Relevance of the chi-squared test when counting alpha emissions

by
D. Crouch[1] and S. Cooke[2]

Submitted:02/01/2024 Accepted: 30/02/2024

Abstract

This paper is intended for technicians and laboratory personnel who perform sample counting on long-lived airborne alpha dust. Hopefully, it will improve their understanding of the performance tests they use, the significance of the results, and how they contribute to confidence in the laboratory work. The paper's audience is radiation safety personnel who may not have had significant exposure to higher mathematics, but the author acknowledges that some maths understanding is necessary.

INTRODUCTION

We use the chi (pronounced kai, with Greek symbol χ) squared test as a statistical test of alpha counting equipment to determine if the variance in the sample is predominantly from random variance and not systematic variance.

Although the procedure is well-documented (Department of Mines and Petroleum, 2010), the rationale behind the χ^2 limits of 3.32 and 16.9 for a ten-count test is seldom elucidated. This raised some initial inquiries for the author:

- When described as "Goodness of fit," what exactly is it fitted to?
- Results must be between 3.32 and 16.9 – why do these values apply regardless of the source activity?; and
- How do factors like count time and the number of counts influence the outcome?

CHI-SQUARED TESTING

The Chi-squared test was discovered twice, once by Helmert in 1875 and again, more formally, in 1900 by Pearson (Brereton, 2015). The underlying premise concerning radioactive source counting is that if a dataset is comprised of independent data points, then the probability of Event A given that Event B has already occurred is the same as the probability of Event A multiplied by the Probability of Event B, mathematically stated as $P(A \cap B) = P(A)*P(B)$ (Greenwood & Nikulin, 1996).

A chi-squared test is used in this context to determine the goodness of fit of a dataset to an expected outcome. If the test determines that counting the radioactive source could generate the dataset from the hypothesis with a reasonable probability, the dataset is from that hypothesis. If it does not pass the test, the chi-squared test can only determine that the hypothesis is wrong, not what the correct hypothesis is.

Degrees of freedom

The following discussion regarding degrees of freedom is a brief overview in the context of radioactive counting statistics, single variable data and goodness of fit testing.

Degrees of freedom (DoF) represent the number of independent values in a dataset that can vary independent of the other values or statistical relationships within the dataset. In this case, a restriction is imposed on the data such as the fixed sample size and known dataset means. When performing a goodness-of-fit test of a series of counts where the population mean is known, the last data point is constrained as it must be a value that gives the population mean.

Calculation:

General formula: DoF = n - k, where:

n is the total number of observations in the dataset.
k is the number of constrained values.

1 Senior Health Physicist, Resources Health & Safety Services, CRSE, MARPS, MAIOH. d.crouch@resourcessafety.com.au
2 Principal Consultant, Unicorn Analytics. s.cooke@unicornanalytics.com.au

Example: A series of ten (10) alpha counting results [9, 11, 8, 5, 11, 11, 12, 10, 10, x] have a mean, or expected value, of 10.1. The last value of the series x must be 14, if the prior hypothesis that the mean is 10.1 is true, thereby making it the last value a constrained (conditional on the hypothesis being true) value, and the series has 9 degrees of freedom.

Expected value of a Poisson distribution

The expected value of a Poisson distribution, denoted by λ (lambda), is both its mean and variance (variance being standard deviation σ squared). It represents the average number of events expected to occur within a specified interval of time given that the events happen independently and at a consistent expected value. In a Poisson distribution, the expected value, mean, and variance are all equal to λ. This property is unique to the Poisson distribution.

Chi-squared

The chi-squared test examines the difference of the observations from the expected value. It does so by calculating each observation minus the expected value, squaring the result, and then dividing by the expected value (Greenwood & Nikulin, 1996). The difference is divided by the expected value to scale it to a proportion of the expected value, i.e., a difference of 10 when the expected value is 20 is more significant than a difference of 10 when the expected value is 1000. This is repeated for each observation, and all the results are summed together. The procedure can be ambiguous when written in English, which is why we use formal mathematic notation:

$$\chi^2 = \sum \frac{(observation - expected)^2}{expected} \quad \text{Equation 1}$$

Things to note:

- Squaring the difference ensures the differences are always positive and don't cancel out,
- Dividing by the expected value scales the difference to make them comparable, also called the coefficient of variance.

