**Normalizing the Frequency of Pendulum Period Observations**

Terence Henriod

TA: Marshall Liddle

Class: PHYS 180L, Section 1010

Thursday, September 20, 2012

**Abstract**

This experiment was designed to demonstrate that with a large enough sample, it is acceptable to approximate the distribution of a random variable, in this case varying observations or measurements, to that of the normal distribution.

The experiment was performed by setting up a simple pendulum and using a stopwatch to time the period of the pendulum for its initial swing, many times. Data was collected, then made to be discrete frequency data, plotted in a histogram, and the normal curve was super-imposed onto the graph to demonstrate how well the normal curve approximates the data that was observed.

The results of this experiment do reasonably support the theory that as the number of observations grows larger, it is increasingly acceptable to assume that the data has a normal distribution.

**Theory**

There were no theories of Physics, per se, involved in this exercise, but rather those of statistical approximation. The theory of interest was that of normal distribution approximation. It was assumed that the measurement of an experiment can be considered a random variable, a random variable whose distribution can be assumed to be approximately normal if enough observations are taken. It should be stressed that this is generally only acceptable if the number of observations (n or N) is large.

Implicit in discussing the normal distribution is defining the parameters of the normal distribution: the mean and the variance. The mean is a measure of central tendency that indicates a result that can be expected should the experiment be repeated, and the variance (or the standard deviation, which is the square root of the variance, and more commonly used) is a measure of the dispersion of the values of a sample or population.

The following are the equations of the mean, standard deviation, and normal probability density function:

Mean: Standard Deviation:

Normal Distribution:

The theory of performing calculations using Significant-Figures and calculating uncertainties was also discussed, but I will not go into depth on this topic as it was not the object of interest for this lab, but the equations for numerical uncertainty (percent difference) and measurement uncertainty are:

The experiment was conducted in the following manner: A simple pendulum was constructed using a workbench, clamp, wood block, string, and a metal bob, and time was measured with a stopwatch. The string was attached to the bob at one end, and the wood block at the other. The block was then clamped to the edge of the work bench.

The pendulum was brought to a starting position in line with the leg of the work bench (the point of reference for the start/finish of the pendulum’s period). The pendulum was released and the stopwatch was started simultaneously. Once the pendulum reached the height of its backswing (becoming even once again with the bench leg), the pendulum and watch were stopped, and the time measurement was recorded. This process was repeated until 40 periods were recorded.

The data was then sorted into time categories that span .05 s and this discrete data set was then plotted on a graph. Then the ideal normal curve that would match the mean and variance of this data set was superimposed on the graph to demonstrate how well (or not well) the data was approximated by a normal distribution.

**Data**

Table 1: Data and calculations for the experiment.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Trial # | Time X (s) | µ (s) | x-µ (s) | (x-µ)^2 (s^2) | σ (s) | n (Normal score approximation) |
| 1 | 1.62 | 1.67 | -0.05 | 0.00 | 0.05 | 15.64 |
| 2 | 1.66 | 1.67 | -0.01 | 0.00 | 0.05 | 22.80 |
| 3 | 1.57 | 1.67 | -0.10 | 0.01 | 0.05 | 4.45 |
| 4 | 1.57 | 1.67 | -0.10 | 0.01 | 0.05 | 4.45 |
| 5 | 1.69 | 1.67 | 0.02 | 0.00 | 0.05 | 20.97 |
| 6 | 1.69 | 1.67 | 0.02 | 0.00 | 0.05 | 20.97 |
| 7 | 1.66 | 1.67 | -0.01 | 0.00 | 0.05 | 22.80 |
| 8 | 1.63 | 1.67 | -0.04 | 0.00 | 0.05 | 18.11 |
| 9 | 1.65 | 1.67 | -0.02 | 0.00 | 0.05 | 21.87 |
| 10 | 1.69 | 1.67 | 0.02 | 0.00 | 0.05 | 20.97 |
| 11 | 1.69 | 1.67 | 0.02 | 0.00 | 0.05 | 20.97 |
| 12 | 1.60 | 1.67 | -0.07 | 0.00 | 0.05 | 10.50 |
| 13 | 1.59 | 1.67 | -0.08 | 0.01 | 0.05 | 8.17 |
| 14 | 1.65 | 1.67 | -0.02 | 0.00 | 0.05 | 21.87 |
| 15 | 1.81 | 1.67 | 0.14 | 0.02 | 0.05 | 0.65 |
| 16 | 1.66 | 1.67 | -0.01 | 0.00 | 0.05 | 22.80 |
| 17 | 1.69 | 1.67 | 0.02 | 0.00 | 0.05 | 20.97 |
| 18 | 1.66 | 1.67 | -0.01 | 0.00 | 0.05 | 22.80 |
| 19 | 1.69 | 1.67 | 0.02 | 0.00 | 0.05 | 20.97 |
| 20 | 1.65 | 1.67 | -0.02 | 0.00 | 0.05 | 21.87 |
| 21 | 1.65 | 1.67 | -0.02 | 0.00 | 0.05 | 21.87 |
| 22 | 1.69 | 1.67 | 0.02 | 0.00 | 0.05 | 20.97 |
| 23 | 1.69 | 1.67 | 0.02 | 0.00 | 0.05 | 20.97 |
| 24 | 1.72 | 1.67 | 0.05 | 0.00 | 0.05 | 14.08 |
| 25 | 1.69 | 1.67 | 0.02 | 0.00 | 0.05 | 20.97 |
| 26 | 1.72 | 1.67 | 0.05 | 0.00 | 0.05 | 14.08 |
| 27 | 1.69 | 1.67 | 0.02 | 0.00 | 0.05 | 20.97 |
| 28 | 1.78 | 1.67 | 0.11 | 0.01 | 0.05 | 2.47 |
| 29 | 1.72 | 1.67 | 0.05 | 0.00 | 0.05 | 14.08 |
| 30 | 1.72 | 1.67 | 0.05 | 0.00 | 0.05 | 14.08 |
| 31 | 1.60 | 1.67 | -0.07 | 0.00 | 0.05 | 10.50 |
| 32 | 1.68 | 1.67 | 0.01 | 0.00 | 0.05 | 22.33 |
| 33 | 1.69 | 1.67 | 0.02 | 0.00 | 0.05 | 20.97 |
| 34 | 1.69 | 1.67 | 0.02 | 0.00 | 0.05 | 20.97 |
| 35 | 1.62 | 1.67 | -0.05 | 0.00 | 0.05 | 15.64 |
| 36 | 1.59 | 1.67 | -0.08 | 0.01 | 0.05 | 8.17 |
| 37 | 1.56 | 1.67 | -0.11 | 0.01 | 0.05 | 3.11 |
| 38 | 1.69 | 1.67 | 0.02 | 0.00 | 0.05 | 20.97 |
| 39 | 1.66 | 1.67 | -0.01 | 0.00 | 0.05 | 22.80 |
| 40 | 1.71 | 1.67 | 0.04 | 0.00 | 0.05 | 16.65 |

