Uncertainty in Inferences We Make from Radiation Measurements: Counting Statistics and Other Uncertainties

Daniel J. Strom
Pacific Northwest National Laboratory
Richland, Washington USA
+1 509 375 2626 strom@pnl.gov

Health Physics Society PEP 2-B July 11, 2004

Work supported by the U.S. Department of Energy under Contract No. DE-AC06-76RLO 1830

Outline

- The problem of terminology
- The *two* "counting problems"
- A discussion of "total propagated uncertainty"
- Individual sample decision levels and the "Great Leap of Inference"
- Compare measurements with decision threshold, not detection level
- Decision strategies (Bayesian and classical)
- Consequences of wrong decisions: do we need better decision rules?
- Detection capabilities
- Classification and misclassification
- Report and record uncensored, un-rounded measurements results with uncertainties
- Utility and limitations of averaging results of many samples
- Probabilistic blank and environmental background subtraction
- Dosimetry, dosinference, and doswaggery
- Non-Bayesian methods of uncertainty evaluation in modeling
- References

NIST Technical Note 1297 (1994)

- Same as 1995 ISO Guide to the Expression of Uncertainty in Measurement (GEUM)
- Doesn't cover
  - the use of measurements in models that have uncertain
    - assumptions
    - parameters
    - form
  - representativeness (e.g., of a breathing-zone air sample)

NIST United States Department of Commerce
National Institute of Standards and Technology

NIST Technical Note 1297
1994 Edition

Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results

Barry N. Taylor and Chris E. Kuyatt

http://physics.nist.gov/cuu/pdf/tn1297.pdf

GEUM Terminology 1

- **measurand**: the unknown value of a physical quantity representing the "true state of Nature"
- **measured result**: result of a measurement made of a measurand
- **error**: the [unknowable] difference between a measured result the actual value of the measurand
- **uncertainty of measurement**: a "parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measureand."
  - a bound for the likely size of the measurement error
- **spurious errors** include **blunders, mistakes**

GEUM Terminology 2

- Uncertainty that is evaluated by the statistical analysis of series of observations is called a "**Type A**" uncertainty evaluation.
- Uncertainty that is evaluated by means *other* than the statistical analysis of a series of observations is called a "**Type B**" uncertainty evaluation.

- *This presentation focuses on Type A uncertainties.*

Terms 3: Error, Uncertainty, Variability

- "The difference between error and uncertainty should always be borne in mind."
- "For example, the result of a measurement after correction can unknowably be very close to the unknown value of the measurand, and thus have negligible error, even though it may have a large uncertainty."
- "Error bars?" No! "Uncertainty bars" is what we should say
- Variability is the range of values for different individuals in a population
    - e.g., height, weight, metabolism

Terms 4: Random and Systematic "Errors"

- Whatever happened to random and systematic "errors"?
- GEUM: There is not always a simple correspondence between the classification of uncertainty components into categories A and B and the commonly used classification of uncertainty components as "random" and "systematic."
- The nature of an uncertainty component is conditioned by the use made of the corresponding quantity, that is, on how that quantity appears in the mathematical model that describes the measurement process.
- When the corresponding quantity is used in a different way, a "random" component may become a "systematic" component and vice versa.

Terminology 5

- Thus the terms "random uncertainty" and "systematic uncertainty" can be misleading when generally applied.
- An alternative nomenclature that might be used is "component of uncertainty arising from a random effect," "component of uncertainty arising from a systematic effect," where a random effect is one that gives rise to a possible random error in the current measurement process and a systematic effect is one that gives rise to a possible systematic error in the current measurement process. In principle, an uncertainty component arising from a systematic effect may in some cases be evaluated by method A while in other cases by method B (see subsection 2.2), as may be an uncertainty component arising from a random effect.

Type A Uncertainty Evaluation

- represented by a statistically estimated standard deviation

$$s_i = \left| \sqrt{s_i^2} \right|$$

- associated number of degrees of freedom = v_i.
- the standard uncertainty is $u_i = s_i$.

Type B Uncertainty Evaluation

- represented by a quantity u_j

$$u_j \approx \text{corresponding standard deviation;}$$

$$u_j = \left|\sqrt{u_j^2}\right|;$$

$$u_j^2 \approx \text{corresponding variance obtained from}$$

an assumed probability distribution

based on all the available information

- Since the quantity u_j^2 is treated like a variance and u_j like a standard deviation, for such a component the standard uncertainty is simply u_j.

Total Propagated Uncertainty (=Combined Standard Uncertainty)

Statistical Criteria for Decision-Making

- *a priori* determinations of detection capabilities
- *a posteriori* decisions of whether radioactivity has been detected in a particular sample
- total propagated uncertainty in measurement results
- sampling strategies

Most Difficult: Alpha-Emitters

- problems are most difficult for alpha-emitting radionuclides
  - ^{230}Th (found in uranium mill tailings)
  - Pu (from reprocessing of irradiated nuclear fuel)
- expense of sampling and laboratory analysis
  - α-spectrometry
  - mass spectrometry
    - Inductively Coupled Plasma Mass Spectrometry (ICPMS)
    - Thermal Ionization Mass Spectrometry (TIMS)
- expense and consequences of incorrect decisions

Uncertainty Propagation Formula

- Combined standard uncertainty

$$u_{\mathrm{c}}^2(y) = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} u(x_i, x_j)$$

$$= \text{Sum of variance terms and covariance terms}$$

- Derived from first-order Taylor series expansion
- Not accurate for large uncertainties (e.g., broad lognormal distributions)
- Covariances usually unknown and ignored

Total Propagated Uncertainty for A Radiochemical Measurement

- Includes
  - counting uncertainty for analyte (A)
  - counting uncertainty for tracer (includes radiochemical recovery and counting yield) (A)
  - uncertainty in tracer calibration (B)
  - uncertainty in tracer volume (A)
  - uncertainty in aliquot volume (B)
  - "system" uncertainty (B) $\approx$ 3%
- We return to *TPU* later

The Two Aspects of the Counting Problem

The Two Counting Problems

- Radioactive decay is a Bernoulli process described by a binomial or Poisson distribution
- The "forward problem"
  - from properties of the process, we predict the distribution of counting results (mean, standard deviation (SD))
  - measurand $\rightarrow$ distribution of possible observations
- The "reverse problem"
  - measure a counting result
  - from the counting result, we infer the parameters of the underlying binomial or Poisson distribution (mean, SD)

    see, e.g., Rainwater and Wu (1947)
  - this is the problem we're really interested in!

Two Kinds of Statistics

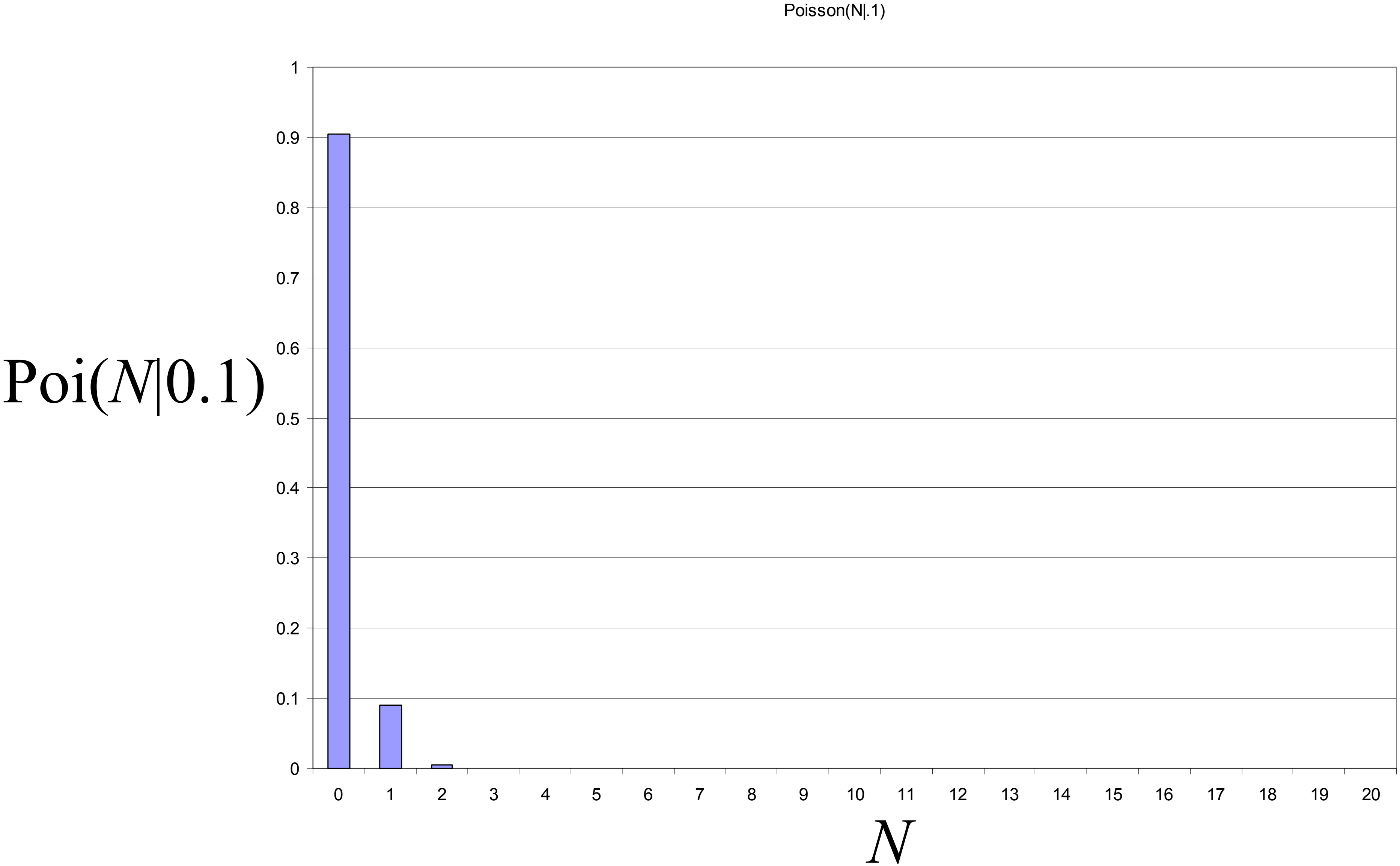
- Classical statistics
  - does the forward problem well
  - does not really do the reverse problem
- Bayesian statistics does the reverse problem using
  - a prior probability distribution
  - the observed results
  - a likelihood function (a classical expression of the forward problem)

The Forward Problem

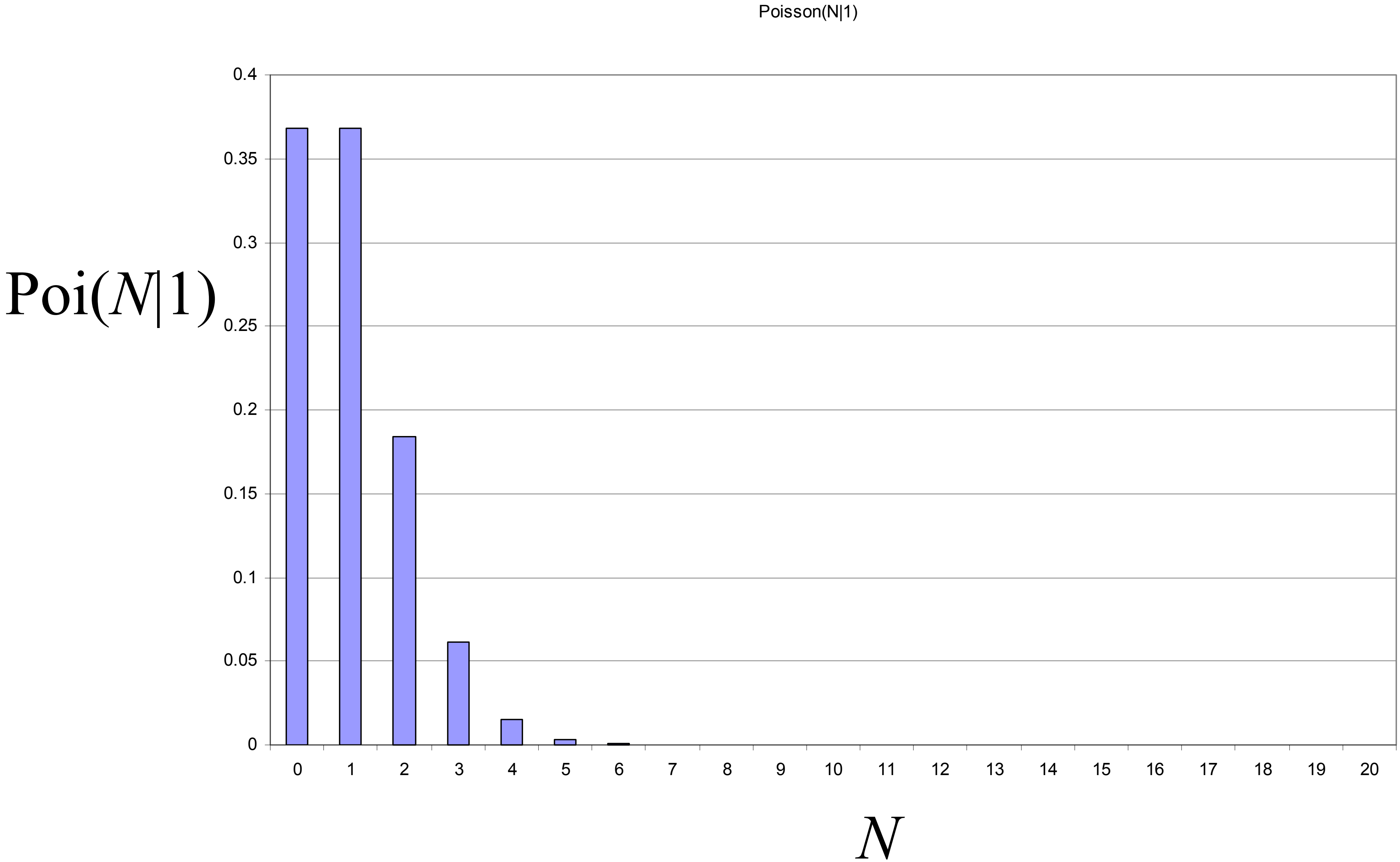
The Forward Problem

- Use Poisson statistics to predict the distribution of observations from a given value of the measurand
- The measurand is best thought of as a count rate ρ
  - otherwise it is difficult to deal with different counting times
- The observable is a number of counts, N, sampled
  - from a Poisson distribution
  - during time t
  - with mean ρt
- $\mathrm{Var}(\mathrm{Poi}(N \mid \rho t)) = \rho t$

