# Bayesian Reasoning and Inference: Assignment

Due date: 11:59pm Friday 3rd May

Students are encouraged to consult each other on the assignment questions, although all submitted work must reflect an individual student's contribution. Note that *all programs* written as part of this assignment must be submitted also. Attempt all questions.

**QUESTION 1**

A monkey called Bungles has been trained for a sideshow at the circus to play a gambling game with a pea hidden under one of three cups which are then shuffled about. However, Bungles takes a somewhat haphazard attitude to his work, and 10% of the time he simply neglects to put a pea under any of the cups.

(a) While engaged in a betting game with Bungles, two of the cups at random are turned over and shown to be empty. Using Bayes theorem, what is the chance now of a pea under the final cup?

(b) Bungles has a sister Wingnut, who is even worse. In addition to forgetting the pea 10% of the time, she sometimes puts two peas under two separate cups 5% of the time. In a game with Wingnut, one cup is turned over and found to be empty. Using a Bayesian approach, what is the chance that the second cup will have a pea?

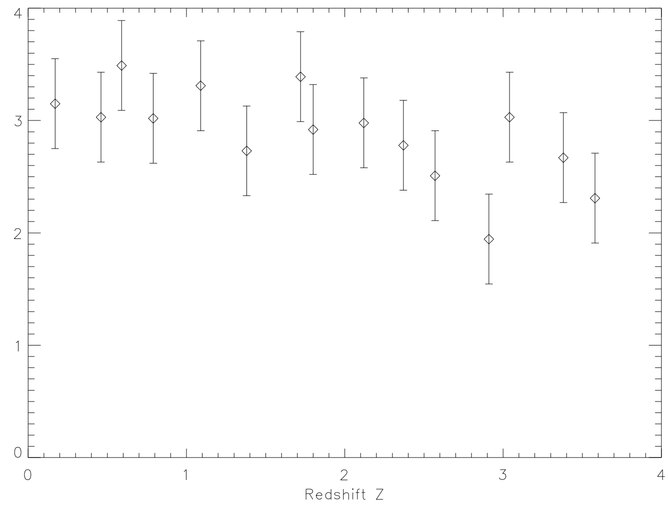
(c) Continuing the same game with Wingnut, the second cup is turned over and also found to be empty. What is the chance that the final cup will have a pea?

**QUESTION 2**

(a) After arriving on a tiny Caribbean island, you notice two cars in the airport taxi rank. The first has license plate number 7 and the second 42. Assuming taxis have license numbers starting from 1, what is the maximum likelihood estimate for the total number of carriages on the island? Plot the posterior probability for the total number N assuming a uniform prior.

**QUESTION 3**

A long time ago, in a galaxy far, far away, astronomers at Corellia University are on the verge of a discovery that could cause a disturbing lack of faith in their previous beliefs. Their experiment consists of careful measurement of the value of $\pi$ by observing circular objects (stars, globular clusters, galaxies) at various distances in space. Their data consists of the ratio of measurements of the circumference to the diameter as a function of Redshift distance Z, as given in the plot below:




(a) Examining these data, Professor Dot Antilles publishes her findings in *Monthly Notices of the Galactic Republic* concluding with the hypothesis $H_0$ that the data are consistent with random Gaussian noise with $\sigma$=0.75. However, a Padawan student at the Jedi University of Coruscant carefully re-analyses the data and claims a better fit with the model $H_1$ - that a signal $\pi_Z = \pi_0 + A \times Z$ has been observed, with the same noise. For this model $\pi_0$=3.14159 and A is the slope of the linear change with time. If true, this would sensationally imply that tiny Midi-Chlorians underpin the circle of life and have created an invisible field of energy that surrounds and penetrates us, binding the galaxy together! Designate this new model, called "The Force" as $H_1$ and use a Gaussian prior for A with zero mean and standard deviation $\delta$=0.2. Assuming a prior odds ratio of 1.0 (no preference for either model) compute the odds ratio for the two models $H_1/H_0$. From this, what is the probability that model $H_1$ is true based on the data. Should we believe in this new "Force"?

(b) Professor Durran at Kessel State University is furious to see Antilles publication, having been working on something similar for years. Anger turns to hate, and hate turns to peer review. He claims that Antilles team have overestimated their uncertainties. He believes the true standard deviation of the errors is $\sigma$=0.45? Re-do the computation assuming these smaller errors. What has happened to your confidence? Why?

