Signal and Noise

Robert L. Belleville (June 2015)

1 Introduction

For quite a long while now, my interest has been in modeling pendulum period running in common household conditions. In order to do this work I collect cycle-by-cycle period measurements from the pendulum under test along with environmental data and a timestamp. Then I use statistical and data analysis techniques to discover relationships between environment and observed period measurements.

Sampled period data does not consist of a stream of constant values. Samples vary from a few to many tens of microseconds or even hundreds of microseconds during earthquakes. In fact, this raw data looks a lot like noise.

With digital sampling there is always an uncertainty of one-half of the resolution of the sampler. In my case, timer resolution is 1 microsecond. So I have a built-in uncertainty of ±0.5 microseconds. This timer is constantly monitored by a GPS-based standard second. The GPS removes timer drift and temperature effects. Clearly observed variations are much greater than simple sampling error.

Variation falls into two groups: short and long term. Short term variation can change from one sample to the next and includes:

1. wind and storm vibration
2. people and vehicles nearby
3. earthquake
4. short term variation in drive

Long term variation can be seen only over a few minutes and includes:

1. temperature change
2. barometric pressure change
3. relative humidity (RH) change
4. aging effects
5. tidal variations in gravity
6. long term variation in drive
7. other factors as yet unknown

This paper dives into logged data to discover its characteristics and tries to determine the best statistical method to separate short and long term effects. I also describe trade-offs involved in averaging cycle-by-cycle data.

2 Experimental Setup

Clocklab [1][3] is my test set-up to explore the performance of various pendulums quickly and easily. A closed box contains the pendulum under test, drive components, environmental sensors, and GPS-based interval timer. Data is streamed on three serial lines to a Linux based portable PC, collected and written to a log file. The log file can be copied from the portable PC at any time, without stopping a test, for analysis on another machine via wireless network. Some tests have run for more than a year in this manner. The data collected includes:

1. pendulum maximum angle
2. drive pulse width
3. period of last cycle
4. GPS second calibration
5. case temperature
6. timer crystal temperature
7. barometer sensor temperature
8. humidity sensor temperature
9. timestamp at the laptop

A previous test [2] was stopped because the pendulum was so light that even modest levels of vibration caused the drive/sense rod to bounce into the sensors and coil a lot like a pinball machine. Last year I revised the gimbal to eliminate this problem and began test *P11d* which is running smoothly and is ongoing. P11d began December 25, 2014 UTC. The log file is over 450 megabytes with almost 6 million cycles collected over more than 100 days. In itself the size of sample data presents problems which must be solved by extracting intervals by date/time and averaging intervals.

