# Chance of random noise in a light curve looking like a transiting exoplanet

John Fabrycky

2 November 2019

## 1 Abstract

Planets travel around other stars. If their orbital ring intersects the line of sight between Earth and the star, the planet will block some of the light of the star from our perspective, creating something called a transit. In this project I ask what the chance is that a given number of random dips in the light curve of a star will have a series of times looking like a transiting planet. For the random dips to resemble a transiting planet the dips in the light curve must be spaced evenly. The dips must also be aligned such that we would not expect to see another transit. These calculations are useful to researchers because it can help them determine the chance that an observation actually means a planet.

## 2 Introduction

I am doing this project because my dad is an astrophysicist and said that, as far as he knows, nobody has used the combinatorics method to compute the probability that what appears to be a planet passing in front of a star in a light curve could actually be just noise. So far, astrophysicists have only used the template method to find planets. They line up the light curve of the star with a template of a planet transiting the star and multiply the values to find planets[2].The problem is more complicated than simply addressing the chance of random dips lining up together and involves nuances such as addressing what happens if the noise at the time of the transit cancels it out. Another thing I address is what happens if there is a dip that looks like it was caused by a transiting planet but is actually just noise. These are two major problems introduced by high noise variation. One important discovery I made is that it is very improbable for random dips to line up to look like 4 or more transits from one planet; this result holds even if noise can add or subtract apparent transits from the data.

# 3 Body

The probability of a random number of dips in a light curve looking like a transiting planet is the number of ways the dips can have times which could be caused by a transiting planet divided by the number of ways they can take on different times in total.

## 3.1 2-transit case

In this project I am not considering the 2-transit case because if we have only two dips they could be from 2 different planets with orbital period long enough such that only one transit of each occurs in our data set. Or, these transits could be only noise. Basically, we can not have any confidence that two large dips by themselves mean a planet, even if they do not imply a third dip. This is widely accepted in the field of astronomy[1].

## 3.2 3-transit case

My first investigation was a very simple case. What is the probability that three completely random dips in the light curve fulfill the requirements of a transiting planet? The requirements of a transiting planet are that the first transit is early enough in time such that there should not be a transit before it, the transits are spaced such that there should not be a fourth transit, and the three transits are all present. We can say that the length of time of the data we have is $t$ integer transit durations that fit in the data set, which is continuous without gaps. We can call the end-point time of the data $T$. We have that $t$ ranges from 1 to $T$. If we call the first passing in front of the star A, the second B, and the third C, and we call their times $a$, $b$, and $c$, then we can see that the orbital period of the transits is $b - a$. Thus we have that $a$ is less than or equal to $b - a$ because there must not be enough space before $a$ such that there should be a $0^{th}$ transit (See Figure 1, Plot 1). We see:

$$a \leq b - a \rightarrow$$

$$2a \leq b \rightarrow$$

$$b \geq 2a. \tag{1}$$

This constraint is shown in Figure 1, Plot 2. We see that $b = 2a$ exactly also fulfills this condition because we see that, since the orbital period is $b - a$, the $0^{th}$ transit would come in at $a - (b - a) = 2a - b = 0$, which is exactly before the data is taken.

We see that for $c$ to fit in $T$ (See Figure 1, Plot 2) we must have that $b + (b - a) \leq T$ because $c = b + (b - a)$. We can simplify to

$$2b - a \leq T \rightarrow$$

$$2b \leq a + T \rightarrow$$

$$b \leq \frac{a}{2} + \frac{T}{2}. \tag{2}$$

We include $2b - a = T$ because if $2b - a = T$ then since $c = 2b - a$, $c$ is at $T$ or the last time data is taken.

We also know that we do not want two orbital periods after $b$ to still be at or before $T$ because that would indicate a fourth transit should be present (See Figure 1, Plot 3). The equation for this condition is:

$$b + 2(b - a) > T \rightarrow$$

$$3b - 2a > T \rightarrow$$

$$3b > 2a + T \rightarrow$$

$$b > \frac{2}{3}a + \frac{1}{3}T. \qquad (3)$$

We do not include the equality $b = \frac{2}{3}a + \frac{1}{3}T$ in figure 2, indicating the strict inequality by a dash, because if this was true then the fourth transit is at the last point at which data is taken but is not beyond where the data is taken, indicating there should be a fourth transit.

From the above, we see that given $a, b$ values correspond to a $c$ value and the $a, b$ values are subject to the above constraints. Thus, we can use the constraints which we discovered to find the number of $a, b$ values given $T$. We can plot these $a, b$ values in a graph where the $y-$axis represents possible $b$ values in time and the $x-$axis represents possible $a$ values in time. Each constraint on the $a, b$ values will be a line on such a graph. The constraints will meet each other to create a shape. The $a, b$ values within the shape are the only ones which are valid.

