Homework 1

Deadline: November 14

Exercise 1

Consider two boxes with white and black balls. Box 1 contains three black and five white balls and box 2 contains two black and five white balls. First a box is chosen at random with a prior probability $p(\text{box} = 1) = p(\text{box} = 2) = 0.5$, secondly a ball picked at random from that box. This ball turns out to be black. What is the posterior probability that this black ball came from box 1?

Exercise 2

The weather in Gothenburg can be summarized as: if it rains or snows one day there is a 60% chance it will also rain or snow the following day; if it does not rain or snow one day there is an 80% chance it will not rain or snow the following day either.

(i) Assuming that the prior probability it rained or snowed yesterday is 50%, what is the probability that it was raining or snowing yesterday given that it does not rain or snow today?

(ii) If the weather follows the same pattern as above, day after day, what is the probability that it will rain or snow on any day (based on an effectively infinite number of days of observing the weather)?

(iii) Use the result from part 2 above as a new prior probability of rain/snow yesterday and recompute the probability that it was raining/snowing yesterday given that it's does not rain or snow today.

Exercise 3

Prove that the Beta distribution

$$\text{Beta}(\mu; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}, \text{ where } \mu \in [0,1], \ \Gamma(a) = \int_0^\infty e^{-x}x^{a-1}dx$$

is correctly normalized, i.e.,

$$\int_0^1 \text{Beta}(\mu; a, b)d\mu = 1 \Longleftrightarrow \int_0^1 \mu^{a-1}(1-\mu)^{b-1}d\mu = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Exercise 4

Consider a variable $x \in 0, 1$ and $p(x = 1) = \mu$ representing flipping of a coin. With 50% probability we think the coin is fair, i.e., $\mu = 0.5$, and with 50% probability we think that is unfair, i.e., $\mu \neq 0.5$. We encode this prior belief with the following prior

$$p(\mu) = \frac{1}{2}\text{Beta}(\mu; 1, 1) + \frac{1}{2}\delta(\mu - 0.5)$$

(i) Assume that we get one observation $x_1 = 1$. What is the posterior $p(\mu|x_1)$? In particular, how does the belief of the fairness of the coin change under this observation?

(ii) Assume that we get one additional observation $x_2 = 1$. What is the posterior $p(\mu|x_1, x_2)$? In particular, how does the belief of the fairness of the coin change under this observation?

(iii) Compute the probability of the coin being fair by defining an event *fair* with the prior probability $p(fair) = 0.5$. Compute $p(fair|x_1, x_2)$ using Bayes' theorem based on the observations $x_1 = 1, x_2 = 1$.

Exercise 5

In this problem, questions are divided into four parts: theoretical questions, programming simulation, visualization, and analysis. In the following lines, you will find a description of a real-world scenario where the solution to this problem can be applied. This description is provided to help build intuition about the problem and to serve as a starting point for the simulation.
Imagine a city's public health department wants to monitor the spread of a disease based on daily counts of new cases reported at a hospital. Let's assume:

- Observed Data (Poisson process): The number of new cases each day (x_i) is random but follows a certain average rate, λ cases per day. This rate λ is unknown, but it reflects how often new cases occur on average over time.

- Prior Belief (Gamma distribution): Based on previous outbreaks or expert knowledge, the department has a prior belief about the rate λ of new cases. They may believe, for example, that typically there are about 2 new cases per day, and they use a Gamma distribution to express this belief.

Data Collection and Bayesian Update:
After monitoring the hospital's reported cases over several days (say 20 days), they record a sequence of counts for new cases each day. These counts serve as the dataset, which allows them to update their belief about λ.
Prior and Posterior Distribution:

Initially, their prior Gamma distribution for λ reflects their pre-existing knowledge or assumption about disease spread. But once they observe real data, they update this belief to get the posterior distribution, which combines both their prior knowledge and the observed data.
In this context:

- **Prior:** The health department's initial guess of disease spread rate (λ) based on past data or expert knowledge, represented by a Gamma distribution with shape and rate parameters α and β.
- **Likelihood:** The observed counts of new cases over the 20 days, assumed to be generated by a Poisson process.
- **Posterior:** The updated belief about the rate of new cases λ after observing the daily counts. As they observe more cases, the posterior distribution sharpens around the true average rate of new cases, giving the department a more accurate understanding of the spread of the disease.

1. **Theoretical questions:**
   - Derive the likelihood of observing a dataset $x = \{x_1, x_2, \ldots, x_n\}$ where each x_i follows a Poisson distribution with unknown rate parameter λ.
   - Assume a Gamma prior on λ with shape parameter α and rate parameter β. Using Bayes' theorem, derive the posterior distribution for λ after observing the dataset x.
   - Show that this posterior distribution is also a Gamma distribution, and specify the updated shape and rate parameters of the posterior.
2. **Programming Simulation:**
   - Generate a synthetic dataset by simulating counts from a Poisson distribution with a true rate parameter λ_{true}.
   - Assume a Gamma prior on λ with initial parameters $\alpha = 2$ and $\beta = 2$. After observing the synthetic data, update the prior parameters to obtain the posterior distribution.
   - Write a Python function that, given data and prior parameters α and β, calculates the updated posterior parameters for λ.
3. **Visualization:**
   - Plot the prior and posterior distributions for λ, clearly showing how observing the data has shifted your belief about the rate parameter.

- Compare the posterior distribution to the true value of λ_{true} used to generate the data. As you increase the number of observations, observe how the posterior distribution becomes more concentrated around λ_{true}.

4. **Analysis:**
   - Explain the impact of observing more data on the posterior distribution. How does the posterior distribution become more concentrated as the sample size grows?