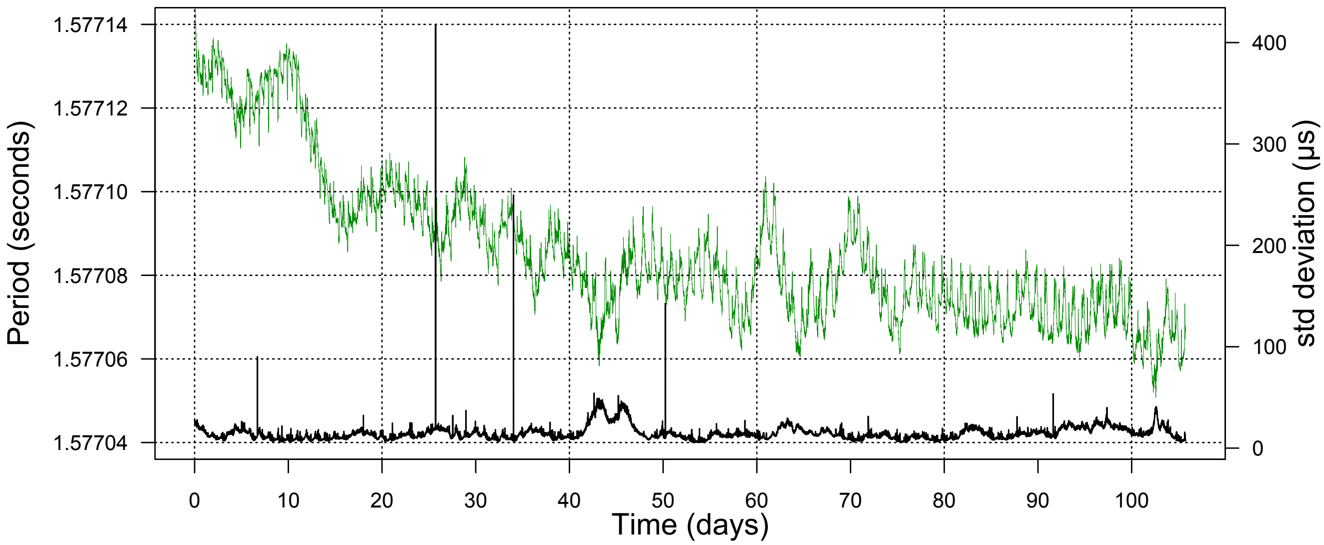

Figure 1: Period and Standard Deviation

Figure 1 shows an overview of 106 days of operation by averaging 10 minutes or 380 cycles. Over 15 thousand samples of 13 variables result. Here two variables are plotted: period in seconds on the left axis with the curve oscillating starting at the upper left and continuing downward to the right. At the bottom using the right axis, I show standard deviation (SD) in microseconds.

As logging started the period curve quickly dropped to about 1.577130 seconds and then continued to speedup to an average period of about 1.577070 seconds. The cause and form of this speed up is still to be discovered, but I generally see it in all my experiments. By day 80 speedup is much slower. Because it takes so much time for a test to begin to stabilize, experimental work is slow.

Most of the large variations (say between day 60 and 75) are due to barometric pressure changes. In fact this is the first of my experiments where normal atmospheric barometric pressure changes can be clearly seen. Since *tempco* is often used for temperature coefficient, *barco* might be used to represent barometric pressure coefficient. Here barco is about 1.3×10^{-6} seconds per mbar.

Turning now to the bottom standard deviation[16] curve, we first note some very large spikes at day 26, 34, and 50 for example. These are earthquakes. Day 50, was on Valentine's day 2015. This quake had magnitude 4.8 and was located about 250 miles east of the pendulum near Beatty NV. A timestamp with each data point is essential in relating spikes to quakes using USGS's database from their exceedingly useful website [7]. I will return to earthquakes later in this paper.

Much smaller 'fuzz' in this curve is household vibration. I know this because just after day 60 the fuzz goes away since we were on vacation and the house was unoccupied. So our movement in the house, opening and closing the garage doors, climbing the stairs, and so on can contribute period variation.

After day 40 there are two humps in the curve caused by a large February storm with gusts of wind and rain. Other smaller humps are from wind as well.

3 Stationarity

Figure 1 contains a lot of information. The period curve trends down but not linearly. Clearly taking the mean or average of all the data is meaningless. The problem is that period lacks *stationarity*. NIST has created an online handbook for engineering statistics [5] which says in section 6.4.4.2:

> Stationarity. A common assumption in many time series techniques is that the

data are stationary. A stationary process has the property that the mean, variance and autocorrelation structure do not change over time.

This is why it is better to look at clock period data in the frequency domain to try to detect cyclic patterns such as tides. But with the exception of tidal forces, none of the environmental factors are truly cyclical.

For analysis a decision will always have to be made concerning how long a sample to average to approximate a stationary process. Figure 1 takes sequential samples of 10 minutes each. Samples taken for an hour or more diminish *resolution*. There will always be a trade-off between noise and resolution. More on this later.

4 Noise Distribution

Statistics has two ideas: that of *population* and of *sample*. Population here is all the period data this pendulum can produce. A sample is some number of period values taken from the population, usually in sequence. The question is always: how well does a sample represent the population? It is impossible to know all the parameters of a population exactly. Statistics has evolved a vast number of mathematical models called *distributions* that attempt to model various real populations.

The poster child for distributions is the ever popular bell shape curve or normal distribution [13]. It tends to describe populations that are noisy and samples that have measurement error. If it turns out that samples of period data can reasonably be seen to have come from a normal distribution then we will be on firmer ground in computing averages and deviations.

I selected two samples of 1000 cycles. The *quiet* sample is near day 80 where SD is about 8 microseconds. The *stormy* sample is near day 43 at the height of February's storm. This storm has a SD of about 32 microseconds. Figure 2 shows the raw data.