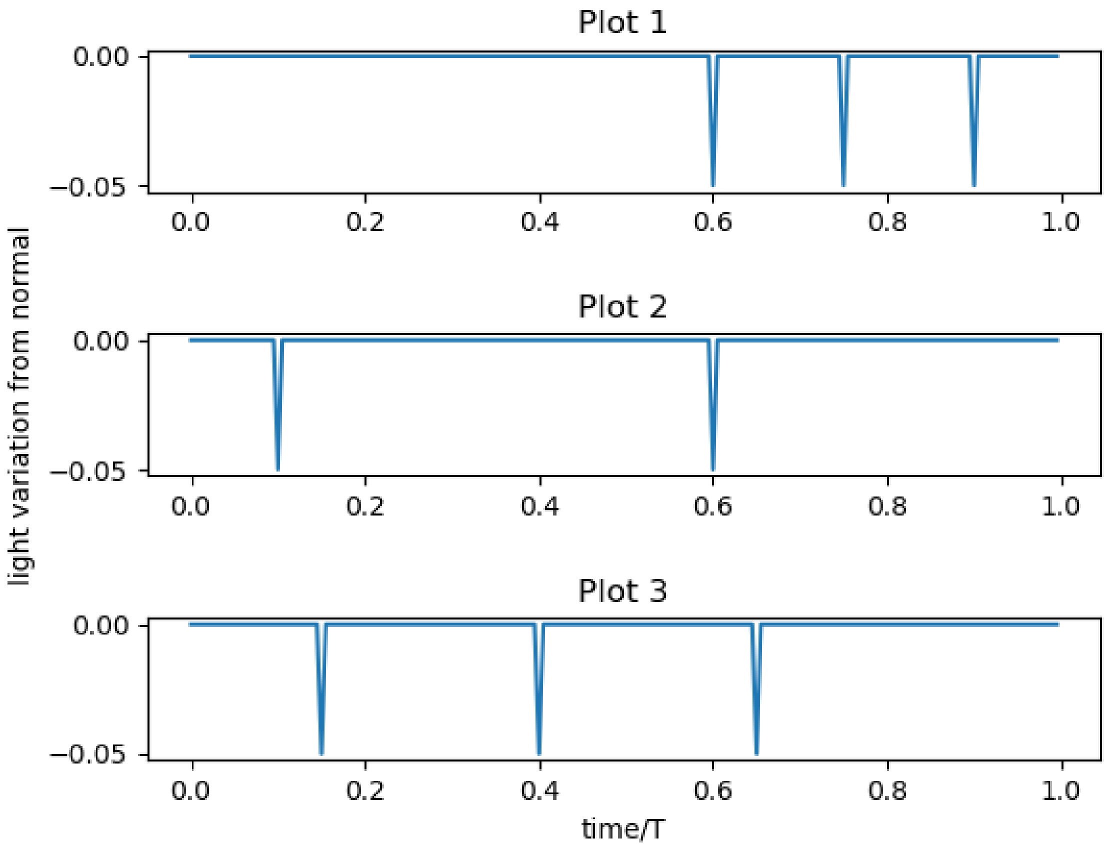



Figure 1:

Invalid arrangements of dips which align:
Plot 1: Invalid: Additional Transit Should Be Detected at time=0.45 but is not
Plot 2: Invalid: there is not enough room for a third transit
Plot 3: Invalid: Additional transit should be detected at time=0.9 but is not

| Region | Area Formula | height | width | area |
|---|---|---|---|---|
| 1 | $\frac{1}{2}\cdot$ height $\cdot$ width | $\frac{T}{6}$ | $\frac{T}{3}$ | $\frac{T^2}{36}$ |
| 2 | $\frac{1}{2}\cdot$ height $\cdot$ width | $\frac{T}{6}$ | $\frac{T}{12}$ | $\frac{T^2}{144}$ |
| 3 | height $\cdot$ width | $\frac{T}{6}$ | $\frac{T}{12}$ | $\frac{T^2}{72}$ |
| 4 | $\frac{1}{2}\cdot$ height $\cdot$ width | $\frac{T}{12}$ | $\frac{T}{4}$ | $\frac{T^2}{96}$ |

Table 1: Calculations to find the areas of the regions

### 3.2.1 Line Intersections

We know the shape of the correct area will be based off of where our bounding lines intersect. It is important to calculate the intersections of the lines so that we can calculate how many possible $a, b$ values there will be given $T$. We can scale $T$ to 1 and add in the $T-$dependence later. We see that $b = 2a$ and $b = \frac{1}{2}a + \frac{1}{2}$ will intersect at $(a, b) = (\frac{1}{3}, \frac{2}{3})$. We see that $b = 2a$ and $b = \frac{2}{3}a + \frac{1}{3}$ will intersect at $(a, b) = (\frac{1}{4}, \frac{2}{4})$. We see that the $y-$intercept of $b = \frac{1}{2}a + \frac{1}{2}$ will be $b = \frac{1}{2}$ and the $y-$intercept of $b = \frac{2}{3}a + \frac{1}{3}$ will be $b = \frac{1}{3}$. We can use these numbers to help calculate the area of the shape. We can divide the area of the shape which represents valid $a, b, c$ values by the total number of possible $a, b, c$ values to get our probability.

### 3.2.2 Calculation to Find Total Possible $a, b$ values

I used Figure 3 to make Table 1 and then used the calculations in Table 1 to find the area of total possible $a, b$ values. Regions 1 to 4 area areas that we do not want to count. We can calculate the area of the square, which is $(\frac{1}{3})^2 = \frac{1}{9}$ and calculate the areas we do not want, regions 1 through 4, and then subtract them from the area of the square to get the desired area. By this reasoning we see that the area of total possible $a, b$ values is $\frac{1}{9} - \frac{1}{36} - \frac{1}{144} - \frac{1}{72} - \frac{1}{48} = \frac{1}{24}$. To scale this area by $T$ we multiply by $T^2$ so we have