(c) Assuming model $H_1$ and the original errors $\sigma$=0.75, compute and plot the likelihood distribution for the free parameter A over the interval 1 to -1. Using the Gaussian prior above with $\delta$=0.2, plot also the posterior for A.

(d) Using results obtained so far, what was the Occam factor which applies to the marginalization operation in part (a)

Hint: This question makes use of the Gaussian Integral:

$$\int_{-\infty}^{\infty} \exp\left(c_0 + c_1 x + c_2 x^2\right)\, dx = \sqrt{\frac{\pi}{-c_2}} \exp\left(c_0 - \frac{c_1^2}{4c_2}\right)$$

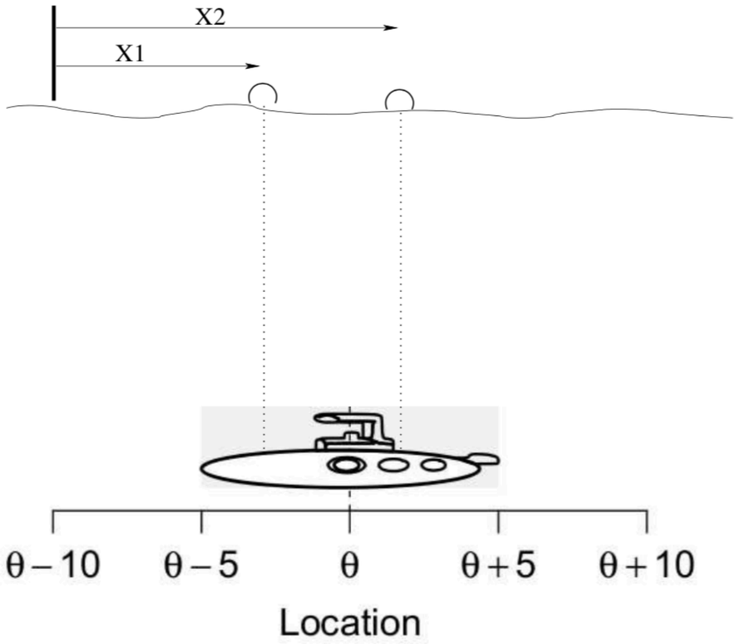
Data for $\pi_Z$ are in redshift bins centered on Z values:

Z = [0.17, 0.46, 0.59, 0.79, 1.09, 1.38, 1.72, 1.80, 2.12, 2.37, 2.57, 2.91, 3.04, 3.38, 3.58]

$\pi_Z$ =[3.15, 3.03, 3.49, 3.02, 3.31, 2.73, 3.39, 2.92, 2.98, 2.78, 2.51, 1.94, 3.03, 2.67, 2.31]

**QUESTION 4**

A research submersible has lost contact with its surface support vessel. Led by Captain Neyman, the crew are desperate to drop a rescue line, which they wish to fall as close as possible to the location of the submersible exit hatch. The submarine is exactly 10 m in total length, and the location of the hatch, θ, is exactly halfway along the hull. The only evidence available to the surface rescue team is the observation of bubbles which are known to originate randomly from any point along the entire length of the submarine with uniform probability. The support vessel measures these locations {x1, x2, ...} with respect to a fixed origin on the surface.




(a) With a single bubble observed at location x1 = 5 m, sketch the likelihood function for θ, the location of the hatch as a function of x.
(b) A second bubble appears at location x2 = 13 m. Now over-plot the new likelihood function for θ.
(c) While getting the rescue line ready, 4 more bubbles appear. Their locations are at x3, .. x6 = {7, 11, 10, 7}. Indicate on your plot the likelihood function that accounts for all six bubbles.
(d) Captain Neyman decides to use the Mean and Standard Deviation of the {x1, .. x6} as his estimator for the location and uncertainty of θ, however he is worried that his data set is too limited to get accurate estimates. Write down two example bootstrap resample distributions of the 6 known bubble locations that illustrate to Neyman how to do a bootstrap analysis.
(e) With time ticking away, Neyman thinks it better to wait for more bubbles so that the standard deviation statistic for θ becomes smaller. Noting that the four recent bubbles were all at locations between the original two x1, x2, what would you tell Neyman about how much better knowledge we have of θ from the last 4 bubbles?

THERE ARE NO MORE QUESTIONS