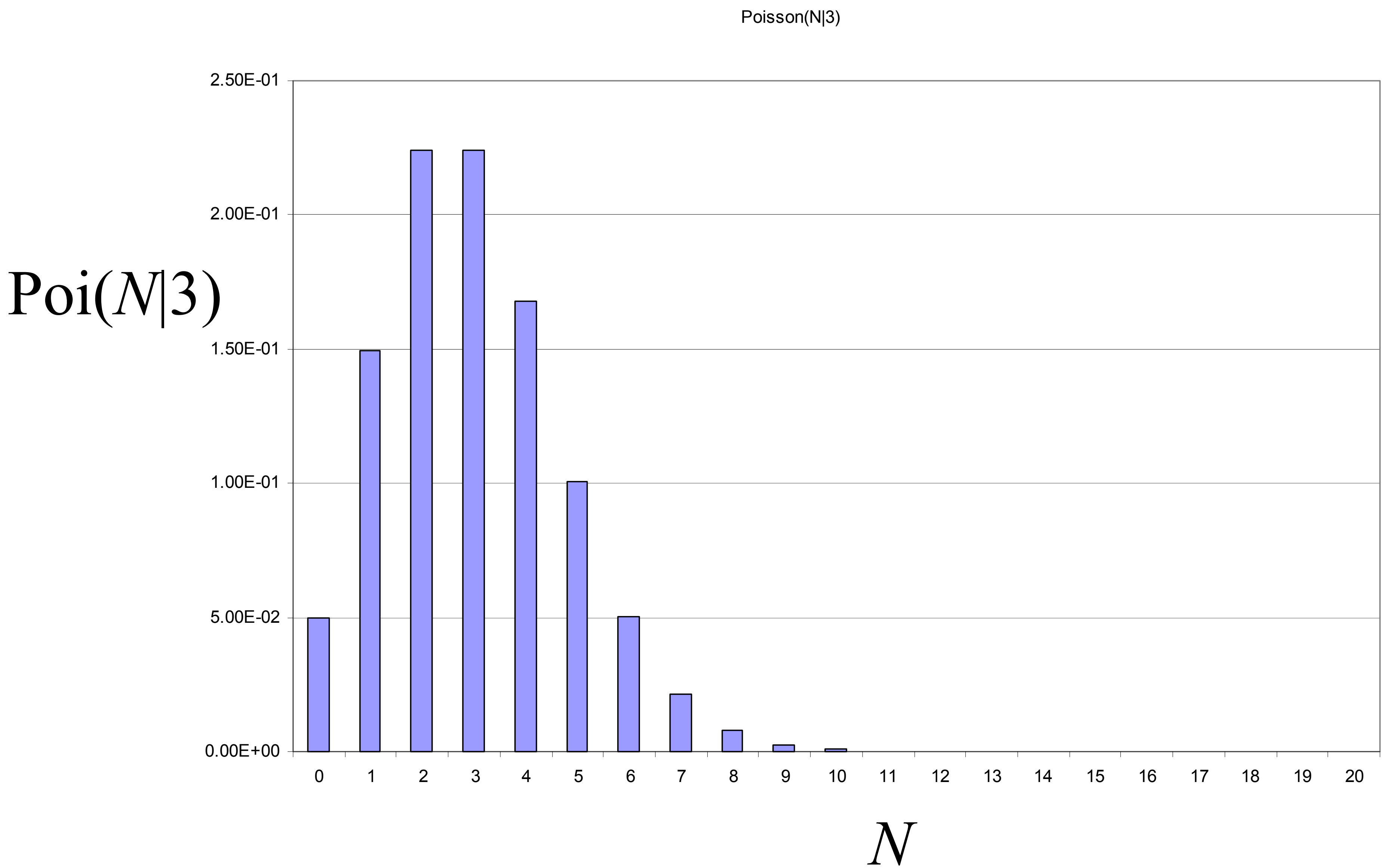
Poisson Distribution, $\mu = \rho t = 0.1$



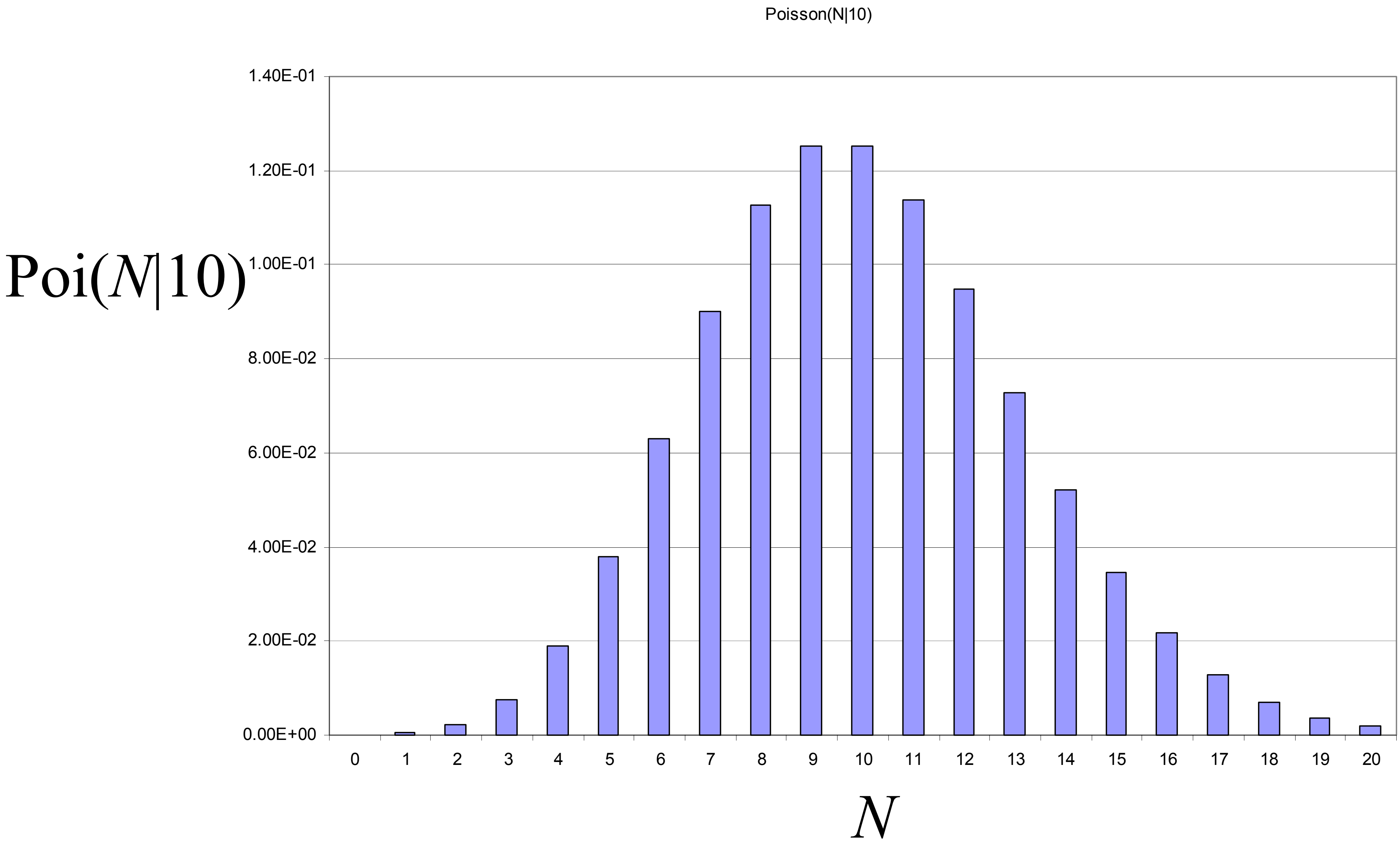
Poisson Distribution, $\mu = \rho t = 1$



Poisson Distribution, $\mu = \rho t = 3$



Poisson Distribution, $\mu = \rho t = 10$



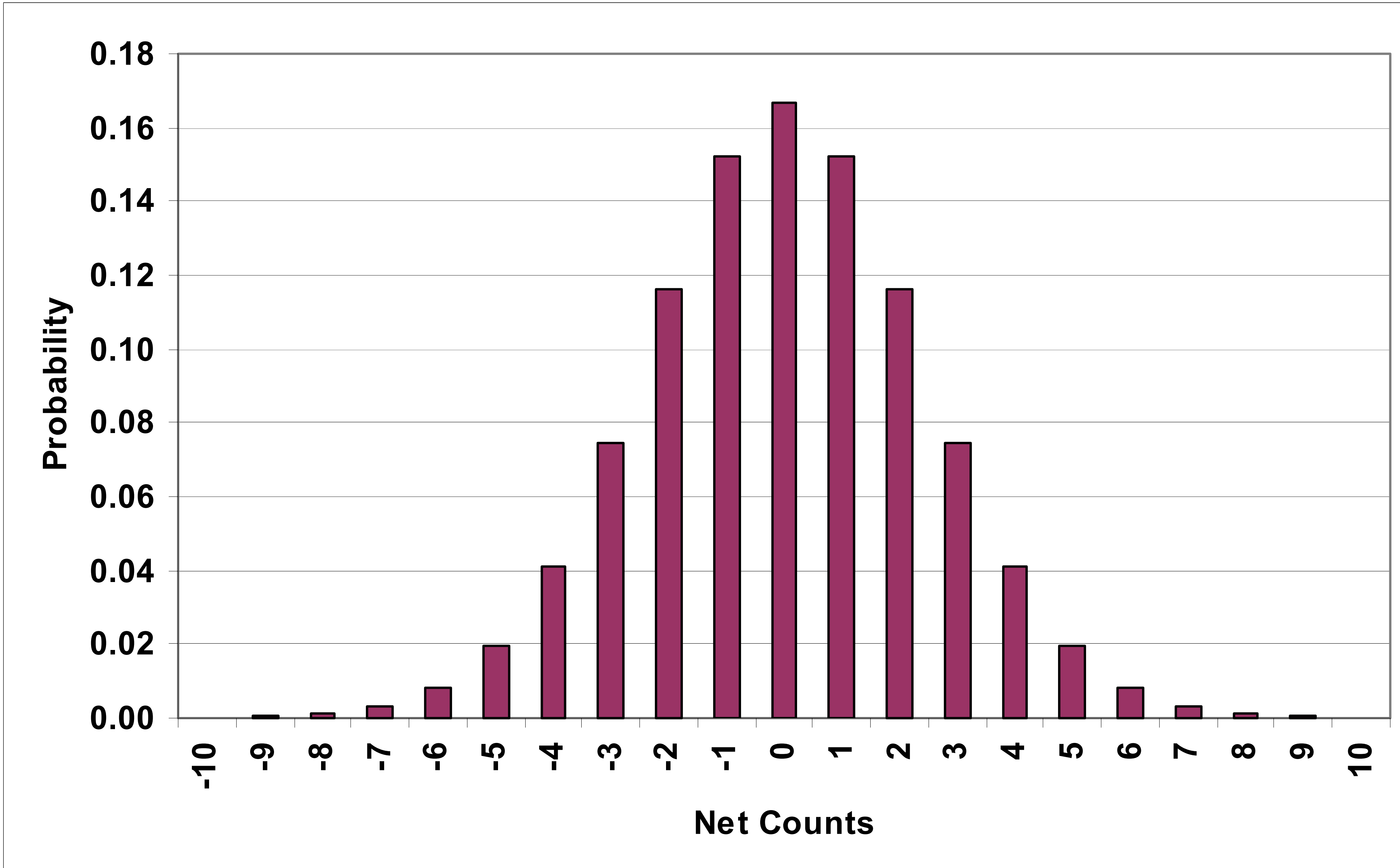
Normal Approximation to the Poisson

- No one tries to approximate a Poisson distribution with a Normal distribution in counting problems
- The normal approximation is applied to the *difference* of two Poisson distributions
  - typically much more symmetric

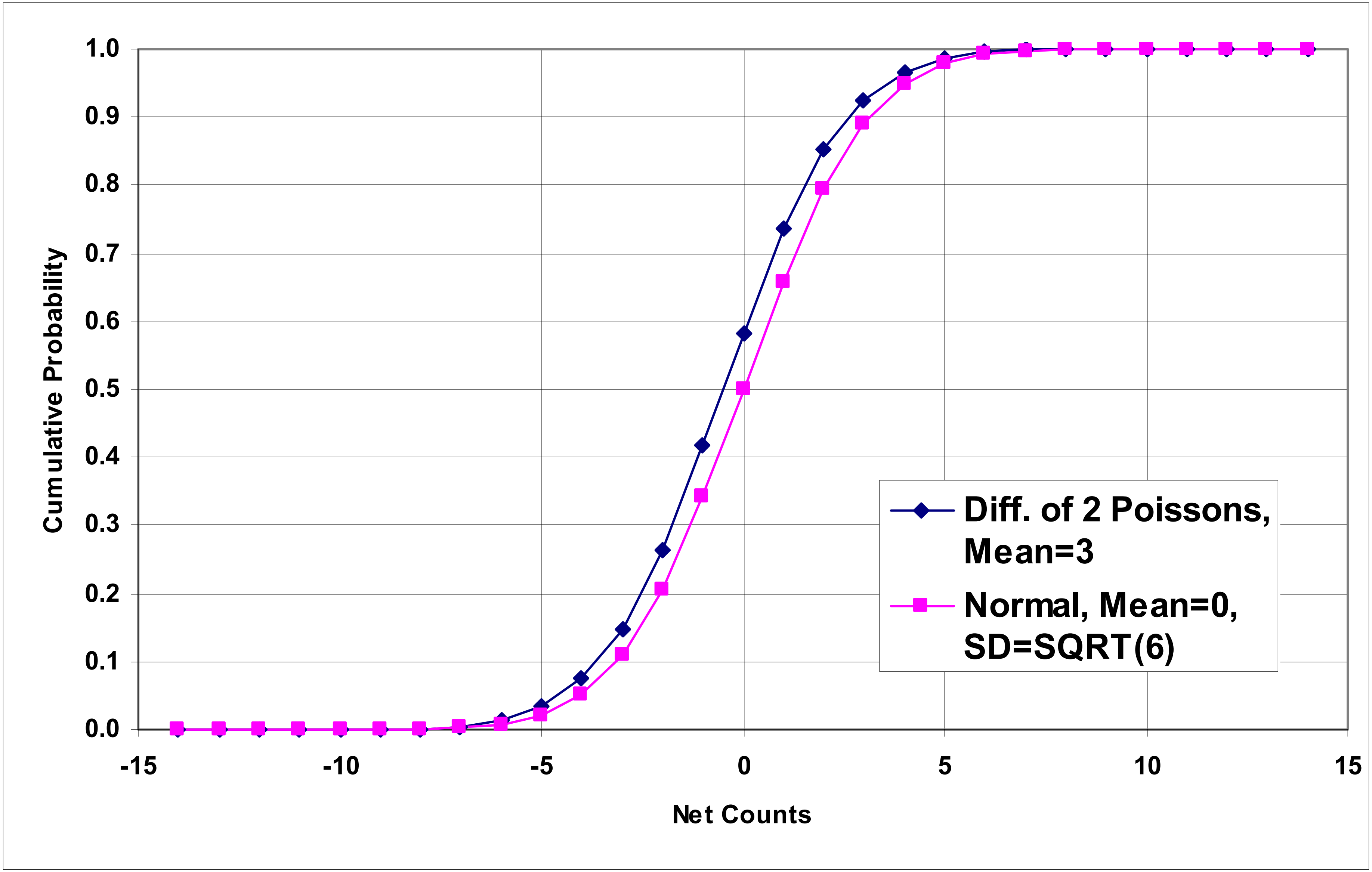
Difference of 2 Poisson Distributions

- When means are equal (e.g., blanks), are symmetric
- Discrete, not continuous
  - For $\mu_b = 3$, $P(N<0) = 0.42$, $P(N\leq 0) = 0.58$
  - For Normal, $P(N<0) = 0.500$, $P(N\leq 0) = 0.500$
- Probabilities on upper tails aren't too much different from Normal

Difference of 2 Poissons with $\mu = \rho t = 3$



Difference of 2 Poissons with $\mu = \rho t = 3$



The Observables
and the Measurands

The Observables

- Same apparatus for blank and sample
- Assume count times known ("time preselection" in ISO parlance)
- Assume no non-Poisson variance
  - sometimes not valid in the real world
  - see, e.g., Kathren 2001, ISO 1995
- *Assume observed count is maximum likelihood estimate and estimate of its variance* ("the Great Leap of Inference")

Notation - 1: Observed Quantities

- Convention: Roman letters denote observed quantities
- N_{b}: number of blank counts observed
- N_{g}: number of gross counts observed
- t_{b}: blank count time (s)
- t_{g}: gross count time (s)
- R_{b}: blank count rate (s^{-1})
- R_{g}: gross count rate (s^{-1})
- R_{n}: net count rate (s^{-1})
- $s(R_{\mathrm{n}})$: standard deviation of net count rate (s^{-1})

Classical Statistics: Traditional Relationships Among Observed Quantities

$$R_{\mathrm{b}} = \frac{N_{\mathrm{b}}}{t_{\mathrm{b}}}; R_{\mathrm{g}} = \frac{N_{\mathrm{g}}}{t_{\mathrm{g}}}$$

$$R_{\mathrm{n}} = R_{\mathrm{g}} - R_{\mathrm{b}} = \frac{N_{\mathrm{g}}}{t_{\mathrm{g}}} - \frac{N_{\mathrm{b}}}{t_{\mathrm{b}}}$$

$$s(R_{\mathrm{n}}) = \sqrt{\frac{\mathrm{Var}(N_{\mathrm{g}})}{t_{\mathrm{g}}^2} + \frac{\mathrm{Var}(N_{\mathrm{b}})}{t_{\mathrm{b}}^2}} \approx \sqrt{\frac{N_g}{t_{\mathrm{g}}^2} + \frac{N_{\mathrm{b}}}{t_{\mathrm{b}}^2}}$$