$$A = \frac{T^2}{24}.$$

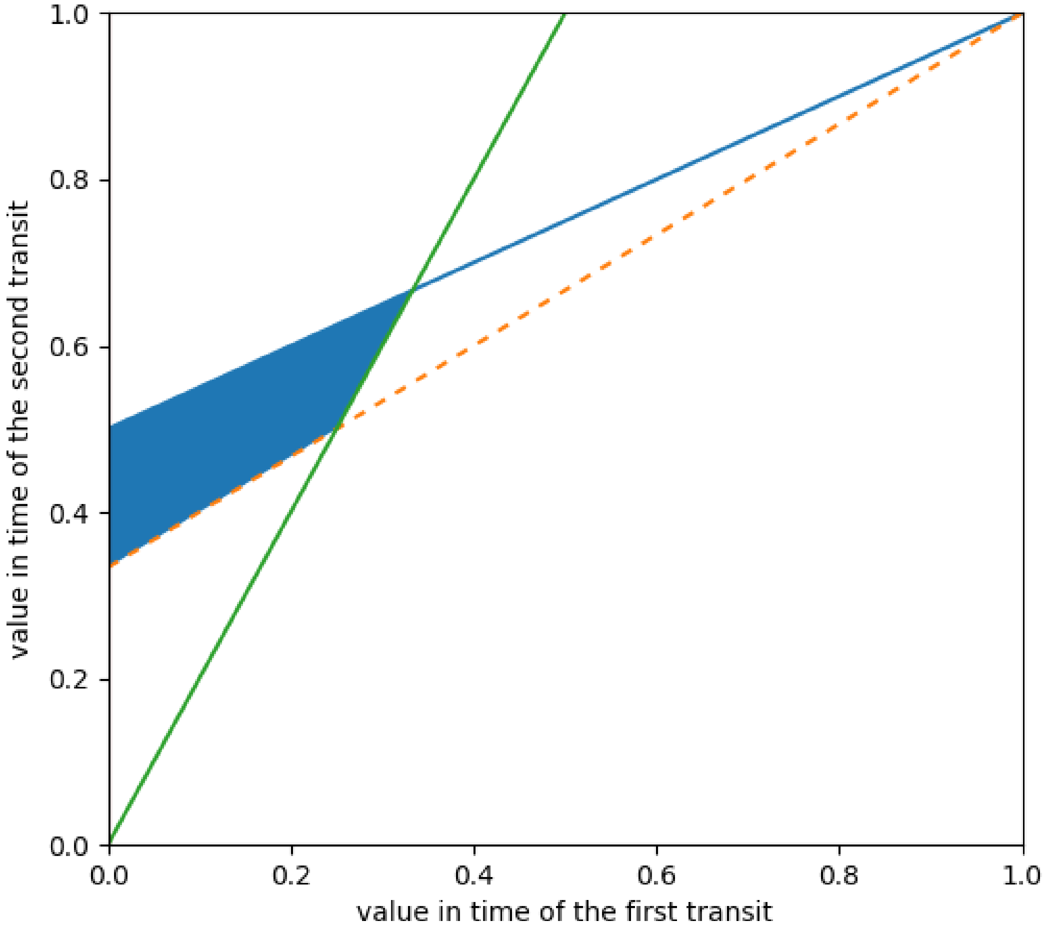



Figure 2: Possible Middle Transit Values Based on the First Transit:

Axes: The total time of the data.
Blue Line: This line puts the third transit at the end. (equation 2)
Orange Line: Puts the fourth transit past the end. (equation 3)
Green Line: Puts the 0th transit at 0 in time. (equation 1)
(If a line is dotted it means that if the middle transit is on the line it is not a valid transit)

(4)

This calculation assumes a continuous light curve whereas in general my paper assumes a discrete light curve but the difference in results between these two calculations is so small for large $T$ that this is fine.

### 3.2.3 Chance calculation

We see that the total number of ways that three numbers $a$, $b$, and $c$ can be ordered such that $a < b < c$ is $\binom{T}{3}$ because each 3 numbers chosen can be ordered in exactly one way, which is from least to greatest, and because we assume that $a, b$ and $c$ are integers. Because the area of possible $a, b$ values is $\frac{T^2}{24}$ we have that the probability of 3 values corresponding to $a, b$ and $c$ to line up is

$$\frac{\frac{T^2}{24}}{\frac{T(T-1)(T-2)}{3!}} \rightarrow$$

$$\textbf{probability} = \frac{T^2}{4T(T-1)(T-2)} \tag{5}$$

.

### 3.3 n-transit case

My second investigation is slightly more complicated because it is not just for three transits in front of the star but for $n$ transits in front of the star. We can call the time of each transit $d_1, d_2, ...d_n$. We can use similar reasoning as in the three-transit example. The distance between two transits is $d_2 - d_1$. Thus, for $d_n$ to be included in the data because we have that $d_n = d_1 + (n-1)(d_2 - d_1)$ we must have $d_1 + (n-1)(d_2 - d_1) \leq T$. This works out to

$$d_2 \leq \frac{d_1(n-2)}{n-1} + \frac{T}{n-1}. \tag{6}$$

We also see that $d_{n+1}$ must be greater than $T$. We see that $d_{n+1} = d_1 + n(d_2 - d_1)$. These combined give us $d_1 + n(d_2 - d_1) > T$ which works out to

$$d_2 > \frac{d_1(n-1)}{n} + \frac{T}{n}. \tag{7}$$

We also see that $d_0$ must come at or before 0 in time. We have that $d_0 = d_1 - (d_2 - d_1)$. We also have that $d_0 \leq 0$. These combined gives us $d_1 - (d_2 - d_1) \leq 0$ which simplifies to

$$2d_1 \leq d_2. \tag{8}$$

Thus we have three bounding lines for $d$:

$$d_2 \leq \frac{d_1(n-2)}{n-1} + \frac{T}{n-1}, d_2 > \frac{d_1(n-1)}{n} + \frac{T}{n}, \textbf{ and } d_2 \geq 2d_1.$$

