Laboratory Session : April 20, 2023

Exercises due on : May 7, 2023

Exercise 1

- study the binomial inference for a study that reports $y = 7$ successes in $n = 20$ independent trial. Assume the following priors:

- a uniform distribution

- a Jeffrey's prior

- a step function:

$$g(\pi) = \begin{cases} \pi & \pi \leq 0.2 \\ 0.2 & 0.2 < \pi \leq 0.3 \\ 0.5 - \pi & 0.3 < \pi \leq 0.5 \\ 0 & 0.5 < \pi \end{cases}$$

- plot the posterior distribution and summerize the results computing the first two moments
- compute a 95% credibility interval and give the results in a summary table
- draw the limits on the plot of the posterior distribution

Exercise 2

- Giardia cysts[1] is a parassite that can contaminate food, water, and surfaces, and they can cause giardiasis when swallowed in this infective stage of their life cycle. Infection occurs when a person swallows Giardia cysts from contaminated water, food, hands, surfaces, or objects.
- a group of researchers, working for a Human Health Department, work for determining the quality of stream water
- they take n = 116 samples of one-liter water from sites identified to have a heavy environmental impact from birds and waterflow.
- out of these samples, $y = 17$ samples contained Giard cystis
- assuming π as the probability that one-liter water sample contains Giardia cystis, find the posterior distribution for π

- assuming a uniform prior distribution

- assuming a Beta $(1, 4)$ prior

- plot the posterior distribution and summerize the results computing the first two moments
- find a normal approximation for the posterior $g\left(\pi \mid y\right)$
- compute a 95% credibility interval both for the original posterior and for the normal approximation, giving the results in a summary table
- add the limits on the plot of the posterior distributions

Exercise 3

- A coin is flipped $n = 30$ times with the following outcomes:

  `T, T, T, T, T, H, T, T, H, H, T, T, H, H, H, T, H, T, H, T, H, H, T, H, T, H, T, H, H, H`

a) Assuming a flat prior, and a beta prior, plot the likelihood, prior and posterior distributions for the data set.

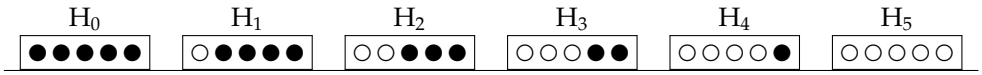
b) Evaluate the most probable value for the coin probability p and, integrating the posterior probability distribution, give an estimate for a 95% credibility interval.

c) Repeat the same analysis assuming a sequential analysis of the data [1]. Show how the most probable value and the credibility interval change as a function of the number of coin tosses (i.e. from 1 to 30).

d) Do you get a different result, by analyzing the data sequentially with respect to a one-step analysis (i.e. considering all the data as a whole) ?

Exercise 4 - Six Boxes Toy Model : inference

- The six boxes toy model is described in reference [2].
- Labeling the boxes as follows:



- write a program in R that performs a simulation of the process

1) after selecting a random box
2) make random sampling from the box
3) prins on the standard output the probability of selecting each box
4) plot the probability for each box as a function of the number of trial

References

[1] `https://www.cdc.gov/dpdx/giardiasis/index.html`

[2] G. D'Agostini, *Probability, propensity and probabilities of propensities (and of probabilities)*, `https://arxiv.org/pdf/1612.05292.pdf`
G. D'Agostini, *More lessons form the six box toy experiment*, `https://arxiv.org/pdf/1701.01143.pdf`

[1] Hint: treat the posterior distribution obtained in step j as prior distribution for step $j + 1$