occurs when the one observes zero background counts (b), as sometimes occurs in α-particle monitoring. In such a case for $\alpha = 0.05$, even as many as 4 gross "sample" counts would not be interpreted as significant; and for a near zero "true" background ($\mu_B$) the minimum detectable signal ($y_D \approx S_D$, in this case) would be ≈ 9.1 counts.

Reconsidering the $^{85}$Kr example for paired counting, we find from the solid curve in Fig. 6, for $\mu_B = 1.7$ counts, that $y_D = 14.6$ counts, corresponding to an increase in the minimum detectable activity to 3.9 mBq $^{85}$Kr. To further illustrate nature of this approach, we show in Fig. 7 (x,y)-projections of bivariate frequency histograms for the null case ($\mu_x = 1.7$ counts, $\mu_y = 1.7$ counts), and for $^{85}$Kr at the detection limit ($\mu_x = 1.7$ counts, $\mu_y = 14.6$ counts). Each of the plots is based on N = 2000 random (Monte Carlo) samples for both variables. Resulting from these numerical experiments were α' and β estimates of 0.0045±0.0015 and 0.048±0.005, respectively -- consistent with the defining expressions ($\alpha' \leq 0.05$, $\beta = 0.05$), and with the significant decrease in false positive risk for very few (expected) background counts.[p]

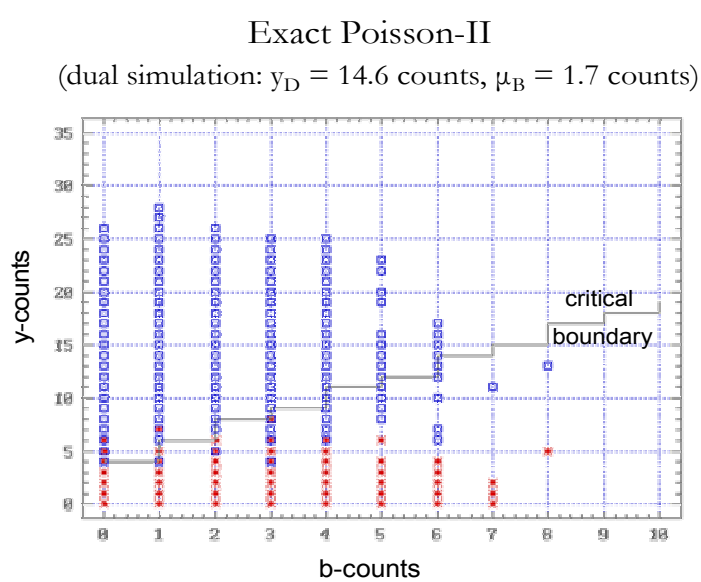



Fig. 7 Exact Poisson-II: Projections are shown for empirical (N=2000) bivariate distributions (α, β = 0.05): 1) for the blank (*) [$\mu_Y = \mu_B = 1.7$ counts] showing false positives: $\hat{\alpha} = 0.0045$; and 2) for the gross count detection limit (□) [$\mu_Y$ = 14.6 counts, $\mu_B$ = 1.7 counts] showing false negatives: $\hat{\beta} = 0.048$.

### *5.2.2 Rare event detection generalized.*

The approach developed in Ref. 3 (Fig. 6) is applicable to a wide array of rare (Poisson) events, beyond radioactive decay -- such as accidents, unusual contamination incidents, equipment or material failures, etc.[6] A recent case in point involves laboratory accidents, where in sequential years the number of accidents reported in a specific laboratory were 2 (x) and 4 (y), respectively. If the Poisson model is valid in this case, y would have had to exceed 8 ($y_C$) for an increase (in expectations) to have been considered

[p] For a 1-sided target-α of 0.05, Ref. 3 shows that α' ranges from 0.034 for μ = 25 counts to 0.012 for μ = 5 counts. Extrapolation for the present example (μ = 2·1.7 = 3.4 counts) gave α' ≈ 0.006. A more precise "experiment" (N=8000) gave $\hat{\alpha} = 0.0050 \pm 0.0008$.

"detected" [Fig. 6: 1-sided test ($\mu_y > \mu_x$; $\alpha$=0.05) or 2-sided ($\mu_y \neq \mu_x$; $\alpha$=0.10)]. If the unknown base level ($\mu_x$) were 2.5, for example, $y_D$ would be 16.6, and the minimum detectable increase ($S_D$) would be 14.1 accidents.