#### 3.3.1 Line Intersections

We can calculate that the intersection of our line $d_2 = \frac{T}{n-1} + \frac{d_1(n-2)}{n-1}$ and $2d_1 = d_2$ is

$$(d_1, d_2) = (\frac{T}{n}, \frac{2T}{n}).$$

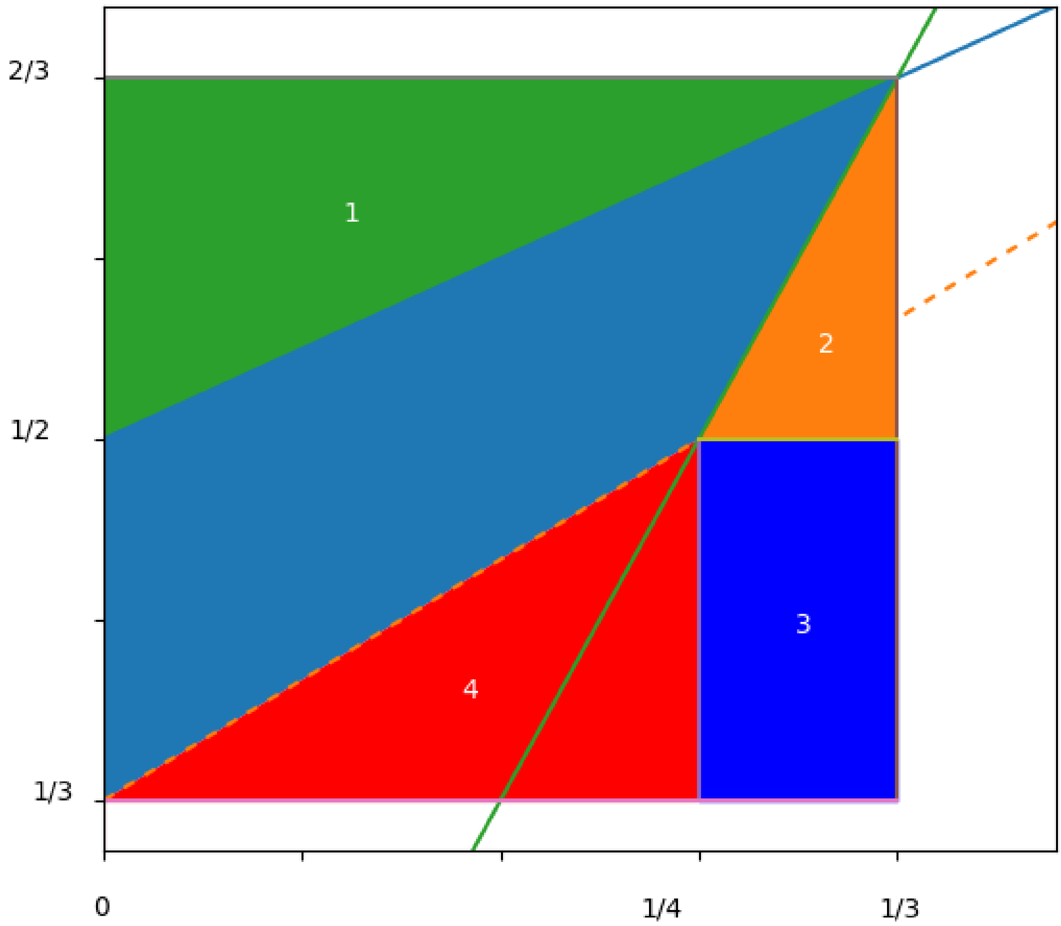



Figure 3: We can subtract areas 1, 2, 3, and 4 from the entire square to get the area of the correct $a, b$ values

The intersection of our line $d_2 = \frac{T}{n} + \frac{d_1(n-1)}{n}$ and $2d_1 = d_2$ is

$$(d_1, d_2) = (\frac{T}{n+1}, \frac{2T}{n+1}).$$

The $y-$intercept of

$$d_2 = \frac{T}{n-1} + \frac{d_1(n-2)}{n-1}$$

is

$$d_2 = \frac{T}{n-1}.$$

The $y-$intercept of

$$d_2 = \frac{T}{n} + \frac{d_1(n-1)}{n}$$

is

$$d_2 = \frac{T}{n}.$$

### 3.3.2 Calculation for Finding the quantity of $a, b, c$ triples given $T$

I used Figure 4 to make Table 2. We can use the calculations in Table 2 to find the area of total possible $a, b$ values using complementary counting. The areas of the regions in the tables are areas we do not want to count. The total area of the square is $(\frac{1}{n})^2 = \frac{1}{n^2}$. Thus, the total area of allowed solutions is

$$\frac{1}{n^2} - a$$

where

$$a = (\frac{T^2}{n^2} - \frac{T^2}{2n^2 - 2n}) - (\frac{T^2}{n^2} - \frac{T^2}{2n^2 - 2n} - \frac{T^2}{n^2 + n} + \frac{T^2}{2n^2 - 2}) - (\frac{T}{n} - \frac{T}{n+1}) \cdot (\frac{2T}{n+1} - \frac{T}{n}) - \frac{T}{n+1} \cdot (\frac{T}{n-1} - \frac{T}{4n})$$

so

$$A = \frac{2T^2}{n(n+1)^2} \tag{9}$$

| Region | Area Formula | height | width | area |
|---|---|---|---|---|
| 1 | $\frac{1}{2} \cdot$ height $\cdot$ width | $\frac{2T}{n} - \frac{T}{n-1}$ | $\frac{T}{n}$ | $\frac{T^2}{n^2} - \frac{T^2}{2n^2-2n}$ |
| 2 | $\frac{1}{2} \cdot$ height $\cdot$ width | $\frac{2T}{n} - \frac{2t}{n+1}$ | $\frac{t}{n} - \frac{t}{n+1}$ | $(\frac{T}{n} - \frac{T}{n+1})^2$ |
| 3 | height $\cdot$ width | $\frac{2T}{n+1} - \frac{T}{n}$ | $\frac{T}{n} - \frac{T}{n+1}$ | $(\frac{T}{n} - \frac{T}{n+1}) \cdot (\frac{2T}{n+1} - \frac{T}{n})$ |
| 4 | $\frac{1}{2} \cdot$ height $\cdot$ width | $\frac{2T}{n+1} - \frac{T}{n}$ | $\frac{T}{n+1}$ | $\frac{T}{n+1} \cdot (\frac{T}{n-1} - \frac{T}{2n})$ |