Notation 2: The Measurands - [Unknown] Population Parameters

- By convention, Greek letters denote population parameters
- These reflect the *measurand*, the "true state of Nature" that we are trying to infer
- ρ_b: long-term blank count rate (s^{-1})
- ρ_n: long-term net count rate (s^{-1}) (due to analyte in unknown)
- ρ_g: long-term gross count rate (s^{-1})

Notation 3: The Measurands - [Unknown] Population Parameters

- Parameters are needed for sampling from population distributions
- μ_b: number of blank counts expected during t_b
- μ_g: number of gross counts expected t_g
- $\sigma(\rho_n)$: standard deviation of long-term net count rate (s^{-1})

Classical Statistics:
Relationships Among Population Parameters

$$\text{number of counts} = \text{rate} \times \text{time}$$

$$\mu_{\mathrm{b}} = \rho_{\mathrm{b}} t_{\mathrm{b}}$$

$$\mu_{\mathrm{g}} = (\rho_{\mathrm{b}} + \rho_{\mathrm{n}}) t_{\mathrm{g}}$$

$$\sigma(\rho_{\mathrm{n}}) = \sqrt{\frac{\rho_{\mathrm{b}}}{t_{\mathrm{b}}} + \frac{\rho_{\mathrm{b}} + \rho_{\mathrm{n}}}{t_{\mathrm{g}}}}$$

The Reverse Problem

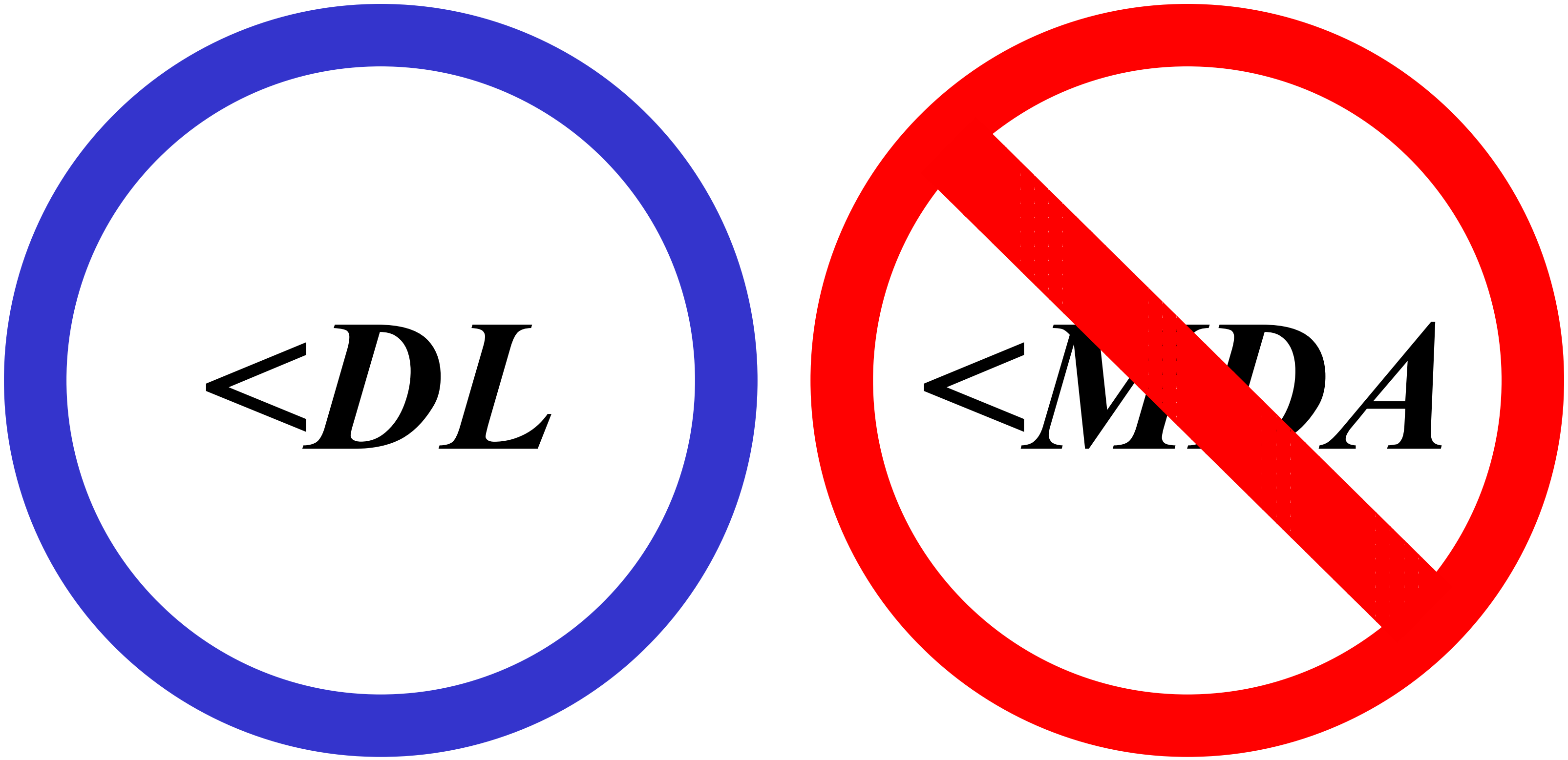
The Reverse Problem: Using Observed Quantities to Estimate Population Parameters (Measurands)

- Classical statisticians
  - use R_n to *estimate* ρ_n
  - use $s(R_n)$ to *estimate* $\sigma(\rho_n)$
    - *ofter a poor assumption for low numbers of counts*
  - every time you make another measurement, you get a new R_n and $s(R_n)$, that is, a new estimate of ρ_n and $\sigma(\rho_n)$
- Bayesian approach shown later

Decision Rules

Terminology Is a Mess! and This Is Just in English!

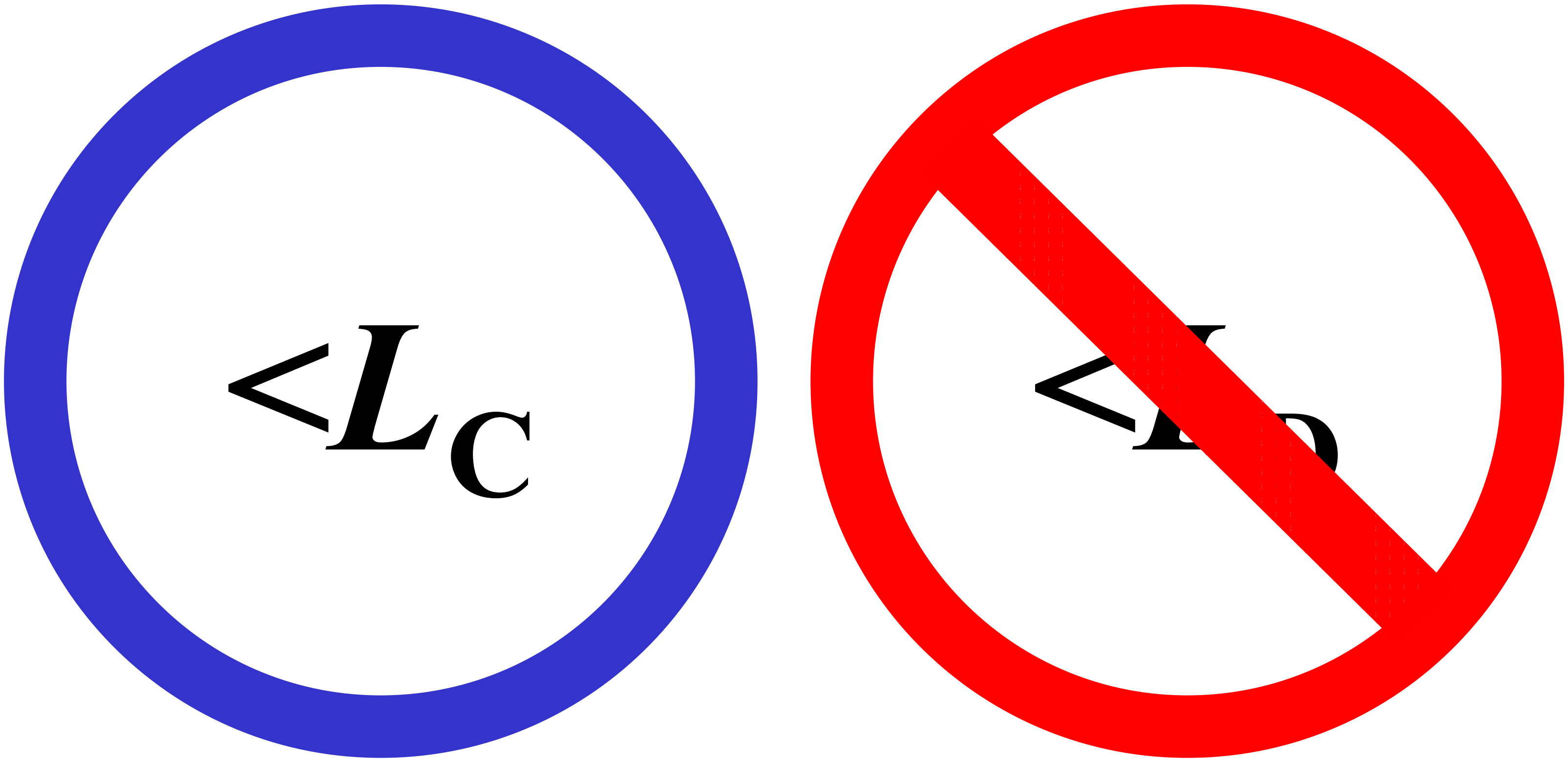
	"DL"	***"MDA"***
Name	decision level	minimum detectable amount
What?	the lowest useable action level	NOT an action level!
Use:	compare measurements to *DL*	Use in planning, advertising or in a statement of work for a contractor: "How much will you charge to provide counting services with this *MDA*?"
When?	*a posteriori*: *after* the measurement is made	*a priori*: *before* the measurement is made (but it does "vary with the nature of the sample" – NUREG-4007)
Defined in	HPS/ANSI N13.30	HPS/ANSI N13.30
Currie's Name	critical level, L_C	detection level, L_D
Ill-defined Names		lower limit of detection, *LLD*; also, un-fortunately, "lower level discriminator," detection limit, limit of detection ("LOD")
Turner's name	"minimum significant measured activity"	"minimum detectable true activity"
ISO 11929 name	"decision threshold"	"detection limit"
Spanish name	umbral de decision	limite de deteccion
MARLAP name	"critical value of []"	"minimum detectable amount" or "minimum detectable concentration"
Strom's name	"false alarm level"	"advertising level" "expected detection capability"



Always compare a result with *DL*

Never compare a result with MDA!

Translation:



Always compare a result with L_C

Never compare a result with L_D!

Measurand versus Measurement Result

- 2 "types" of errors (wrong decisions)

		Is anything there? (Is any activity present [above blank]?)	
		Yes	No
Did I detect anything? (Was the result above the decision level?)	Yes	• I made the correct decision (no error)	• False alarm • False positive • I've committed a Type I error
	No	• The alarm should have gone off, but didn't • False negative • I've committed a Type II error	• I made the correct decision (no error)

Error Terminology

- A **Type I error** (wrong decision) is falsely concluding there's activity present when no activity is present
- A **Type II error** is falsely concluding there's no activity present when activity is present
- The **probability of a Type I** error is called α
- The **probability of a Type II** error is called β
- The number of standard deviations above zero on the standard normal distribution having a probability of α or β of being higher is known as the "**standard normal deviate**," k_α or k_β
  - these are $k_{1-\alpha}$ or $k_{1-\beta}$ in ISO notation
- For $\alpha = 0.05$ (a 5% chance of making a Type I error), $k_\alpha = 1.645$
- For $\beta = 0.05$ (a 5% chance of making a Type II error), $k_\beta = 1.645$

Characteristics of Many Decision Rules

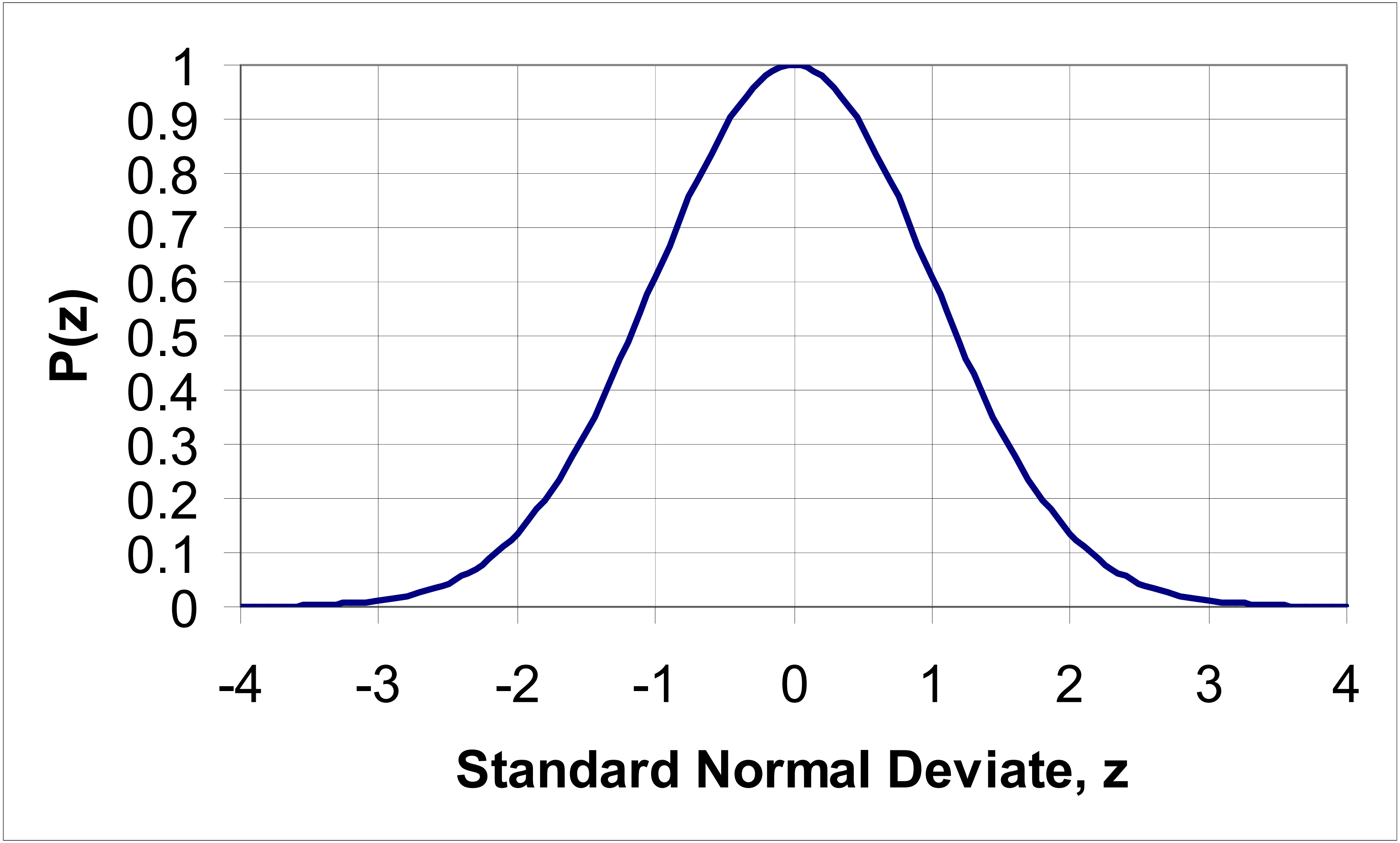
	Assumes Var(μ) = N	Assumes or permits Var(μ) > N	Exact Method or Binomial Distribution
Uses Blank (background) Counts Only	ISO 11929-1 2000 Currie (1968) ANSI N-13.30-1996 Altshuler & Pasternak (1963) Nicholson D_2 (1963)	DL_{N+1} (in Strom & MacLellan 2001)	
Uses Blank and Sample Counts	Nicholson D_1, D_3 (1963)	"Stapleton's decision criterion" (in Strom & MacLellan 2001)	Nicholson D_e (1963) Sumerling and Darby (1981)