Confidence

When considering if a dataset is made of random data points, we must define how far from the expected value is considered acceptable. For instance, all the data being at or around the expected value is relatively likely, whilst values being at extremes from the expected value are highly unlikely.

A good example is rolling three fair 6-sided dice. There are 216 potential outcomes, but when the dice are totalled, the results form a probability curve, as shown in Figure 1.

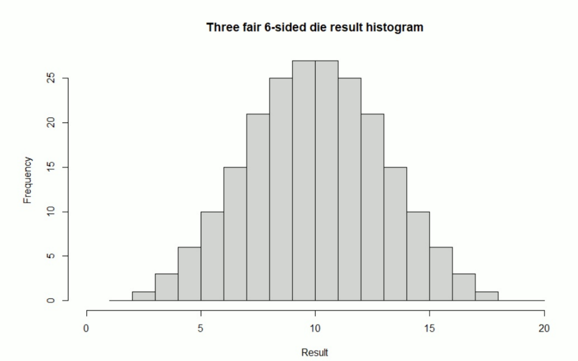


Figure 1. *Frequency of 3 fair 6-sided die results*

From Figure 1, the most likely outcome on each die, around 3.5 times three, gives the most likely outcome of all die rolled and summed. The only way to get a result of 3 is if all three dies must be 1. Likewise, to get a result of 18, all three dies must be 6. Conversely, 27 separate permutations give a combined result of 10.

We replace the discrete dice values (1:6) with a continuous variable along a standard normal distribution in a chi-squared test.

Conventionally, we define anything with less than a 5% probability of occurring as evidence that the hypothesis should be rejected. Having defined a confidence limit of 5%, if the result falls between the 5th percentile and 95th percentile of possible values, the dataset is considered to contain independent and normally distributed elements. Any result with less than 5% probability, less than 5th percentile and more than 95th persentile, of occuring is said to fail the chi squared test.

Using the same data from example 1 and the Microsoft Excel ChiSq Test function, it returns a probability of 0.81 of seeing these results given the expected outcome above the 5% confidence threshold. Changing the expected value from 10.1 to 15, the chi squared value becomes 19.5 and has a probablility of the dataset occuring of 2%. This does not pass the chi squared test and the hypothesis that the dataset came from a distribution with an expected value of 15 must be rejected.

General usage

The Chi-squared test assumes data varies from expected values as a standard normal distribution (equation 2). If the assumptions are that each sample is:

a) random and independent and
b) taken from the same normal distribution,

the further the observations deviate from the expected value, the less likely that either assumption *a* or *b* is true.

$$PDF(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{Equation 2}$$

PDF = Probability Density Function

μ = mean and

σ = standard deviation.

At a base level, the chi-squared test assumes the hypothesis (H_0) that the dataset will vary from the expected values via a normal distribution around the expected value.

How the chi-squared test results indicate goodness-of-fit

To better understand the chi-squared test results, we can perform the test with different expected values on the same dataset and how more than one hypothesis may pass the chi-squared test if the dataset has high variance.

For this example, our dataset is a series of numbers drawn from a random normal distribution around a specific mean with a specific standard distribution, both indicated. Our hypothesis will change such that the expected value of the mean takes on some integer value. In the case where the expected value is 7, it is noted as H_0=7.

Tables 6, 7 and 8 show several fabricated datasets with known means and standard deviations and the chi-squared test based on different hypotheses for the expected value. The Chi-squared 5% confidence limits for 9 degrees of freedom are 3.32 and 16.9 respectively, so any chi-squared value outside this range leads us to reject the hypothesis.

Table 6 shows the chi-squared test results of a dataset with a mean of 10 and a standard deviation of 1. The chi-squared results for each hypothesis (expected mean value) have been calculated and show that the most likely mean = 10, and a different mean is less likely, the further it is from 10 i.e., a hypothesis with an expected mean of 12 could give the dataset, but it is less likely than expected mean = 10.