Here raw data looks noisy with some out of pattern values such as a dip after cycle 800. Next step

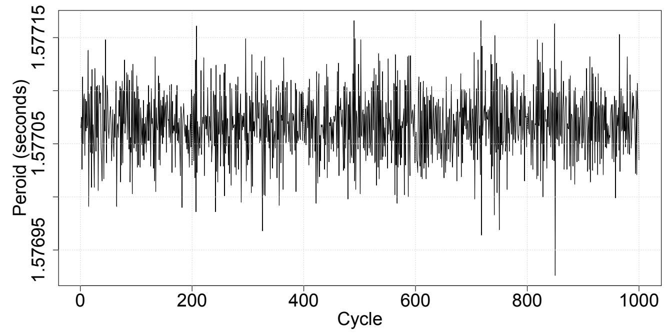

Figure 2: Raw cycle data for stormy sample.

is to make a box plot [9] of both stormy and quiet to see what we see.

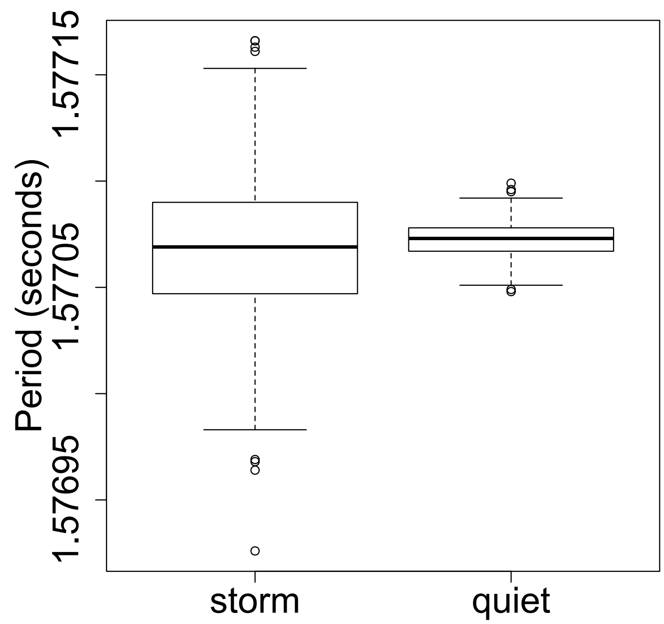

Figure 3: Box plot for stormy and quiet samples of 1000 cycles.

Figure 3 shows a lot. This is a Tukey[11][6] style plot [9]. Each line in the middle of a *box* is the median of the dataset. Most of the samples fall between top and bottom which are the 1st and 3rd quartile [15]. Lines at the end of dotted lines are *whiskers*. These are set in this plot to encompass most all of the data. Circles outside the whiskers are likely *outliers*. Outliers tend to bias averages away from the most useful mean value. Clearly stormy's box is bigger because greater noise spreads samples apart more than with quiet data.

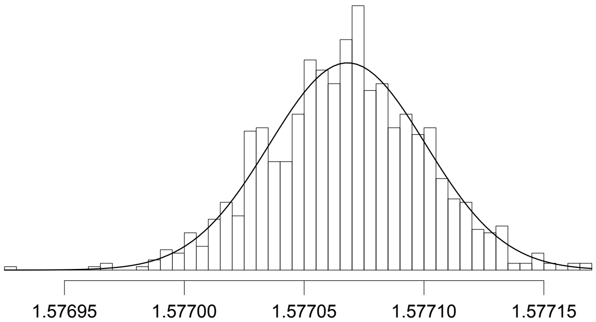

Figure 4: Period histogram with normal curve at a stormy time.

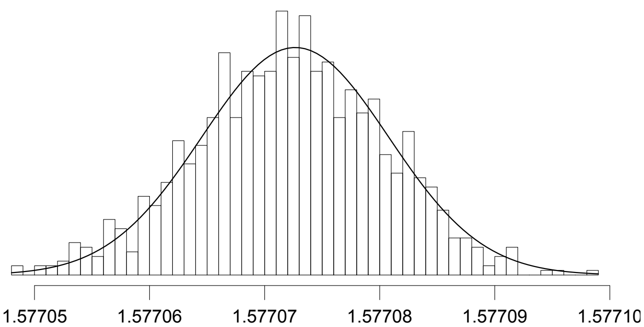

Figure 5: Period histogram with normal curve at a quiet time.

Next we look at histograms[10] of these two samples. A smooth curve shows a normal curve for the mean and standard deviation (SD) of the sample. This is best evidence yet that the population distribution is perhaps normal. Notice that the range of period on stormy is about 150 microseconds while the range on quiet is more like 30 microseconds. In neither case do the histogram bars perfectly match the curve. This is because the normal curve tells only the likelihood of a period value. Actual samples will vary a bit.

