

Loaded Dice Activity

Dr Martin Archer, School of Physics & Astronomy, Queen Mary University of London

An activity on hypothesis testing and signals vs backgrounds suitable for students aged 14+ and at public events. Developed around Particle Physics, it could easily be applied to any topic which relies on statistical analysis.

Introduction

This activity was developed to discuss how particle physicists use statistics to identify new particles like the Higgs Boson through their decay products (the signal) whilst considering the likelihood of getting the same decay products from other known channels (the background). A video detailing this further through applying the activity can be found [here](#).

The activity uses a mixture of loaded (the signal) and fair (the background) dice alongside either an application or spreadsheet to test the significance of there being any unfair dice present from the results of dice rolls. While we have custom-made some weighted dice that are only slightly unfair (as shown in the video), these require large numbers (several thousands) of rolls in order to guarantee 5 σ significance (a 1 in 1,744,278 probability of being down to chance). However, there are a number of more heavily weighted dice along with identical fair dice available commercially e.g. [The Dice Shop](#), [Smiffy's](#), or [Funny Man](#). Note: some loaded dice do not contain weights inside but simply have the same number displayed on multiple faces. Two to three of these heavily weighted dice should be sufficient to arrive at a significant result for a classroom-sized audience.

What You'll Need

2-3 loaded dice

15-30 fair dice

Computer (at least one)

How to Play

Depending on your IT capabilities, either everyone can participate individually on multiple computers or the activity can be done collectively on just one computer. The basics of how to play remain the same:

1. Each person will need at least one dice to start with
2. Upon rolling they either enter which number came up on their computer (either in the application or the spreadsheet) or alternatively if doing collectively they call out their result.
3. Ideally everyone would change dice after every roll. If that is not practical let everyone roll their dice for 10-15 seconds before either by everyone passing their dice on to their neighbour or pooling all the dice together again and picking a new one from the pile.
4. Repeat this for 3-5 minutes of play.

If everyone was participating individually you can ask to see if anyone has a significant result? Then by pooling all the data together you can see if a result emerges by having more dice rolls.

Application

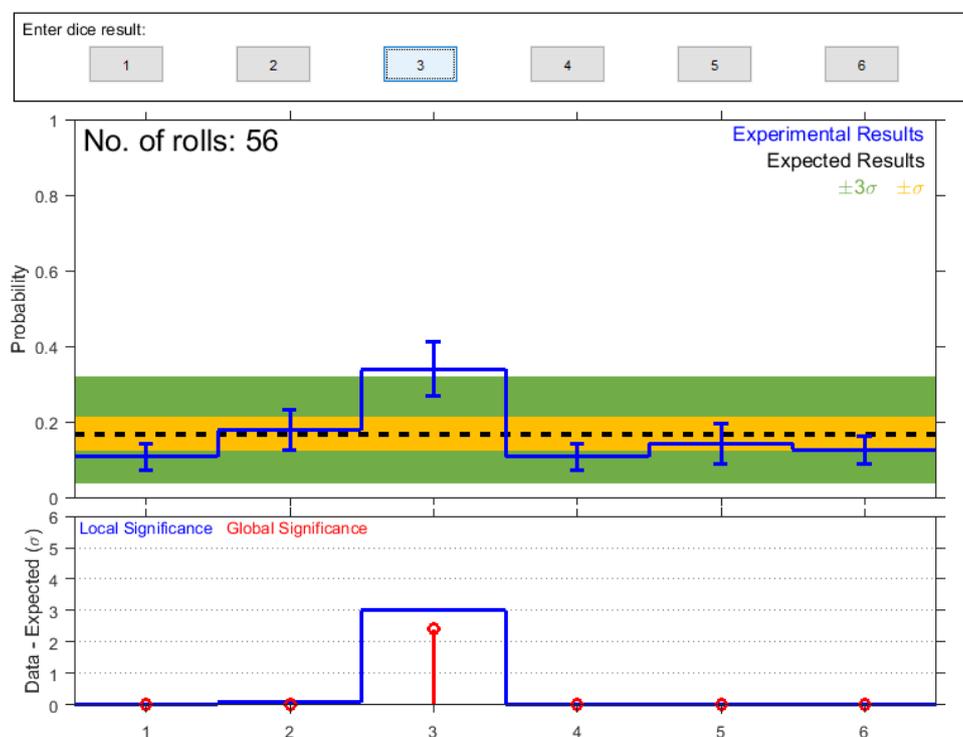
There are two applications, the User and the Master programmes. To run either of these you will need to have the MATLAB runtime installed, the correct versions have been included (note unfortunately only Windows is currently supported).

