

Statistics Worksheet

Motivation

During the experiment, you designed a mouse trap racer, and ran your design on the floor for a total of ten trials. For each trial, you were able to generate a table of acceleration versus time values. For each time value, you averaged ten acceleration readings, and used the resulting data in analysis. A natural question you may have is

- **Why is it necessary to gather and average the readings from 10 trials, when it is easier to perform just one trial and use the resulting data?**

The reason we need to gather and average data for several trials has to do with the fact that our acceleration sensor is very sensitive to vibrations that can be caused by rough floors, irregularities in the tires, and even local seismic activity! So even though the acceleration versus time graph does not change each time we run the same mouse trap racer, our **recorded** readings change due to any one of these (and other) disturbances. Averaging a set of acceleration readings for a given time instant therefore allows us to obtain a more accurate **estimate** of the actual acceleration at that time.

Gaussian Probability Function

Suppose we know that at a given instant, the mouse trap racer is accelerating at 4.0 m/sec^2 . It is reasonable to expect that most of the time, our accelerometer will give readings close to the true value, for example 4.3 m/sec^2 , and only rarely give readings far from the true acceleration, for example 7.2 m/sec^2 . We can describe each accelerometer reading as Gaussian with mean 4.0 m/sec^2 and some unknown variance, as shown in the Figure 1, where we assume a variance of 1.

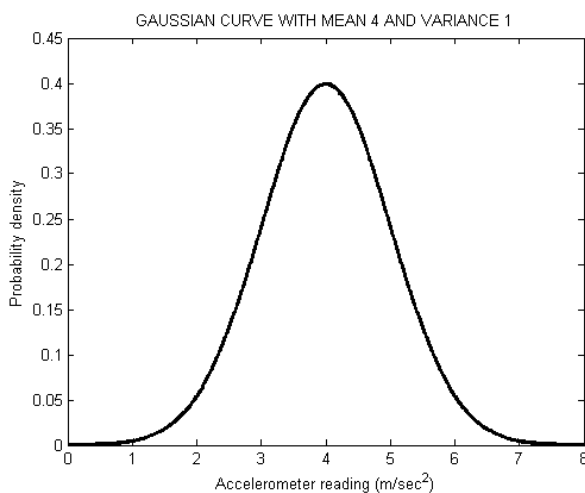


Figure 1: An accelerometer reading probability function, assuming actual acceleration of 4.0 m/sec^2 and variance of 1.

We can estimate a mean of N Gaussian acceleration readings using the following formula

$$\hat{m} = \frac{1}{N} \sum_{i=1}^N a_i$$

where \hat{m} is an estimate of the mean. To estimate the variance of these N readings, we can use the expression

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (a_i - \hat{m})^2$$

Exercise 1

Let's say we run our mouse trap for $N = 3$ trials. At a given time, we know that the true acceleration is 4.0 m/sec^2 , but our obtained readings to three significant digits are

4.18, 3.80, 4.77.

- Estimate the mean \hat{m} of these readings. _____
- Is this estimate close to the true value? _____

Exercise 2

This time we run the experiment for $N = 10$ trials, and obtain the following readings

3.41, 6.18, 3.86, 4.11, 5.07, 4.06, 3.90, 3.17, 4.29, 2.66.

- Estimate the mean \hat{m} of these readings. _____
- Is this estimate closer to the true value than the one in Exercise 1? _____

Variance of the Average

The exercises above gave us some idea of how averaging more readings can improve the estimate of the true acceleration. Now we want to find out what happens to the variance of the estimate. The variance of an average of N readings is related to the variance of each reading by the following formula

$$\sigma_{\text{avg}}^2 = \frac{\sigma^2}{N}$$

where σ^2 stands for the variance of each accelerometer reading. For example, we assumed $\sigma^2 = 1$ in Figure 1. We see that, as we increase N , the variance of the estimate decreases. Therefore, even though it is possible for an average of many readings (large N) to be far from the true value, this will happen much less often compared to the average of just a few readings.

