Problem Set 2

D. Jason Koskinen
koskinen@nbi.ku.dk

Advanced Methods in Applied Statistics
Feb - Apr 2024

Photo by Howard Jackman

Info

- The submission is:
  - A write-up as a PDF document, which includes any plots, diagrams, tables, pictures, and explanations
  - In a separate "file", submit all code used to derive the results
    - Tarball, zipped directory, lots of individual files w/ self-explanatory titles, etc.
    - Do NOT include lines of code in your write-up. If results are dependent on coding choices then include those comments in the write-up.
- Include any original data files or how the data was accessed
  - If you use a internet scraping tool, note the date when you retrieved the data
  - If you can save the data to a file, do so and submit the data file. There is no need to change the format, e.g. HTML, XML, txt, JSON…

Problem 1 (3.5 pts. total)

- In 1974, Stephen Hawking predicted that black holes should radiate elementary particles and lose mass (evaporate)
  - This process is called Hawking radiation
  - The enthusiasts can find more information [here]
- Neutrinos could be one of the radiated particle species
  - Though this has not been experimentally confirmed yet, future detectors might be sensitive to neutrinos from evaporating black holes (EBHs)
  - The energy distribution of neutrinos depends on the EBH mass
  - In this problem, we will be given neutrino energy data (i.e. what the signal would look like if it were detected) and asked to deduce the black hole mass

Problem 1a (0.5 pt)

- The **unnormalised** PDF of neutrino energies E_ν is

$$f(E_\nu) = \frac{e^{E_\nu/T} - 1}{e^{E_\nu/T} + 3} \cdot \frac{1}{e^{E_\nu/T} + 1}$$

  - T is the black hole temperature, which is related to the black hole mass M_{BH} via $T \simeq \dfrac{1.057 \times 10^{13}}{M_{\mathrm{BH}}}$. Here, the black hole mass is expressed in grams, and the resulting temperature is in GeV (which is also the units for neutrino energy in this problem).

- Plot the **normalised** PDFs of neutrino energies for black hole masses of $2.5 \cdot 10^{11}$ g; $4 \cdot 10^{11}$ g; and $9 \cdot 10^{11}$ g.

Problem 1b (0.5 pt)

- The simulated neutrino data, assuming a certain BH mass, is available at https://www.nbi.dk/~koskinen/Teaching/data/neutrino_energies.csv
- Read the data and make a histogram of the simulated neutrino energies.
    - Put the histogram on the same plot as the PDFs from exercise 1a, and make sure they show up on the same scale
    - Do it by either by showing the density of the histogrammed data, or by rescaling the normalised PDF according to the number of data points and the histogram bin width
- Just by looking at the plot (w/o any fitting or minimizing or chi-squared tests) what is your 'eye-ball best guess' for the true value of the black hole mass?

Problem 1c (1 pt)

- Make a 1D raster scan of the unbinned ln-likelihood of the neutrino energy data for a reasonable range of the black hole masses
  - Use your intuition from Problem 1b to define what is 'reasonable'
  - The end result should be a plot of the LLH vs BH mass
- What is the maximum likelihood estimate of the BH mass based on your scan?
  - Produce the PDF with the black hole mass equal to the MLE and overlay it with the histogrammed data
- What are the values of the LLH at your MLE estimate and the three proposed BH masses from Problem 1b?

Problem 1d (1.5 pts)

- Use any of the minimiser packages (Minuit, scipy, or any other minimiser of your choice) to find the black hole mass that best describes the neutrino data
  - I.e., repeat problem 1c, but using a minimiser instead of the raster scan
- Discuss how well the MLE from your raster scan and the MLE from your minimiser agree.
- What is the uncertainty on the black hole mass estimate?

Problem 2 (2.5 points)

- Bayesian statistics
  - The likelihood will be the PDF for a gamma function $\Gamma(\alpha, \beta)$, which is defined as:

$$f(x|\alpha,\beta) = \frac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$$

  - For this problem $\alpha = 2$ and $\beta = 2$
  - The prior will be a step function defined as:

$$f(x) = \begin{cases} 1/N, & 0 < x \leq 1.6 \\ 2.66/N, & 1.6 < x \leq 2.1 \\ 0, & x \leq 0 \text{ or } x > 2.1 \end{cases}$$

for a value of N which normalizes $f(x)$

Problem 2 continued

- Plot the likelihood, prior, and bayesian posterior on the same plot over the range of $0 < x < 4$
  - All 3 functions should be properly normalized
  - All 3 plots should use a different color scheme than the Matplotlib default. I.e. no usage of blue, orange, or green even if you aren't using matplotlib for plotting
  - All axes should have correct labels and the plot should include an appropriate legend
- Using the bayesian posterior, what is the most likely value of x, i.e. $\hat{x}$?
- If $g(x)$ is the bayesian posterior, what is the value of $g(\hat{x})$?