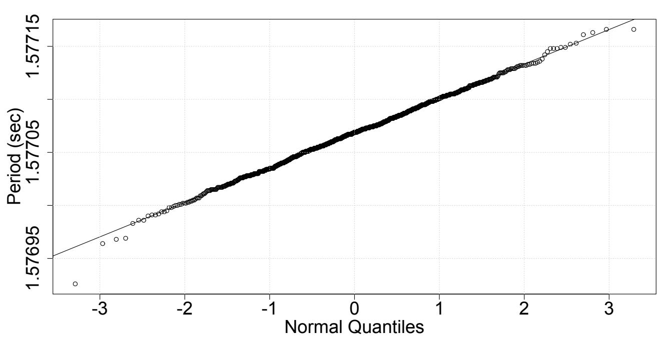

Figure 6: QQ-norm plot for stormy.

Greater confidence comes from making a Q-Q plot[14]. This compares sampled data to a normal curve. If data is just a straight line the data is perfectly normal. In our case, toward the ends of the curve, points begin to diverge from normal. These are outliers shown in figure 3.

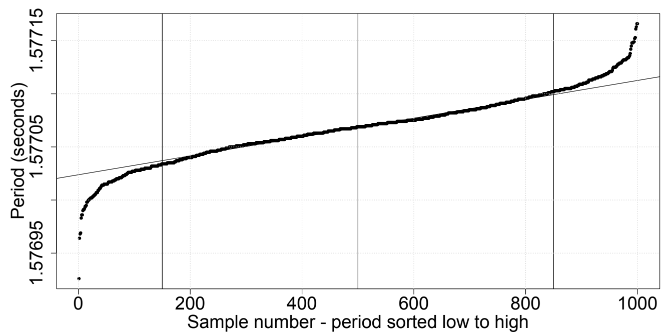

Figure 7: Sorted period data for stormy sample.

Finally figure 7 shows sorted period data. At the ends there is a good deal of divergence from a nearly straight curve of most of the data near the middle. In itself this isn't so bad but the extreme points can bias an average. Vertical lines show 15% of the sample and a center line shows the median[12].

At this point we can cautiously accept that short term variation is from something like the normal distribution and press on to discover the best way to compute an *average* in order to see long term variations.

5 Median, Mean, and Trimmed Mean

A *mean* is just the sum of values in a sample divided by the number of values. To get the *median* first you sort the sample from low to high and take the value in the middle. If the number of values is even then one can take the mean of the two center values. For a *trimmed mean,* a given percent of sorted sample values on the high and low end are simply ignored and a mean taken of the rest. [8][12][18]

Since a plain mean is subject to outliers, we shouldn't really use it to estimate an average of a sample of clock data. This leaves the median and trimmed mean.

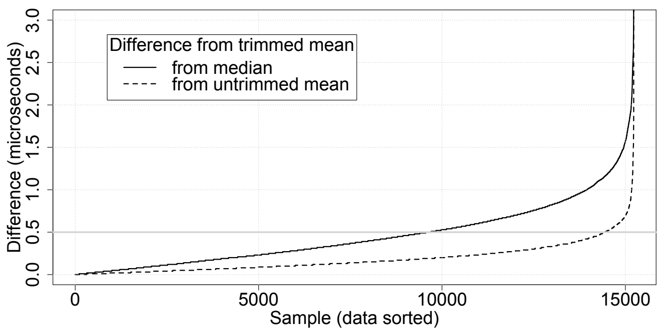

Figure 8: Difference between trimmed mean and either median or untrimmed mean.

Figure 8 shows the absolute values of the differences between a 15% trimmed mean and the median and untrimmed mean for all 15,229 10 minute samples (380 cycles) of the full dataset shown in Figure 1. For about 14,000 samples the difference for an untrimmed mean is less than 0.5 microseconds so my vote goes to using the trimmed mean because it removes outliers that show here as a quick rise at right. Note that this data contains everything from the best to the worst standard deviations recorded during earthquakes.

How much should we trim? About 92% of Figure 8 is below the 0.5 microsecond level so a trim of 4% should be fine or perhaps 5% to stay in round numbers. Quickly rerunning with a 5% trim shows that the trimmed mean stays under 0.25 microseconds for nearly the first 15,000 samples. We lose little from an untrimmed mean and remove outliers - a rare statistical win win.