The shaded chi-squared test results are those with more than 5% likelihood and show the distribution of hypotheses that would pass the chi-squared test. At the hypothesised expected means of 6 and 15, there is insufficient confidence that the actual mean was below 6 or above 15.

Table 7 shows the same dataset as Table 6 but with an offset of +2. Adding a number (offset) to a normal distribution moves that distribution and the hypothesised distributions reflect that means with sufficient confidence have also moved +2.

Table 8 shows the chi-squared test and confidence for a much wider normal distribution to a range of hypothesised means. While the highest confidence is still around 10-11, the confidence is barely above the acceptable 5% (0.05) threshold. Intuitively, this makes sense as it is difficult to make any statement of a dataset with a high degree of variance.

Overall, this experiment demonstrates the sensitivity of the chi-squared test to changes in expected values and variance, highlighting the importance of careful hypothesis selection.

Table 6. *Ten normally distributed N(10,1) samples with different hypothesis chi tests and 5% confidence test.*

Mean = 10 SD = 1 Offset = 0	x_i	1	2	3	4	5	6	7	8	9	10
	Data	9.6	10.1	8.9	9.5	10.7	10.7	9.7	10.3	9.8	9.6
	H_O=6	H_O=7	H_O=8	H_O=9	H_O=10	H_O=11	H_O=12	H_O=13	H_O=14	H_O=15	H_O=16
ChiSq	25.53	**12.41**	**5.07**	**1.59**	**0.80**	**1.97**	**4.62**	**8.39**	**13.06**	18.43	24.39
P(ChiSq)	0.002	**0.191**	**0.828**	**0.996**	**1.000**	**0.992**	**0.866**	**0.495**	**0.160**	0.030	0.004

Table 7. *Ten normally distributed N(10,1) samples with +2 offset with different hypothesis chi tests and 5% confidence test.*

Mean = 10 SD = 1 Offset = 2	x_i	1	2	3	4	5	6	7	8	9	10
	Data	11.6	12.1	10.9	11.5	12.7	12.7	11.7	12.3	11.8	11.6
	H_O=6	H_O=7	H_O=8	H_O=9	H_O=10	H_O=11	H_O=12	H_O=13	H_O=14	H_O=15	H_O=16
ChiSq	58.30	34.57	19.27	**9.60**	**3.86**	**0.98**	**0.25**	**1.17**	**3.39**	**6.64**	**10.74**
P(ChiSq)	0.000	0.000	0.023	**0.384**	**0.920**	**0.999**	**1.000**	**0.999**	**0.947**	**0.675**	**0.294**

Table 8. *Ten normally distributed N(10,6) samples with different hypothesis chi tests and 5% confidence test.*

Mean = 10 SD = 6 Offset = 0	x_i	1	2	3	4	5	6	7	8	9	10
	Data	5.6	12.7	4.8	9.6	18.4	11.7	7.0	4.7	11.8	10.6
	H_O=6	H_O=7	H_O=8	H_O=9	H_O=10	H_O=11	H_O=12	H_O=13	H_O=14	H_O=15	H_O=16
ChiSq	50.44	34.18	24.47	19.15	16.89	16.86	18.50	21.43	25.37	30.12	35.52
P(ChiSq)	0.000	0.000	0.004	0.024	0.050	0.051	0.030	0.011	0.003	0.000	0.000

Additionally, the shaded results above 5% likelihood offer valuable insights into the hypotheses that could pass the chi-squared test, providing a nuanced understanding of the dataset's distribution.

RADIATION COUNTING

When applied to radiation counting systems, the chi-squared test provides the technician with confidence (or lack thereof) that the variability in the counting system is only from the random fluctuations associated with radioactive decay.

We've seen how an offset and the distribution of the dataset can affect the chi-squared test result. Utilising the statistical characteristics of radioactive decay, we can test if the counting fluctuations are within the required confidence that they could have been derived from a Poisson distribution.