User

This programme is for entering dice rolls, either individually or collectively.

When setting this programme up for users please change the contents of the file 'datalocation.txt' which sets where the all the results will be written. Ideally you should set this to a writable network folder for easy pooling of the data in the Master programme. In the event the programme can't find the right location, it will default to the 'Data' subfolder. Each person's data will be written to a text file named after their username.

Upon opening the programme you'll be asked for number of faces the dice you're using has (the default is 6) and then you should be shown a display like this:



The buttons on the top are for recording the result of each dice throw.

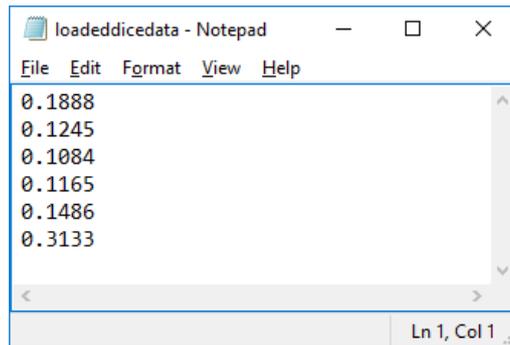
In the top panel, the blue histogram depicts the probabilities of each dice outcome from the results so far along with 1σ error bars. The black dashed line is the expected result for fair dice along with the 1σ (yellow) and 3σ (green) regions around it. Therefore, a likely signal would be where the blue pops above the green.

The significance of any results is made more quantitatively in the bottom panel, depicting the likelihood of at least that number of rolls purely by chance as measured in standard deviations. The blue histogram here is the 'local significance', which is the straightforward comparison, whereas the 'global significance' in red takes into account that we're testing all 6 numbers and therefore must reduce the probabilities of a result. Therefore, the true significance of any signal is given in red.

Master

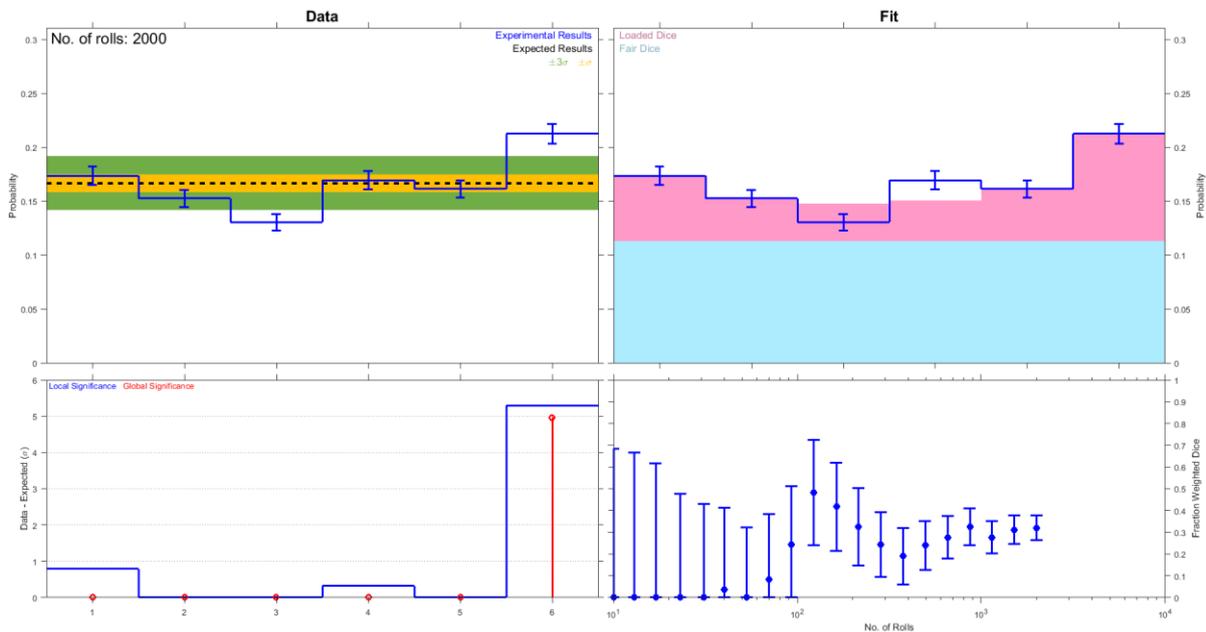
This programme pools all the dice roll data together and animates the results as you add more and more data.