6 Standard Error of the Mean

It is possible to quantify how much error there may be in a mean we have calculated from a sample. The equation is [17]

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

where s is the standard deviation of the sample and n is the number of items in the sample. Units here are microseconds because SD is in microseconds. SEM units will always be the same as the units of the underlying SD.

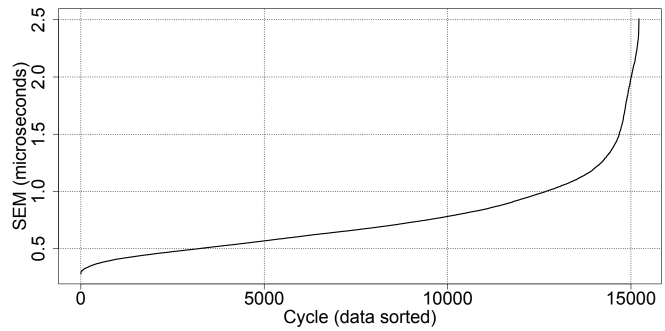

Figure 9: Standard error of the mean computed for all samples in figure 1.

In our 10 minute sample dataset there are 15,229 samples. Only 8 of these have a SD greater than 50 microseconds. These were trimmed from this plot so that a more useful scale could be shown. About 12,500 of all samples resulted in an error of less than ±1 microseconds.

	Sample size				
SD	200	500	1000	2000	5000
5	0.354	0.224	0.158	0.112	0.071
10	0.707	0.447	0.316	0.224	0.141
20	1.414	0.894	0.632	0.447	0.283
40	2.828	1.789	1.265	0.894	0.566
60	4.243	2.683	1.897	1.342	0.849

Table 1: Standard Error of the Mean (microseconds

Table 1 shows a few values of $SE_{\bar{x}}$ for different standard deviations and sample sizes. Clearly bigger samples are better (but these reduce resolution) than small ones (which show more noise). Smaller SD is better in all cases.

7 Resolution

Next we need to look at an interval of days from the dataset, at modest SD, to show graphically how resolution is reduced as sample size increases.

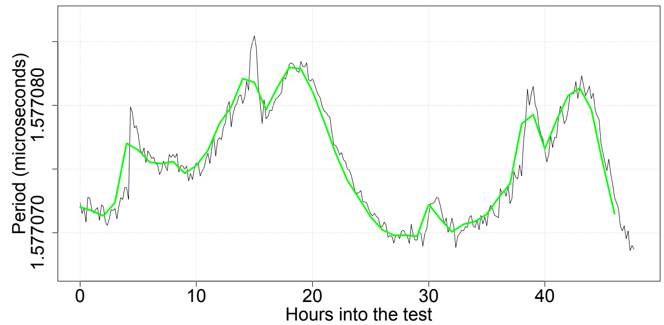

Figure 10: Two days with 10 minute and 60 minute samples.

In figure 10 a curve from 10 minute samples is shown with the same data but with 60 minute samples. The longer samples in light gray (green in the .pdf) are smoother. While 10 minutes takes 380 samples, 60 takes 2282 (because of rounding.) SD was about 14.6 microseconds for both. For 380 samples $SE_{\bar{x}}$ is 0.78 microseconds and for 2282 samples 0.32; both are less than 1 microsecond. Peaks clipped at hour 4 and 15 are about 3 microseconds higher than the 1 hour curve.

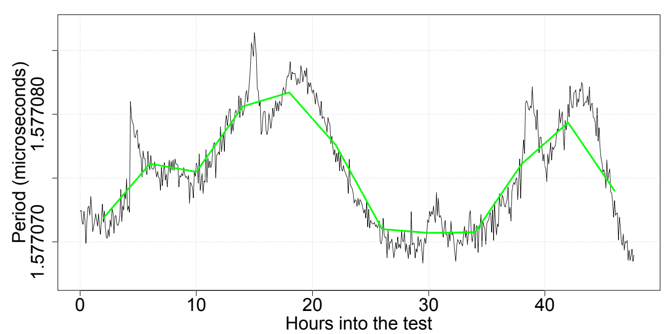

Figure 11: Two days with 5 minute and 240 minute samples.

