STDSR-2023-Assignment 2

BS-2 and BS-4

April 2023

Task 1 (for both BS2 and BS4)

In a research program on human health risk from recreational contact with water contaminated with pathogenic microbiological material, the National Institute of Water and Atmosphere (NIWA) instituted a study to determine the quality of NZ stream water at a variety of catchment types. This study is documented in McBride et al. (2002) where $n = 116$ one-liter water samples from sites identified as having a heavy environmental impact from birds (seagulls) and waterfowl. Out of these samples, $x = 17$ samples contained Giardia cysts. Let θ denote the true probability that a one-liter water sample from this type of site contains Giardia cysts.

1. What is the conditional distribution of X, the number of samples containing Giardia cysts, given θ?
2. Before the experiment, the NIWA scientists elicited that the expected value of θ is 0.2 with a standard deviation of 0.16. Determine the parameters α and β of a Beta prior distribution for θ with this prior mean and standard deviation. (Round α and β to the nearest integer).
3. Find the posterior distribution of θ and summarize it by its posterior mean and standard deviation.
4. Plot the prior, posterior and normalized likelihood.
5. Find the posterior probability that $\theta < 0.1$.
6. Find a central 95% posterior credible interval for θ.

Task 2 (for BS2 only)

(a) Read the tutorial on Thompson Sampling (moodle-link)

(b) Implementation

  i. One of the Basic methods (greedy, UCB1)

  ii. Implement Thompson Sampling for Bernoulli reward

  iii. Implement both (UCB and TS) in case of Gaussian rewards

(c) Simulation

  i. Apply Thompson Sampling method for Correlated Travel Times (see the Example 4.2 on page 22 of the Tutorial)

  ii. Alternatively, for the simulation you can use one of the examples from the Section 7 (page 48 of the Tutorial)

(d) Write a report (1-2 pages) + provide the source code (moodle or github).

Task 2 (for BS4 only)

(a) Read the document about Simulated Annealing (moodle-link)

(b) Download the location data for cities in Russia;

(c) Find the optimal traveling salesman path using SA for 30 most populated cities (*remember that the dataset contains geodesic coordinates, calculating the distance would require to apply a suitable transformation*);

(d) For your selected annealing schedule, track the speed of convergence for three different values of the annealing rate (*try fast cooling, slow cooling, and some middle value*). Compare the optimization result;

(e) Write a report about your findings.

(f) Produce a visualization of your optimization process. Use documentation as a guideline. (*Bonus points if you overlay the travelling path with the country outline*)

(g) Upload your code to github, write a short description, and provide the link to repository in the report.