0.1 1b) Conceptual

When specifying a Bayesian model, we use our domain knowledge to establish certain distributions, and then we use computation to find other ones. Which of the following do we establish using our domain knowledge? Pick all that apply.

(a) Prior

(b) Likelihood

(c) Marginal distribution of the data

(d) Posterior

Type your answer here, replacing this text.

0.1.1 1c (i)

Fix `alpha_value = 5`, and experiment with different values of `beta_value`. As we increase `beta_value`, what happens to the mode of the distribution?

Type your answer here, replacing this text.

0.1.2 1c (ii)

Fix `beta_value = 5`, and experiment with different values of `alpha_value`. As we increase `alpha_value`, what happens to the mode of the distribution?

Type your answer here, replacing this text.

0.1.3 1c (iii)

Set `alpha_value = beta_value = 1`, increase their value such that `alpha_value=beta_value`. What happens to the *variance* of the distribution?

Type your answer here, replacing this text.

0.1.4 1e (i)

Start by setting $k = 0$, and then steadily increase the value of k. Record your observations using **2-3 sentences** in the space below. Be sure to address the following questions: 1. At $k = 0$, where do the data points lie relative to the horizontal and diagonal lines? Why? 2. As we increase the value of k, do the points move towards or away from the horizontal line? Why?

Type your answer here, replacing this text.

0.1.5 1e (ii)

As you increase k, some points move faster than others. Which points move faster, the larger or smaller data points? Explain why this is the case in **1-2 sentences**.

Type your answer here, replacing this text.

0.1.6 1e (iii)

Imagine that we let $k \to \infty$. How do you think the two graphs above will look in the limit $k \to \infty$? Limit your response to **1-2 sentences**.

Type your answer here, replacing this text.

0.1.7 1e (iv)

Fill in the blank in this sentence with either "small" or "large", and explain your answer in **1-2 sentences**:

If we're very sure that the true SAR is close to $\frac{1}{3}$*, we should choose a* __________ *value of* k*.*

Type your answer here, replacing this text.

Note that the shape of `thetas` is (N x M). What are N and M, and what does each mean?

Type your answer here, replacing this text.

0.2 2b) Using the output of PyMC

Now that we've run our sampler, we now have access to the posterior distributions of *all* the random variables we defined in PyMC. Using these empirical distributions, we can now calculate the posterior means for each θ_i. But before we do that, let's visualize the samples we got back.

Generate a histogram of all 2,000 posterior samples for θ_2 (the SAR for Study 2). Use the `sns.histplot` function with `stat='density'`.

In [ ]:
```
# TODO: Create histogram of posterior samples
...
```

How do the samples compare to the two different estimates you saw in Question 1?

Type your answer here, replacing this text.

0.2.1 2d (i)

Compare the curve of the theoretical distribution with the histogram of samples from the empirical posterior. Are they similar or different?

Type your answer here, replacing this text.

0.2.2 2d (ii)

Compare the two figures corresponding to 'weak' prior $\theta_i \sim Beta(2, 4)$ and 'strong' prior $\theta_i \sim Beta(20, 40)$. How are they different? Explain why.

Type your answer here, replacing this text.

Notice that the `trace` now contains samples for both `theta` and `A`!

Plot a histogram of the posterior estimates for A if $\alpha = 5$ and $\beta = 10$.

```
In [ ]: model, trace = approximate_inference_asympotmatic_MCMC(5, 10)
        ...
```

Assuming the model we defined is correct, what can you conclude about the asymptomatic rate A based on the studies and the model?

Type your answer here, replacing this text.