Table 2: Calculations to find the areas of the regions

### 3.3.3 Chance Calculation

For n-transits, we will have $\binom{T}{n}$ choices for $n$ random dips. Now we divide the area of the correct possibilities by the total number of possibilities and we get

$$\frac{\frac{2T^2}{n(n+1)^2}}{\binom{T}{n}} \rightarrow$$

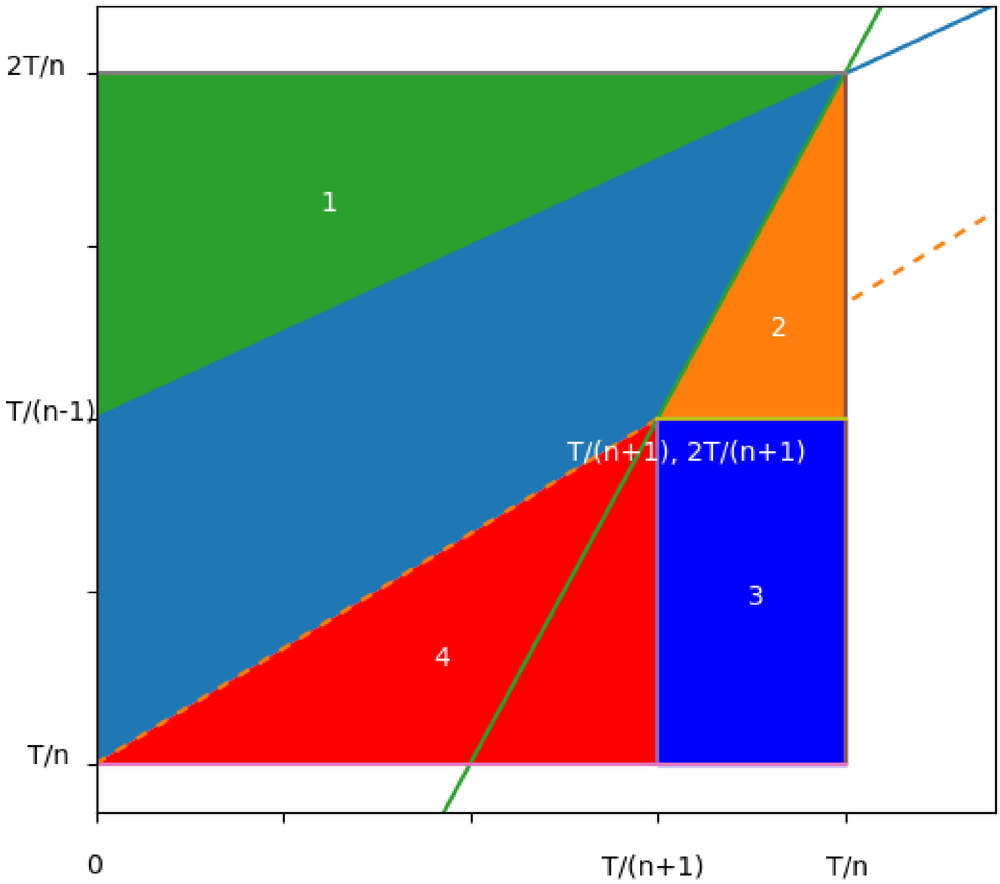
$$\textbf{probability} = \frac{2T^2}{n(n+1)^2\binom{T}{n}} \tag{10}$$

where probability is the chance of the $n$ dips lining up correctly. We see that for large $T$ this is approximately equal to

$$2T^{2-n}\frac{(n-1)!}{(n+1)^2}.$$

## 3.4 When there are missing dips

Sometimes the noise in the data can be so large that it completely cancels out a transit so we don't detect it (figure 5). I simulated this by creating a light curve where one transit is cancelled by high noise variation. I made a sample light curve with a gausian distribution and then at each of the transit times I subtracted 0.05 from the value of the light curve at that time. The transit dips varied because sometimes the noise that I imposed the transit dips on was above average and sometimes it was below average. Then I tested if at least one of




Finding Area of the First and Second Transits

Figure 4:

We can subtract areas 1, 2, 3, and 4 from the whole square to get the area of correct $d_1, d_2$ values.

the transit dips was 0.02 less deep than the deepest non-transit dip. If this was satisfied, the code showed the figure with the light variation. This can cause what appears to be missing dips in the data. If there are missing dips, then if we see $n$ dips that line up correctly with $m$ gaps then we have $\binom{n+m}{n}$ different ways to choose which dips we see and which we don't see. The "area" that those different possibilities will occupy in an $n-$dimensional graph will be $\frac{2T^2}{(n+m)(n+m+1)^2}$ when we substitute $n+m$ in equation 9 and the different possibilities for the ways $n$ can line up will be $\binom{T}{n}$. Now we have that the probability is

$$\frac{\left(\frac{2T^2}{(n+m)(n+m+1)^2}\right)\binom{n+m}{n}}{\binom{T}{n}} \rightarrow$$

$$\textbf{probability} = \frac{2T^2\binom{n+m}{n}}{(n+m)(n+m+1)^2\binom{T}{n}}. \tag{11}$$

This, for large $T$, is approximately equal to

$$2T^{2-n}\left(\frac{\binom{n+m}{n}}{(n+m)(n+m+1)^2 n!}\right).$$

This is the probability of a random $n$ dips having times which are evenly spaced but have $m$ gaps where transits are predicted. It is important because researchers want to know if transits with gaps should be considered reliable as predicting a planet.