Problem 3 (4 points)

- Inspired by the Anatoly Dneprov short story "Crabs on the Island" we will look at population evolution. Imagine a perfectly circular island with a 5 km radius inhabited by metal crabs, each with a mass of 1 kg. The metal crabs only consume other metal crabs.
- All crabs move in individual random directions (0-2π), once per day until:
  - they move 200 meters, or
  - they reach the edge of the island
- At the end of the day, any metal crabs within a certain distance of another crab will try and eat each other or defend from being eaten. Details are on later slides.

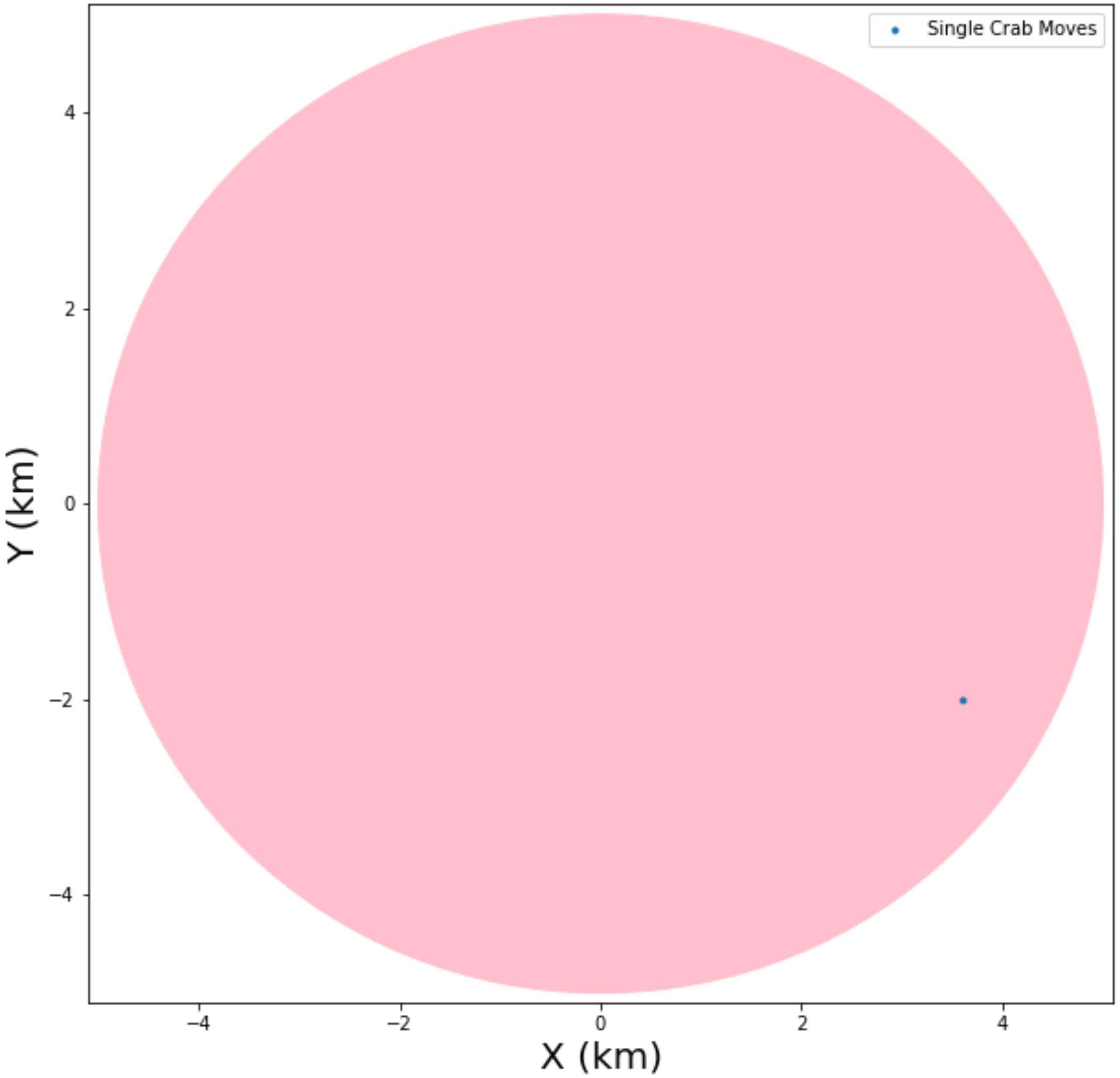
*outdoorsonparade.com

Crab Movement

- Consider the crabs to move instantaneously (for the simplicity of this Monte Carlo simulation).
- If a crab reaches the island edge they will continue to move the next day, i.e. they do not stop forever upon reaching the island's edge.
- A crab will move up to 200 m if possible. For example, if it is 50.2 m away from the island boundary and choses a direction directly towards the island boundary, the crab will move 50.2 m that day.
- The center of the circular island is at (0 km, 0 km).

Problem 3a (0.5 points)

- For simplicity, we consider only a single crab on the island for this part of the problem. No other cannibal metal crabs… yet.
- Show a plot of a single crab starting at (3.6 km, -2.0 km) and moving for 200 days.
  - Put a single point for each stopping position of the crab
  - Example figure ONLY includes the starting point, i.e. day=0
  - There should be a total of 200 points, or 201 if you include the starting position.



Problem 3b (1 point)

- For simplicity, we continue considering only a single crab on the island, and simulate the total distance travelled before reaching the edge of the island.
- Run 501 pseudo-experiments with a single crab starting at the location (3.6 km, -2.0 km) and make a histogram of the 501 distances the crab travels before arriving at the edge of the island.
  - The histogram x-axis range axis range should go from 0-200 km, and the bin widths should be 2 km.
  - This is the total travel distance of the crab, and NOT the straight line distance between the starting point and the edge of the island where the crab ultimately stops.

Problem 3 Crab Battle Details

- Crabs that end the day within 175 m of each other will battle. The larger crab will consume the smaller crab with odds of $\frac{M^2_{larger}}{1kg} : \frac{M^2_{smaller}}{1kg}$.
  - For example, a 5 kg crab eating a 2 kg crab has 25:4 odds, i.e. 25/(25+4)=86.2%. The smaller crab has a 4/(25+4)=13.8% of surviving, but might get eaten the next day (life is difficult for small crabs).