An even more recent example of generalized rare event detection, reported in the press [Washington Post, 17 October 2006] relates to the diversion of aircraft as a result of smoke, especially in the cockpit. According to US Federal Aviation Administration records, the rate of such rare diversions in 2005, for example, was 3.2 per $10^5$ flights in the US, for a total of 320 for the year. It was reported also that such diversions had become sufficiently rare that they were now considered "random," but that during a 2 day period in late September 2006, there were "several such incidents." In the context of Fig. 5 ($\alpha$, $\beta$ = 0.05), we may ask: How large must "several" be, to be considered significantly different from a baseline rate (320/365)·2 per 2 days, for a Poisson process? The result ($y_C$) is 4, and the corresponding detection limit ($y_D$) would 9.15.[q] (In this particular case, taking the 95% confidence interval for the baseline rate per 2 days, to be 1.56 to 1.95, would still place it in the $y_C$=4 row of the Fig. 5 table.)

To illustrate another important application of Fig 6 for planning low-level radioactivity measurements, considering x, y as background and gross counts with means $\mu_x$, $\mu_y$, one can pose the question: How many background counts are necessary to ensure a 95 % chance of detection when the (net) signal/background ratio is 5:1? This is quickly determined graphically, by looking for the intersection with the $y_D$ curve of a line from the origin with slope $y_D/\mu_B$ = 6.0. The requisite background count ($\mu_B$) is 2.93 counts.

## 6. SOME CLOSING REMARKS AND OBSERVATIONS

Key observations from of our exploration of low-level (Poisson) detection issues are: (1) that moderately rare events may be treated adequately down to ≈ 2 to 5 counts ($\mu_B$) using the Poisson-normal approximation, with central and non-central t statistics used in connection with $S_C$, and $S_D$, respectively. The reported failure of the classic (Poisson-normal) test to control false positives in this range resulted primarily from the way in which the test was performed, using the *same* background observation (b) in the numerator and denominator of the experimental t-ratio; (2) that rigorous treatment for the detection of "rare" events (e.g., $\mu_B$ down to a fraction of a count) was found in documents published long ago (as early as the 1930s). Two specific "exact Poisson" approaches were reviewed and illustrated, for the well-known blank and for paired sample-blank counting, using published data on $^{85}$Kr measurement capabilities from the NIST low-level gas counting laboratory. Related issues that deserve emphasis follow.

*Validity of the Poisson hypothesis*. Apart from the initial cautionary tale of significant non-Poisson and even non-normal blank distributions [Section 2], the entire treatment given here rests on the validity of the Poisson model. That must be tested in each context, especially for backgrounds and blanks. Failure of the Poisson approximation of the binomial distribution for short-lived radioactivity is well known, as is use of the index of dispersion to test for equality of the mean and variance. Low-level counting backgrounds, however, have been found to be especially vulnerable to certain

[q] These results are intended strictly to illustrate an approach to an actual rare event problem. To address the "real" problem, many assumptions would need to be specified and adequately tested.

subtle artifacts, as manifest, for example, by deviations from the exponential distribution of inter-arrival times[26]. An excellent source for guidance on the general topic is the monograph by Cox and Lewis[6]. Especially pertinent and instructive, also, is the detailed investigation of Poisson model validity[2] that preceded its application to dodder seeds.[r]

*Brief perspective on detection capabilities.* Two important comments are in order: (1) Detection limits (minimum detectable signals $S_D$, disintegration rates, concentrations, ...) are particularly relevant as measurement process performance characteristics, useful for planning, method comparison, and advance assessment of method adequacy for specific applications. Use of $S_D$ as a critical (decision) value should be discouraged, as well as its use as a surrogate upper limit for "non-detects," which suppresses new information. (2) Experimental detection limits are necessarily estimates, having uncertainties that can be relatively large, especially when few degrees of freedom are involved.

*Rare nuclear events-I: Dominant null state and multiple detection decisions.* In low-level monitoring of rare events the spectral or environmental baseline is apt to be largely "empty," consisting primarily of background noise. As a result there will be many null decisions and a corresponding increase in the overall false positive risk. To control the overall risk to 0.05 for *n* null decisions, for example, the critical value must be increased such that $(1-0.05) = (1-\alpha)^n$, or $\alpha \approx 0.05/n$.[17] Such a procedure has been proposed by De Geer in connection with the extensive nuclear monitoring program of the Comprehensive Nuclear Test Ban Treaty Organization[22]. In the case of gamma-ray detection, De Geer found also that the spectral baselines were sufficiently smooth that the increase in $S_C$ could be minimized by increasing baseline width well beyond that of gamma-ray peak width. As a consequence, peaks having very few counts could be assessed rigorously using the well-known background Poisson treatment (Fig. 5). Consistent with false positives that had been experienced, and with the naming of 84 nuclides relevant to the detection of a nuclear weapon test, $\alpha$ was reduced to 0.0005. For a baseline ($\mu_B$) of 1.20 counts, for example, $y_C$ would be increased from 3 counts (for $\alpha$ = 0.05) to 6 counts (for $\alpha$ = 0.0005); and the minimum detectable net peak area ($S_D$) would then be equal to (11.84-1.20) counts.[s]