Current Decision Level

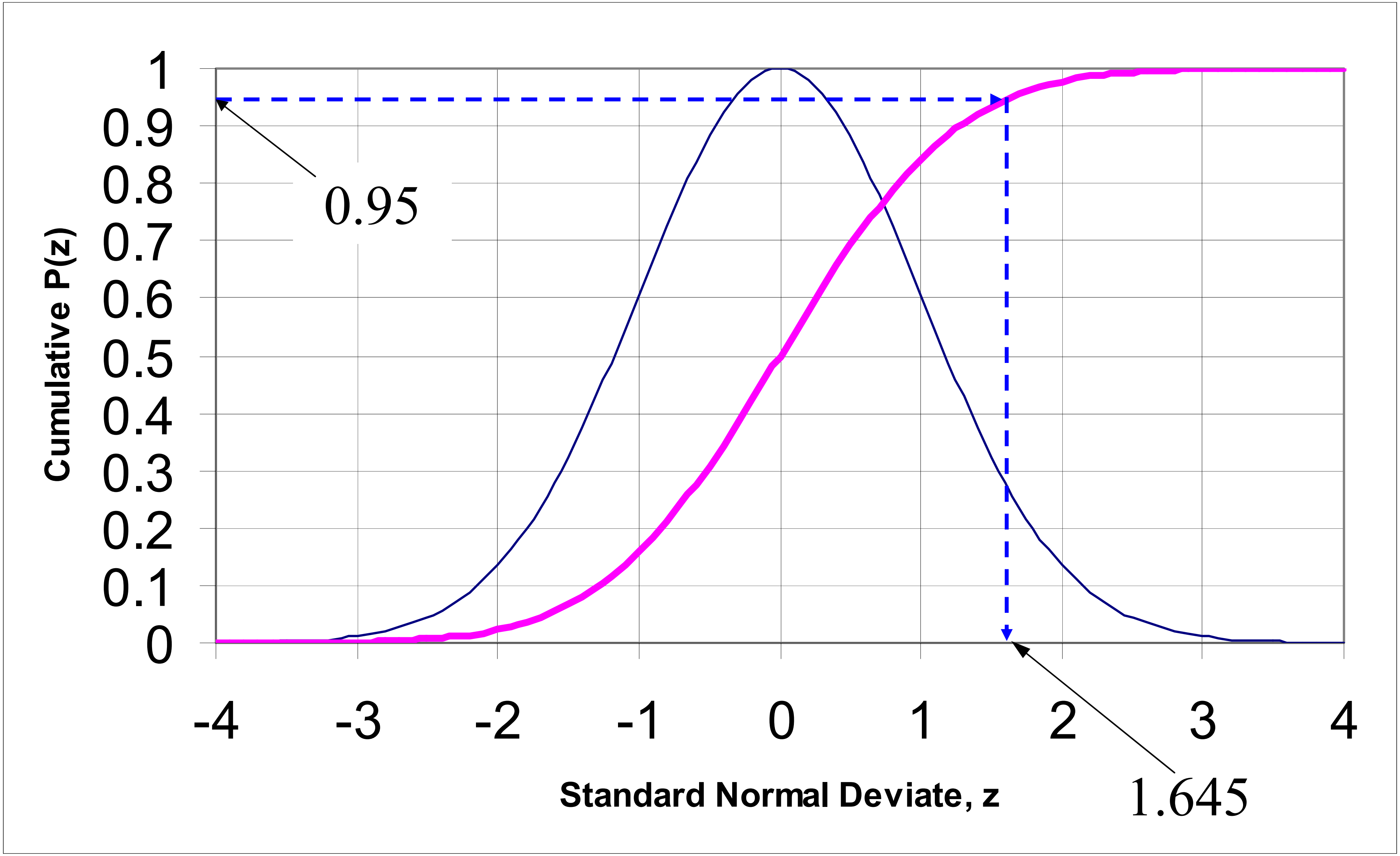
(a.k.a. Critical Level)

- α: acceptable probability of making wrong decision (Type I error): false alarm or false positive
  - α is often taken to be 0.05
- k_{α}: value of standard normal deviate for area 1-α
  - $k_{0.05}$ is 1.645
- ignore non-Poisson uncertainty for simplicity

Standard Normal Distribution, $\mu = 0$, $\sigma = 1$



Cumulative Standard Normal Distribution



Current "N13.30" Decision Rule

- Nicholson's (1963) D_2 rule; Currie's (1968) rule; ANSI/HPS N13.30-1996; MARSSIM; Equation 15a, Table 1 of ISO 11929-1:2000

$$DL_{\mathrm{N13.30}}(N_{\mathrm{b}},\alpha) = k_{\alpha} s_0 = k_{\alpha}\sqrt{2N_{\mathrm{b}}}$$

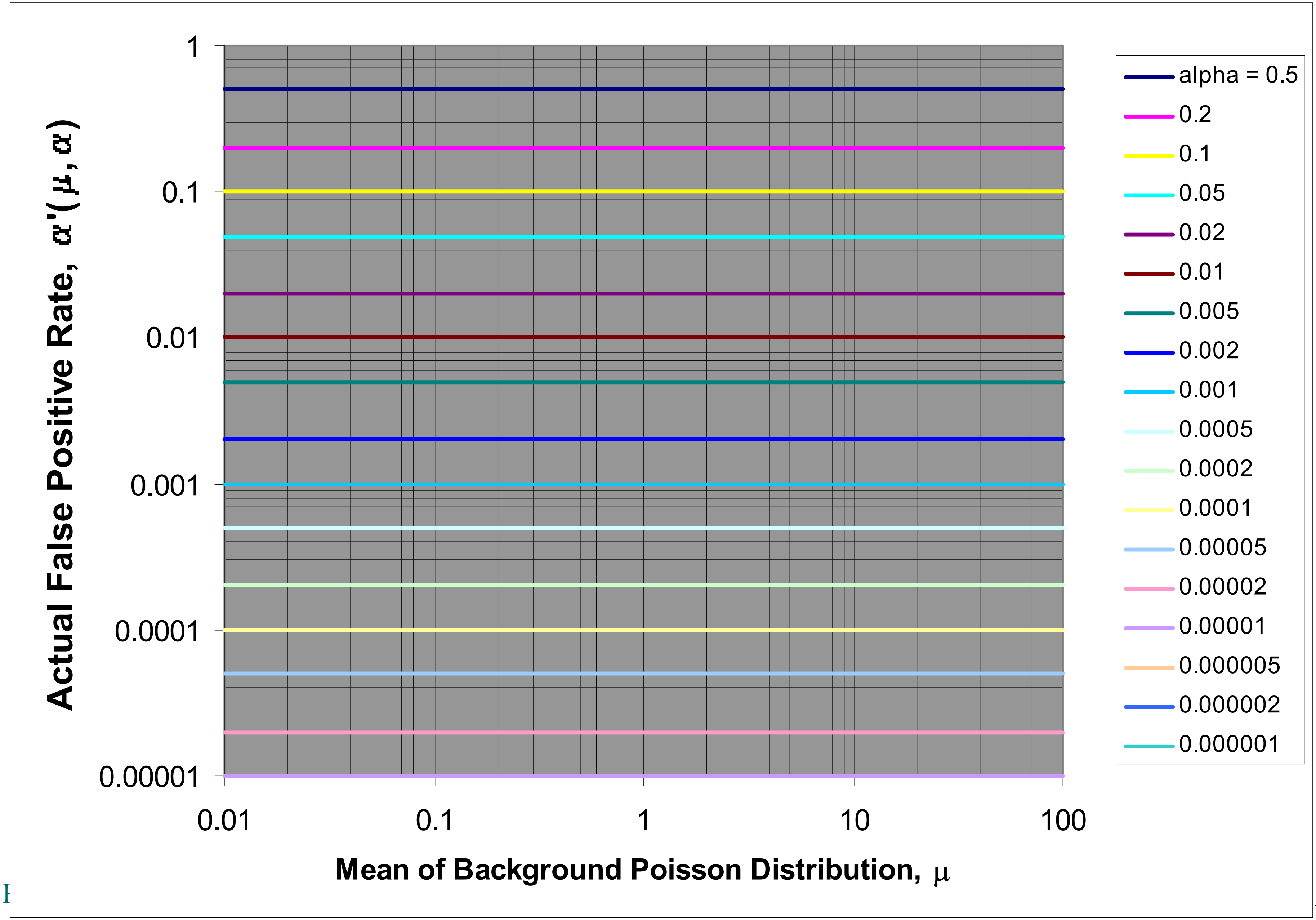
- For $\alpha = 0.05$

$$DL_{\mathrm{N13.30}}(N_{\mathrm{b}},0.05) = 1.645\sqrt{2N_{\mathrm{b}}} = 2.329\sqrt{N_{\mathrm{b}}}$$

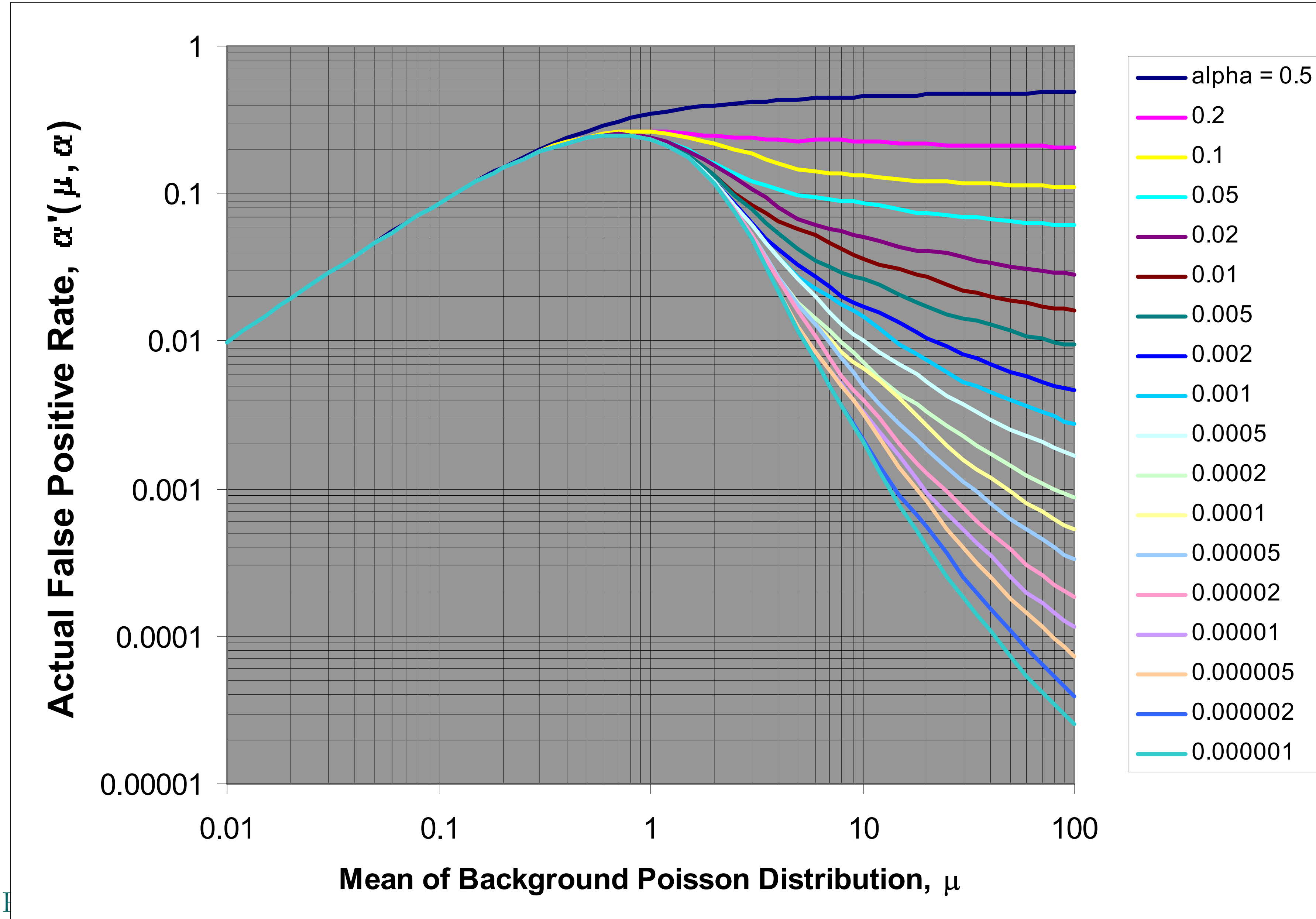
- Expressed as a rate, for non-paired blank:

$$DL_{\mathrm{N13.30}}(R_{\mathrm{n}},\alpha) = k_{\alpha}\sqrt{\frac{N_{\mathrm{b}}}{t_{\mathrm{b}}}\left(\frac{1}{t_{\mathrm{b}}}+\frac{1}{t_{\mathrm{g}}}\right)}$$

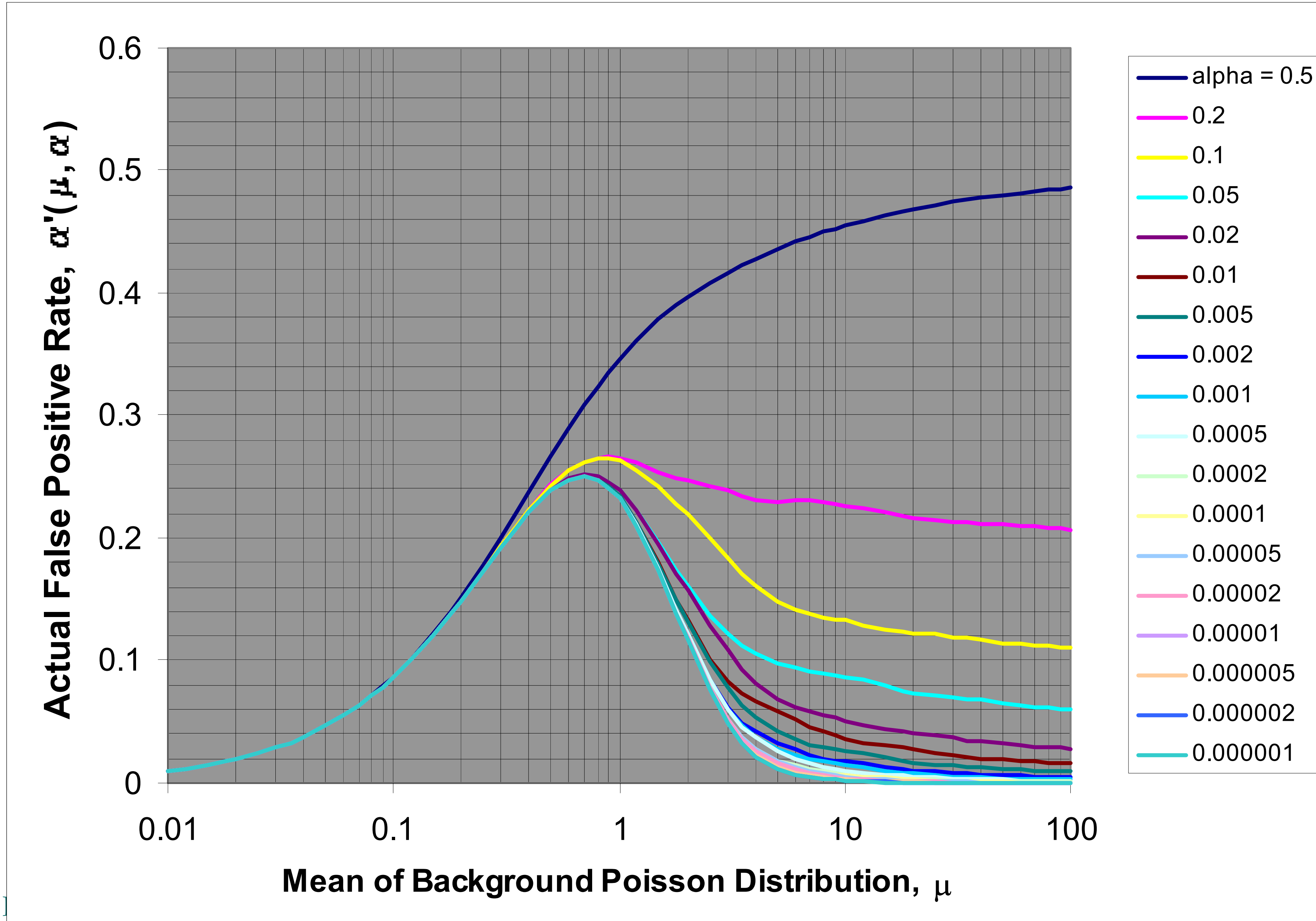
Ideal (Impossible) *DL*



$$DL(\alpha, N_b) = k_\alpha \sqrt{2N_b}$$



$$DL(\alpha, N_b) = k_\alpha \sqrt{2N_b} \text{ (on a linear vertical scale)}$$



Problems with the "N13.30" Decision Rule

- Should be horizontal lines at $\alpha' = \alpha$
- 25% wrong decisions at $\mu_b \approx 0.7$ count, regardless of α
- Actual false positive rate α' is *independent* of α at very small numbers of counts

$$\mu_b = \rho_b t_b << 1$$

- Even at $\mu_b = 10$, only asymptotically approaches α for larger values
- For very small α, no good even at $\mu_b = 100$!