Upon opening the programme you'll again be asked for number of faces the dice you're using has (the default is 6). The programme will then ask you whether you'd like to perform a fit to the data. This fit will estimate what fraction of loaded dice there were, but it requires you to know the probabilities associated with whatever loaded dice you use. These results should be stored in 'loadeddicedata.txt' as below.



```
loadeddicedata - Notepad
File Edit Format View Help
0.1888
0.1245
0.1084
0.1165
0.1486
0.3133
Ln 1, Col 1
```

The programme will then ask you to navigate to the directory where all the data is stored. This should either be the network folder you set up for the User programme or a folder where you've copied all the data across to. It will then load data from all the text files in this directory and analyse the results.

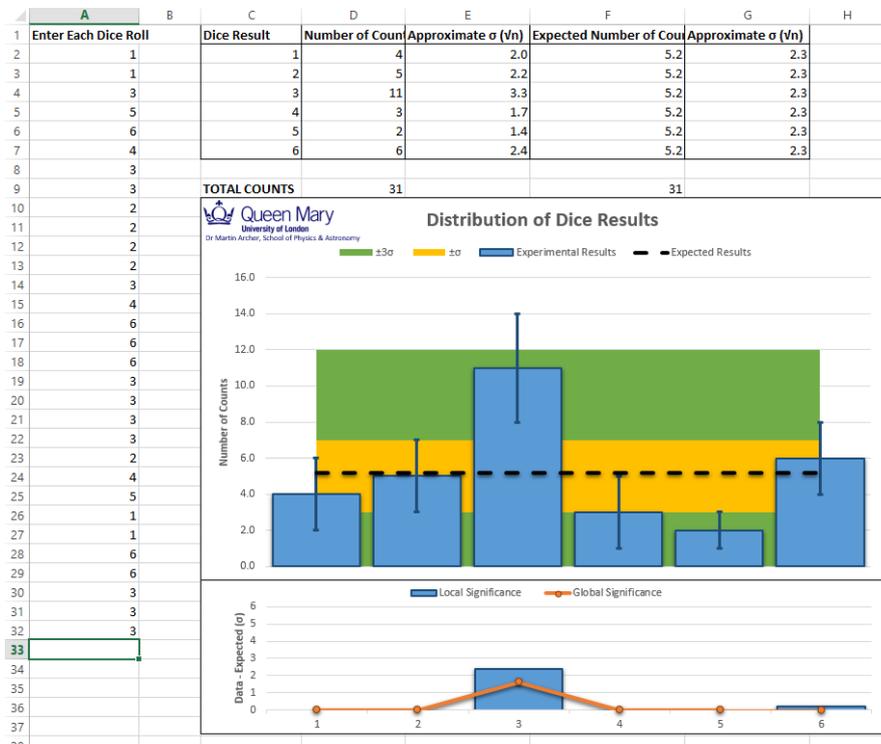


The eventual output will look something like the above. While the left hand side is nearly identical to the User programme, if you selected the fit option you'll also get the graphs on the right hand side. The top right shows the same results and error bars in blue but the coloured areas represent the estimated proportions of fair (light blue) and loaded (light red) dice. The bottom right hand panel shows the estimated proportion of loaded dice along with 1σ uncertainties as a function of the number of dice rolls. You can see in this example that initially the estimate was zero but with large uncertainty and as the number of rolls increased did it arrive with a non-zero estimate and smaller uncertainties.

Spreadsheet

If you are unable to use the applications, most of the same functionality has been implemented in Excel spreadsheets. If you are using Excel 2007 or earlier you should use the 'legacy' version.

Dice roll results should be added to the first column. All other cells are locked down for editing with the password 'martinarcher'. The results displayed are in a similar format to the User application.



The 'macro' version of the spreadsheet has the ability to pool together data from multiple spreadsheets together. Make sure macros are enabled, press the 'Import Data' button and navigate to the folder where all the spreadsheets are located. Their data will then be copied into the 'macro' version displaying the results. Unfortunately it is not possible within Excel to do the fitting procedure to estimate the proportion of loaded dice.