*Rare nuclear events-II: "Dββ0ν"* We close with a glimpse at one of the most heroic low-level experiments in fundamental particle physics today: the detection of neutrinoless double beta decay. (Ray Davis' experimental discovery of solar neutrinos, resulting in a Nobel Prize, was perhaps the first such heroic experiment.[23]) The *Dββ0ν* work published in 2005 by R. Arnold (and 47 coauthors)[9] represents an extremum of rare nuclear event detection research, with a background rate (in the critical energy window) of ≈ 1 count per 125 days, using a 0.932 kg target of $^{82}$Se, deep underground in Fréjus, France. (The Fréjus *Laboratoire Souterrain de Mondane* [LSM] has a depth of 4800 meters of water equivalent [m.w.e], and a surface muon shielding factor in excess of

---

[r] *Historical Footnote*: The dodder seeds publication of Przyborowski and Wilenski[3] concludes with a poignant footnote (p. 319), where the editor of *Biometrika*, E.S. Pearson, indicated that he had made "certain modifications and additions to the paper since it was received for publication at the beginning of July 1939, [because]... circumstances have unfortunately made communication with the authors impossible." [September 1939 marked the beginning of World War II, with the invasion of Poland.]

[s] Because of the discrete nature of the Poisson distribution, the actual value $\alpha'$ (0.00025) for $\mu_B = 1.20$ is less than 0.0005. ($\beta$ is unaffected.) Note that an extended range for $\alpha$ is given in Table 2 of Ref 22; a table similar to that in Fig. 5 may be derived also through the use of Eq. (12). (See also Section 5.1.)

$6 \times 10^6$.) Phase-I of the study, completed in September 2004, was characterized by 0.993 kg-y exposure, with a (partly modeled) background of 3.1±0.6 counts, and a gross signal of 5 counts observed from the $^{82}$Se target. Neutrinoless decay was not detected in phase-I, but limits were set for the half-life (>$10^{23}$ a) and effective neutrino mass (<4.9 eV, shell model). Using Eq. (12) we can set confidence limits for the gross counts; and using a bounding technique for $\mu_B$ we can use Fig. 5 to form a somewhat conservative estimate for the detection limit -- i.e., if $\mu_{B+}$ is taken as 3.7 counts, the corresponding critical value is 7 counts, and $S_D$, which may be useful for planning, is roughly 9.5 counts. The investment is huge; phase-II of the research, initiated in December 2004, is scheduled to continue for a 5 year period.

## ACKNOWLEDGMENTS

Special thanks go to the following colleagues for important suggestions, discussion, and/or review of the manuscript: I. Anderson, R. Cavanagh, J. Filliben, R. Fletcher, S. Glover, R. Hanlen, K. Inn, R. Lindstrom, A. Rukhin, and D. Simons.

## REFERENCES

1. A. HALD, Statistical Theory with Engineering Applications (Ch. 22, The Poisson distribution) John Wiley & Sons, New York, 1952.
2. J. PRZYBOROWSKI, H. WILENSKI, Statistical principles of routine work in testing clover seed for dodder, Biometrika 27 (1935) 273-292.
3. J. PRZYBOROWSKI, H. WILENSKI, H. Homogeneity of results in testing samples from Poisson series, Biometrika 31 (1939) 313-323.
4. MARLAP (2004) Multi-agency radiological laboratory analytical protocols manual, vol. 20: detection and quantification capabilities [www.epa.gov/radiation/marlap/manual.htm]
5. T.M. SEMKOW, Theory of overdispersion in counting statistics caused by fluctuating probabilities, Appl Radn Isot 51 (1999) 565-579. [See also, L.A. CURRIE, The Limit of Precision in Nuclear and Analytical Chemistry, Nucl. Instr. Meth 100 (1972) 387-395.]
6. D. COX, P. LEWIS. The Statistical Analysis of Series of Events, Methuen & Co., London, 1966.
7. L.A. CURRIE., Some case studies of skewed (and other ab-normal) data distributions arising in low-level environmental research, Fresenius J. Anal. Chem. 370 (2001) 705-718.
8. M. SCHWAIGER, R. EDELMAIER, E. HOLM, S. JEROME, (Eds.) Low level radionuclide measurement techniques-ICRM, Applied Radiation & Isotopes (2004) v. 61 (2, 3).
9. R. ARNOLD, et al., First results of the search for neutrinoless double-beta decay with the NEMO-3 detector, Phys. Rev. Lett. 95 (2005) 182302.
10. IUPAC Recommendations: Nomenclature in evaluation of analytical methods, including detection and quantification capabilities. Pure and Applied Chemistry 67 (1995) 1699-1723.
11. ISO Standard 11843-1,2 Capability of Detection, Geneva (1997f).