The Bayesian Approach to the Reverse Problem

The Reverend Thomas Bayes 1702-1761

- *Probability is that degree of confidence dictated by the evidence through Bayes's theorem. -- E.T. Jaynes*

Bayes's Rule (Simple form)

$$P(B \mid A) = \frac{P(A \mid B) P(B)}{P(A)}$$

- Names:

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Normalizing Factor}}$$

Bayesian Approach:
The Prior Probability 1

$$P(B_i \mid A) = \frac{L(A \mid B_i)P(B_i)}{\sum_{\text{all } j} L(A \mid B_j)P(B_j)}$$

- Some form of prior probability is required!
- The prior probability is what you know before you start
- The prior can have more or less effect on the posterior, depending on the precision of the data
- The prior can be subjective
- The prior is the topic of unresolvable arguments

Bayesian Approach: The Prior Probability 2

$$P(B_i \mid A) = \frac{L(A \mid B_i)P(B_i)}{\sum_{\text{all } j} L(A \mid B_j)P(B_j)}$$

- The prior can be "nothing"
  - even "nothing" can take several forms
  - "uniform," "flat," or "uninformative" prior: all values of *B* are "equally probable"
  - "vague" prior: all values of ln(*B*) are equally probable…
- The prior can be hard to nail down
  - "small values of blank are more likely than large ones"

Philosophical Statement of Bayes's Rule

$$P(\text{measurand} \mid \text{evidence}) = \frac{L(\text{evidence} \mid \text{measurand})P(\text{measurand})}{\text{normalizing factor}}$$

- The measurand or "state of nature" (e.g., count rate from analyte) is what we want to know
- The "evidence" is what we have observed
- The likelihood of the "evidence" given the measurand is what we know about the way nature works
- The probability of the state of nature is what we believed before we obtained the evidence

Probability Density

- probability density is the probability that x lies in an interval between x and $x + \mathrm{d}x$

$$P(x)\,\mathrm{d}\,x = \text{probability that } x \text{ lies in the interval between } x \text{ and } x + \mathrm{d}x$$

- probability density is a continuous function

Bayes's Rule: Continuous Form

- P's are probability densities

$$P(\mu \mid N) = \frac{L(N \mid \mu)P(\mu)}{\int_0^\infty L(N \mid \nu)P(\nu)\,\mathrm{d}\nu}$$

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Normalizing Factor}}$$

- We want to determine the posterior probability density

Use of the Posterior Probability Density

Expectation value of $\mu = <\mu> = \int_{-\infty}^{\infty} \mu' P(\mu' \mid N) \,\mathrm{d}\,\mu'$

Probability that μ lies between μ_1 and μ_2 =

$$P(\mu_1 \le \mu \le \mu_2) = \int_{\mu_1}^{\mu_2} P(\mu' \mid N) \,\mathrm{d}\,\mu'$$

Bayesian Approach for Blank Only

- Assume "uniform," "uninformative," or "flat" prior probability density
- Assume the likelihood probability density is a Poisson

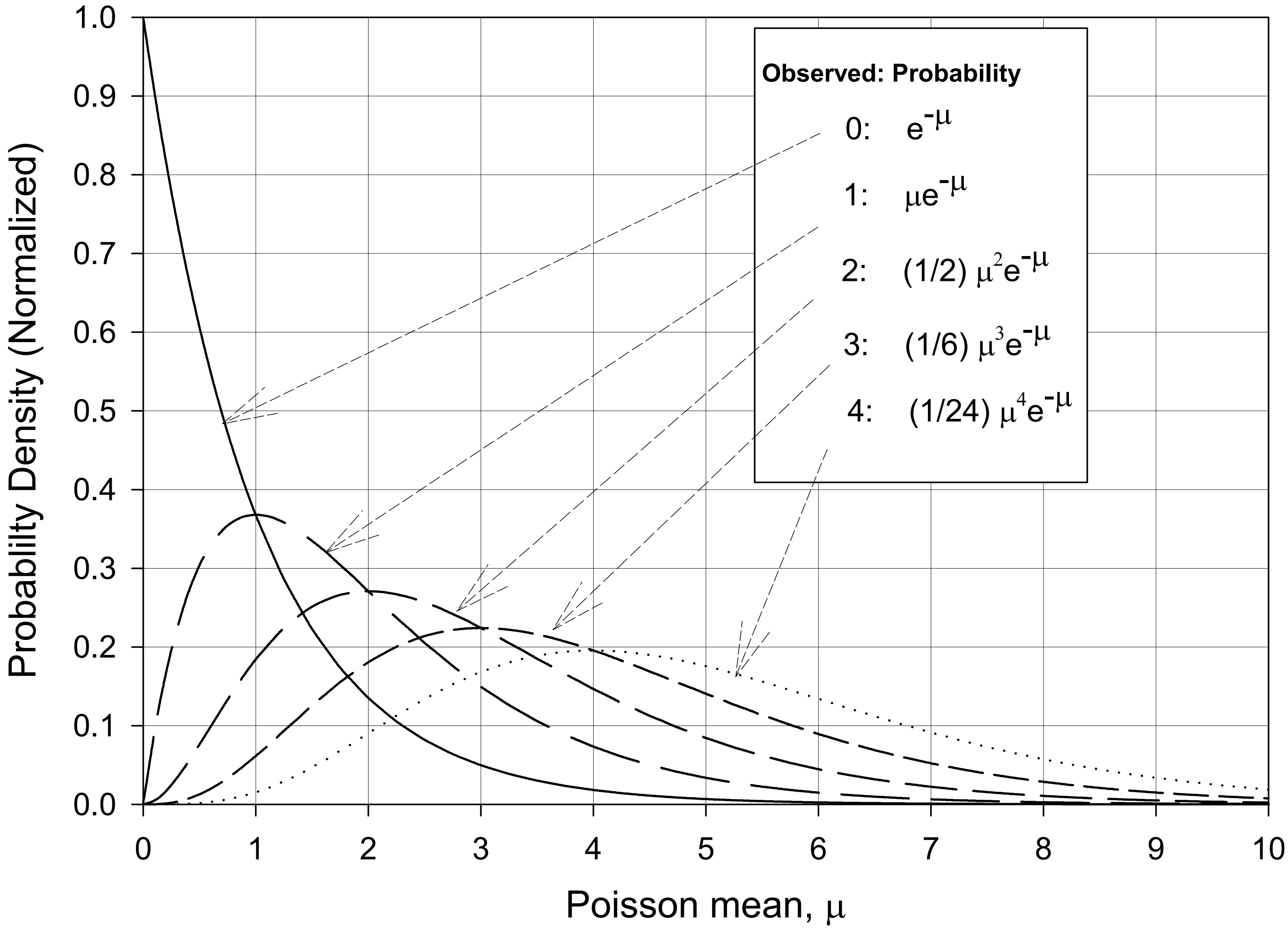
$$L(N \mid \mu) = \frac{e^{-\mu} \mu^{N}}{N!}$$

Bayesian Approach for Blank Only

- With a uniform prior, Bayes's rule inverts the likelihood to yield the posterior
- μ becomes a function of N, instead of N being a function of μ
- Posterior probability density:

$$P(\mu \mid N) = \frac{e^{-\mu} \mu^{N}}{N!}$$

Posterior Probability Densities for μ
(conditional on observed values)



Bayesian Approach

- Assuming uniform "flat" *prior* probability distribution: any value of N is equally likely
- If N counts observed
  - N is maximum likelihood, but $N + 1$ is expectation value:

$$< N_b > = N_b + 1$$

  - variance and standard deviation are simple:

$$\mathrm{Var}(< N_b >) = N_b + 1$$

$$s(< N_b >) = \sqrt{N_b + 1}$$

Ancient References for Bayesian *N*+1 Result Using a Flat Prior

Rainwater, L.J.; Wu, C.S. Applications of Probability Theory to Nuclear Particle Detection. Nucleonics 1(October):60-69; **1947.**

Friedlander, G.; Kennedy, J.W.; Miller, J.M. Nuclear and Radiochemistry. 2nd edition. New York: John Wiley & Sons, Inc.; **1955** & **1963**. The 1963 reference has a section on "Statistical Inference and Bayes' Theorem" (pp. 178-181).

Stevenson, P.C. Processing of Counting Data. NAS-NS-3109. Livermore, California: National Academy of Sciences -- National Research Council; **1966**.

Little, R.J.A. The Statistical Analysis of Low-Level Radioactivity in the Presence of Background Counts. Health Phys. 43(5):693-703; **1982**.

Flat Prior?

- “True” Bayesians are offended by a flat prior
  - “You always know more than nothing”
- Strom’s arguments for the flat prior
  - it is the best of both worlds, classical and Bayesian
    - inverting the prior gives a probability distribution that’s not available from classical methods
  - it obeys *Bohr’s correspondence principle*: ‘Any new theory must correspond to the old theory in the regime in which the old theory is known to be valid.’
  - it’s what you use on the first experiment
  - it does not require one to postulate that everything is drawn from the same population

Why the N13.30 Decision Rule Fails at Very Low Background Rates

- an observed background count causes a decision of "detected" ($\propto \mu_b$)
- unless a gross count is observed as well ($\propto \mu_b^2$)
- *independent* of α !
- so, for $\mu_b < 0.3$, $\alpha' \propto \mu_b - \mu_b^2$
- reason: false assumption that observed values N_b and $N_b^{1/2}$ are good estimates of the mean and standard deviation of background

McCroan/MARLAP/ISO Decision Rule

- Generalization of Altshuler & Pasternak

$$DL_{\mathrm{McCroan}}(R_{\mathrm{n}},\alpha) = \frac{k_\alpha^2}{2t_{\mathrm{b}}} + \frac{k_\alpha}{2}\sqrt{\frac{k_\alpha^2}{t_{\mathrm{b}}^2} + 4R_{\mathrm{b}}\left(\frac{1}{t_{\mathrm{g}}} + \frac{1}{t_{\mathrm{b}}}\right)}$$

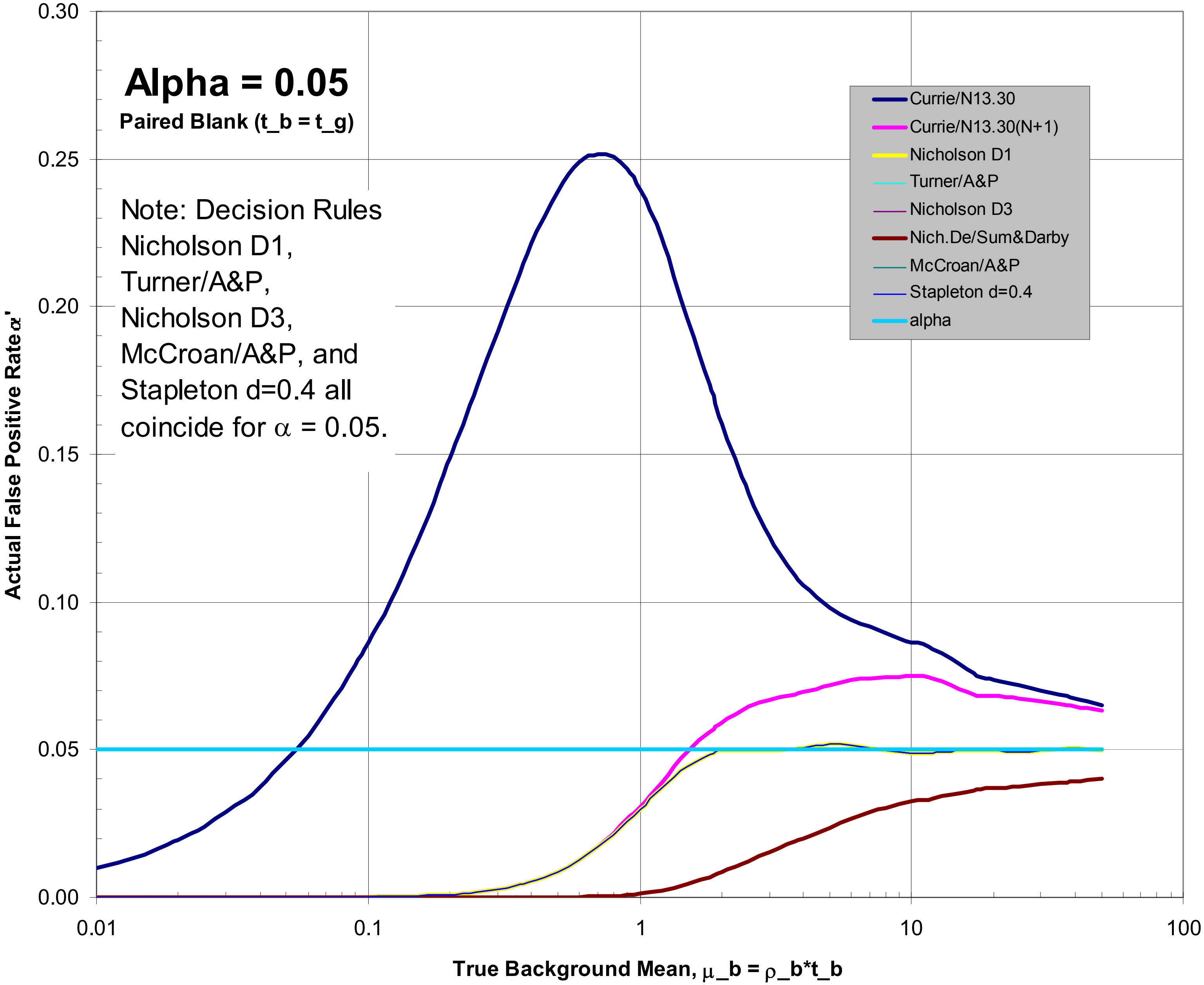
- MARLAP July 2001 *Draft*; same as ISO (*ISO notation*):

$$DL_{\mathrm{ISO}}(R_{\mathrm{n}},\alpha) = R_{\mathrm{n}}^* = \frac{k_{1-\alpha}^2}{2t_0}\left(1 + \sqrt{1 + \frac{4R_0 t_0}{k_{1-\alpha}^2}\left(1 + \frac{t_0}{t_{\mathrm{S}}}\right)}\right)$$

- Only differs from A&P when count times differ
- notation problem: Strom uses k_α where ISO uses $k_{1-\alpha}$

An Obvious Argument?

