Elementary Statistics Using the Geiger Counter

Carolyne Pickler, Neil Edelman

April 13, 2007

Abstract

The Geiger counter ran for several intervals creating various mean distributions. The chi-squared of each of these values was taken and using this information, it was determined whether a Poisson or Gaussian standard form best fit the distribution. Upon analysis of the χ^2 values, the Poisson form was the best fit for the distribution created by the Geiger counter till a large mean distribution, one of about 20, was reached.

1 Introduction

When a radioactive source is brought into contact with the Geiger counter, a random distribution occurs. By calculating the χ^2 value for each measured point, we were able to attempt to fit the distribution with a Gaussian or Poisson form. The period of measurement was changed to allow us to examine various mean distributions and to see how these changes affected which form was the best fit. Our goal is to determine which standard form best fits the random radioactive decay.

2 Theory

To be able to determine which standard form, Gaussian or Poisson, best fits the data, we took the mean of each replica and column. The mean was calculated using equation 1.

$$\bar{x} = \frac{1}{n}(\sum_{i=1}^{n} x_i) \tag{1}$$

In equation 1, x_i represents each value and n is the total amount of values. The variance of each replica and column was also calculated with equation 2.

$$\sigma^2 = \frac{1}{n}(\sum_{x=1}^{n} x_i - \bar{x})^2 \tag{2}$$

In this equation we sum up the square of all values, x_i, and subtract the square of the mean then divide by the total number of values. These values, the mean and the variance, allow us to calculate the Gaussian and Poisson. To calculate the Poisson, we used the following equation.

$$P(v) = \frac{u^v}{v!}(e^{-u}) \tag{3}$$

In equation 3, the v is the bin number and u can represent the mean or variance because in this case they should be equivalent. When we calculated Poisson, the replica mean was used for u. To calculate the Gaussian, equation 4 was used.

$$P(v) = \frac{1}{\sigma\sqrt{2\pi}}(e^{-\frac{1}{2}}(\frac{v-u}{\sigma})^2) \tag{4}$$

In equation 4, u represents the replica mean, σ is the square root of the variance, which is also referred to as the standard deviation, and v is the bin number. To determine which is a better fit, the χ^2 test was used.

$$\chi^2 = \frac{1}{n}(\sum_{i=0}^{n} \frac{(O_i - E_i)^2}{\sigma_i^2}) \tag{5}$$

In equation 5, O_i represents the observed values, which means the measured values from the Geiger counter and σ_i^2 are the replica variance values. E_i represents the expected values, which is where we compare the measured values to the expected Gaussian or Poisson. For comparison to the Poisson, E_i would be the expected Poisson value determined from equation 3. The same is true when comparing to the Gaussian except we determine the E_i value using equation 4. Equation 4 is how to calculate the Gaussian but we must take into consideration the values that are negative. This is done using the error function seen in equation 6.

$$\mathrm{erf}(y) = \frac{2}{\sqrt{\pi}}(\int_0^y e^{-x^2}dx) \tag{6}$$

With some manipulations, equation 4 is put into equation 6. We plug in the exponent of equation 4 in for $-x^2$ and then take its integral. This allows us to make sure we take into account all points when calculating the Gaussian. This is extremely important when looking at low mean distributions.

3 Experimental Methods

A small ^{137}Cs gamma source was placed near the Geiger counter. To get various mean distributions, we varied the period over which we measured. For larger periods, we got large mean distributions and for smaller ones, we got smaller means.

4 Results

We calculated the Gaussian and Poisson form for three main distributions, low (2), medium (10), and large (22). To able to determine which form was the best fit, the data was compressed. This compression is crucial because it allows us to improve the quality of our data by eliminating some of the noise. The importance of this compression can be seen in the following three figures.