12. ISO Standard ISO 11929-1 through –4 Determination of the Detection Limit and Decision Threshold for Ionizing Radiation Measurements, Geneva (2000f).
13. ASTM Standard D 6091-97. Standard Practice for 99 %/95 % Interlaboratory Detection Estimate (IDE) for Analytical Methods with Negligible Calibration Error (1997).
14. ASTM Standard D 6512-00, Standard Practice for Interlaboratory Quantitation Estimate (2000).
15. ISO-3534 Statistics -- Vocabulary and Symbols (ISO, Geneva, Switzerland, 1993.)
16. L.A. CURRIE. Limits for qualitative detection and quantitative determination, application to radiochemistry, Anal Chem 40 (1968) 586-593.
17. L.A. CURRIE. Detection: International Update, and Some Emerging di-Lemmas involving Calibration, the Blank, and Multiple Detection Decisions, Chemometrics and Intell. Lab. Systems 37 (1997) 151-181.
18. F. GARWOOD. Fiducial limits for the Poisson distribution, Biometrika 28 (1936) 437-442.
19. J.O. IRWIN. The frequency distribution of the difference between two independent variates following the same Poisson distribution, J Royal Statistical Society 100 (1937) 415-416.
20. W.E. POTTER. Neyman-Pearson confidence intervals for extreme low-level, paired counting, Health Physics 76 (1999) 186-190.
21. L.A. CURRIE. The measurement of environmental levels of rare gas nuclides and the treatment of very low-level counting data, IEEE Trans. Nucl. Sci. NS-19 (1972) 119-126. (See also L.A. CURRIE. Lower Limit of Detection ... (Table 7) NUREG/CR-4007 Nuclear Regulatory Commission, 1984).
22. L.-E. DE GEER, Currie detection limits in gamma-ray spectroscopy, ICRM Conf on Low-level radioactivity measurement techniques, Applied Radiation and Isotopes 61 (2004) 151-160.
23. R. DAVIS, JR., D.S. HARMER, K.C. HOFFMAN. A search for neutrinos from the sun, Phys. Rev. Lett. 20 (1968) 1205-1209.
24. L.A. CURRIE, G.A. KLOUDA. Detection and quantification capabilities for $^{85}$Kr with the NIST low-level gas counting system: impacts of instrumental and environmental backgrounds, J. Radioanal. Nuclear Chem. 248 (2001) 239-246.
25. M.G. NATRELLA. The relation between confidence intervals and tests of significance, The American Statistician 14 [1] (1960) 20-22.
26. L.A. CURRIE, E.M. EIJGENHUIJSEN, G.A. KLOUDA. On the validity of the Poisson Hypothesis for low-level counting; investigation of the distributional characteristics of background radiation with the NIST Individual Pulse Counting System, Radiocarbon 40 (1998) 113-127.

# APPENDIX

Powerpoint presentation at the International Conference on Methods and Applications of Radioanalytical Chemistry (MARC VII, Kona HI, April 2006) and the National Institute of Standards and Technology (NIST, Gaithersburg MD, May 2006). [Updated, December 2006.]

# International Topical Conference (MARC VII)

Kailua-Kona, Hawai'i (April 2006)

NIST
National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

# On the Detection of Rare and Moderately Rare Events

OR:

What do dodder seeds & $^{85}$Kr atoms have in common?

L. A. Currie
Chemical Science and Technology Laboratory

NIST
National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

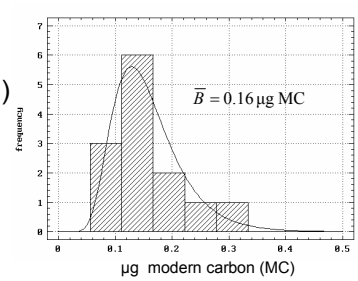
# Topics

- Perspective: Impact of the blank (B)
- Detection capabilities: Poisson-normal approximation
  - Asymptotic expression; large numbers of counts
  - "Moderately rare" (background ≈ 5 - 50 counts)
- Exact Poisson treatment (some ancient history)
  - Special issues: $H_o$ dominance; global vs local B estimates
  - Well-known background
  - Paired observations (y, b -- counts)
- Summary



