Analysis of Normality in the Difference of Two Poisson Random Variables

Parag Magunia

130 Filly Dr
N Wales, PA 19454
pmagunia@sharpermath.com

ABSTRACT

To test H_0: $\lambda_1 = \lambda_2$, using $\hat{\lambda}_1$ and $\hat{\lambda}_2$, which are sample means from two Poisson random variables, the normal approximation test statistic, z, will yield a p-value influenced by values of $\hat{\lambda}_1$, $\hat{\lambda}_2$, and sample size n. To understand the nature of this relation by means of an empirical study, computer simulations controlling for λ and n were used to build datasets while also noting the detected non-normality. Analysis supports the conclusion that detected normality in samples generated from Poisson random variables actually tends to decreases with an increase in sample size but detected normality in samples tends to increase with larger lambda values. Also, the difference between $\hat{\lambda}_1$ and $\hat{\lambda}_2$, even taking sample size into consideration, is not correlated with the outcome of a null hypothesis testing the equality of lambda parameters of two Poisson random variables. Keywords: Shapiro-Wilk, normality, hypothesis test

METHODS AND MATERIALS

Possible values for λ_1 and λ_2 were taken from the set {1,3,5,10,15,20} and only including those values which lead to nonzero differences between λ_1 and λ_2. Please note that sometimes different pairs of lambda values lead to identical differences.

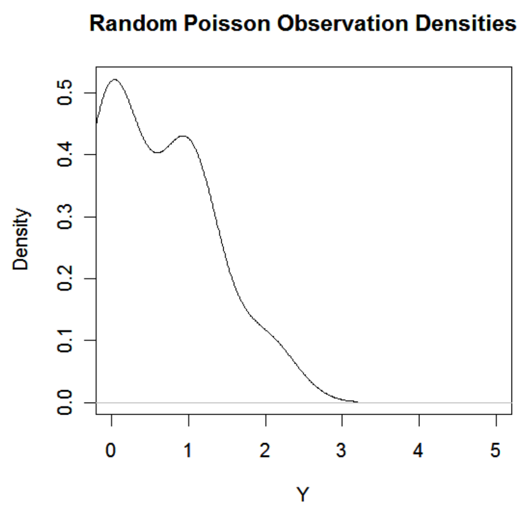
For each pair of λ_1 and λ_2, samples were generated using values of n = 10, 15, 20, and 25 from the two separate Poisson distributions, X_1 and X_2, with parameters λ_1 and λ_2, respectively. Each sample from X_1 was paired with a sample from X_2 of identical sample size. A hypothesis test was then performed to test the eqaulity of λ_1 and λ_2. This test employed the above mentioned z statistic. Significance levels were α = 0.01, 0.0001, and 0.000001. Detected nonnormality was also recorded as the averaged p-value of the Shapiro-Wilk test for each of the paired samples. The outcome of the null hypothesis was tallied along with the detected nonnormality and recorded in a table under the variable names *pSW* and *propRej*.

For example, if $X_1 \sim Poisson(\lambda_1 = 1)$ and $X_2 \sim Poisson(\lambda_2 = 3)$, two hundred samples would be generated from a distribution with p. d. f. $f(x_1)$ and two hundred samples would also be generated with p. d. f. $f(x_2)$, each of sample size ten. The outcome of the test with null hypothesis that $\lambda_1 = \lambda_2$ would be tallied with the detected nonnormality. Once two hundred tests had been performed and the variables had been tallied, averaged, and recorded, the algorithm would move on to the next sample size of n = 15. Then after the samples sizes had been exhausted, this process would be repeated for the next pair of λ_1 and λ_2.

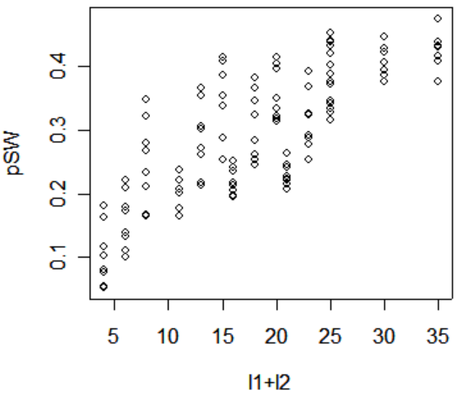
ANALYSIS AND CONCLUSION

One relationship that exists in all three of the datasets is that as the sample size increases for each pair of λ_1 and λ_2, the W-statistic, representing the average p-value of the Shapiro-Wilk test, always decreases. In other words, detected nonnormality increases with an increase in sample size. This relationship appears to be counterintuitive and inverse when taking the Central Limit Theorem into consideration. However, this disparity can be attributed to the increased sensitivity to nonnormality of the Shapiro-Wilk test as sample size is increased.

Noting that the sum of two Poisson distributions, X_1 and X_2, is distributed $Y = X_1 + X_2$ and keeping in mind that such a Poisson distribution, Y, with a small lambda parameter is positively skewed with a small variance, it would not be suprising to see all the data massed at but two points which would have a S-W p-value well below the 0.01 significance level implying departure from normality as can be seen in the following plot of an empirical density of random Poisson sample with λ=0.50.



This is evident in the data-pairs with small λ_1 and λ_2. This data tended to have proportionately small p-values from the Shapiro-Wilk test. Since λ_1 and λ_2 are small, this would imply that they have small variances as noted above. Small λ_1 and λ_2 parameter values mean that most of the mass in p. d. f. of Y would be near 0, an "asymptote" for the Poisson distribution. Since there is an asymptote, the data will be positively skewed near this point for small Y. The effect on the p-value can be seen in the plot below. When $\lambda_1 + \lambda_2$ are small, the p-value from the Shapiro-Wilk test is proportionately small.



