Signal vs. background problem set

You can work through this problem set however you like. Jupyter notebooks, matlab, analytic calculations, etc. But please document your thinking clearly in words and plots, as well as coming up with an answer.

For the first few problems we're going to play with a Poisson distribution. Tools for throwing random poisson distributions and calculating cumulative probability functions can be found in scipy.stats or matlab's Statistics and Machine Learning Toolbox.

So let's imagine that we have an experiment that counts events, and the average number of events is well determined to be 50.1. You can imagine this as something that counts radiative decays, with a background decay rate of 50.1 in your measurement interval. Or it could be counting cosmic rays coming from a particular direction on the sky.

1) Simulate what this distribution will look like, and plot on a semilogy histogram plot. 1e4 or 1e5 points should be fine. This is a set of realizations of the background.

[It turns out in matlab plotting a semiology histogram is needlessly finicky. For matlab the following magic incantations help make semilogy histograms:

```
[n, xout] = hist(data));
bar(xout, n, 'barwidth', 1, 'basevalue', min);
%where min is smallest value to show on the log scale. 1e-1 or 1e-5 for example.
set(gca,'YScale','log')
```
]

2) Turn your measurements of the background into a measured probability distribution. The result should integrate to 1.

3) Now imagine that you are looking for a source (something that will add signal counts to your background), and you measure 80 events. What is the probability of your detection? [Hint: your simulation does not have enough realizations, so an analytic calculation is needed.]

4) Now imagine you are looking for a sink (something that blocks the background). For cosmic rays the moon is a great example as it blocks celestial cosmic rays and appears as a hole in the sky. If you measure 30 events what is the probability of your detection?

5) Convert your answers from 3 & 4 into 'sigma'.

6) Note that both the signals 3 & 4 were about the same distance from the mean. Did they have the same significance? Explain/discuss.

For the next few problems, we are imagining that we are concerned with the length of a vector, where the measurements of both coordinates are independent. To simulate this start with two Gaussian random variables A & B, both with zero mean and the same variance, then calculate realizations of the magnitude measurement.

7) Plot the resulting magnitude pdf.

8) Now make the variance of one of the variables much larger than the other, and replot the magnitude pdf.

9) Now let's use the result of 8 as the basis for a different problem. In this case you can measure the full vector. Create a 2D histogram of the simulation from 8 (axis are length of A & B respectively).

Two brain teasers (no need to turn in)

10) Think about how one would extend the idea of 'more signal like' from problems 3 & 4 to this two dimensional example. You have the answer for a magnitude. What if the signal had a particular direction in the AB plane?

11) In problem 3 your simulations did not have enough throws to calculate the significance. How many throws would be needed for the simulated background to accurately determine the significance in this case?