# B as: Baseline, Blank, Background

- Issue-1: Non-Poisson error (systematic, random) components must not be ignored
- Issue-2: Such B's are often positively skewed, but limited observations restrict the ability to define tails of the distributions, e.g.,
  - NIST-WHOI, AMS blanks (C-$^{14}$C)
  - recommendation:
    paired observations
    [central limit theorem]



# Part-1

- Detection capabilities: Poisson-normal approximation
  - Asymptotic expression; large numbers of counts
  - "Moderately rare" (background ≈ 5 - 50 counts)
  - The "false positive" dilemma



# Detection: Poisson-normal approximation

- <u>Defining relations</u> (ISO, IUPAC, MARLAP) [S=net signal=(y - b)]

  detection decision: $\Pr(S>S_C|\ \mu_S=0) \le \alpha$ (default: 0.05)

  detection limit: $\Pr(S\le S_C|\ \mu_S=S_D) = \beta$ (default: 0.05)

- <u>Poisson-normal approximation</u> (paired (η=2); counts)

$$S_C \approx z_{1-\alpha}\sigma_o = z_{0.95}\sqrt{(2\sigma_B^2)} = 1.645\ \sqrt{(2\mu_B)} = 2.326\ \sqrt{\mu_B}$$

$$S_C \approx t_{\nu,1-\alpha}s_o = t_{\nu,1-\alpha}\sqrt{(2s_B^2)}$$

where $s_B^2$ = replication-$s^2$ ($\nu$ = n-1, d.f.)

or, $s_B^2$ = b (Poisson-$s^2$) ($\nu$ = 2b)

# Poisson-normal: false positive ($S_C$) dilemma

[y=gross counts, b=background counts, S=net counts]

- "commonly used formula"

  $S_C \approx 2.33\sqrt{b} = z_{1-\alpha}\sqrt{(2b)}$, or:

  $$S_C/s_o = z_{1-\alpha} = (y-b)_C/\sqrt{(2b)}$$

- better:

  $(y-b)_C/\sqrt{(2b')} = t_{\nu,1-\alpha}$

  $(y-b)_C/\sqrt{(b'+b'')} = t_{\nu,1-\alpha}$

  where b, b', b'' are independent background observations

"commonly used" problems

- $t_{\nu,1-\alpha}$ more appropriate
- $S_C$, $s_o$ dependence

  (non-normal ratio; excessive false positives)
- the "zero catastrophe"

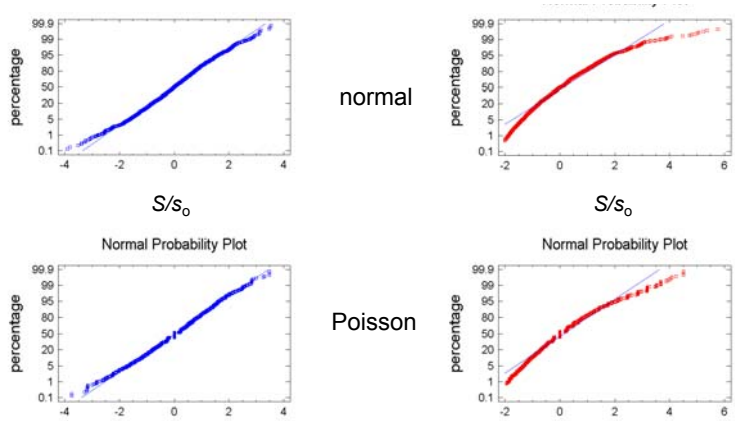
  (minimize Pr(b)=0, by limiting application to $\mu_B$> 6.9 (or 3.4) counts) [also: granularity]



# Simulation-1 [$\mu_B$ = 8.52 counts] Test Ratios: ($S/s_o$)

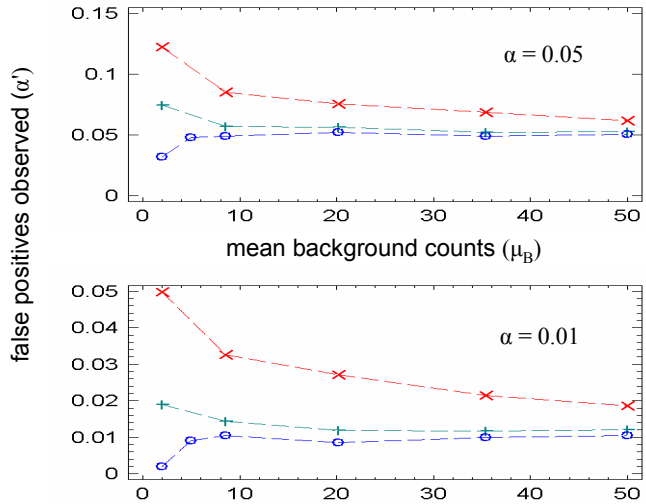
$$\hat{S}/s_o = (y-b)/\sqrt{2b'}$$

independent (b≠b') dependent (b=b')



