STATISTICAL BEHAVIOR OF NUCLEAR COUNTING EQUIPMENT

In my original article (*1*), I described tests for statistically evaluating the operational reliability of nuclear counting equipment. The objective was simply to point out that such tests were available, that the tests were relatively simple to perform, and that the tests could be conveniently implemented by any laboratory regardless of the degree of sophistication of calculating aids available. The useful information and criticisms provided by Ms. Gerson's letter [this issue, p. 217] add to the discussion of quality assurance testing of nuclear counting equipment. However, her letter contains several misconceptions and perhaps some factual errors. In order to respond adequately, it will be necessary to go into considerably more detail than originally intended. My comments are fully referenced so that serious readers can refer to the literature and develop their own attitudes.

The main issues as I understand them are nomenclature and the appropriateness of using small sample statistics for the calculation of the value of the chi-square statistics. I will attempt to clear up this confusion by pointing out that the nomenclature used is variable, depending upon the textbooks consulted, and that the use of small sample statistics is based upon a philosophical approach as well as mathematics.

Nomenclature

Textbook advice is not always correct, but the use of the chi-square statistic defined in Eq. 6 of my original article as a goodness-of-fit test is, I believe, established terminology in medical physics and radiotracer methodology when applied to the statistical behavior of nuclear counting equipment. For example, Hendee comments that "the chi-square test is used to determine the 'goodness of fit' of measured data to a Poisson probability density function" and goes on to describe the same chi-square statistic (*2*). Evans discusses the same statistic for measuring the randomness of G–M counting data (*3*). This statistic can be derived from the chi-square test for the homogeneity of variance. Snedecor and Cochran imply that the chi-square variance test and the chi-square test used as a goodness-of-fit measure are different entities, but when applied to the specific problem of radionuclide counting data, the distinction is an arbitrary one and is not made by all authors (*4*). Hoel also includes the statistic under question (*5*). The following will shed additional light on the subject.

A statistic is a rule or algorithm for obtaining a number (*6*). A probability density function is the distribution of that number (or variable) under discussion (*7*). Thus, a chi-square statistic can be defined (*8*) as

$$\chi^2 = \frac{(n-1)\,S^2}{\sigma^2}\ , \qquad (1)$$

where n is the number of observations, S the sample standard deviation, and σ the presumed or true standard deviation of the population from which the sample was obtained.

This statistic is sometimes called the chi-square variance test or a test for the homogeneity of variance and is used to test the null hypothesis that $S = \sigma$.

If samples of size n are drawn from a Gaussian distribution with a standard deviation of σ, for which the chi-square value is calculated, then a sampling distribution is obtained. This distribution is called the chi-square distribution (*9*) and is given by

$$Y = Y_c\, \chi^{\nu-2} \exp\left(-\tfrac{1}{2}\chi^2\right), \qquad (2)$$

where ν is the number of degrees of freedom.

It is well known that the statistical law which governs radioactive decay is the Poisson distribution, in which the standard deviation of the distribution is identically equal to the square root of the mean. Further, when the number of counts obtained per unit time interval is large (100 or more), the Poisson distribution can be approximated with negligible error by the Gaussian distribution with a standard deviation equal to the square root of the mean.

Theoretically, one can conceive of a distribution which is either leptokurtic or platykurtic and/or skewed and still have a standard deviation equal to the square root of the mean. However, when one is dealing with replicate counts of a radioactive source this is highly improbable if the instrument is working properly and an obvious statistical test is to compare the sample standard deviation to the square root of the mean of the sample. By substituting $\bar{X}$ for σ^2 and the numerical definition of S^2 into Eq. 1, the following results:

$$\chi^2 = \sum_{i=1}^{n} \frac{(X_i-\bar{X})^2}{\bar{X}}\ . \qquad (3)$$

The confidence intervals for Eq. 3 are determined from Eq. 2. If the calculated chi-square value falls within predetermined confidence intervals, then the conclusion is that the population sampled indeed belongs to a single Poisson distribution, where $S^2 = \bar{X}$. However, if the chi-square value falls outside the predetermined confidence interval, then the conclusion is that the individual counts obtained belong to more than one Poisson distribution (not that the counts obtained do not follow a Poisson distribution).

The statistic described by Eq. 3 has been termed the Poisson Index of Disperson by Hoel (*10*), a term also used by Price (*11*). However, a goodness-of-fit test is simply a test in which an observed frequency distribution is compared with a theoretical distribution such as a Poisson distribution, a Gaussian distribution, a binomial distribution, etc. (*4,12*). This is precisely what the statistic in Eq. 3 is used for in this application.

