# STDSR-2023-Assignment 2

Danis Alukaev
d.alukaev@innopolis.university
B19-DS-01

May 5, 2023

## Task 1

Let $n = 116$ be a number of one-liter water samples from sites identified as having a heavy environmental impact from birds, $x = 17$ be a number of samples contained Giardia cysts, and $\theta$ denote the true probability that a one-liter water sample from this type of site contains Giardia cysts.

**Problem 1.1.** What is the conditional probability of $X$, the number of samples containing Giardia cysts, given $\theta$?

*Solution.*

Primarily, let's note that in a given setting each one-liter water sample can either contain or not pathogenic microbiological material, which in a broader sense reflects successful or failed trial. Thus, the number of samples containing Giardia cysts $X$ could be summarized through binomial distribution (1) with the number of trials $n$ and probability $\theta$ of trial to be successful.

$$X \sim \text{Binomial}(n = 116, p = \theta) \tag{1}$$

Following the definition of binomial distribution the probability to observe $k$ samples with Giardia cysts out of $n = 116$ is defined by (2).

$$P(X = \hat{x}|\theta) = \binom{116}{\hat{x}} \theta^{\hat{x}}(1-\theta)^{116-\hat{x}} \tag{2}$$

**Problem 1.2.** Before the experiment, the NIWA scientists elicited that the expected value of $\theta$ is 0.2 with a standard deviation of 0.16. Determine the parameters of $\alpha$ and $\beta$ of a Beta prior distribution for $\theta$ with this prior mean and standard deviation. (Round $\alpha$ and $\beta$ to the nearest integer).

*Solution.*

Recall that for a Beta distribution $\theta \sim \text{Beta}(\alpha, \beta)$ the expected value and variance are given by (3) and (4) respectively.

$$E(\theta) = \frac{\alpha}{\alpha + \beta} \tag{3}$$

$$Var(\theta) = \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{4}$$

Therefore, the parameters $\alpha$ and $\beta$ of a distribution $\theta$ with a prior mean $E(\theta) = 0.2$ and standard deviation $\sigma = 0.16$ could be derived from a system of equations (5).

$$\begin{cases} \frac{\alpha}{\alpha+\beta} = 0.2 \\ \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0.16^2 \end{cases} \tag{5}$$

Solving (5) yields $\alpha = 1.05 \approx 1$ and $\beta = 4.2 \approx 4$. The prior distribution $P(\theta)$ is thus given by $\text{Beta}(1, 4)$.

**Problem 1.3.** Find the posterior distribution of $\theta$ and summarize it by its posterior mean and standard deviation.

*Solution.*

Following the Bayes' theorem (6) the posterior distribution can be expressed through the likelihood and prior distribution. Note that $P(X)$ is a constant and can be considered as a scaling factor.

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} \propto P(X|\theta)P(\theta) \tag{6}$$

For the likelihood $P(X|\theta) \sim \text{Binomial}(n = 116, p = \theta)$ (see problem 1.1) and prior distribution $P(\theta) \sim \text{Beta}(1, 4)$ (see problem 1.2) for posterior distribution holds (7).

$$\begin{aligned} P(\theta|X = 17) \propto P(X|\theta)P(\theta) &\propto \text{Binomial}(n = 116, p = \theta) \cdot \text{Beta}(1, 4) \\ &\propto \binom{116}{17} \theta^{17}(1-\theta)^{116-17} \frac{\Gamma(1+4)}{\Gamma(1)\Gamma(4)} \theta^{1-1}(1-\theta)^{4-1} \\ &\propto \theta^{18-1}(1-\theta)^{103-1} \end{aligned} \tag{7}$$

Note that the beta function is a scaling factor for a distribution to be valid. Thus, let's consider the obtained kernel $\theta^{18-1}(1-\theta)^{103-1}$, which appear to belong to another beta distribution with parameters $\alpha = 18$ and $\beta = 103$. Therefore, Beta distribution is a conjugate prior for binomial distribution and the posterior distribution of $\theta$ takes the form of (8).

$$P(\theta|X = 17) \sim \text{Beta}(18, 103) \tag{8}$$

The obtained Beta distribution is characterised by mean $\text{E}(\theta|X = 17) = 0.1488$, variance $\text{Var}(\theta|X = 17) = 0.001038$, and standard deviation $\sigma(\theta|X = 17) = 0.0322$.

