DATA 221

Trimble/Nussbaum

Due: Thursday 2023-01-12 midnight

Reading: MacKay Chapter 2 and 3

1. Henry Newson reported a series of measurements of the decay rates of ^{17}F produced by deuteron bombardment of oxygen (Henry W. Newson. "The Radioactivity Induced in Oxygen by Deuteron Bombardment." Phys. Rev. 48, 790 (1935) doi:10.1103/PhysRev.48.790) procuding the following measurements of decay rate as a function of time since the accelerator was turned off.

time (min)	Decay rate (arbitrary)
0.161	87.1
0.578	69.7
1.113	51.5
1.584	38.3
2.226	25.3
3.061	15.8
4.324	7.6
6.229	2.5

   We are interested in estimating the decay constant.

   If we belive the function should be

   $$r(t) = Ae^{\frac{-\ln(2)t}{t_{1/2}}} + B$$

   where A, $t_{1/2}$, and B are constants to be determined, how should you weight the errors at different points? find the maximum a posteriori value (the value corresponding to the maximum of the posterior density) for $t_{1/2}$ by optimization.

2. [Simplifed version of Problem 3.3 form Mackay, p 47] "Unstable particles are emitted from a source and decay at a distance x, a real number that has an exponential probability distribution with characteristic length λ." In other words, x is exponentially distributed. Let us imagine a magical counter that can measure decays near x=0 and can measure decays infinitely far away from the source. The counter observes six events, $x_n = 1.5, 2, 3, 4, 5, 12$ and is certain no events occurred outside of those six.

   Using a prior density that is proportional to $\frac{d\lambda}{\lambda}$ (which is the appropriate prior density for a scale parameter), find

   (a) The most likely value (MAP estimate) for λ?

   (b) A 95% confidence interval for λ.

   Plot the posterior density for λ.

3. [Exercise 3.1 from MacKay p.47]

   A die is selected at random from three twenty-faced dice on which the symbols 1–10 are written with nonuniform frequency as follows.

Symbol	1	2	3	4	5	6	7	8	9	10
NumberoffacesofdieA	6	4	3	2	1	1	1	1	1	0
NumberoffacesofdieB	3	3	2	2	2	2	2	2	1	1
NumberoffacesofdieE	2	2	2	2	2	2	2	2	2	2

   A randomly chosen die, from A or B is rolled 7 times, with the following outcomes: 5, 3, 9, 3, 8, 4, 7.

4. A randomly chosen die from these three is rolled 8 times, with the following outcomes: 5, 3, 9, 3, 8, 4, 7, 10.

   What are the probabilities that the die is die A, B, or E?

5. What does the zero probability for die A to return a 10 mean for inferences ?

   How many rolls on average would you need to establish 99:1 confidence between B and E?

   Hint: there is a theoretical answer (sums over things) and an attack by simulation.

6. Given a distribution of bigrams in English text, the distribution of initial letter given final and that of the final letter given the initial are different.

   Reproduce the three figures in Fig 2.1 and 2.3 in MacKay (Hinton diagrams or heatmaps are fine) using the novella Carmilla by Joseph Sheridan Le Fanu `https://www.gutenberg.org/files/` . Count the one-letter tokens. Split the text into two-letter tokens and count them. Make a two-dimensional visualization of the bigrams, the row-marginal, and column-marginal variants.