Exercise 3

- Estimate the variances $\hat{\sigma}^2$ of the datasets in Exercises 1 and 2. _____
- Are these values close to 1? _____
- Assuming $\sigma^2 = 1$, what is the variance of the average with $N = 10$? _____
- Assuming $\sigma^2 = 1$, what is the variance of the average with $N = 3$? _____
- Which estimator has the smaller variance? _____

Graphical Example

Let's illustrate our approach using a specific acceleration versus time graph shown in Figure 2.

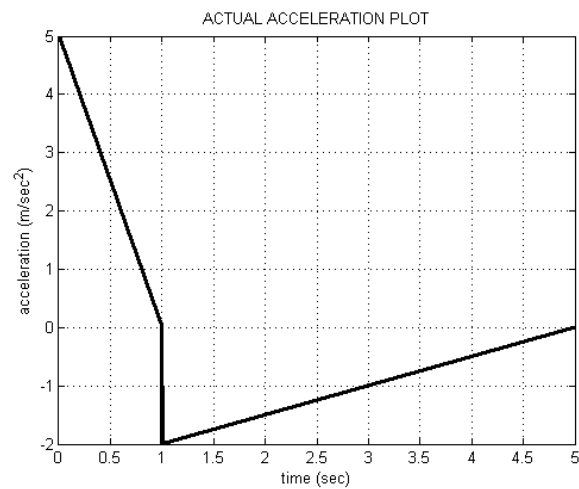


Figure 2: Example actual acceleration plot for a mouse trap racer.

Due to accelerometer disturbances, our recorded data is noisy. Suppose we carry out 10 trials. For a given trial, we may observe the following acceleration versus time plot.

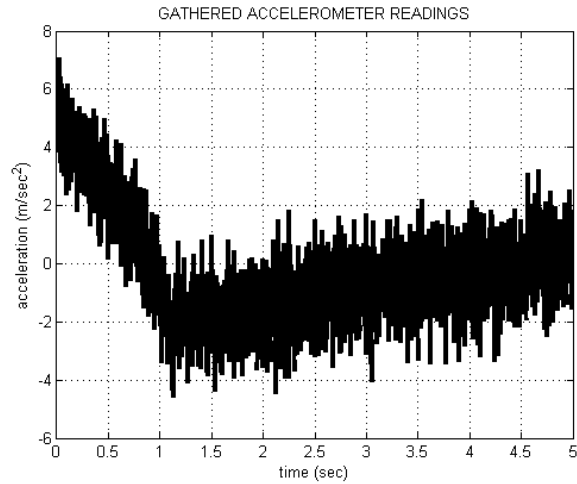


Figure 3: Accelerometer readings for one trial.

After averaging over 10 trials, we obtain a more accurate estimate for the acceleration profile.

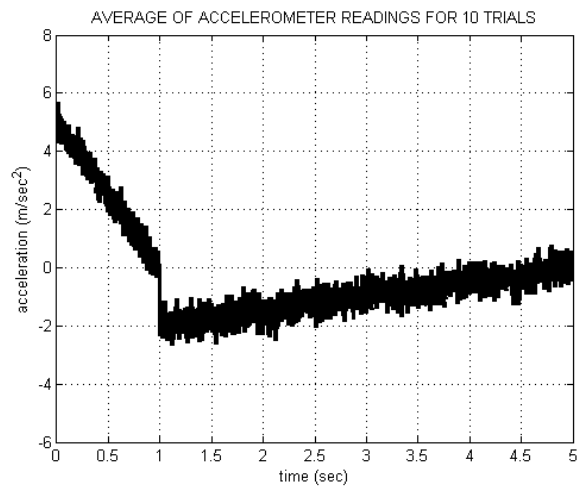


Figure 4: Accelerometer readings averaged over 10 trials.

Exercise 4:

- Compare Figures 3 and 4 to Figure 2. Which is most similar to Figure 2? _____
- In Figure 4, what is the value of N ? _____
- Assuming $\sigma^2 = 1$, compute the variance of the estimates in Figure 4? _____

