Homework 2

Question 1:

Generate the coin flipping model for the following observed data:

trials = 4

theta_real = 0.35 # *unknown value in a real experiment*

data = pz.Binomial(n=1, p=theta_real).rvs(trials)

Compare the following priors:

1. $\theta \sim$ Beta(20, 20)
2. $\theta \sim$ Uniform(0, 1)
3. $\theta \sim$ Uniform(-1, 2)

a. Create graphic models (Kruschke diagrams) for all the options.

b. Compare with the results of using *pm.model_to_graphviz.*

c. Are there models that don't run or have errors? Explain.

d. Compare the posteriors.

Question 2:

a. Load the chemical shift data you saw in the lecture.
   You can use this line to load the chemical shift data:

```
data =
np.loadtxt("https://github.com/aloctavodia/BAP3/raw/refs/heads/main/code/da
ta/chemical_shifts.csv")
```

b. We will model the data using a normal distribution.
   The prior for σ will be a halfnormal distribution with $\sigma_\sigma = 5$.
   The prior for μ will be a normal distribution with $\mu_\mu = mean(data)$.
   We will try a few values for σ_μ (the standard deviation of the prior of μ).

   - Draw the model
   - Write the model equations

c. Try a few values for σ_μ.
   How robust/sensitive are the inferences to these changes? Include posterior predictive checks in your analysis.

d. What do you think of using a Gaussian, which is an unbounded distribution (goes from − inf to inf), to model bounded data (between 0-100) such as this?

Question 3:

Model the data using the Student's t-distribution instead of a normal distribution. Compare the results to those from Question 2.

Question 4:

We will define outliers as any value more than 1.5 IQR (Interquartile range) beyond the IQR.

a. Compute the mean and standard deviation of the chemical shift with and without the outliers.
b. Compare those results to the Bayesian estimation using the Gaussian and Student's t-distribution. What do you observe?
c. Add more outliers to the data and compute new posteriors for the Student's t model and one of your normal models. What do you observe?

Question 5:

In both the tutorial and the lecture, we looked at the Tips dataset.

a. Load the data and run the model as we did.
b. Explore the InferenceData object.

    1. How many groups does it contain?
    2. Inspect the posterior distribution of the parameter μ for a specific day using the sel method.
    3. Compute and plot the distribution of the mean differences between Thursday and Sunday. What are the coordinates and dimensions of the resulting DataArray?