**Problem 1.4.** Plot the prior, posterior and normalized likelihood.

*Solution.*

Figure 1 shows the prior, posterior, and normalised likelihood. The plot was generated using *scipy* and *matplotlib* packages in Python language. For more details, please refer to this notebook.

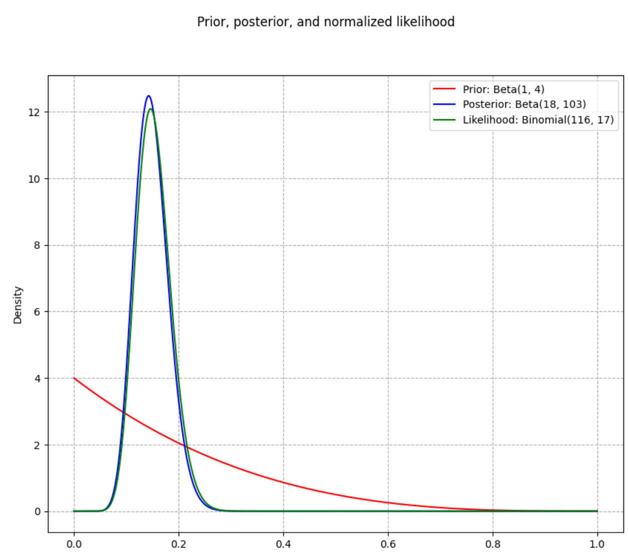


Figure 1: Prior, posterior, and normalised likelihood.

**Problem 1.5.** Find the posterior probability that $\theta < 0.1$.

*Solution.*

The posterior probability that $\theta < 0.1$ is 0.0531. It was computed in *scipy* package by using corresponding value of $\theta$ as argument for cumulative distribution function of beta distribution. For more details, please refer to this notebook.

**Problem 1.6.** Find a central 95% posterior credible interval for $\theta$.

*Solution.*

Central 95% posterior credible interval for $\theta$ is $[0.0914, 0.2171]$. It was computed in *scipy* package using percent point function of beta distribution. For more details, please refer to this notebook.

# Task 2

The traveling salesman problem was solved using simulated annealing optimisation. The code is available in this repository and in Moodle submission. The implementation is well-documented and contains comments on architectural decisions and obtained results.

| | name | rate | shortest_path | avg_shortest_path |
|---|---|---|---|---|
| 0 | Slow annealing 0.99 | 0.99 | 20583.153168 | 21814.608733 |
| 1 | Moderate annealing 0.8 | 0.80 | 20539.501711 | 23551.080347 |
| 2 | Fast annealing 0.6 | 0.60 | 24655.975499 | 25652.931033 |

Figure 2: Shortest paths and average shortest paths for fast, slow, middle annealing.

Simulated annealing with rates 0.99 and 0.8 converged to compatible paths of approximately 20.5k, although the latter performed slightly better. This indicates that annealing with higher rates might converge to a better solution due to increased chance to explore alternatives. Annealing with the rate of 0.99 on average converges to shorter paths compared to rates 0.8 and 0.6. This observation also emphasises that with a higher pace in decreasing entropy the simulated annealing tends to converge to a non-optimal solution.

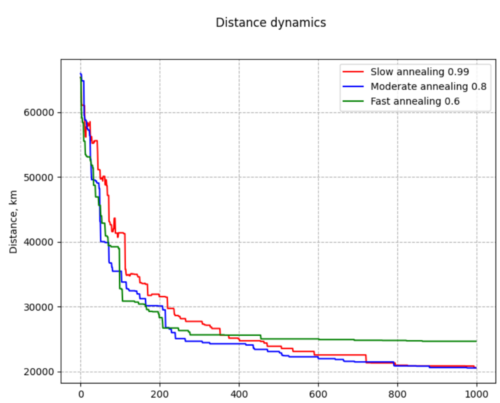


Figure 3: Distances of fast, slow, middle annealing with respect to iteration.

Visualization of shortest distances with respect to iterations for annealing schedules shows that the annealing with rate 0.6 converges faster, which can be accounted to stucking in a local optima. The annealing with rate 0.99 in our observation converges to nearly same value as annealing of rate 0.8, but does it more gradually. Remarkably, the version with a rate of 0.6 contains several plato-like intervals and demonstrates the lowest result.