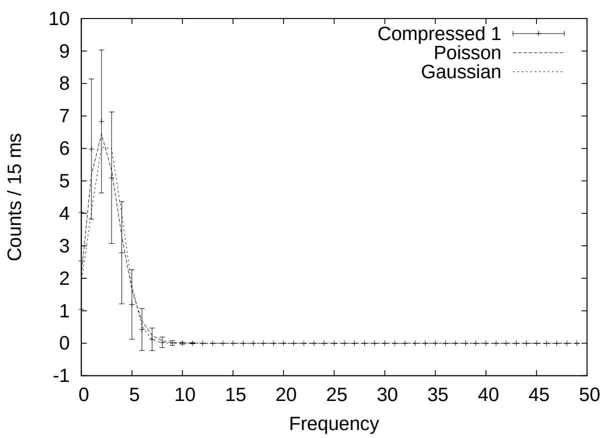

Figure 1: The Poisson and Gaussian Fits for the Low Mean Distribution with a Compression Ratio of One

As seen in this figure, it is difficult to distinguish which form is the best fit when we only have a compression ratio of one. This can also be seen when examining the failure rates, which are 8 percent for the Gaussian and 7 percent for the Poisson.

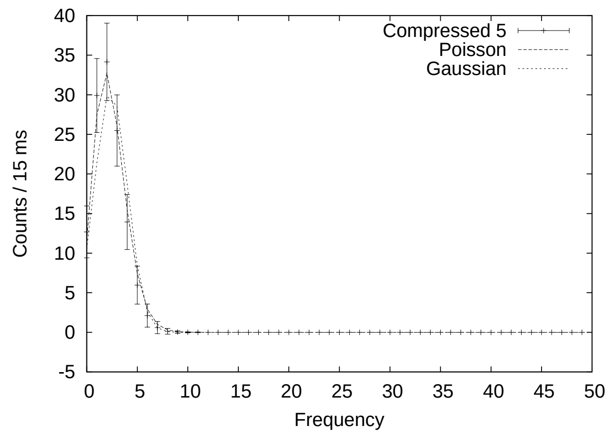

Figure 2: The Poisson and Gaussian Fits for the Low Mean Distribution with a Compression Ratio of Five

In Figure 2, we can see that as we compress our data by a ratio of 5, it is easier to distinguish which form best fits. In this case for the low mean distribution, the Gaussian has 22 percent of failures and the Poisson has 8 percent of failures.

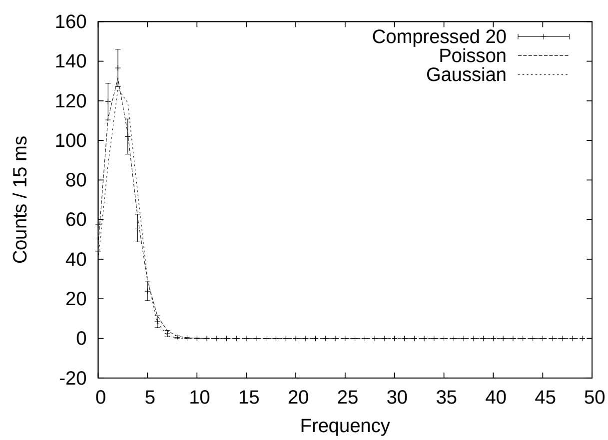

Figure 3: The Poisson and Gaussian Fits for the Low Mean Distribution with a Compression Ratio of Twenty

With a compression ratio of 20, we can really distinguish the Poisson from the Gaussian, as seen above. We can see that the Gaussian has a large failure rate of 86 percent while the Poisson only has a failure rate of 13%. To determine which form was the best fit for the low mean distribution, we plotted the following figure.