# False Positive Functions

[x: Dependent (b=b'),$z_{1-\alpha}$ +: Indep.(b≠b'),$z_{1-\alpha}$ o: Indep.(b≠b'),$t_{1-\alpha,\nu}$]

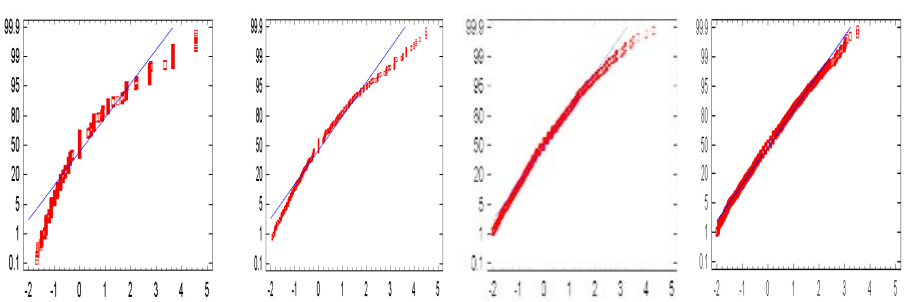





# Approach to normality

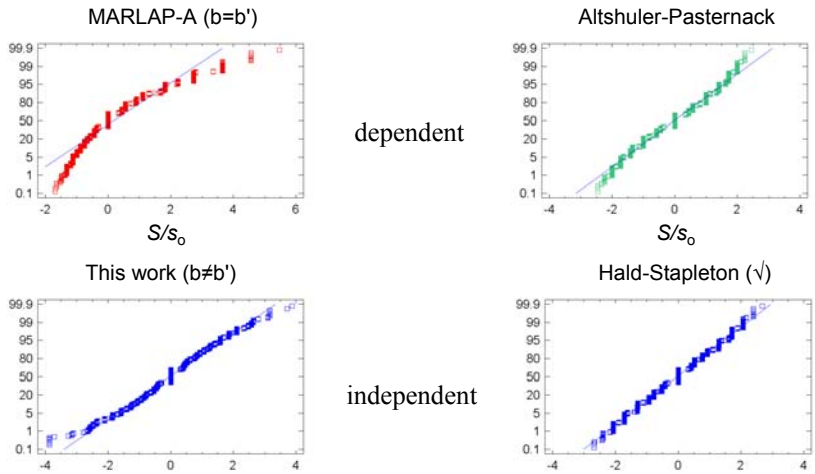
[Dependent ratio: b = b'] [Asymptotic $k_{0.95}$ = 1.645]

$\mu_B$=2.0 c $\mu_B$=8.52 c $\mu_B$=20.2 c $\mu_B$=50.0 c



$k_{0.95}$ ≈ 2.7 2.2 2.0 1.8

# Near the Breaking Point ($\mu_B$=2.00 counts)






# Part-2

- Exact Poisson treatment (some ancient history)
  - Special issues: $H_o$ dominance; global vs local B estimates, e.g.: gamma-ray peak detection by the Comprehensive Nuclear Test Ban Treaty Organization
  - I: Well-known background (y - counts, $\mu_B$)
  - II: Paired observations (y, b -- counts)

# Extreme Poisson Counting ('rare events')

## some ancient history (graphical, tabular solutions)

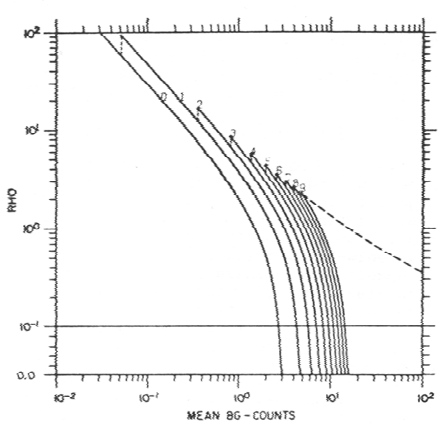
- $y_C$, $y_D$ directly from defining relations
- I: y (Poisson), $\mu_B$; <u>well-known blank</u> (Currie: 1972, 1984)
- II: y (Poisson), b (Poisson); <u>paired counts</u>

