PHYS 509C Assignment 5

Callum McCracken, 20334298

December 10, 2022

Code for this assignment is here:
`https://github.com/callum-mccracken/PHYS-509C-A5`

1 t Distribution

If a light object with mass m is placed on a scale with “fat tails”, the reading on the scale follows a Student’s t-distribution with 3 degrees of freedom centered at m with a standard deviation of $1\,\mathrm{mg}$.

A. Write down the likelihood function for obtaining a reading of x milligrams on this scale.

We know the standard t-distribution has mean 0 and variance $\frac{N}{N-2}$.

$$f(t|N) = \frac{\Gamma\left(\frac{N+1}{2}\right)}{\sqrt{N\pi}\Gamma\left(\frac{N}{2}\right)}\left(\frac{1}{1+\frac{t^2}{N}}\right)^{\frac{N+1}{2}}$$

So if we have $N = 3$,

$$f(t|3) = \frac{2}{\sqrt{3}\pi}\frac{1}{\left(1+\frac{t^2}{3}\right)^2}$$

And then if we want a mean of m and a variance of 1, we must translate / scale: re-label $x = s(t-m)$ where $s = \sqrt{\frac{N}{N-2}} = \sqrt{3}$.

$$\begin{aligned} P(x) &= sf(s(x-m)) \\ &= \sqrt{3}\frac{2}{\sqrt{3}\pi}\frac{1}{\left(1+\frac{(\sqrt{3}(x-m))^2}{3}\right)^2} \\ &= \frac{2}{\pi}\frac{1}{\left(1+(x-m)^2\right)^2} \end{aligned}$$

(in the first line the factor of s out front comes from normalization)

B. You use this scale to measure an object and obtain a reading of 0.5 mg. Assuming a flat prior on m, calculate a Bayesian 90% upper limit on the object's mass.

First, calculate the posterior:

Likelihood: $P(x|m)$ (t-distribution from earlier)

Prior: $P(m) = A$ for $m \geq 0$ (some constant, flat prior), else 0 (physically, mass can't be negative).

For $m < 0$, the prior means $P(m|x = 0.5) = 0$. Otherwise,

$$\begin{aligned}
P(m|x = 0.5) &= \frac{P(x = 0.5|m)P(m)}{\int P(x = 0.5|m)P(m)dm} \\
&= \frac{P(x = 0.5|m)A}{\int P(x = 0.5|m)Adm} \\
&= \frac{P(x = 0.5|m)}{\int P(x = 0.5|m)dm}
\end{aligned}$$

We want to find the a such that

$$\begin{aligned}
0.9 &= \int_0^a P(m|x = 0.5)dm \\
&= \int_0^a \frac{2}{\pi} \frac{1}{(1 + (0.5 - m)^2)^2} dm
\end{aligned}$$

That's going to get messy analytically, see the code for details but numerically we find $a = 1.592$.

C. Now calculate the Feldman-Cousins 90% confidence interval on m for a reading of 0.5 mg.

This one was done exclusively through code, and we get $a = 1.843$.

2 Chi-squared fits with systematics.

The file data.txt contains 100 pairs of x, y values you wish to fit to a straight line. The nominal uncertainty in the y measurements is 0.5, while the x values are known perfectly. However, there are clearly outliers in the data.

A. Fit the data to a straight line using the "Tukey's biweight" version of an M-estimator, and report the slope and intercept that you get.

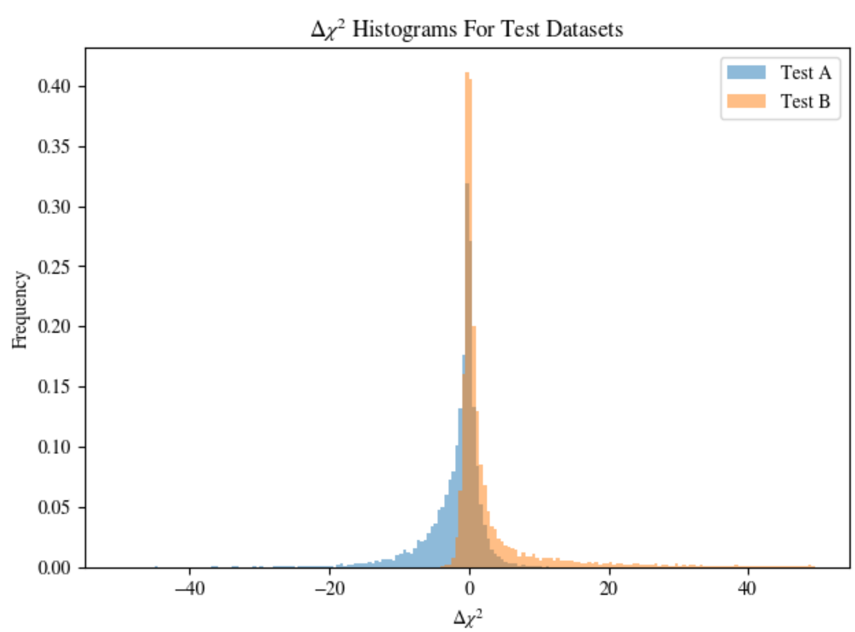
See the code for implementation details, we get $m = 0.50073, b = 4.80973$.

B. Using those best-fit values, calculate the residuals between the data and the fit. Then use these residuals in the bootstrap method to obtain uncertainty estimates for the fit you did in Part A. For the bootstrap assume that the residuals are independent of x.

Again, see code for details. $\sigma_m = 0.00268, \sigma_b = 0.15696$.

3 In this problem you will use the delta(chi2) method to form a test statistic to discriminate between two types of events (let's call them type A and type B). See Lecture 17, slide 20 for the definition of this test statistic. You are given two large training sets of type A events and type B events. For each event two quantities have been measured: x (first column in the file) and y (second column), and there are 10,000 events in each training set. Use these training sets to define a delta(chi2) statistic analogous to what is shown on the slide referred to above. You are also given two testing sets of type A and type B events. Apply your delta(chi2) statistic to each testing set and plot the distributions of the test statistic for each type of events on the same plot. For a given event with x=2.5, y=-0.5, what fraction of type A events have a higher value of this test statistic? What fraction of type B events have a higher value? Given your results, can you calculate the probability that this particular event is of type A?

- Here's that plot (see code for how it was made):



- For a given event with x=2.5, y=-0.5, what fraction of type A events have a higher value of this test statistic? What fraction of type B events have a higher value?

$$f_A = 0.0823, f_B = 0.3099 \text{ (again, see the code for details).}$$

- Given your results, can you calculate the probability that this particular event is of type A?

  No, since we don't know how likely we are to sample from type A vs type B – e.g. if you have a 10% chance of seeing this event in a sample of all As, and A happens once a day, and you have 1% chance of seeing that in a sample of all Bs but B happens once a second, then if you see a random event (could be type A or B), it's more likely for the event you saw to be a B even though you have a higher f_A.