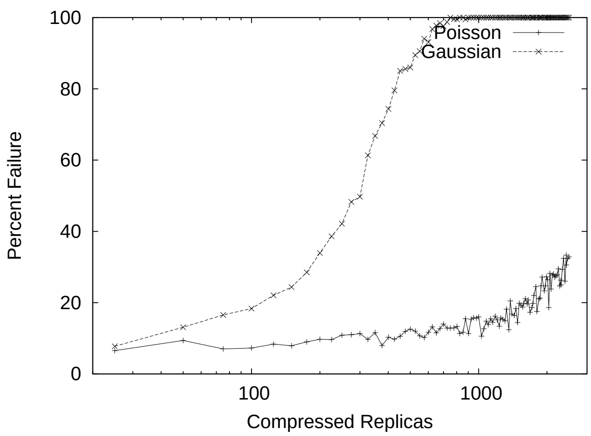

Figure 4: The Gaussian and Poisson Fits for the Low Mean Distribution

When comparing the χ^2 values of this case, which are illustrated in Figure 4, we get 22 percent of failures for the Gaussian and 8 percent of failures for the Poisson. This is with a compression ratio of 5. Compressing the data is important because it allows us to improve the quality of our data by eliminating some of the noise. The failures of the χ^2 test indicate the amount of times the calculated χ^2 value for the Poisson and Gaussian are greater than the expected χ^2 value at a 10% failure rate. This means that for the Gaussian or Poisson to pass the test, which means that they could represent the distribution, it has to have only 10% or less of its χ^2 values greater than the expected χ^2 value for that particular form. These expected χ^2 values are dependent on the degrees of freedom. For the medium mean distribution,

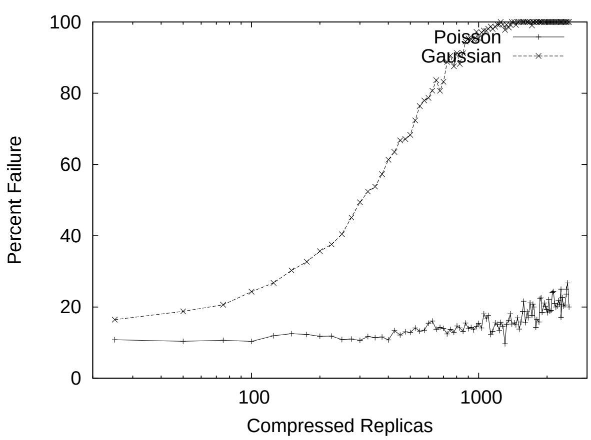

Figure 5: The Gaussian and Poisson Fits for the Medium Mean Distributions

When comparing the χ^2 values of this case, illustrated in Figure 5, we get 16 percent of failures for the Gaussian and 10 percent of failures for the Poisson. This is for a compression ratio of 5. For the high mean distribution,

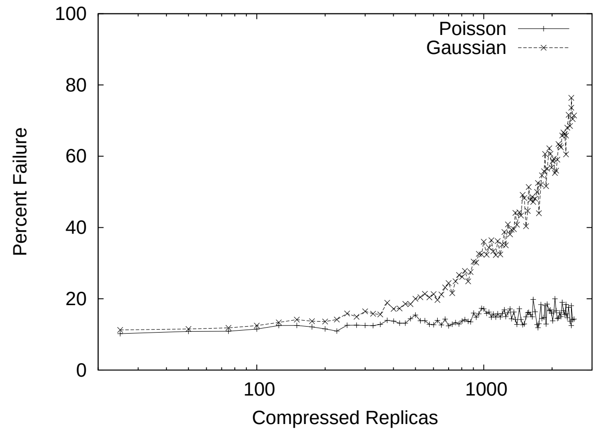

Figure 6: The Gaussian and Poisson Fits for the High Mean Distributions

Upon examination of this case's (Figure 6) χ^2 values, we find that the Gaussian and Poisson have 13 percent of failures. Poisson has a neccecary condition that the column variance is equal to its mean, which we checked using the following figures.

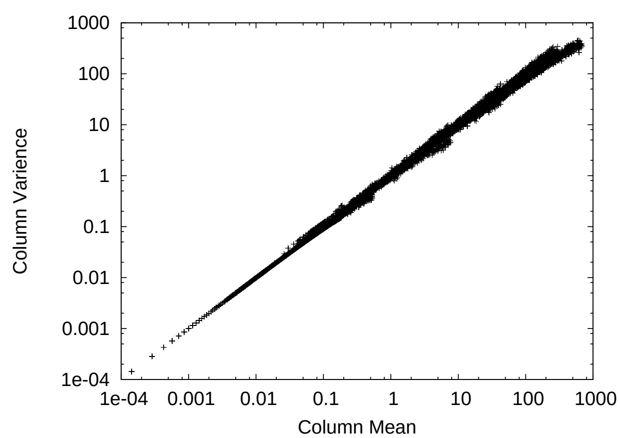

Figure 7: The Logarithm of the Column Variance as a Function for the Column Mean for Low Values