Poisson distribution

In the context of radioactive decay counting, when we refer to "the count," we are talking about the number of observed radioactive decay events within a specific timeframe or in a particular sample. When this count exceeds 10, it implies that a sufficient number of events have been observed, allowing the statistical distribution to approximate a normal distribution. This is important for making certain types of statistical inferences or analyses.(Knoll, 2010).

In practice our expected value is the mean of a series of counts. In this case, each radiation count results would have the following Poisson distribution properties:

- Only positive integers/counts in each counting period,
- all positive integers/counts are possible in each counting period; however, the probability may become negligible at high counts,
- at expected values λ more than ~10, the

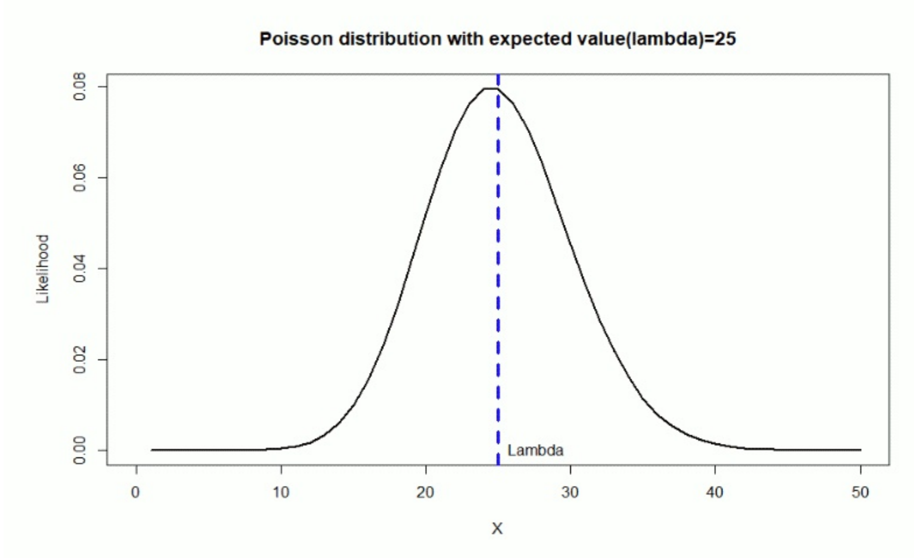


Figure 2. *Poisson Distribution approximating a normal distribution.*

expected value is approximately the mean,

- at expected values λ more than ~10, the variance (standard deviation squared) is approximately the expected value.
- 50% of the likely counts are below the expected value, and 50% are above.

The distribution of successive counts should have a standard deviation that is the square root of the expected value. Note the slight skewness compared to a normal distribution of the distribution in Figure 2 due to the underlying Poisson characteristics of the curve.

Another characteristic of the Poisson distribution is that as the expected value becomes bigger, the standard deviation becomes a smaller proportion of the expected value, i.e., as λ increases σ/λ decreases as can be seen in Table 9. This has the consequence that as expected values become smaller with low activity samples, the uncertainty becomes large.

Using the Poisson and normal distribution attributes

Using the attributes of the normal distribution inherent in radiation decay (i.e., when the expected value is greater than ~10), if each individual count in a series has the expected value (mean) subtracted, then divided by the standard distribution of the dataset, dataset becomes the standard normal dataset, with a mean (μ, sometimes called $\overline{x}$, x-bar) of 0 and variance of 1.

$$x_i' = \frac{x_i - \mu}{\sigma} \quad \text{Equation 3}$$

μ = mean and
x'_i = normalised value

We can now put any suitable source into our detector and perform a series of counts over an appropriate counting time. Using equation 3, we can reduce our counting dataset from a Poisson distribution centred on the expected value (or

Table 9. *Example Poisson distributions*

Poisson expected value (λ)	1	4	9	16	25	100	65536
Mean ($\mu = \lambda$)	1	4	9	16	25	100	65536
Standard deviation ($\sigma = \sqrt{\mu}$)	1	2	3	4	5	10	256
$\frac{\sigma}{\mu}$	1	0.5	0.33	0.25	0.2	0.1	0.0039

mean) with a standard deviation of the square root of the expected value, $N(\mu, \sqrt{\mu})$, to a normal distribution centred on 0 with a standard deviation of 1; $N(0,1)$.