Looking at the problem in a slightly different way, consider the following test statistic:

$$T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} , \tag{4}$$

where k is the number of classes, O_i the observed number of observations in class i, and E_i the expected number of observations in class i.

Conover notes, "The exact distribution of T is difficult to use, so the large sample approximation is useful. The approximate distribution of T, valid for large samples, is the chi-square distribution with (k−1) degrees of freedom" (*13*). Thus, Eq. 4 can be written as

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_1)^2}{E_1} . \tag{5}$$

The chi-square statistic defined in Eq. 5 is called the chi-square test and is used to compare an observed frequency distribution with a theoretical frequency distribution, i.e., it is used as a measure of goodness of fit. At this point it is important to distinguish between the concept of a goodness-of-fit test and the procedure by which we measure the goodness of fit. For example, in a series of replicate measurements of a radioactive source it can be shown (*14,15*) that the E_i for all classes can be considered equal with an expectation value of $\bar{X}$. Thus, Eq. 5 can be written as

$$\chi^2 = \sum_{i=1}^{k} \frac{(X_i - \bar{X})^2}{\bar{X}} , \tag{6}$$

which is the same as Eq. 3. Thus, the same chi-square statistic is derivable from either the chi-square variance test or the chi-square test. Note that Dixon and Massey (*15*) also include this statistic as a test for goodness of fit.

The information presented here, I believe, adequately justifies calling the chi-square statistic presented as Eq. 6 in my original article a goodness-of-fit test. I find no basis for criticism of this terminology. Notice that I did not use the term chi-square dispersion test as that term was not used in the statistical texts consulted. After reviewing the index in several other statistical texts, my impression remains that the term chi-square dispersion test is not commonly used. Thus, if technologists choose to refer to the test statistic described in my article as a goodness-of-fit test, they need not feel like the Lone Ranger.

Small Sample Statistics

Ms. Gerson has questioned the appropriateness of using small sample statistics for calculating the chi-square statistic. She comments, "If the experimental work is worth doing then it is also worth using fully the information thus obtained...." Certainly I would not argue against rigor—that would be like arguing against motherhood and for sin. The use of small sample statistics is sometimes "better felt than telt." My purpose for pointing out the application of the range as an estimate of the standard deviation was not to discourage technologists from using more rigorous calculations, but rather to encourage technologists to use the chi-square statistic more often. I believe that Ms. Gerson's attitude on the universal availability of electronic calculators is somewhat provincial.

Adequate justification exists for the use of inefficient estimators of statistical parameters (*16–19*) and the interested reader should consult the references cited and draw his or her conclusions. A particularly good discussion is found in Dixon and Massey (*15*). In further review of the literature, I ran into another reference where small sample statistics were applied to the chi-square test (*20*), which had escaped my attention previously. It may be worthwhile to point out that a statistically demonstrated difference is not necessarily the same as a practical difference. Someone has said that "a difference is a difference only if it makes a difference." I have used this quick estimate for several years and have found it robust enough for my use.

The Kolmogorov-Smirnov Test

Ms. Gerson further comments that "when properly carried out the chi-square dispersion test is more powerful than the Kolmogorov–Smirnov test in most situations...." It is not clear from my previous comments about the essential equality of the chi-square test and the chi-square variance test and the comments published in several statistical texts that Ms. Gerson's comment is true.

Note: "The Kolmogorov–Smirnov test possesses several advantages over the simplest chi-square test of fit" (*21*); "In almost all cases the Kolmogorov–Smirnov test of goodness of fit is a more powerful test than the χ^2 test" (*22*); and "The Kolmogorov test may be preferred over the chi-square test for goodness of fit if the sample size is small, because the Kolmogorov test is exact even for small samples, while the chi-square test assumes that the number of observations is large enough so that the chi-square distribution provides a good approximation of the distribution of the test statistic" (*23*). I do not want to get into a discussion of the meaning of the term observation here. I would agree that, in my experience, the chi-square test is simpler and operationally more convenient to calculate but I would reserve judgment on whether it is the obvious method of choice.