We can use this calculation to find some results (table 1). We say that each row represents the total number of transits, seen and predicted, starting at 3 transits at the top and increasing to 9 transits at the bottom. Each column represents the total number of dips seen. The number in each box is calculated by using the formula we just derived and plugging in the given numbers. Let's do this calculation for Kepler's data because that is what most scientists use. We can calculate $T$ for Kepler's data. If we say that the length in time of one data point is 1 hour, and there are approximately 25 hours in a day, 365 days in a year, and Kepler took data for 4 years, then we can see that $T \approx 36500$. We say that the

Table 3: Reliability of Data With Missing Dips: Kepler viewed 200,000 or $2 \cdot 10^5$ light curves, so if the chance of a certain configuration happening is greater than $10^{-6}$ the chance of a false detection of a planet orbiting a star is greater than $\frac{1}{5}$ so we do not want to count these. If a probability is shown in green, it means this configuration is good and if we find it we can deduce a planet.

| $n+m$ | $n=3$ | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|
| 3 | 6.90e-06 | | | | | | |
| 4 | 1.32e-05 | 3.65e-10 | | | | | |
| 5 | 1.84e-05 | 1.01e-09 | 2.80e-14 | | | | |
| 6 | 2.25e-05 | 1.86e-09 | 1.03e-13 | 2.84e-18 | | | |
| 7 | 2.59e-05 | 2.85e-09 | 2.36e-13 | 1.30e-17 | 3.60e-22 | | |
| 8 | 2.86e-05 | 3.95e-09 | 4.36e-13 | 3.60e-17 | 1.99e-21 | 5.49e-26 | |
| 9 | 3.09e-05 | 5.11e-09 | 7.06e-13 | 7.79e-17 | 6.45e-21 | 3.56e-25 | 9.81e-30 |

$n+m =$ total number of dips, seen and predicted,
$n =$ total number of dips seen.

data is reliable only if the chance such an arrangement will happen randomly is less than $10^{-6}$ because Kepler observed 200,000 stars, so if the chance is greater than $10^{-6}$ the chance of a false positive in this data set is greater than $\frac{1}{5}$ which is too large.

### 3.5 When there are extra dips detected

Sometimes there will be exceptionally large dips in the data purely due to noise (figure 6). These can cause extra dips that do not line up properly with the rest of the data. I stimulated this by creating a sample light curve with a gaussian distribution and then imposing three transit dips on it by subtracting 0.05 from each transit time. After this, I wrote the code to test if there was a dip which was deeper than a transit dip and one of the transit dips was less than 0.04. If both of these were fulfilled, the figure was shown. We can say that there are $n$ dips detected that line up properly and $l$ dips detected total. Then the total number of possibilities for how these $l$ dips can be arranged will be the number of possibilities for how the $n$ dips can line up properly and for each of those possibilities there are $\binom{T-n}{l-n}$ ways to choose where the random $l-n$ extra dips will go. Thus, the total number of correct ways

