PHYS 509C Assignment 1

Callum McCracken, 20334298

September 22, 2022

Code for this assignment is here:

`https://github.com/callum-mccracken/PHYS-509C-A1`

It's in a bit of a strange format since I make it write the LaTeX file that I use for making the document you're reading, but here are the highlights:

- Open the file with `numpy.loadtxt()`
- Get the mean with `numpy.mean()`
- Get the standard deviation with `numpy.std()`
- Get the correlation coefficient with `numpy.corrcoef()`
- Get the skew with `scipy.stats.skew()`
- Use `scipy.stats.chi2.pdf()` for the chi-squared PDF
- Integrate using `scipy.integrate.quad()`

1 Here are three easy applications of Bayes' theorem:

A. Suppose the rate R of visible galactic supernovae is unknown but that supernovae follow a Poisson distribution. In the past 10 centuries astronomers have observed four supernovae in our galaxy. Assuming a uniform prior for the rate R, use Bayes' theorem to calculate the probability distribution for R. Now repeat the calculation, assuming this time that the prior for R is uniform in $\log_{10}(R)$ (i.e. it's equally probable that the true rate is between 0.02 and 0.2 as it is that it is between 0.2 and 2.0). Plot the resulting posterior probability distribution for R in both cases.

put plots here

B. Measurements are drawn from a uniform distribution spanning the interval (0, m). The probability of getting a measurement outside of this range is zero. The endpoint m is not well-known, but a prior experiment yields a Gaussian prior of m = 3 +/- 1. You take three measurements, getting values of 2.5, 3.1, and 2.9. Use Bayes' theorem to calculate and plot the new probability distribution for m.

put plot here

C. Suppose that an unmanned rocket is being launched, and that at the time of the launch a certain electronic component is either functioning or not functioning. In the control centre there is a warning light that is not completely reliable. If the electronic component is not functioning, the warning light goes on with probability 1/2; if the component is functioning, the warning light goes on with probability 1/3. At the time of launch, the operator looks at the light and must decide whether to abort the launch. If she aborts the launch when the component is functioning well, she wastes \$2M. If she doesn't abort the launch but the component has failed, she wastes \$5M. If she aborts the launch when the component is malfunctioning, or if she lets the launch proceed when the component is working normally, there is no cost. Suppose that the prior probability of the component failing is 2/5. During launch the warning light doesn't go on. From a costs standpoint, should she abort the mission or not? Compute and compare the expected cost of launching to the expected cost of aborting, given that the light didn't go on.

She should abort the mission. See code for now this was derived.

2 COVID-19 Study

- Sample size: 3330 people in Santa Clara County, California
- $N_{+,t} = 50$ positive test results in test group
- Control group 1: 3324 people, definitely negative
- $N_{+,c1} = 16$ positives in control 1
- Control group 2: 157 people, definitely negative
- $N_{+,c2} = 130$ positives in control 2 (27 false negatives)

Calculate the Bayesian 95% central interval on the fraction of people in Santa Clara County who actually had antibodies for COVID-19, marginalizing over the false positive and false negative rates. Assume flat priors on all parameters.

Submit a plot of the posterior distribution for the true incidence rate as well as your code or calculation.

3 CO_2 Meter

Assume that the data follows an exponential plateau, approaching some steady state value C.

Do a maximum likelihood fit for this level, and determine the uncertainty on it.

Note that you have not been given the uncertainties on the measured values—instead, assume that all measurements have the same uncertainty level, and fit for it as one of the parameters in your fit.

Submit your code, the functional form you fit, and your result for the steady state value (with uncertainty).

What uncertainty value per data point did you get?

How much of that uncertainty can be attributed to the time binning (measurements are only reported to the nearest minute)?

4 Retirement investments.

A. The percentage yield on an investment has a Gaussian distribution with mean of 8% and standard deviation (SD) of 15%. (A yield of 8% would mean the amount of money increases by a factor of 1.08 in a year. A yield of -8% would mean multiplying by 0.92 instead.) Suppose that you put \$3000 into a retirement account investing in this item on January 1st of every year, starting in 2018. What is the mean amount of money you will have in the account on Dec 31, 2047? Show a plot of the distribution of the amount of money on that date for 1000 trials of the "experiment". What is the SD? Hand in your code or equivalent documentation.

how to deal with dec 31st vs jan 1?

B. Suppose now that the retirement account contains three classes of investments: Canadian stocks, foreign stocks, and bonds. The yields on these three investments each vary randomly but with some correlation. Here is the yield information for each investment: $\mu_C = 8\%, \sigma_C = 15\%, \mu_F = 8\%, \sigma_F = 15\%, \mu_B = 5\%, \sigma_B = 7\%, \rho_{CF} = 0.50, \rho_{CB} = 0.20, \rho_{FB} = 0.05$. On January 1 of each year you put \$1000 into each class of investment. Show the distribution of the total amount of money in your account on Dec 31, 2047. What are the mean and SD?

C. Now suppose we add a procedure called "rebalancing". On January 1 of each year we contribute a total of \$3000 to the account, but at the same time we redistribute the total amount of money in the account evenly between the three investments. How does this change the total amount on Dec 31, 2047? Show a plot of the distribution, and report the mean and SD as well.

Commentary: You should find that although the total amount of money with rebalancing is slightly lower than in Part B, the SD is significantly smaller. The rebalancing procedure effectively forces you to sell investments when they're high and buy more of an investment when it's low. This effect is even more important when one accounts for the fact that the yield in one year is not completely independent of the yield in the next year—there is a small correlation that decreases with time which is not modelled in this problem. The net effect of this correlation is often that periods of high yield tend to be followed by periods of lower yield, and vice versa.

Retirement planning is in general a matter of optimizing the allocations to multiple types of investments in such a way as to yield the maximum probability of having ”enough”. High yield is of course good, but so is low variance. Having twice as much money as you need is a very different beast than having half as much as you need!

Ideally what you would want would be high yield investments with as little correlation as possible, so as to lower the variance. An investment strategy only slightly modified from the basic one in this problem, implemented with passively managed index funds, will generally outperform most actively management portfolios!