  (Przyborowski & Wilenski, dodder seeds: 1935, 1939)
- Special issue: low-level monitoring (DeGeer, 2004: 'global' bg, $H_o$ dominance, multiple detection decisions [$\alpha \rightarrow \alpha/n$])
- Example: NIST low-level gas counting [$\mu_B$=1.7 counts; $^{85}$Kr]



# Extreme Poisson-I (well-known blank, 1972+)

## graphical & tabular critical levels ($y_C$) and detection limits ($y_D$)

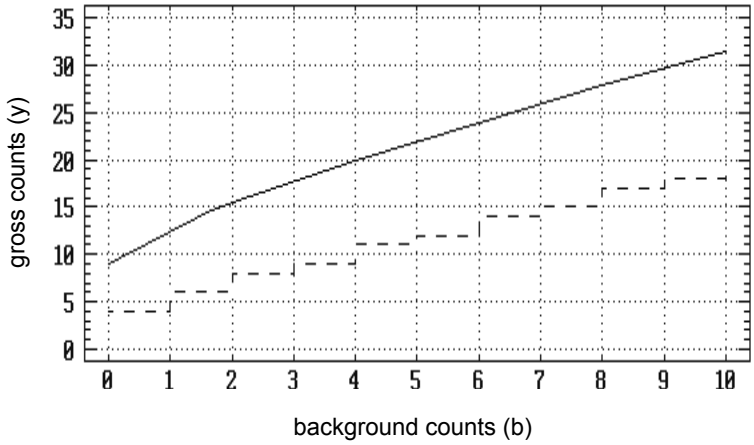



| $\alpha$ (minimum)[a] | $\mu_B$ (range) | $y_C = S_C + \mu_B$ ($\alpha \leq 0.05$) | $y_D = S_D + \mu_B$ ($\beta = 0.05$) |
|---|---|---|---|
| -- | 0 - 0.051 | 0 | 3.00 |
| 0.0013 | 0.052 - 0.35 | 1 | 4.74 |
| 0.0060 | 0.36 - 0.81 | 2 | 6.30 |
| 0.0089 | 0.82 - 1.36 | 3 | 7.75 |
| 0.013 | 1.37 - 1.96 | 4 | 9.15 |
| 0.016 | 1.97 - 2.60 | 5 | 10.51 |
| 0.018 | 2.61 - 3.28 | 6 | 11.84 |
| 0.020 | 3.29 - 3.97 | 7 | 13.15 |
| 0.021 | 3.98 - 4.69 | 8 | 14.44 |
| 0.022 | 4.70 - 5.42 | 9 | 15.70 |

a. For each $\mu_B$ range, $\alpha$ varies monotonically from $\alpha_{min}$ to 0.05

# Extreme Poisson-II (paired counting, 1935+)

**Critical Boundary** (dashed) **and Detection Limits** (solid)



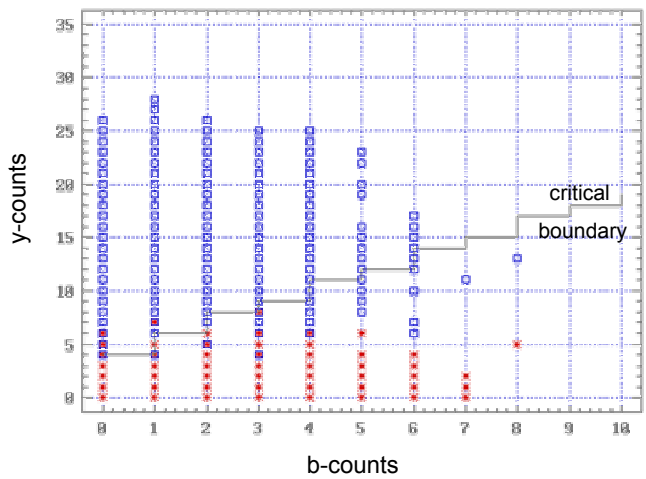



# Extreme Poisson: NIST example

(85 min screening experiment for $^{85}$Kr)