  - If the smaller crab defends, then both crabs continue to exist.
  - Crabs of equal mass have 1:1 odds of one crab consuming the other. It is arbitrary which one is considered 'larger'.
- If the larger crab eats the smaller crab its mass increases equal to the smaller crab it consumed. The position on the island of the 'winning' crab remains the same as where it stopped for the day, despite the battle. In the case of the smaller crab surviving, then both crabs stay in the same position as where they stopped for the day.
- Although rare, multiple crabs can end their day within 175 m of multiple other crabs. In such scenarios, the battles occur in order of shortest distance to largest distance.

Problem 3 Crab Battle Tips

- Make sure that the sum of metal crab mass on the island is always 20 kg. If it's ever not equal to 20 kg, then there's a bug in your code.
- Eaten crab cannot battle. If a crab is consumed, somehow keep track that it cannot battle anymore. There are many, many, many ways to do the 'book keeping' for this. For example,
    - Change the mass of the consumed crab to 0 and the position to -infinity.
    - Remove the eaten crab from the array or list after the battle in which they are eaten.
    - Include a data element in the crab array, list, or class which notes whether it has 1) battled already that day, and/or 2) whether it has been eaten.

Problem 3c (1.5 points)

- There are 20 metal crabs on the island with the starting positions provided in the file at https://www.nbi.dk/~koskinen/Teaching/data/CrabStartPositions.txt.
  - The first column is the x position and the second column is the y position.
  - Each row is the starting position of an individual crab.
- After 200 days:
  - What is the mostly likely number of individual crabs that remain alive?
  - What is the most likely mass of the largest crab?

Problem 3d (1 points)

- We will continue using the same 20 crab on the island with the starting positions provided in the file at https://www.nbi.dk/~koskinen/Teaching/data/CrabStartPositions.txt
- Show the distribution of days it takes until only 10 crabs remain alive
- What is the 1σ confidence interval on the number of days at which only 10 crabs remain?
  - If the distribution is asymmetric, then calculate the lower 1σ bound as the value where the lower tail contains (100%-68.27%)/2 of the distribution. Same thing for the upper 1σ bound calculated using the higher tail.
  - N.B. review the lecture *Parameter Estimation and Uncertainty* for a reminder about confidence intervals.