In addition, a positive linear or at least monotonic increasing relationship is evident in the plot. As the value of $\lambda_1 + \lambda_2$ is increasing, the averaged p-value of the Shapro-Wilk tests tend to also increase. This increase would indicate smaller detection of nonnormality. Also, the Pearson correlation coefficient between $\lambda_1 + \lambda_2$ and the W-statistics with $\rho \cong 0.80$ in each of the three datasets, supports this conclusion about the positive linear relationship between these two measures. A correlation coefficient value 0.80 is well above 0.0 and indicates a strong linear relationship between $\lambda_1 + \lambda_2$ and the W-statistic.

Shifting to the analysis of the proportion of null hypotheses rejected, the proportion of rejected null hypothesis did repsond to changes in sample size. Given the fact that power and the ability of a hypothesis test to detect differences between means appreciates with *n*, increases in sample size for a paired λ_1 and λ_2 led to increases in the proportion of null hypothesis rejected. It can be seen that for every increase in sample size for a paired λ_1 and λ_2, the proportion of null hypotheses rejected also increased.

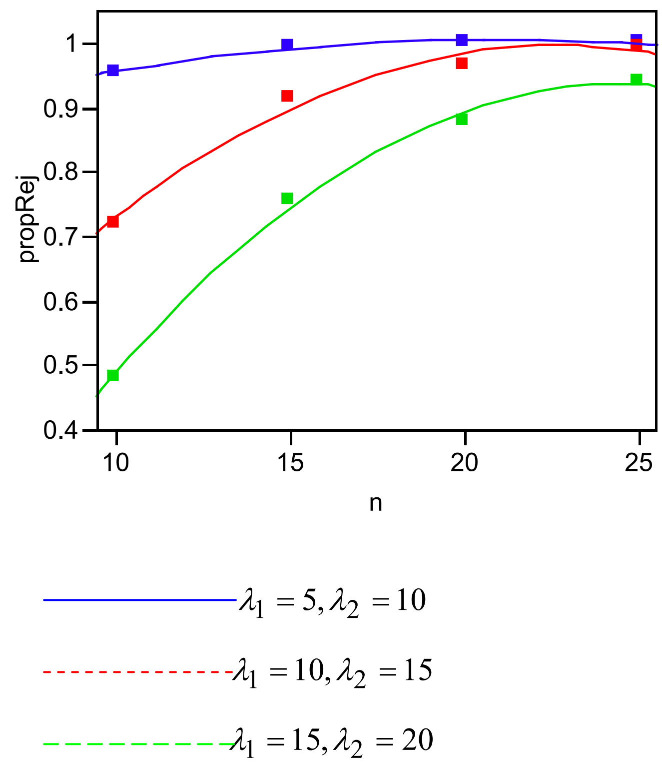
It would be thought that the difference between λ_1 and λ_2 would be an indicator to the proportion of null hypothesis rejected. However, taking the correlation coefficient between $\lambda_1 - \lambda_2$ and *propRej* leads to a value of 0.003 in the dataset created with the $\alpha = 0.01$ and values of -0.001 and 0.003 in the second and third datasets which seem to be inconsistant values for these variables since it is known that large differences between λ_1 and λ_2 lead to a greater proportion on null hypothesis rejected.

Pearson's ρ	$\lambda_1 - \lambda_2$
α	
0.01	0.003
0.0001	-0.001
0.000001	0.003

This discrepancy can be resolved considering the fact that the *absolute value* of the difference between $\lambda_1 - \lambda_2$ leads to a rejection of the null hypothesis in a two-sided alternative hypothesis test. The above mentioned values which are approximately zero in each case seem to be appropriate given the fact that the datasets were designed with λ_1

less than λ_2 and λ_1 greater than λ_2 an equal number of times. The correlation coefficient between $|\lambda_1 - \lambda_2|$ and *propRej* is more plausible with ρ = 0.458.

However, let us consider samples where the $|\lambda_1 - \lambda_2|$ are equal. In the plot below, these values are grouped by sample size and plotted against their respective *propRej* values with dashed lines differentiating paired λ_1 and λ_2 paramaeter values.



We can see that despite the fact $|\lambda_1 - \lambda_2| = 5$ in all three cases, *propRej* assumes three different values. The proportion of null hypothesis rejected with $\lambda_1 = 5$ and $\lambda_2 = 10$ is greater than either the other pair of lambda values. This is apparent not only when $n = 10$ but also the other sample sizes albeit to a lesser degree as sample size increases. This can be attributed to the Poisson distribution having a mean and variance dependent upon one another. It is true that $|\lambda_1 - \lambda_2| = 5$ in each instance, but the expression used to calculate the standard error for the z test statistic, $\sqrt{\frac{\lambda_1}{n} + \frac{\lambda_2}{n}}$, would be larger for the pair λ_1 = 10, λ_2 = 15 than λ_1 = 5, λ_2 = 10. A larger standard error would decrease the z test statistic thus making the null hypothesis less likely to be rejected.

In summary, the generated data support evidence that the Shapiro-Wilk test increases in sensitivity to nonnormality as sample size is increased, and that the two parameters, lambda and nonnormality, are positively correlated. Furthermore, it is $|\lambda_1 - \lambda_2|$ that must be considered when correlating the difference between means of a sample and the outcome of hypothesis test with a two-sided alternative using a Poisson distribution. Also, when testing equality of a mean parameter, in this case λ, where the mean and variance are not independent, identical differences between pairs of sample means will not necessarily imply identical outcomes of hypotheses tests.