Table 2: The data in discrete form.

|  |  |
| --- | --- |
| Bin Upper Limits (s) | Frequency of Observation |
| 0.00 | 0 |
| 1.50 | 0 |
| 1.55 | 0 |
| 1.60 | 5 |
| 1.65 | 5 |
| 1.70 | 23 |
| 1.75 | 5 |
| 1.80 | 1 |
| 1.85 | 1 |
| 1.90 | 0 |
| 1.95 | 0 |

Table 3: How the data would lie (numerically) if it truly followed a normal distribution. Also a display of the numerical error found by comparing the data with the ideal normal scores.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Ideal X scores (s) | µ (s) | x-µ (s) | (x-µ)^2 (s^2) | σ (s) | Ideal Normal Scores | Actual Scores | Uncertainty (%) |
| 1.425 | 1.667 | -0.242 | 0.058 | 0.054 | 0.001 |  |  |
| 1.475 | 1.667 | -0.192 | 0.037 | 0.054 | 0.038 |  |  |
| 1.525 | 1.667 | -0.142 | 0.020 | 0.054 | 0.690 |  |  |
| 1.575 | 1.667 | -0.092 | 0.008 | 0.054 | 5.300 | 5.00 | 5.65 |
| 1.625 | 1.667 | -0.042 | 0.002 | 0.054 | 16.983 | 5.00 | 70.56 |
| 1.675 | 1.667 | 0.008 | 0.000 | 0.054 | 22.723 | 23.00 | 1.22 |
| 1.725 | 1.667 | 0.058 | 0.003 | 0.054 | 12.693 | 5.00 | 60.61 |
| 1.775 | 1.667 | 0.108 | 0.012 | 0.054 | 2.960 | 1.00 | 66.22 |
| 1.825 | 1.667 | 0.158 | 0.025 | 0.054 | 0.288 | 1.00 | 246.89 |
| 1.875 | 1.667 | 0.208 | 0.043 | 0.054 | 0.012 |  |  |
| 1.925 | 1.667 | 0.258 | 0.067 | 0.054 | 0.000 |  |  |
|  |  |  |  |  | Average Uncertainty: | | 75.19 |

Figure 1: Graphical representation of the experiment results in discrete form. Note: The normal curves had to be added by hand due to the difficulty of combining chart types in Excel. The solid line represents a true normal curve, while the dashed one represents the skewed curve that actually represents the data.

**Computations**

I performed all computations using Microsoft Excel. I will display a few sample calculations to demonstrate the process (Note: it is customary to use symbols like x bar in place of mu and s in place of sigma when performing computations for sample data, but for simplicity in using this word processor, I did use the population data symbols for the computation of sample values):

Mean: = 1.67s

Difference of observation and mean, squared (ideally used to simplify calculation of standard deviation) = (x-µ)2 = (1.67s) – (1.67s))2 = 0 s

Standard Deviation: = = 0.05s

Normal Distribution: = = 23

= 5.65%

**Results**

As can be seen in Figure 1, the observations of this experiment do support the theory that a large enough sample for a random variable has an approximately normal distribution. It can also be reasonably inferred that had more observations been collected from this experiment that the distribution of the data might become even more normal.

The average numerical uncertainty for this experiment was rather high, but the numerical uncertainty might not be the best measure of error for this experiment. After all, the uncertainty can be ascertained by viewing the difference between the ideal normal curve and the curve that was created using the experiment data, which intuitively appears to be rather close. As I left the lab I was left with the impression that we were using the curve to show a rough approximation, not to precisely compare our results with accepted values.

When interpreting the results, it can be seen that the experiment data obtained is approximately normal, but has a heavier left tail than the right one. It might be said that this curve is skewed to the right, and had “bins” containing smaller ranges been chosen, I believe the rightward skew-ness would become even more apparent. It can also be seen that this data has a rather low rate of dispersion, when viewed at the scale of 0.05 s. Again, had smaller “bins” been chosen, this would likely increase the variance/standard deviation of the sample.

This experiment was suitable for achieving its aim: it allowed for a random variable, the experimenter’s accuracy in timing the period of a pendulum, to be observed many times in a short period of time, allowing for a simple demonstration of the theory that most random variables have an approximately normal distribution as the number of observations grows large. The only improvements that I would recommend would be to simply point out that not all random variables will have an approximately normal distribution and that there are more requirements to a variable being normally distributed than simply being random and having a large number of observations.