

DATA 221

Homework 1 (rev 2)

W. Trimble

Due: Friday 2022-04-08

Reading: MacKay Chapter 2 and 3

1. [Problem 3.3 from 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 . Decay events can be observed only if they occur in a window extending from $x = 1$ cm to $x = 20$ cm. N decays are observed at locations x_1, \dots, x_N . Plot the posterior probability density of λ for the outcomes of a single event at $x_n = 5$ and for six events at $x_n = 1.5, 2, 3, 4, 5, 12$
2. Let us imagine a magical counter that can measure decays near $x=0$ and can measure decays infinitely far away from the source (the $x=1$ cm and $x=20$ cm window around our observations is gone). Now our counter has observed the same six events, $x_n = 1.5, 2, 3, 4, 5, 12$ but is certain no events occurred outside of those six.

Using a prior density that is proportional to $\frac{d\lambda}{\lambda}$, find

- (a) The most likely value (MAP estimate) for λ ?
- (b) A 95% confidence interval for λ .

Hint: this is a numerical integration problem.

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. Reproduce the three figures in Fig 2.1 and 2.3 in MacKay by downloading The Frequently Asked Questions Manual for Linux (or any other text with more than 30k words), splitting it into two-letter tokens, and counting them. Make a two-dimensional visualization of the bigrams, the row-marginal, and column-marginal variants. This graph in MacKay is called a "Hinton diagram" but heatmaps or other symbol- or color- encodings to communicate the 2d-histogram are acceptable.
7. Perform an eigenvalue decomposition of one of the 27×27 letter occurrence probability matrices and find the eigenvector corresponding to the largest eigenvalue. This eigenvector is special. Plot the values of this eigenvector along with sums or probabilities from the bigram density.