Sampling over 4 hours as shown in figure 11 is compared to 5 minute samples. Here 5 minutes starts to show significant noise while 4 hours completely distorts the curve. Clearly 4 hour samples violate stationarity.

8 Earthquake

If I were to just open the clock case, stick my hand in and wiggle the pendulum, I could make the period be any value I wanted. This is what an earthquake does.

I have recorded dozens of earthquakes from all over the world. Figure 12 is typical of period variation. It looks very much like a seismograph trace with period values ranging milliseconds away from average. This quake was offshore 250 miles northwest from my home.

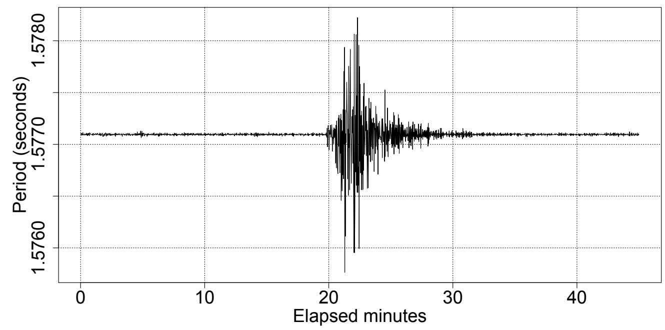

Figure 12: Magnitude 5.7 earthquake west of Ferndale, CA (Jan 28, 2015.)

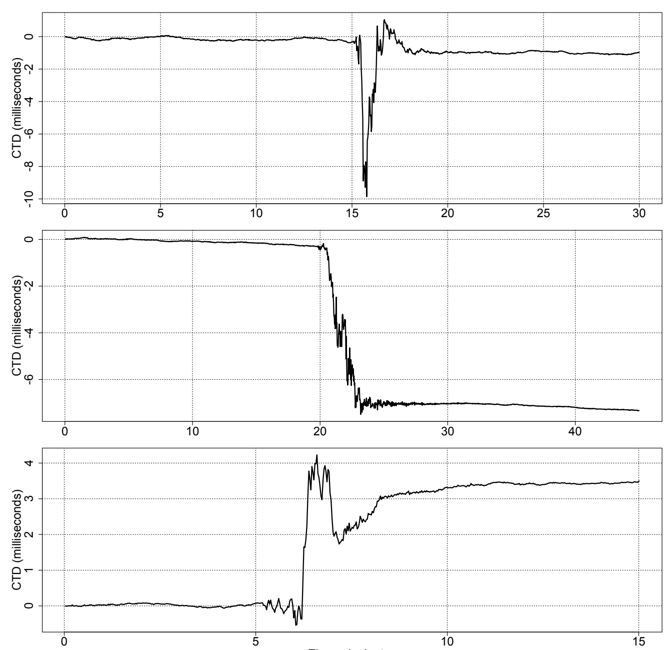

Figure 13: Cumulative time deviation for three earthquakes.

Figure 13 shows from top to bottom a 4.4 magnitude quake 100 miles southeast (Jan 20, 2015), the Ferndale quake shown above, and a Beatty quake (actually in Death Valley) 250 miles east at magnitude 4.8 (Feb 14, 2015).

Each plot is created by taking a trimmed mean of the first 180 cycles of normal pendulum operation before each quake as reference. Then this mean is subtracted from each cycle and the results accumulated by summing one by one to create a cumulative time deviation (CTD) chart[4]. If the earthquake had no net effect on time keeping these plots would have returned to zero after any disturbance - but they did not. Taking another trimmed mean of last 180 cycles doesn't show any difference from the first mean. This shows that the pendulum wasn't permanently changed. Basically the earthquake just added or subtracted a few milliseconds to the accumulated time. Namely Jan 20 -1 milliseconds, Jan 28 -7 milliseconds and Feb 14 + 3.5 milliseconds.

My clocklab experiments have been disturbed by earthquakes as far away as the Japanese quake and tsunami, quakes in South America, and a large number of quakes within 1000 miles of the pendulum. While living only a few miles from the San Andreas fault isn't the best place for a pendulum, large earthquakes can affect clocks far from active faults. These will affect timekeeping by milliseconds in a completely random way.

9 Summary, Notes and Future work

Recording cycle-by-cycle values of pendulum period along with a timestamp, environmental data, and drive information is essential to actually understanding variation in period and hence period modeling for the purpose of correcting a clock.