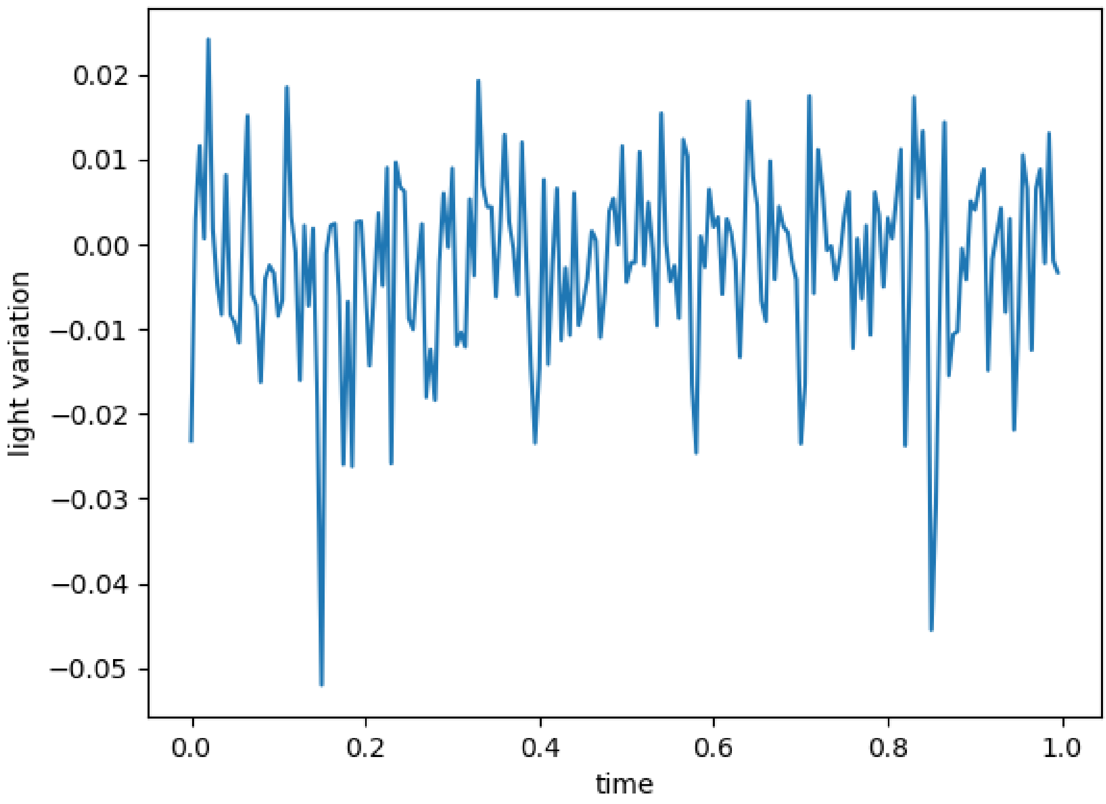



Figure 5: Possible transit cancelled by noise

the $l$ dips can occur is $\frac{2T^2}{n(n+1)^2} \cdot \binom{T-n}{l-n}$. The total number of ways the $l$ dips can occur is $\binom{T}{n+m}$. Now, we divide the correct possibilities by the total number of possibilities and get that the probability of $l$ dips taking on times with $n$ of them lining up without gaps is

$$\frac{\frac{2T^2}{n(n+1)^2} \cdot \binom{T-n}{l-n}}{\binom{T}{l}}$$

or

$$\textbf{probability} = \frac{2T^2\binom{T-n}{l-n}}{n(n+1)^2\binom{T}{l}}. \tag{12}$$

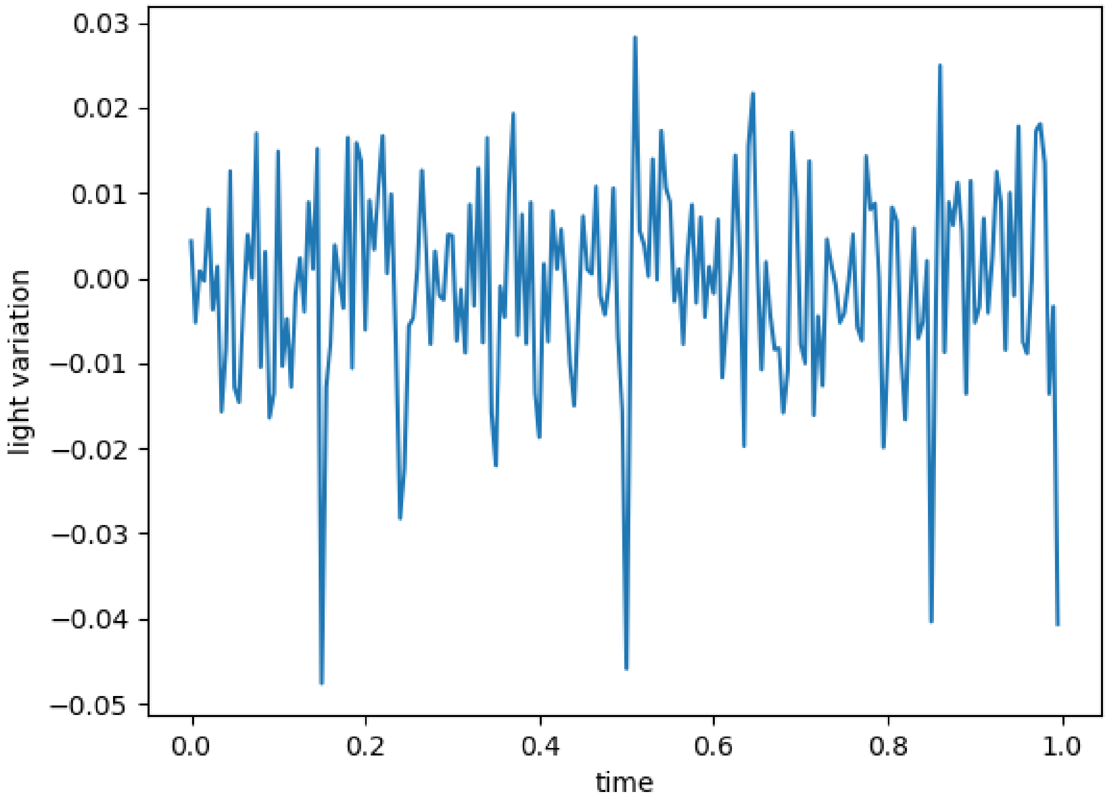
The probability here is the chance of $l$ dips lining up with $n$ of them lining up correctly to look like a transiting planet and the rest having random times. We can approximate this equation for large $T$ as

$$2T^{2-n}\frac{l(l-1)(l-2)...(l-n+1)}{n(n+1)^2}$$

We can use this to find some results. This calculation is assuming similar things to the last calculation ($T = 36500$, it is safe to count aligning dips as a planet as long as how they align has a probability of less than $10^{-6}$). For this calculation, each row represents a certain number of dips which align, starting at 3 dips aligning at the top and going to 9 dips aligning at the bottom. Each column represents a total number of extra dips, starting at 3 extra dips and going to 9 extra dips. We see that the data is reliable if there are at least 4 dips which align even with some extra dips which don't align. (table 2).