Using the Poisson distribution attribute that the expected value (λ) is the same as the mean (μ) and the square of the mean is the standard deviation ($(\sqrt{\mu}=\sigma)$), the technician can make the following equivalence between the chi-squared test formula and results of source counting:

$$\chi^2 = \sum \frac{(observation - expected)^2}{expected} = \sum \frac{(x_i - \mu)^2}{\mu} = \sum \frac{(x_i - \mu)^2}{\sigma^2} = \sum \left(\frac{x_i - \mu}{\sigma}\right)^2 = \sum_1^k [N(0,1)]^2$$

Equation 4

Chi-squared testing of source counting

Using the demonstrated equivalence between the Poisson distribution of a series of radioactive source counts and the chi-squared distribution, the technician can make the following hypothesis:

> If n successive counts (convention says n=10) of a radioactive source return a chi-squared result between the 5% and 95% limits (3.32 and 16.9 respectively) of a chi-squared distribution of κ degrees of freedom, the source count variance is predominantly from random fluctuations of the isotope decay and only limited systematic variation comes from the counting system.

The chi-squared test of successive radioactive decay counts is also a goodness-of-fit test. As seen from the previous discussion, the goodness-of-fit is not goodness to some arbitrary 'fit', but specifically a **goodness-of-fit to a Poisson distribution with an expected value equal to the mean of the successive decay counts**.

REJECTING THE 'ONLY RADIOACTIVE DECAY VARIANCE' HYPOTHESIS

The Poisson distribution of radioactive decay provides only one variable that can vary, the expected value. This variable determines the mean AND variance of the distribution of results when counting. We reject the *only radioactive decay variance* hypothesis for one of two reasons.

Successive counting exhibits more variance than predicted by the chi-squared distribution

A common fault with radiation detection systems is the failure of the light-tight requirement. Detectors that incorporate scintillators or passivated implanted planar silicon (PIPS) generate a high noise level when the detector or housing allows visible light to reach the detector.

As shown in Table 7 the chi-squared test result can be influenced by adding an offset of c to Poisson distributed data around an expected value lambda, or if the standard deviation of the dataset around the expected value increases. The hypothesis that the data is from a distribution around its true expected value is less likely to be accepted.

Rejecting the hypothesis that all the variance is from random decay implies the counting system has some systematic signal noise that the technician must remove before using the equipment.

Successive counting exhibits more variance than predicted by the chi-squared distribution

A true Poisson distribution defines the maximum and minimum permissible variance. In cases where the dataset has less than μ variance (remembering the variance $\sigma^2=\mu$) , the hypothesis that the variance is from the random decay events is rejected as the chi-squared result is below the cut-off.

In practice, very few scenarios would result in a chi-square test rejecting the hypothesis due to the low variance. This scenario requires the mean of the dataset to be significantly higher than the variance μ {σ} ^ {2 which means the distribution is not a Poison distribution .

CHI-SQUARED BY BRUTE-FORCE

A technician can look up tolerance limits for chi-squared tests of k degrees of freedom from various sources. Alternatively, someone proficient in maths can integrate the Poisson distribution equation between A and B values, such that the accumulated probability densities between A and lambda and lambda and B equals 0.90. Refer to Appendix A for 5% and 95% tolerance limits for chi-squares distributions of 1 to 10 degrees of freedom.

It is also possible to create computer code to generate the chi-squared distributions as histograms of a very large number of trials ($>10^5$)with the following procedure:

- sampling k random numbers from a standard normal distribution,
- square all values in the dataset,
- sum all squared values in the dataset.

The result is a random chi-squared test with k-degrees of freedom. The program will repeat this process many times, and a histogram created of the result of each trial. The 5^{th} and 95^{th} percentiles are determined from the collection of 1 million results. Appendix B has several histograms for various degrees of freedom chi-squared tests. These histograms ran 1 million trials each, and, as can be seen from the smoothness of the histogram curve, they are relatively good approximations of calculus-derived values.