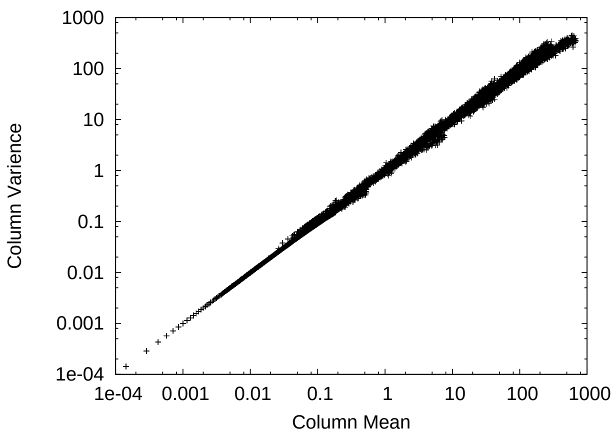

Figure 8: The Logarithm of the Column Variance as a Function for the Column Mean for Mid Values

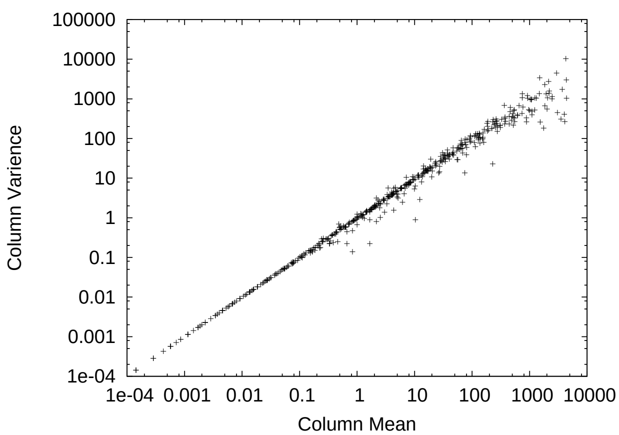

Figure 9: The Logarithm of the Column Variance as a Function for the Column Mean for High Values

5 Discussion

As seen in Figure 1 through 3, compressing our data allowed us to eliminate some noise and determine which form best fits the data. As Figure 1 illustrates, without compression, we wouldn't be able to determine whether the Gaussian or Poisson best fit the data. By examining Figure 4, with a compression ratio of 5, and the amount of failures of the χ^2 test for the Poisson and Gaussian forms, we can conclude that the Poisson form is the best fit for this low mean distribution. The same can be said for the medium mean distributions when examining Figure 5 and its failure rates. On the other hand, when examining the high mean distributions illustrated in Figure 6 and its failure rates, it is difficult to draw the same conclusions. This occurs because as explained in Introductory Mathematical Statistics for large mean distributions, the Gaussian and Poisson forms become very similar. In these cases, the Poisson is well approximated around the peak by a Gaussian [1]. This explains why at large means, it is difficult to determine whether the Gaussian or Poisson is the better fit. We begin to see the difficulty of using a Poisson form in Figure 9. This illustrates how the mean and variance are beginning to differ at high mean distributions. As seen in equation 3, this isn't the case for the Poisson form, where the mean and variance are always equivalent. We can really see that the Poisson form is a good fit by examining Figures 7 and 8. These figures shows a log-linear relationship between the column mean and variance. This means that the column mean and variance are equivalent. This is true for the Poisson form, as seen in Equation 3.

6 Conclusions

These results lead us to the conclusion that compression is necessary to eliminate noise and improve the quality of our data. It can also be concluded that for low to medium mean distributions, a Poisson form is the best fit. The same cannot be said when examining large mean distributions. For these cases, one cannot determine whether the Gaussian or Poisson is the better fit.

7 References

1 - E.Kreyszig. "Introductory Mathematical Statistics: Principles and Methods" John Wiley and Sons, Inc., (1970), pg. 97.