- using *both* the background and gross sample measurements to estimate the background increases the power of the test



Alpha = 0.05
Paired Blank (t_b = t_g)
Note: Decision Rules Nicholson D1, Turner/A&P, Nicholson D3, McCroan/A&P, and Stapleton d=0.4 all coincide for α = 0.05.
Currie/N13.30
Currie/N13.30(N+1)
Nicholson D1
Turner/A&P
Nicholson D3
Nich.De/Sum&Darby
McCroan/A&P
Stapleton d=0.4
alpha
Actual False Positive Rate α'
True Background Mean, μ_b = ρ_b*t_b
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.01
0.1
1
10
100

Results when $t_b = t_g$, $N_b < 10$

- Nicholson D_1, Turner/A&P, Nicholson D_3, and McCroan/ISO all coincide when $t_b = t_g$
- Nicholson D_2/Currie/N13.30/MARSSIM is poorest
- "$N + 1$" rule is much better, but not adequate
- Stapleton's rule is best, followed by the quartet, followed by D_e/S&D
- No rule is good below $N_b = 3$; smaller α is worse
- Need further work for different count times, $t_b \neq t_g$
- ANSI/HPS N13.30 under revision; so is MARLAP

Comparison of Decision Rules for $\alpha = 0.05$

$R_n t_g$	Currie (1968)	Altshuler & Pasternak Eq15	Sumerling and Darby (1981)	Currie using (N+1)	Nicholson (1963)
0	0 (1)	3	5	3	5
1	3	5	6	4	6
2	4	5	7	5	7
3	5	6	7	5	7
4	5	7	8	6	8
5	6	7	8	6	8
6	6	8	9	7	9
7	7	8	9	7	9
8	7	9	10	7	10
9	7	9	10	8	10
10	8	9	10	8	10

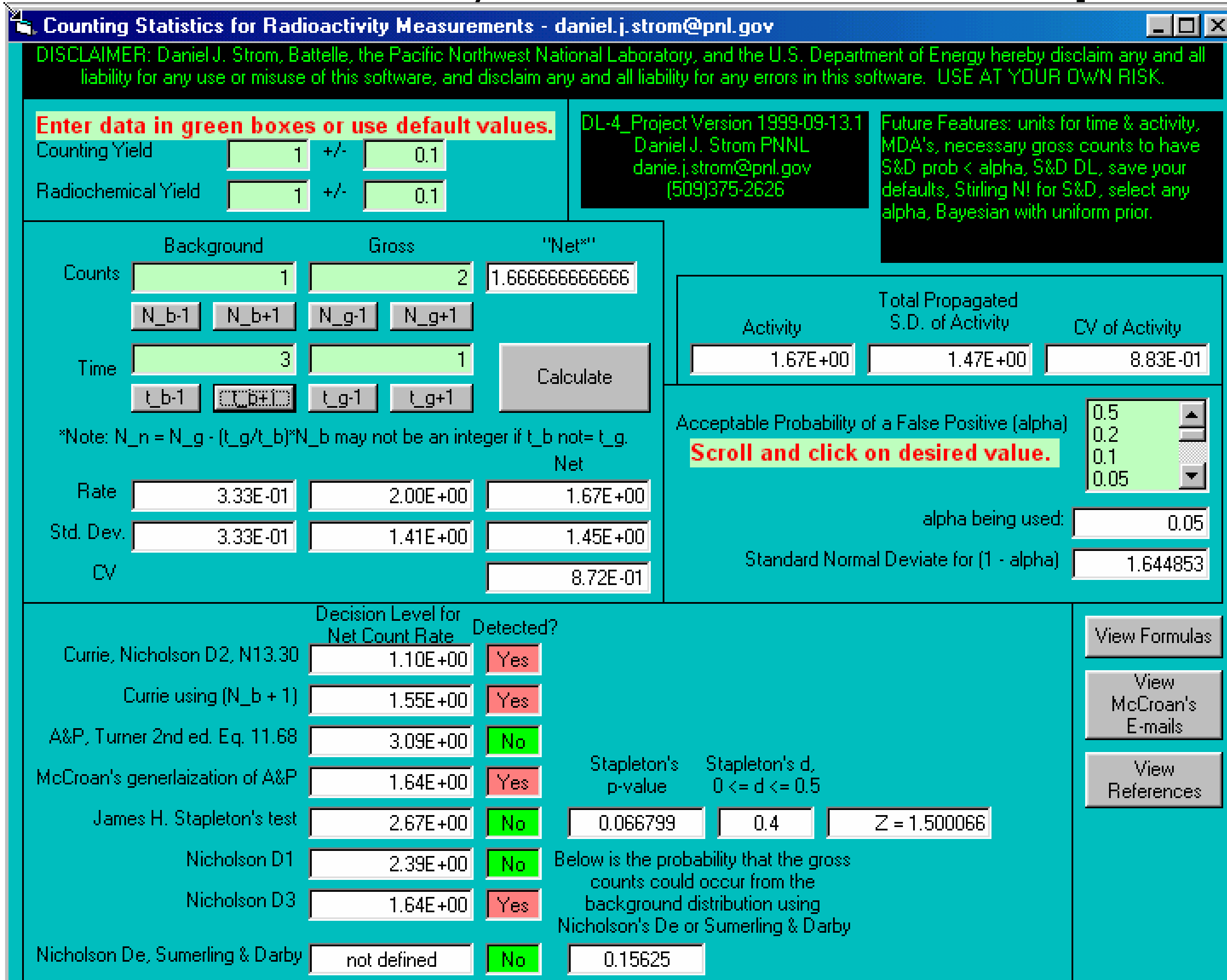
Software Utility under Development

- Freeware Windows 9x/2000/XP 32-bit GUI application
- Shows decisions for all 8 rules for any N_b, N_g, t_b, t_g.
- Not for public release yet – beta available
- Handles unequal background and gross count times
- Shows the amazing diversity of the decision rules
- Shows weakness of current "N13.30" rule

Software Utility under Development

[PNNL Counting Statistics Utility]

Software Utility under Development



Monte Carlo Proofs

- Crystal Ball is an add-in to Microsoft Excel
  - www.decisioneering.com
- Poisson simulation

Reporting and Recording of Measurement Results

“Censoring” of Data

- Censoring data means changing measured results from numbers to some other form that cannot be added or averaged or analyzed numerically
- Examples of data censoring
  - Left-censoring
    - changing results that are less than some value to zero
    - changing results that are less than some value to “less than” some value
  - Right-censoring
    - changing values from the measured result to “greater than” some value
  - Rounding

Why should censoring of data be avoided?

- Censoring means changing the numbers
- In a sense, it is dishonest
- If results are ever
  - summed,
  - averaged, or
  - used for some other aggregate analysis such as fitting a distribution,

  censoring makes this
  - difficult,
  - impossible, or
  - simply biased.

Censoring Examples

- Five results for discharge from a pipe taken over 1 year
  - uncensored results: –2, –1, 0, 1, and 2
  - sum = 0 (total discharge for the year is 0)
  - average = 0 (average discharge for the year is 0)
- Example 1: Set negative values to zero
  - censored results: 0, 0, 0, 1, and 2
  - sum = 3 (i.e., total discharge for the year is 3; this is not true)
  - average = 0.6 (i.e., average discharge for the year is 0.6; false)
- Example 2: Suppose $L_C = 2$. Set all values < 2 to "<"
  - censored results: <, <, <, <, and 2
  - sum = ? (total discharge for the year cannot be determined)
  - average = ? (average discharge for the year cannot be determined)

But Negative Activity Is Meaningless…

- No, it's not meaningless
- Just like money, subtracting a big number from a small number gives a negative value
  - You have 100€, you charge 200€, you owe 100€
  - 100€ – 200€ = –100€ (your net value)
  - this doesn't mean you can find a bank note for –100€
  - stocks go up and down; the end of the year value includes all changes, positive and negative
- Negative activity only means that random statistical fluctuations resulted in a negative number
- If negative, zero, or less-than values are suppressed, the sum is biased.

More Reasons Not to Censor

- Upper confidence limits of negative, zero, or less-than values
  - may be small positive numbers
  - needed for some applications (e.g., probability of causation)
- Censoring is prohibited by many standards and regulations
  - ANSI N13.30-1996: "Results obtained by the service laboratory shall be reported to the customer and shall include the following items …quantification using appropriate blank values of radionuclides whether positive, negative, or zero"
  - Many U.S. Department of Energy regulations require reporting raw data, calculated results (positive, negative, or zero), and total propagated uncertainties
  - Decision on actions can be made with uncensored data

Rounding Is Censoring

- Rounding a number is
  - changing its value
  - biasing the value
  - censoring
- Rounding often "justified" by claiming uncertainty
  - Uncertainty does not justify changing the answer
  - Explicitly state the uncertainty
- Beware of converting units of a rounded number and then rounding again!
- Intermediate results and laboratory records should never be rounded
- The only time to round is in presentations or communications

Report and Record All Measurements with No Censoring and Minimal Rounding

Exact Numbers, Imprecise Numbers, and Rounding of Numbers

2 Kinds of Numbers in the World

- *exact* numbers and *imprecise* numbers
- While “round” numbers are easy to deal with, rounding of exact numbers can cause difficulties

Exact Numbers Are Found in…

- counting (a catch limit of 3 trout, 14 coins)
- definitions (12 items in a dozen, 1000 m/km, 37,000,000,000 Bq/Ci)
- geometry (6 faces on a cube, π, φ[Golden mean])
- mathematics (the “2” in $E = mc^2$, e, √2)
- addresses (1600 Pennsylvania Avenue, www.pnl.gov, 192.234.201.101, 1-800-555-1212, a serial number, a Social Security number)
- laws (voting at age 21 and over)
- regulations (50 mSv per year, 1/3 WL, 4 WLM)

Perceiving "Exactness" or Precision

- the way a number is represented affects our perception of its exactness
- attempting to express the exact rational number 1/3 as a decimal leads only to successive approximations, depending on how many figures one uses: 0.3, 0.33, 0.333, etc.
- some exact numbers defy expression: e.g., π and $\sqrt{2}$
- exact numbers are arbitrarily precise and have no uncertainty

Imprecise Numbers

- measurements
- values inferred from measurements using models or theories
- approximate numbers or bounding numbers
  - "He's at least 170 cm and 80 kg."
- poorly recalled numbers
  - "I think she lives in the 300 block of Colley Avenue."
- estimates
- qualitative expressions of quantitative ideas
  - early, hot, fast, energetic, nearly full
- scientific wild assumption guesses (SWAGs)
  - "My dad is a million times stronger than your dad."

Imprecise Numbers

- It is important to separate conventions that are appropriate for results of measurements, e.g.,
  - the use of significant figures
  - rounding
  - explicit statements of uncertainty, such as 5.6 ± 1.3)

  from use with exact numbers

Expressing What We Know About a Quantity

How Many "Significant Figures" in

- 15?

  $= 1111_2 \Rightarrow 4$ significant figures (base 2)

  $= F_{16} \Rightarrow 1$ significant figure (base 16)

- MDCCCLXXXVIII?

  $= 1888_{10} \Rightarrow 4$ significant figures (base 10)

- 1/3?

  $= 0.1_3 \Rightarrow 1$ significant figure (base 3)

  $= 0.33333333\ldots_{10} = 0.\overline{3}_{10} \Rightarrow \infty$ significant figures (base 10)

  $= 0.01010101\ldots_2 = 0.\overline{01}_2 \Rightarrow \infty$ significant figures (base 2)

How Many "Significant Figures" in

- 1,073,741,824 ?

  $=2^{30}$

  $=100,000,000,000,000,000,000,000,000,000_2$

  $\Rightarrow$ 1 significant figure (base 2)

- $2^{30} \neq 10^9$

uppercase

IEC Prefixes for Binary Multiples

Binary Prefix	Symbol	Value	Value (base 10)
kibi	Ki	2^{10}	1,024
mebi	Mi	2^{20}	1,048,576
gibi	Gi	2^{30}	1,073,741,824
tebi	Ti	2^{40}	1,099,511,627,776
pebi	Pi	2^{50}	~1.1259E+15
exbi	Ei	2^{60}	~1.1529E+18

- International Standard IEC 60027 2, 2nd ed., 2000
- "Although these prefixes are not part of the SI, they should be used in the field of information technology to avoid the incorrect usage of the SI prefixes." http://physics.nist.gov/Pubs/SP330/sp330.pdf

“Significant Figures”

- The concept of significant figures is out of date in the computer age
- Abridging numbers (rounding) is fine for simplifying communication of quantitative values
- We must disaggregate what we know about the precision and accuracy of a number from how we represent the number
- There are 2 reasons not to round in scientific work:
  - rounding causes inaccuracies when intermediate results are used in further calculations
  - rounding causes dual values when doing unit conversions

Rounding Causes Inaccuracies

- ICRP 30 gave 1 “significant figure” for dose coefficients ‘because the values are not well-known’
- Rounding 1.49 down to 1, or 1.51 up to 2 doesn’t help! It only creates bias

What Could "1 Significant Figure" Mean?

	Range on 1 sig. fig. log scale	±1 in the last digit
1	0.95 – 1.49	+100% or – ∞%
2	1.5 – 2.49	± 50%
3	2.5 – 3.49	± 33%
4	3.5 – 4.49	± 25%
5	4.5 – 5.49	± 20%
6	5.5 – 6.49	± 16.7%
7	6.5 – 7.49	± 14.3%
8	7.5 – 8.49	± 12.5%
9	8.5 – 9.49	± 11.1%

Significant Figures

- As far as the phone company is concerned, my phone number is 5,093,752,626
- However, if the ICRP is calling, they can just dial 5E9 (rounded to 1 significant figure)
- Q: "What's the difference between 5001 millirems and 4999 millirems?"
- A: "6 months in jail and $10,000 per day."

Avoidable Errors: Rounding Causes Dual Values When Doing Unit Conversions

^{222}Rn	PAEC (WL)	EEC, traditional units (µCi/mL)	EEC, SI units (Bq m^{-3})
Exact value	1/3	3.3E–8	1.23E3
10 CFR 835 Appendix A published value	1/3	3E–8	1E3
Exact value is	-	11.1% higher than Appendix A value	23.3% higher than Appendix A value

- Values found in various locations in 10 CFR 835, DOE's Occupational Radiation Protection

Rounding Is Sometimes Okay

- For communication
- For simple conversions
  - 4×8 sheet of plywood is about 122×244 cm or even 1.2 × 2.4 m
- There is no need to express an uncertainty to unreasonable number of digits; usually 2 or 3 is sufficient

How We Represent Numbers

- We must abandon the notion that how we represent numbers is inevitably tied to their precision or accuracy uncertainty
- Example
  - computers usually represent numbers as single or double precision "floating point" numbers (about 7 or 14 digits with a sign and a power of 10) or short or long integers ($\pm (2^{15} - 1)$ or $\pm (2^{31} - 1)$)
  - The computer's internal representation of numbers as 1s and 0s makes no consideration of the precision or accuracy or even exactness of the numbers.
  - That doesn't make computers wrong

So, What Should We Do?

- Aggregating the value of a number with the uncertainty in the number is no longer sensible
  - especially for intermediate results that will be used in further calculations
  - e.g., dose coefficients (Sv/Bq)
- One way to express uncertainty:
  - elementary charge, *e*
  - 1.602 176 462 (63) 10^{-19} C
  - [fractional uncertainty] 3.9 10^{-8}

What We Really Need to Know About a Number

1. the quantity
2. the unit
3. how the number was obtained (measurement, calculation using 1 or more measurements, model, estimate, …)
4. the value
5. the uncertainty
6. the kind of uncertainty (standard deviation, geometric standard deviation, range)
7. how the uncertainty was obtained (e.g., repeated measurements, calculations, models, estimate…)

Example of 7-Vector for 2 Numbers

Component of 7-Vector	External Irradiation	Intake
Quantity	Deep Dose Equivalent	Committed Effective Dose
Unit	mSv	mSv
How obtained	OSL Dosimeter	Serial Pu/Am Urine & Fecal Bioassay; ICRP 68 Models; IMBA
Value	0.13	42
Uncertainty	0.03	3.5
Type of Uncertainty	S.D.	G.S.D.
How uncertainty obtained	ISO T.P.U.	Multi-tracer method; Resampling Statistics

Back to the Real World…
Decision Strategies

Additional *CV*s

- CV(tracer calibration) = $s_{\text{tracer}}/A_{\text{tracer}}$
- CV(tracer volume) = $s_{\text{tv}}/V_{\text{tracer}}$
- CV(aliquot volume) = $s_{\text{aliqout vol.}}/V_{\text{aliquot}}$
- CV(Type B system performance) = 0.03

Uncertainty Propagation Formula

- Simple form of uncertainty propagation ignoring covariances:

$$CV_{\text{total}} = \sqrt{\sum_{i=1}^{N} (CV_i)^2}, \text{ where } CV_i = \frac{u(x_i)}{x_i}$$

- "Total Propagated Uncertainty," *TPU*:

$$TPU = A \times CV_{\text{total}}$$

- Hanford uses $2\times TPU$ as a starting point for decisions

Practical Decision Making: A Multi-step Process Using Classical Statistics

- If a bioassay result is **unexpectedly** > 2×*TPU*, several possibilities
  - if result not excessively elevated, **recount** sample for 4× as long (i.e., 10,000 min for Pu α-spec at Hanford)
  - if recount result > 2×*TPU*, obtain new sample (if possible and meaningful) and count new sample
  - if new sample result > 2×*TPU*, decide "detected"
- Result: negligible false positive rate
- Result: minimization of false negative rate
- Moral of the story: you don't have to decide on only one measurement of one sample!