CLOSING REMARKS

The author hopes this constructive method of determining goodness-of-fit and the chi-squared test is intuitive without requiring higher-level maths and helps technicians identify and problem-solve issues when a chi-squared test rejects the goodness-of-fit of detection equipment.

REFERENCES

Brereton, R. G. (2015). The chi-squared and multinormal distributions. Journal of Chemometrics, 29(1), 9-12.

Department of Mines and Petroleum. (2010). Managing naturally occurring radioactive material (NORM) in mining and mineral processing - guideline. NORM-3.4 Monitoring NORM - airborne radioactivity sampling.

Greenwood, P. E., & Nikulin, M. S. (1996). A guide to chi-squared testing (Vol. 280). John Wiley & Sons.

Knoll, G. F. (2010). Radiation detection and measurement. John Wiley & Sons.

Microsoft. (2023). CHISQ.TEST function. Retrieved 20/07/2023 from https://support.microsoft.com/en-au/office/chisq-test-function-2e8a7861-b14a-4985-aa93-fb88de3f260f.

National Institute of Standards and Technology. (n.d). Critical Values of the Chi-Square Distribution. Retrieved 20/07/2023 from https://www.itl.nist.gov/div898/handbook/eda/section3/eda3674.htm.

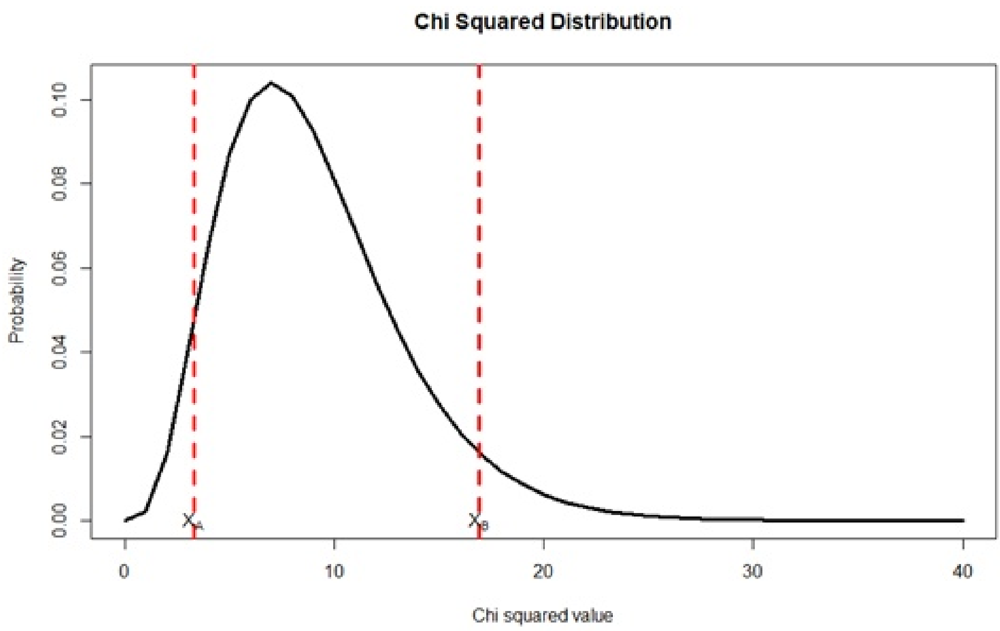
APPENDIX A CHI-SQUARED PROBABILITIES

	X_A	X_B
df	**0.95**	**0.05**
1	0.004	3.841
2	0.103	5.991
3	0.352	7.815
4	0.711	9.488
5	1.145	11.07
6	1.635	12.592
7	2.167	14.067
8	2.733	15.507
9	3.325	16.919
10	3.94	18.307

(National Institute of Standards and Technology, n.d)

X_A is the value where 95% of the area under the curve is to the **right**.

X_B is the value where 95% of the area under the curve is to the **left**.



APPENDIX B: BRUTE FORCED CHI-SQUARED DISTRIBUTIONS WITH 5% AND 95% LIMITS.