Winter Quarter 2019 – UCSB Physics 129L
Homework 4
Due Friday, February 15, 5 pm

Read carefully the instructions on the website on how to prepare your homework for turning it in to the TAs.
If your last name starts with A through N, send the homework to Jenny. Otherwise send it to Francesco.
The emails of the TAs are on the website.
Put the instructor in cc to the email.

Make sure to have your rpi updated to pick up the latest/greatest example programs from class.

Exercise 1

The binary file
`/home/pi/physrpi/campagnari/python/dataSet.npy`
stores the value of a 1D np.array x which I saved with this python command:
`np.save("dataSet.npy", x)`
Write a python program to read the file to recreate the x array. Then make a histogram of x and display it. Make sensible choices about the binning and the axes (linear? semilog? log-log?). How would you want to display these data? Make sure to label the horizontal and vertical axes. Also, include the statistics box. To that end you can use the code in
`/home/pi/physrpi/campagnari/python/ccHist.py`.
Hint: for the npy files, look at
https://docs.scipy.org/doc/numpy-1.13.0/reference/routines.io.html
In this case the binary `.npy` file is 3 times smaller than the equivalent `.txt` file.

Exercise 2

In class we went over an example of calculating the p-value for the background in a Poisson counting experiment. The code is
`/home/pi/physrpi/campagnari/python/genPoisson.py`
Now imagine that there is an uncertainty on the background. This needs to

be accounted for. Here is how we can do this.
Instead of "throwing" N toy experiments, all with the same (Poisson) mean, we will let the mean of the Poissonian from which we draw the counts fluctuate from one toy experiment to the next. The fluctuations will reflect the uncertainty on the bakground. This is the point were things can become a bit hazy, because we need to assign a pdf to the mean background. In most cases the uncertainty on the background is systematic in nature, and we do not really know its pdf. Nevertheless we continue.....
A common procedure is to assume a Gaussian pdf for the mean background. Suppose we say that the expected background is $b_0 \pm \delta b$. We proceed as follows:

- Pick b fron a Gaussian of mean b_0 and $\sigma = \delta b$;
- Pick a number of counts n_{counts} from a Poissonian of mean b;
- Repeat N times.

You have to be careful because negative values of b are meaningless. (Poissonians of negative means do not make sense). Therefore if you find $b < 0$ you should discard it and pick a different one. This means that in practice b is from a truncated Gaussian.
To get away from using truncated Gaussians, which can be problematic, people often use LogNormal distributions, which can never go negative. For comparisons of Gaussians and LogNormals see
`/home/pi/physrpi/campagnari/python/compareLogNormal.py`
Play around with the (hardcoded) parameters to get a feeling for what you have, and look at the code to see how it is to be used.
The exercise then consists of the following:

- Start from `genPoisson.py`;
- Give it a different name;
- Add a switch `-u` to specify the uncertainty on the mean of the Poisson, *e.g.*, `-u 1.1` will set the uncertainty to 1.1;
- Add a (boolean) switch, such that if you specify `-g` it will use a Gaussian and if you do not specify anything it will use a LogNormal. (See `exampleOfCommandInput.py` for how to set up a boolean switch);

- Draw the mean of the Poisson from the Gaussian or LogNormal;
- Draw the number of counts n_{counts} from the Poisson;
- Repeat N times;
- Calculate the p-value and make a plot (this is already done for you in `genPoisson.py`).

Exercise 3

Perform a MC integration of

$$\int_2^4 dy \int_0^1 dx \;\; f(x,y)$$

with $f(x,y) = (x+2y)(x+y)$
Use 100K MC points. (Note: the analytic result is 47).

Exercise 4

Rewrite $f(x,y)$ from Exercise 3 as $f(x,y) = g(x,y)p(x,y)$ where $p(x,y) = (x+y)/7$ and $g(x,y) = 7(x+2y)$. Note that $p(x,y)$ is a properly normalized pdf over the interval of integration.
Generate a Markov Chain of (x,y) pairs according to $p(x,y)$. Use a symmetric proposal function, pick something reasonable, and use 10% of the chain for "burn in". Then use this chain to calculate the integral of Exercise 3. Use a 100K long chain.

Exercise 5

Consider the following function

$$f(x) = \int_0^{+\infty} e^{-(x+y)}(x+y)^N \; G(y \mid \mu, \sigma) dy$$

where $G(y \mid \mu, \sigma)$ is a Gaussian pdf for y of mean μ and standard deviation σ, properly normalized so that its integral from 0 to ∞ is 1. Take $N = 5$, $\mu = 3$, $\sigma = 0.5$.
Plot an approximation of $f(x)$ between -3 and 15. This effectively means

calculating many integrals over y, (say 100?) each at a fixed value of x, and then plotting these integrals vs. x.
We can do these integrals with the MC method:

- For each fixed value of x, pick many (say 1000?) values of y according to a truncated Gaussian distribution $G(y|\mu, \sigma)$
- For each y calculate $g(y) = e^{-(x+y)}(x + y)^N$;
- Add alll the $g(y)$ and divide by the number of y-picks at this x.

We will encounter this type of functions in the near future.