Bayesian Decision Process

- Depends on whose prior is used
  - "Management Prior" or "Optimist's Prior:" There's no way that there's activity in this sample (zero and low values are heavily weighted)
  - "Pessimist's Prior:" Of course there's activity in this sample, and it's bigger than you think (high values weighted)
  - "Ignorant (or Uniform) Prior:" Anything is possible
- Bayesians can recount and resample, too!
- More information is better, but takes time and €€

Classification and Misclassification Statistics

Definitions of Terms for Correct Classification and Misclassification

- A test, e.g., a nasal smear, may be for whether a person experienced an intake of radionuclides
- Do nasal smears correctly predict intakes and non-intakes?
- Wrong prediction (wrong classification) is a *spurious error*

		Classified State		
		+	–	Total
Actual State	+	*a*	*b*	*a*+*b*
	–	*c*	*d*	*c*+*d*
Total		*a*+*c*	*b*+*d*	*n*

Definitions of Terms for Correct Classification and Misclassification

$$prevalence = (a+c)/n$$

$$sensitivity = a/(a+b)$$

$$specificity = d/(c+d)$$

$$positive\ predictive\ value,\ PPV = a/(a+c)$$

$$negative\ predictive\ value,\ NPV = d/(b+d)$$

$$false\ negative\ rate = 1 - sensitivity$$

$$false\ positive\ rate = 1 - specificity$$

Use of Statistics to Evaluate Nasal Swabs

- Obtain real data from a plutonium facility
- Bioassay triggered by workplace indicators
  - Soon after suspected intake, perform nasal swabs
  - Follow up with bioassay samples (fecal)
  - Bioassay = Actual State
  - Nasal Swab = Classified State
- Are nasal swabs a good indicator of intake as confirmed by bioassay?
- Acknowledgement: Terry A. Brock did this work

Nasal Swab and Bioassay Data

Bioassay Results	Nasal Swab Positive (≥*DL*)	Nasal Swab Negative (<*DL*)	Total
Positive (≥*DL*)	36	181	217
Negative (<*DL*)	24	82	106
Total	60	263	323

Data courtesy Terry A. Brock, Ph.D.

Are Nasal Swabs Worse than Nothing?

Statistic	**Results**	Lower 95% CI	Upper 95% CI
Sensitivity	**0.166**	0.116	0.215
Specificity	**0.774**	0.694	0.853
Positive PV	**0.600**	0.476	0.724
Negative PV	**0.312**	0.256	0.368
Prevalence of Intakes	**0.672**		
False Negative Rate	**0.834**		
False Positive Rate	**0.226**		

Nasal Swab Conclusions

- If nasal swab is "very high," there probably was an intake
  - Highest 10 nasal swabs were all associated with intakes
- If nasal swab is zero or other than "very high," it means essentially nothing
  - How can this be?
  - No inhalation intake occurred, but
    - particles stopped in nose or mouth, were removed by swab
  - There was an intake, but
    - nose blow cleared particles
    - no particles deposited in nose or mouth
    - nasal swab didn't capture particles

Consequences of Wrong Decisions

Do We Need Better Decision Rules?

- False positive bioassays
  - needlessly place worker on work restriction
  - needlessly alarm worker
  - needlessly spend money on unneeded re-sampling and analyses
  - needless investigations
- False negative bioassays
  - not protective of worker
  - if later discovered, can destroy trust and communication

Do We Need Better Decision Rules - 2?

- False positive environmental samples
  - unnecessary, costly cleanup
  - needlessly alarm public
  - needlessly spend money on unneeded resampling and analyses
  - needless investigations
  - political consequences
- False negative environmental samples
  - not protective of public and environment
  - if later discovered, can destroy trust and communication
  - political consequences

Averaging to Improve Detection Capabilities

Applicable Under Some Circumstances

- Examples
  - Monthly, weekly, or daily samples of discharges from a pipe or stack that must meet an annual limit
  - Multiple samples from a "risk analysis unit"
  - Any time a standard applies to a volume that is large compared to sample volume
  - Evaluation of "missed dose" from pooled bioassay samples
- Why it works: mass or number of samples scales with n, counting uncertainty scales with $\sqrt{n}$

Air Sample Example (NUREG-1400)

- n is number of samples
- T_b is blank count time
- T_g is gross (sample) count time
- R_b is background count rate
- E is filter efficiency
- T_S is sampling time
- F is air flow rate; FT_S is volume sampled
- K is counting efficiency (counts s^{-1} Bq^{-1})
- $MD\overline{C}$ is "minimum detectable *average* concentration"

Air Sample Example (NUREG-1400)

$$MD\overline{C} = \left(\frac{\sum_{i=1}^{n} T_{S,i}}{\sum_{i=1}^{n} E_i F_i K_i T_{S,i}^2} \right) \left(\frac{3}{\sum_{i=1}^{n} T_{g,i}} \right) + 3.29 \sqrt{\frac{\sum_{i=1}^{n} \frac{R_{b,i} \left(\frac{1}{T_{g,i}} + \frac{1}{T_{b,i}} \right)}{E_i^2 F_i^2 K_i^2}}{\left(\sum_{i=1}^{n} T_{S,i} \right)^2}}$$

- and if all efficiencies, flow rates, sample times, etc. are equal,

$$MD\overline{C} = \frac{3}{nEFKT_S T_g} + 3.29 \frac{\sqrt{R_b \left(\frac{1}{T_g} + \frac{1}{T_b} \right)}}{\sqrt{n} EFKT_S}$$

Method Works to Extract Signal from Noise

- Summing γ-Spectra
  - 46-keV (4%) ^{210}Pb γ-photon not visible or detected by IGe in vivo counting of 20 uranium miners
  - summing 20 spectra revealed a 46-keV peak (Palmer 1984)
  - can't tell which miner it's in, but can tell what average and total is!
- $1/\sqrt{n}$ principle is widely used
  - kinetic phosphorimetry for chemical determination of U
  - ion cyclotron resonance mass spectroscopy for proteomics
  - any multiple-interrogation technique
- Can add counts of same sample made at different times
  - e.g., recount

Probabilistic Background Subtraction

The Problem: 2 Kinds of "Background"

- Question 1: Is there activity in the sample above blank? If "no," stop.
- Question 2: If the answer to Question 1 is "yes," how much of the activity is due to ambient levels and how much is due to a particular source (e.g., the workplace, release from a nuclear power plant)?

Problem Radionuclides: ^{137}Cs and ^{nat}U

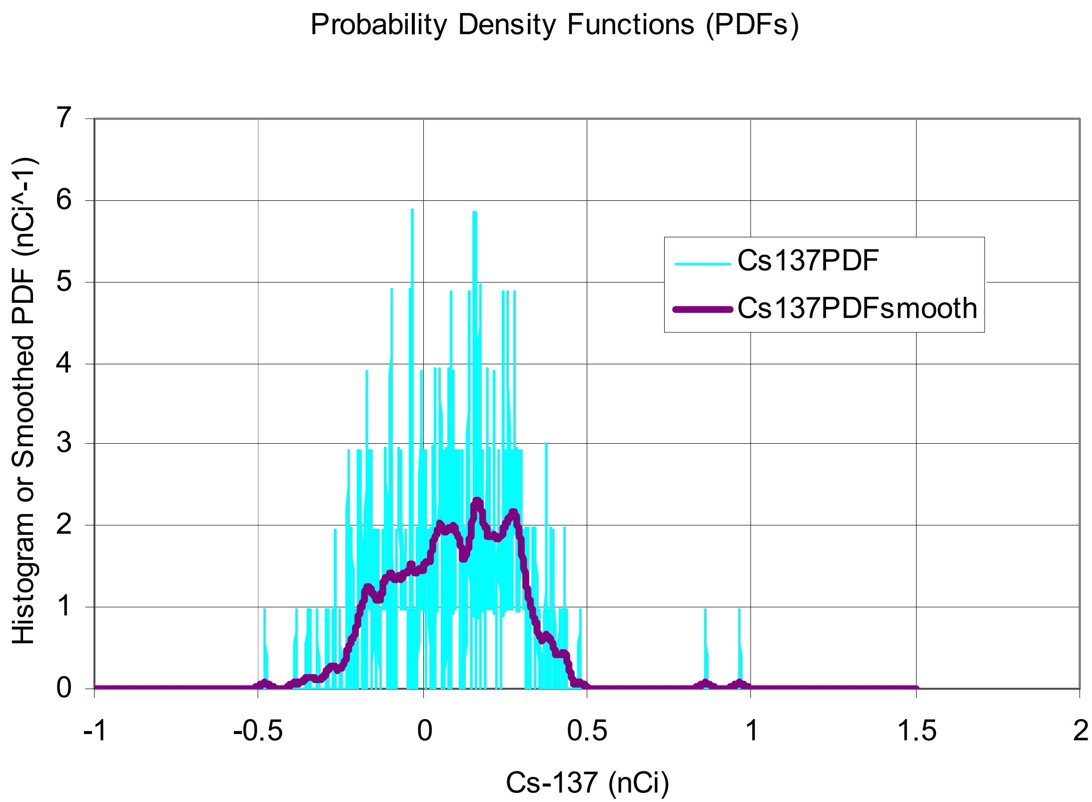
- Your job: analyze environmental or bioassay samples and determine what is "natural" and what is occupational or anthropogenic (human-made)
- ^{137}Cs is ubiquitous (fallout and orphan source meltings at steel mills) as is ^{nat}U
- Solution(s)
  - tricks like ^{236}U as a tracer for "recycled" U
  - baseline measurements (not possible after the fact)
  - analysis of nearby environment or similar workers
  - probabilistic background subtraction (makes the most of what you've got)

Example: ^{137}Cs in a Worker

- Intrinsic germanium (IGe) in vivo chest count
- ^{137}Cs is a tracer for Hanford tank waste, a mixture of fission products, traces of U and transuranics (TRU)
- ^{137}Cs produces much less dose than TRU
- *Question: Is the ^{137}Cs from fallout or the workplace?*
- Answer: make a probabilistic statement about the measurement

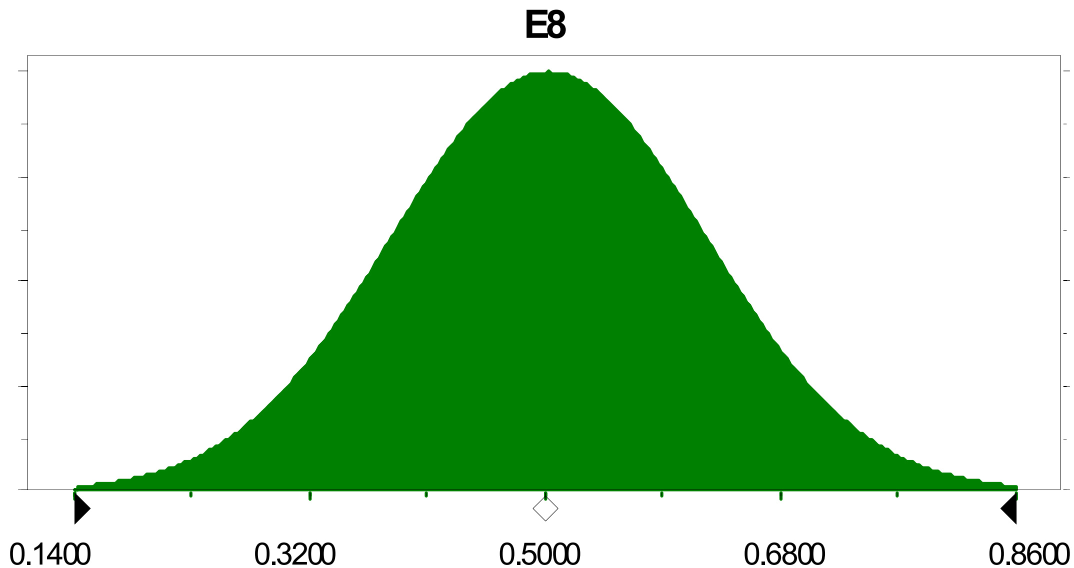
Whole-Body Counts of 409 Unexposed Workers

- Average 3.46 Bq (0.094 nCi)
- Chomentowski & Kellerer (2000) smoothing shown
- For a given worker result with its stated uncertainty, one can subtract "background" 409 different times (not shown)

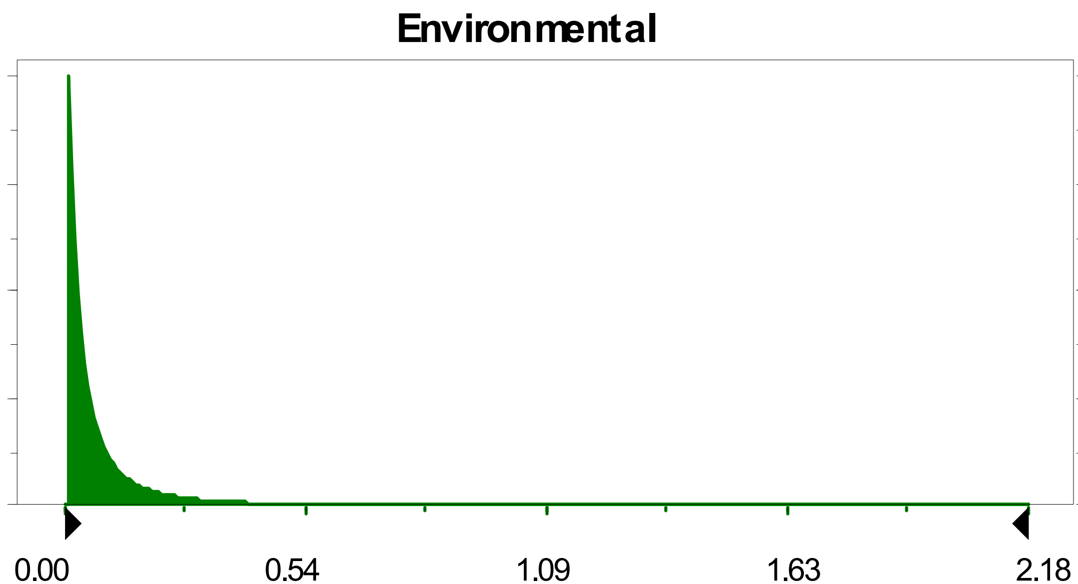


2 Distributions: Worker and Environmental

- Worker: Normal distribution, 0.50 ± 0.12 dpm/d

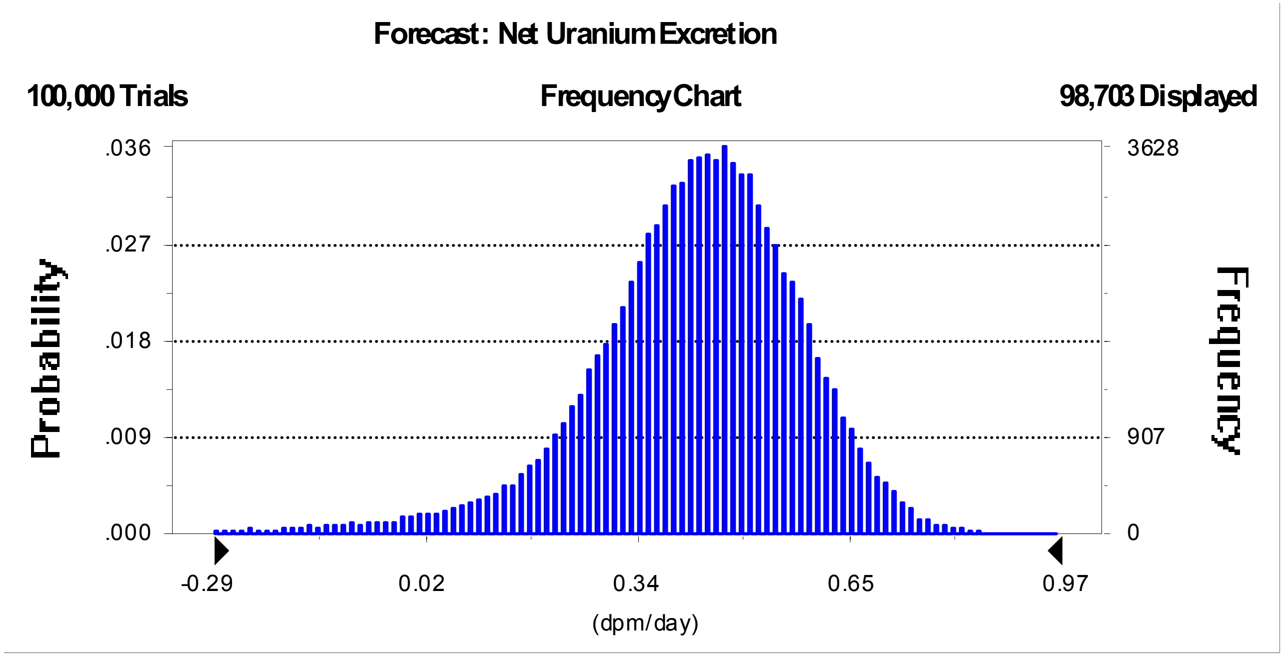


- Environmental: lognormal dist., geo. mean = 0.037 dpm/d, GSD = 3.9



Probabilistic Subtraction of Environmental U from Single Worker Bioassay

- mean 0.407 ± 0.241 dpm/d
- median = 0.436 dpm/d
- 95th %ile = 0.655 dpm/d
- 99th %ile = 0.695 dpm/d



The Future: A Personal Forecast

- Distributions are the way of the future
- In USA, probability of causation is being tied to 95th %ile; this will spread to other fields
- Regulatory bodies, competent authorities, will begin to require compliance to upper percentiles, not means or medians, as "protective" of public, workers, and the environment
- Probabilistic methods will become the norm

Non-Bayesian Methods for Evaluating Uncertainty in Complex Models

Non-Bayesian Methods for Evaluating Uncertainty in Complex Models

- A variety of methods are available for evaluating Type B uncertainty (e.g., model suitability)
- Biokinetic models have been prototyped, but environmental models are similar
- Methods include
  1. A "resampling" method
     - doesn't mean "obtaining another sample," but rather, it means "sampling subsets of all data"
  2. Updating: Recalculating every time you get another data point
  3. Use of multiple radionuclides one at a time
- Strom 2003

Inferring Dose from Measurements, Models, Assumptions, or None of the Above

Dosimetry

- "dose" + metry
  - root is *metron* (Greek: to measure)
- current usage: any dose number is presumed to be the result of "dosimetry"
- thesis
  - If measurement or observation is the dominant activity, and
  - uncertainties in results are predominantly due to measurement uncertainty,

  use the word "dosimetry." Otherwise, maybe new terms would be more appropriate!