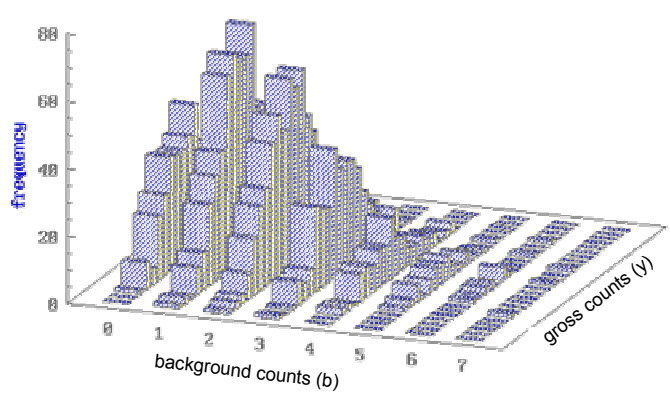
- 5 mL gas counter; bg = 1.2 counts/hour; Eff = 0.65
- well-known blank, t=85 min: $\mu_B$ = 1.70 counts
  - $y_C$ = 4 counts; $S_D$ = (9.15-1.70) = 7.45 c = 2.25 mBq $^{85}$Kr
  - $y_{obs}$ = 1 count: 90 % CI ($\mu_y$) = ½ ($\chi_{2,0.05}^2$, $\chi_{4,0.95}^2$) = (0.051, 4.74) counts: equivalent to an upper limit of 0.92 mBq
- paired observations, t=85 min
  - $S_D$ (for $\mu_B$ = 1.70 c) = (14.6-1.7) = 12.9 c = 3.9 mBq $^{85}$Kr

## Extreme Poisson-II: (paired counting)
## Count Contours at $y_D$ (□, 14.6) and $\mu_B$ (*, 1.7)






## Extreme Poisson-II
## 3D histogram (at $y_D$ = 14.6 counts)



## Concluding Observations

- The "false positive" problem can be avoided
- Extreme low-level counts addressed in the "ancient literature" (1939; 1972)
- Modern example: well-known bg, and paired counting "screening" limits for $^{85}Kr$ (NIST low-level gas counter)
- Finally: dodder seeds, and a footnote on history (1939)



## Extreme Poisson-II: historical footnote on the work of Przyborowski and Wilenski


Herbal "olfaction": dodder finds its way to host (tomato plant) through volatile chemical cues [J.B. Runyon, et al., *Science* 313 (2006) 1964].

A. dodder seeds as discrete, rare objects in clover, analog to trace radioactivity -- both described by the Poisson distribution.
B. 1939: editor completes revision, because communication with authors impossible -- beginning of WWII.

# Selected references


- Przyborowski, J. and Wilenski, H. (1935). Statistical principles of routine work in testing clover seed for dodder, *Biometrika* 27, 273-292.
- Przyborowski, J. and Wilenski, H. (1939). Homogeneity of results in testing samples from Poisson series, *Biometrika* 31, 313-323.
- Cox, D. and Lewis, P. (1966). *The Statistical Analysis Of Series Of Events* (Methuen & Co., London).
- Davis, Jr. R., Harmer, D.S. and Hoffman, K.C. (1968). A search for neutrinos from the sun, *Phys. Rev. Lett.* 20, 1205-1209.
- Currie, L.A. (1972). The measurement of environmental levels of rare gas nuclides and the treatment of very low-level counting data, *IEEE Trans. Nucl. Sci*. NS-19, 119-126.
- IUPAC (1995). Nomenclature in evaluation of analytical methods, including detection and quantification capabilities (prepared by L.A. Currie) *Pure & Applied Chem.* 67, 1699-1723.
- Currie, L.A., et al. (2000). Low-level (submicromole) environmental $^{14}C$ metrology, *Nucl. Instrum. Meth. Phys. Res.,* B172, 440-448.
- ISO (2000f). Standard ISO 11929-1 through –4 Determination of the detection limit and decision threshold for ionizing radiation measurements, Geneva.
- Currie, L.A. and Klouda, G.A. (2001). Detection and quantification capabilities for $^{85}Kr$ with the NIST low-level gas counting system: impacts of instrumental and environmental backgrounds, *J. Radioanal. Nuclear Chem.* 248, 239-246.
- Currie, L.A. (2001). Some case studies of skewed (and other ab-normal) data distributions arising in low-level environmental research, *Fresenius J. Anal. Chem.* 370, 705-718.
- MARLAP (2004). Multi-agency radiological laboratory analytical protocols manual, vol. 20: detection and quantification capabilities [www.epa.gov/radiation/marlap/manual.htm]
- Currie, L.A. (2004). Detection and quantification limits: basic concepts, international harmonization, and outstanding ("low-level") issues, *Applied Radiation and Isotopes* 61, 145-149.
- DeGeer, L-E. (2004). Currie detection limits in gamma-ray spectroscopy, ICRM Conf on Low-level radioactivity measurement techniques, *Applied Radiation and Isotopes* 61 (2004) 151-160.
- Currie, L.A. (2006) Some perspectives on nuclear detection and the blank, Ch 3 in ACS Sympos Ser 945, *Applied Modeling and Computations in Nuclear Science* (Eds: T. Semkow, S. Pommé, S. Jerome, D. Strom).