Numerical Calculation of the Chi-Square Statistic

Ms. Gerson has presented a useful numerical formula for calculating the value of the chi-square statistic. The formula which she presents is derivable from an expansion of Eq. 6 (see Croxton for some of the mathematical details) (*25*) and can be written as

$$\chi^2 = \frac{\Sigma X_i^2}{\bar{X}} - \frac{(\Sigma X_i)^2}{\bar{X}n} \quad . \qquad (7)$$

Since $\bar{X} = \Sigma Xi/n$, Eq. 7 can be further reduced to

$$\chi^2 = \frac{\Sigma X_i^2}{\bar{X}} - \Sigma X_i \, . \qquad (8)$$

Equation 8 is often used in the calculation of chi-square values in certain classes of contingency tables but has not been detailed in any discussion of the use of the chi-square test in quality control of nuclear counting equipment which I have read.

The relevancy of Ms. Gerson's comment that "this method of calculating S^2 should not be used in circumstances where n is large and S is much smaller than $\bar{X}$..." is unclear. In the first place, the calculation pertains to a chi-square calculation, not S^2. Secondly, for a presumed Poisson distribution $S = (\bar{X})^{1/2}$, which is always much smaller than $\bar{X}$ for practical counting situations. Lastly, since we are dealing with mathematical identities, the chi-square value calculated from Eq. 8 is the same as the chi-square value calculated from Eq. 6, as the interested reader can easily verify.

Conclusions

There are other comments that could be made but this letter has already consumed more space than intended. In closing I would like to acknowledge the letter I received from Stephen L. Walaski of the Veterans Administration Hospital, Little Rock, AR. Mr. Walaski has pointed out that Eq. 8 in my original article is incorrect. It should be as follows:

$$\chi^2 = \frac{(K*R)^2 *(n-1)}{\bar{X}} \quad . \qquad (9)$$

The square sign was left off of the manuscript sent to the *JNMT* and this error was not picked up. The calculations remain correct.

Mr. Walaski further points out correctly that Tc-99m is not a suitable radionuclide for determining equipment counting reliability. It should be apparent that an analyst should not measure equipment reliability and radioactive decay at the same time.

Finally, Mr. Walaski suggests using a "closer" chi-square acceptance range. He uses a chi-square probability range of 0.2–0.8 for dual-probe detectors. I believe that such a close range of acceptance may unduly penalize equipment operating performance, but each laboratory must make its own choice.

JOHN R. PRINCE
Research Medical Center
Kansas City, MO

References

1. Prince JR: Goodness-of-fit tests for describing the statistical behavior of nuclear counting equipment. *J Nucl Med Technol* 5: 41–45, 1977

2. Hendee WR: *Medical Radiation Physics.* Chicago, Yearbook Medical Publishers, Inc., 1970, p 377

3. Evans, RD: Statistical tests for goodness-of-fit. In *The Atomic Nucleus.* New York, McGraw-Hill, 1955, pp 781–783

4. Snedecor GW, Cochran WG: *Statistical Methods,* 6th ed. Ames, IA, Iowa State University Press, 1967, p 237

5. Hoel P: Testing goodness of fit. In *Introduction to Mathematical Statistics,* 3rd ed. New York, John Wiley & Sons, 1962, p 244

6. Conover SJ: *Practical Nonparametric Statistics.* New York, John Wiley & Sons, 1962, p 244

7. Crow EL, Davis FA, Maxfield MW: *Statistical Manual.* New York, Dover, 1960, p 8

8. Ref. 7, p 70

9. Spiegel MR: *Schaum's Outline of Theory and Problems of Statistics.* New York, McGraw-Hill, 1961, p 190

10. Ref. 5, p 257

11. Price WJ: *Nuclear Radiation Detection,* 2nd ed. New York, McGraw-Hill, 1964, p 64

12. Dixon WD, Massey FJ: *Introduction to Statistical Analysis,* 3rd ed. New York, McGraw-Hill, 1969, p 243

13. Ref. 6, p 187

14. Ref. 4, p 232

15. Ref. 12, p 249

16. Ref. 12, p 127

17. Ref. 4, p 46

18. Simon J, Hicks E, Pavlovec R, et al: Quick estimation of standard deviation by use of a nomogram. *Clin Chem* 22: 1758–1859, 1976

19. Dean RB, Dixon WJ: Simplified statistics for small numbers of observations. *Anal Chem* 23: 636–638, 1951

20. Ref. 3, p 904

21. Blum JR, Rosenblatt JI: *Probability and Statistics.* Philadelphia, W. B. Saunders Co, 1972, p 414

22. Hays WL, Winkler RL: *Statistics: Probability, Inference and Decision, vol II.* New York, Holt, Rinehart and Winston, Inc., 1970, p 224

23. Ref. 6, p 295

24. Ref. 12, p 28

25. Croxton FE: *Elementary Statistics with Applications in Medicine and the Biological Sciences.* New York, Dover, 1959, p 279