Measuring the Quantity of Radiation

- observation of biological response (e.g., erythema, chromosome aberrations)
- cloud chambers
- film blackening
- appearance or sound of bubbles in superheated liquids
- analysis of activation or fission product yield
- scintillations
- Cherenkov radiation (light)
- thermoluminescence (TL) or optically stimulated luminescence (OSL)
- observation of radiation damage (e.g., chemically etching damage in film, radiochromic changes, thermal and electrical conductivity changes)
- chemical changes as quantitated by light absorption or nuclear magnetic resonance
- measurement of electric charge or current in solids (Ge and Si) or gases such as xenon, P10, or air, and
- calorimetry

Dosimetry for External Irradiation

- most measurements are *outside* of the human body
- want to know dose inside or at surface
- external irradiation: few inferential steps
  - absorption
  - albedo
  - spectrum changes
  - based on types, energies, directions of incident radiation
  - assumptions about person wearing dosimeter
  - neutrons still a challenge
- irradiation following intake or ontake of radioactive material
  - surgical implantation of dosimeters? no.
  - inference

Dosinference for Internal Irradiation

- blend of "dose" + "inference" (Strom 2002)
- uncertainties associated with inferential steps dwarf uncertainties of measurement steps
  - exceptions: ^{3}H and alkali metals, e.g., ^{137}Cs
- measurements tend to be of dose-rate like quantities, rather than dose-like quantities
  - rate of photon emission from regions of body (in vivo counting)
  - count rate or numbers of atoms (TIMS, ICP-MS) in excreta
  - count rates from air samples
  - exception: chromosome aberrations
- infer activity (and its uncertainty) in organs and tissues from measurements and biokinetic models

Non-Measured Inputs to Dosinference

- knowledge or guesses of time course and route(s) of intake
- identity of all radionuclides and proportions in a mixture
- particle size distribution and transportability for inhalation
- gastrointestinal (GI) tract absorption
- chemical and physical form for ingestion, injection, wound, or dermal absorption from an ontake
- true daily excretion rate for in vitro bioassay (non 24-h samples)
- biokinetic models
  - Reference Man usually used, not individual data
  - individual chest wall thickness and ^{40}K corrections
  - site-specific solubility, e.g., Y-12's Class Q uranium
- air sample data, stay time, respiratory protection data, respiratory tract model

What's Uncertain When Inferring Intake?

- Circumstances
  - time or time course of intake
  - route(s)
- Material characteristics
  - radionuclide mixture
  - particle size and shape
  - chemical form(s) and transportability (S, M, F, or real)
- Measurements
  - counting or measurement uncertainty
  - 24-h sample? simulated? adulterated or contaminated?
- Biological variability
  - availability and validity of model(s)
  - systematic differences between individual and models
  - among bioassay samples or measurements
- Interpretation
  - interference from environmental exposures
  - prior intakes

Dosinference from Radon Progeny

- short-lived decay products of radon & thoron
  - particle size
  - equilibrium factor
  - unattached fraction
  - smoking
  - nose breathing
  - level of exertion
  - diurnal variations
- ICRP (1995) “dose conversion convention”
  - 5 mSv/WLM rather than 12.5, based on epidemiology

Doswaggery

- blend of "dose" + "swag" (Strom 2002)
  - root is acronym for *scientific wild assumption guess* (US popular usage)
  - examples of *swags*
    - predicting the weather two weeks in advance
    - predicting the value of the stock market in a year
- uncertainties in assumptions dwarf even the uncertainties in the inferential steps, much less the uncertainties in the measurements
- may not rely on measurement at all, or may rely on measurements only tenuously associated with individual for whom a dose is being inferred

Imputed Values

- to “impute” has taken the meaning to “make up a number”
  - Reissland (1982) used the term “notional dose” for what today is termed an “imputed dose”
- lost or damaged external dosimeter, spoiled bioassay or air sample
- imputation commonly done for regulatory compliance
  - interview worker & colleagues, dose rates, time-in-area
  - average preceding and subsequent dosimeter results
- can be very accurate
  - CARI-6 for air travel

  http://www.cami.jccbi.gov/AAM-600/610/600Radio.html
- can be done for “less than detectable” results

Doswaggery to Impute Doses

- not all imputed doses are doswags
  - production lines
  - radiology department with steady caseload
  - careful dose reconstructions such as RERF DS02
- examples of doswaggery:
  - assigning historical uranium miners potential alpha energy exposures (J h m^{-3} or WLM) based on measurements in similar mines
  - historical dose reconstruction for litigation in U mining, milling, refining in absence of any concurrent workplace measurements
  - some projections of future (50-year “committed”) doses
  - population doses from high level waste repositories

Uncertainty Is Not *Necessarily* Error

- "The result of a measurement (after correction) can unknowably be very close to the value of the measurand (and hence have a negligible error) even though it may have a large uncertainty. Thus the uncertainty of the result of a measurement should not be confused with the remaining unknown error." – ISO (1995)
- a doswag *may* be accurate but *is* highly uncertain
- long-range weather forecasts are sometimes correct!

Word Choice Based on Uncertainty

Term	Typical Dominant Uncertainty				Ratio of 97.5%ile to 2.5%ile of Inferred Dose
	Measure-ments	Models	Model Parameters	Imputed Data	
Dosimetry	✓✓	✓	~	~	**1.01 to 2**
Dosinference	✓	✓✓	✓✓	~	**2 to 20**
Doswaggery	~	✓	✓	✓✓	**>20**

✓ denotes important; ✓✓ denotes very important; ~ relatively trivial

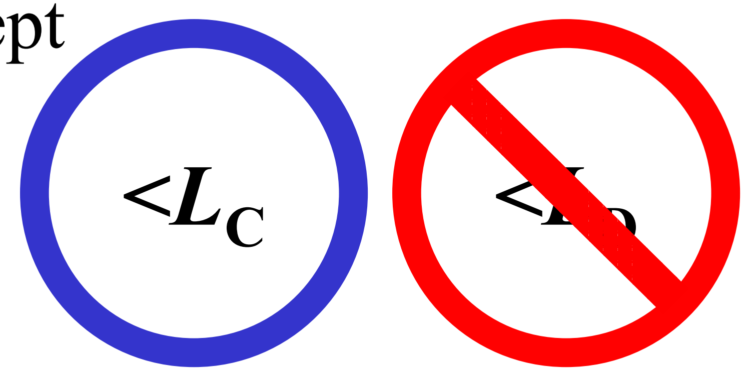
Calling a Spade a Spade…

- maybe it's time to choose different words when the dose in question is measured, inferred, or essentially assumed
- *dosimetry* when measurement uncertainty predominates
- *dosinference* when model parameter uncertainty predominates
- *doswaggery* when assumption or imputed value uncertainty predominates

Conclusions and Recommendations

Conclusions 1

- There are severe problem of terminology
  - many names for the same concept
  - concepts mis-named
  - concepts mis-used
  - compare measurements with decision threshold L_C, not detection level L_D



- There are 2 "counting problems"
  - Forward problem done well by classical statistics
  - Reverse problem only done by Bayesian methods
  - Bayesian result: expectation value is N+1, so uncertainty is greater

Conclusions 2

- There are at least 8 decision rules
  - Most use a "Great Leap of Inference:" N is a good estimate of ρt & Var(ρt)
  - False positive rates are not as claimed for small numbers of counts
- ANSI N13.30/MARSSIM/Currie/etc. *DL* formula
  - yields too many false positives
  - fairly good only if counts > 100

Conclusions 3

- There are much better approximate & exact solutions
  - $N + 1$ rule is better, but not correct
  - Altshuler & Pasternak/Turner better (??? stay tuned for MARLAP)
  - Detailed Bayesian method works if a flat prior is acceptable
  - Nicholson's (1963) and Sumerling's & Darby's (1981)exact (classical) solution can be implemented - probably the soundest theoretical foundation, but very few false positives
  - McCroan/ISO 11929 ok for equal count times
  - Stapleton's rule performs the best
  - Rigaud (2003) not yet evaluated
  - **No rule works well at very low counts!**

Conclusions 4

- MARLAP is eagerly awaited
- More work to be done on software – this shouldn't be so hard!
- Consensus needed
- Bayes's methods need to be mainstreamed

Conclusions 5

- Decision strategies (Bayesian and classical)
  - You can re-count a sample and decide again
  - You can count the other half of a split or re-sample
  - Combinations of these lead to essentially no false positives
- An enumeration of the consequences of wrong decisions shows that we need the best decision rules we can get
- Detection capabilities: a decision on decision rules needed first
- Report and record uncensored, un-rounded

  positive, zero, and negative measurements

  results with uncertainties

  to avoid falsely exaggerating levels

Conclusions 6

- There is great utility in **averaging results** of many samples when appropriate
  - If a regulatory guideline applies to a year, average samples over a year
- **Probabilistic blank and environmental background subtraction** maximizes information content while distinguishing human-made signal from environmental background
- There exist several **non-Bayesian methods of uncertainty evaluation** in modeling

References

- Visit http://bidug.pnl.gov/ (DOE Bioassay and Internal Dosimetry Users Group)
- Visit http://www.pnl.gov/bayesian for many Bayesian links, including Los Alamos National Laboratory (LANL) site and its links

Acknowledgement

- Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle under Contract DE-AC06-76RLO 1830
- Contact: strom@pnl.gov

The opinions expressed in this presentation are those of Daniel J. Strom may or may not represent those of Battelle, the Pacific Northwest National Laboratory, or the U.S. Department of Energy

References

- Altshuler B and BS Pasternack. 1963. "Statistical Measures of the Lower Limit of Detection of a Radioactivity Counter." Health Physics 9:293-298.
- Berry DA. 1996. *Statistics: A Bayesian Perspective*. Wadsworth Publishing Company, Belmont, California.
- Brodsky A. 1992. "Exact Calculation of Probabilities of False Positives and False Negatives for Low Background Counting." Health Physics 63(2):198-204.
- Currie LA. 1968. "Limits for Qualitative Detection and Quantitative Determination. Application to Radiochemistry." Analytical Chemistry 40(3):586-593.
- d'Agostini G. 1999. *Bayesian reasoning in high energy physics. Principles and applications*. CERN Yellow Report 99-03. http://www-zeus.roma1.infn.it/~agostini/index.html, Dip. di Fisica Università "La Sapienza" and INFN, Rome.
- Gilbert RO. 1987. *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold, New York.
- Hickey EE, GA Stoetzel, DJ Strom, GR Cicotte, CM Wiblin, and SA McGuire. 1993. *Air Sampling in the Workplace*. Final Report. NUREG-1400; PNL-7814, U.S. Nuclear Regulatory Commission, Washington, DC.

References

- Howson C and P Urbach. 1991. "Bayesian Reasoning in Science." Nature 350:371-374.
- International Organization for Standardization (ISO). 1995. *Guide to the Expression of Uncertainty in Measurement. Corrected and Reprinted, 1995*. International Organization for Standardization, Geneva.
- Jaynes ET. 2003. *Probability Theory : The Logic of Science*. Cambridge University Press, New York.
- Kathren RL. 2001. "Everything But the Counting Statistics: Practical Considerations in Instrumentation and Its Selection and Use." Chapter 17 in *Radiation Instruments*, ed. H Cember, pp. 361-384. Medical Physics Publishing, Madison, Wisconsin.
- Little RJA. 1982a. "The Statistical Analysis of Low-Level Radioactivity in the Presence of Background Counts." Health Physics 43(5):693-703.
- Little RJA. 1982b. "Hypothesis Tests Based on Counts from Blank and Radioactive Samples." Health Physics 42(2):230-231.
- Nicholson WL. 1963. *Fixed Time Estimation of Counting Rates with Background Corrections*. AEC Research and Development Report HW-76279, Hanford Works, Richland, Washington.

References

- Ott WR. 1995. *Environmental Statistics and Data Analysis*. CRC Press LLC, Boca Raton, Florida.
- Palmer HE. 1984. "In-Vivo Counting of Uranium." in *Biokinetics and Analysis of Uranium in Man*. Proceedings of a Colloquium held at Richland, Washington, August 8-9, 1994. USUR-05 HEHF-47, ed. RH Moore, p. I-1-I-29. Hanford Environmental Health Foundation, Richland, Washington.
- Strom DJ and JA MacLellan. 2001. "Evaluation of Eight Decision Rules for Low-Level Radioactivity Counting. PNNL-SA-32868." Health Physics 81(1):27-34.
- Strom DJ. A Novel Method of Evaluating Uncertainty in Intake and Effective Dose from Bioassay Measurements Using Existing Computer Codes. Presented at the 49th Annual Radiobioassay and Radiochemical Measurements Conference, Jackson, Wyoming, October 12-16, 2003. PNNL-SA-39669. Richland, Washington: Pacific Northwest National Laboratory; 10-12-2003. http://www.pnl.gov/bayesian/strom/strombio.htm

References

- Taylor BN and CE Kuyatt. 1994. *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*. NIST Technical Note 1297 http://physics.nist.gov/cuu/pdf/tn1297.pdf, National Institute of Standards and Technology, Physics Laboratory, Gaithersburg, MD 20899.
- Thomas J. 1963. *Interpretation of Low-Activity Counting*. Risø Report No. 70, Danish Atomic Energy Commission, Risø, Denmark.
- Turner JE. 1995. *Atoms, Radiation, and Radiation Protection*. John Wiley & Sons, New York.
- U.S. Environmental Protection Agency (EPA), U.S. Nuclear Regulatory Commission (NRC), U.S. Department of Energy (DOE), and U.S. Department of Defense. 2003. *Multi-Agency Radiological Laboratory Analytical Protocols Manual. Draft*. http://www.epa.gov/radiation/marlap/about_marlap.htm, U.S. Environmental Protection Agency, Washington, DC.
- U.S. Nuclear Regulatory Commission (NRC), U.S. Environmental Protection Agency (EPA), U.S. Department of Energy (DOE), and U.S. Department of Defense. 2000. *MultiAgency Radiation Survey and Site Investigation Manual (MARSSIM; http://www.epa.gov/radiation/marssim/)*. NUREG-1575, U.S. Government Printing Office, Washington, DC.