STDSR Assignment 2

Konovalova Oksana B21-DS-02

o.konovalova@innopolis.univesity

Task 1 is also made in *ipynb* file.

Task 1

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?

The conditional distribution of X, given θ, is a binomial distribution with parameters $n = 116$ and θ, denoted as $X \sim Bin(116, \theta)$.

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).

The prior distribution of θ is a Beta distribution with parameters α and β, denoted as $\theta \sim Beta(\alpha, \beta)$. The mean and standard deviation of the prior distribution can be expressed in terms of α and β as follows:

(1) $$E(\theta) = \frac{\alpha}{(\alpha+\beta)} = 0.2$$

(2) $$Var(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0.16^2$$

Find α and β:

(3) $\beta = 4\alpha$ from (1)

(3) → (1):

$$\frac{4\alpha^2}{(5\alpha^2)^2(5\alpha+1)} = 0.16^2$$

With using wolfram alpha, α equals:

Real solution

$\alpha \approx 1.01451$

Therefore, $\underline{\alpha = 1 \text{ and } \beta = 4}$

Answer: $\alpha = 1$ and $\beta = 4$

3. Find the posterior distribution of θ and summarize it by its posterior mean and standard deviation.

The posterior distribution of θ is also a Beta distribution, which can be expressed as:

$$\theta|X \sim Beta(\alpha + x, \beta + n - x),$$

where $x = 17$ is the observed number of samples containing Giardia cysts and $n = 116$ is the total number of samples.

The posterior mean and standard deviation of θ can be calculated as follows:

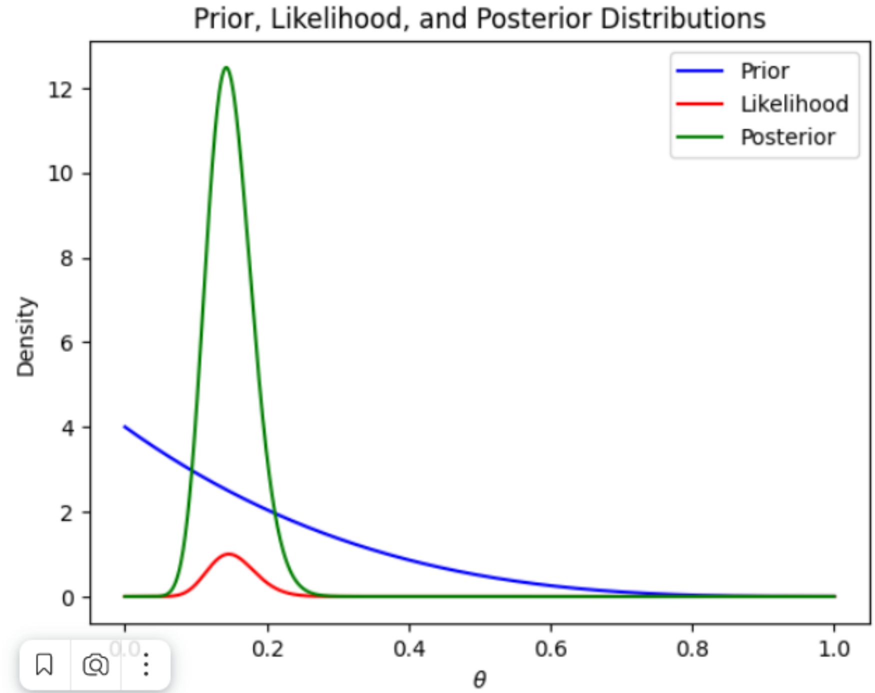
$$E(\theta|X) = \frac{\alpha + x}{\alpha + \beta + n} = \frac{1 + 17}{1 + 4 + 116} = 0.149$$

$$Var(\theta|X) = \frac{(\alpha + x)(\beta + n - x)}{(\alpha + \beta + n)^{2(\alpha+\beta+n+1)}} = \frac{(1 + 17)(4 + 116 - 3)}{(1 + 4 + 116)^{2(1+4+116+1)}}$$
$$= 0.001038$$

$$SD(\theta|X) = 0.0322$$

Answer: $E(\theta|X) = 0.149 \ and \ SD(\theta|X) = \ 0.0322$

4. Plot the prior, posterior and normalized likelihood.



5. Find the posterior probability that $\theta < 0.1$.

The posterior probability that $\theta < 0.1$ can be calculated as:

$$P(\theta < 0.1 \mid X) = pbeta(0.1, \alpha + x, \beta + n - x)$$
$$= pbeta(0.1, 18, 103) = 0.053$$

Answer: 0.053

6. Find a central 95% posterior credible interval for θ.

A central 95% posterior credible interval for θ can be calculated using the quantile function of the Beta distribution:

$$[qbeta(0.025, \alpha + x, \beta + n - x), qbeta(0.975, \alpha + x, \beta + n - x)] =$$
$$[qbeta(0.025, 18, 103), qbeta(0.975, 18, 103)] = [0.091, 0.217]$$

Answer: $[0.091, 0.217]$