## 3.6 When both occur

What if both of the above complicating factors to our original problem happen in the same data set? Then we will have missing transits and extra transit-like features that do not line up correctly. We can say that the total number of dips below a certain level is $l$ and $n$ of




Extra Dip Detected

Figure 6: Sometimes randomly there will be a dip in the data that is so low that it is lower than one of the three transits, like the last dip here. (The line represents the amount of light detected)

Table 4: Extra Dips

| $n$ | $l-n=3$ | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|
| 3 | 1.38e-04 | 2.41e-04 | 3.86e-04 | 5.79e-04 | 8.28e-04 | 1.14e-03 | 1.52e-03 |
| 4 | 1.28e-08 | 2.56e-08 | 4.60e-08 | 7.67e-08 | 1.21e-07 | 1.81e-07 | 2.61e-07 |
| 5 | 1.57e-12 | 3.53e-12 | 7.06e-12 | 1.29e-11 | 2.22e-11 | 3.60e-11 | 5.61e-11 |
| 6 | 2.38e-16 | 5.96e-16 | 1.31e-15 | 2.62e-15 | 4.87e-15 | 8.52e-15 | 1.42e-14 |
| 7 | 4.32e-20 | 1.19e-19 | 2.85e-19 | 6.17e-19 | 1.23e-18 | 2.31e-18 | 4.11e-18 |
| 8 | 9.06e-24 | 2.72e-23 | 7.06e-23 | 1.65e-22 | 3.53e-22 | 7.06e-22 | 1.33e-21 |
| 9 | 2.16e-27 | 7.02e-27 | 1.96e-26 | 4.91e-26 | 1.12e-25 | 2.39e-25 | 4.77e-25 |

$n =$ the number of dips which line up,
$l =$ the total number of dips used.

them line up with $m$ gaps where transits are predicted. One factor of the chance of this happening is $\binom{n+m}{n}$ because of the $n+m$ transits only $n$ of them show. Another factor is $\binom{T-n-m}{l-n}$ and is the number of ways the $l-n$ extra dips can be placed. Another factor is $\frac{2T^2}{(n+m)(n+m+1)^2}$ and is the area of the first two dips in a sequence of $n+m$ transits because all the other dips are based on the first two. The total number of ways that this can happen can be obtained by multiplying these factors together, so we have

$$\binom{n+m}{n}\binom{T-n-m}{l-n}\frac{2T^2}{(n+m)(n+m+1)^2}.$$

Now we have that the total number of ways that $l$ dips happen is $\binom{T}{t}$. Thus, we have that the final probability is

$$\frac{\binom{n+m}{n}\binom{T-n-m}{l-n}\frac{2t^2}{(n+m)(n+m+1)^2}}{\binom{T}{l}} \rightarrow$$

$$\textbf{probability} = \frac{2T^2\binom{n+m}{n}\binom{T-n-m}{l-n}}{(n+m)(n+m+1)^2\binom{T}{l}}. \tag{13}$$

This probability is the chance that $l$ dips will have times with $n$ of them lining up correctly with $m$ extra transits predicted. This is important because researchers want to know if transit dips are reliable even if there are missing dips and additional non-aligned dips. We can approximate this for large $T$ as

$$2T^{2-n}\frac{\binom{n+m}{n}l(l-1)(l-2)...(l-n+1)}{(n+m)(n+m+1)^2}.$$

Let us find some results (table 3). We are using the same assumptions as in the last 2 tables. Our whole table is for level 10, or when we use the lowest 10 dips. Each row represents the number of dips detected which align, which is 4 in the top row and 6 in the bottom row. We can deduce the number of extra dips which are predicted from the number of dips which are seen because the total number of dips is always 10 in this graph, so it would be 6 extra in the top row and 4 extra in the bottom row. Each column represents

Table 5: Missing Dips and Extra Dips ($l = 10$ dips detected total)

| n | $m = 0$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 4 | 7.67e-08 | 2.13e-07 | 3.91e-07 | 5.99e-07 | 8.28e-07 | 1.07e-06 | 1.33e-06 | 1.60e-06 |
| 5 | 7.06e-12 | 2.59e-11 | 5.95e-11 | 1.10e-10 | 1.78e-10 | 2.64e-10 | 3.70e-10 | 4.96e-10 |
| 6 | 5.96e-16 | 2.74e-15 | 7.57e-15 | 1.63e-14 | 3.04e-14 | 5.11e-14 | 7.98e-14 | 1.18e-13 |

$n$ = dips detected which align,
$m$ = dips which are predicted but not seen.

the number of transit dips which are predicted but not seen. The left-most column is for 0 predicted transit dips and the right-most is for 7 extra predicted transit dips. We see that if we have 4 dips which align with 5 or more predicted dips and 6 extra dips which don't align then we cannot deduce a planet but otherwise we can as long as we don't have too many extra dips predicted.

# 4 Conclusion

In conclusion, we discovered it is highly unlikely that noise would randomly look like 4 or more transits caused by a planet but that it is much more likely for 3 transits. We also discovered that the probability that random dips will line up to look like a transiting planet is surprisingly insensitive to whether we allow missing dips or extra dips.

At the end of this project I was much better at coding than at the beginning of the project because I had to code all of the figures in this project. I also became much more confident with probability and how to count. I learned how to better apply probability to real-world problems, in places where people tend to use brute force methods.

Some research that I could do to extend this problem is to not only calculate the chance of random dips having times in certain forms but to also calculate the chance of noise in a data table cancelling out some transit dips and adding extra dips, but this is more statistics and my project was strictly about probability.

# 5 Bibliography

My dad, who inspired this project.

1. Jenkins, J. M., Chandrasekaran, H., McCauliff, S. D., Caldwell, D. A., Tenenbaum, P., Li, J., et al. (2010). Transiting planet search in the Kepler pipeline. *Proc. SPIE 7740, Software and Cyberinfrastructure for Astronomy*, 77400D; https://doi.org/10.1117/12.856764

2. Batalha, N.M., Rowe, J.F., Bryson, S.T., Barclay, T., Burke, C.J., Caldwell, D. A., et al. (2013). Planetary Candidates Observed By Kepler. III. Analysis of the First 16 Months of Data. *The Astrophysical Journal Supplement Series* 204:24