Using a trimmed mean will remove short term noise from 10 to 30 minute samples without degrading the fidelity of period signal too much. Even most earthquake vibration is filtered out by this process even if timekeeping is disturbed.

Changes in sample standard deviation lowers confidence in a calculated mean as SD increases.

It seems likely that all pendulums are unstable to some degree over time through aging and material changes. Statistically, pendulums are not stationary processes. This implies that they don't actually have a fixed average period about which variations occur. In the case of my experiments, some now 12 years in length, while change slows it doesn't completely go away.

In this paper I have avoided the term *statistical significance*. It is possible to use statistical inference testing to say that one sample mean is different from another with a certain level of confidence; however, I don't think this is particularly useful in pendulum research.

Clearly isolating a pendulum from vibration is a good goal. My setup isn't great.

Earthquakes will likely introduce an error of several milliseconds in the accumulated time.

If you have access to the .pdf[3] for this paper you can just click on the links in References to go to that web page. You can also click any number in square brackets to go to the References page.

All of the concepts in this paper are quite elementary statistics that any first term student would encounter. There are lots of books with the same information as Wikipedia but simple statistical information is neatly packaged there.

Microsoft Excel remains a rough and ready platform for data analysis. I use it all the time. For more serious work *R* and its companion *RStudio* is a freely available statistics and data analysis environment with far more power than Excel. Trimmed mean is available in R (in mean as parameter tr=%) and Excel (=TRIMMEAN(cells,percent)).

To write this paper I have used a much improved working environment made up of Texmaker, MikTex, R, RStudio, my own C code, and Excel. pdfLaTex (the typesetter) isn't that hard to use and I can merge .pdf figures from R directly into the source text with ease. R's base graphics is fiddly but I have learned to cope. Also the bibliographic system is relatively easy to use. Live links to URLs are a big win for readers of the .pdf.

Special thanks to Jim Hansen for carefully proofreading/editing this paper and contributing many improvements. Thanks too to CWB for her careful reading and notes.

Using this work I'll continue to model environmental factors. There is a great deal more work to do.

References

[1] Robert L. Belleville. What I've been up to. *Horological Science Newsletter*, 3:11, 2006.

[2] Robert L. Belleville. Research pendulum - p9vh, fused quartz. `http://www.thirdandlark.com/showArticle.php?article=posts/2013_04_21_p9vh/article.html`, April 2013. Description of a fused quartz pendulum with silk flexure.

[3] Robert L. Belleville. Horological science newsletter archive. `http://www.thirdandlark.com/showArticle.php?article=posts/HSN_Archive/article.html`, 2015. An archive of all my HSN papers in PDF format for download with a brief description.

[4] Robert L. Belleville. Tidal effects on pendulums at various latitudes. *Horological Science Newsletter*, 2:21, 2015. Apendix on cumulative time deviation.

[5] NIST/SEMATECH. e-handbook of statistical methods. `http://www.itl.nist.gov/div898/handbook/`, October 2013. Extensive coverage with many case studies and useful commentary.

[6] John W. Tukey. *Exploratory Data Analysis.* Addison-Wesley, 1977.

[7] USGS. Latest earthquakes. `http://earthquake.usgs.gov/earthquakes/map/`. Dynamic map shows worldwide earthquake information.

[8] Wikipedia. Average. `http://en.wikipedia.org/wiki/Average`.

[9] Wikipedia. Box plot. `http://en.wikipedia.org/wiki/Box_plot`.

[10] Wikipedia. Histogram. `http://en.wikipedia.org/wiki/Histogram`.

[11] Wikipedia. John Tukey. `http://en.wikipedia.org/wiki/John_Tukey`. One of the great figures in real world statistics.

[12] Wikipedia. Median. `http://en.wikipedia.org/wiki/Median`.

[13] Wikipedia. Normal distribution. `http://en.wikipedia.org/wiki/Normal_distribution`.

[14] Wikipedia. Q-q plot. `http://en.wikipedia.org/wiki/Q%E2%80%93Q_plot`.

[15] Wikipedia. Quantile. `http://en.wikipedia.org/wiki/Quantile`.

[16] Wikipedia. Standard deviation. `http://en.wikipedia.org/wiki/Standard_deviation`.

[17] Wikipedia. Standard error. `http://en.wikipedia.org/wiki/Standard_error`.

[18] Wikipedia. Truncated mean. `http://en.wikipedia.org/wiki/Truncated_mean`.