**Documentation: Bayesian Statistics with Beta Distribution**

**1. What is Bayesian Statistics?**

Bayesian statistics is a mathematical framework that allows us to update our beliefs about an unknown parameter (e.g., the proportion of people supporting a policy) based on prior knowledge and observed data. It is based on **Bayes' Theorem**, which states:

* **Prior**: Represents our initial belief about the parameter before observing any data.
* **Likelihood**: Represents the probability of observing the data for a given value of the parameter.
* **Posterior**: Combines the prior and likelihood to give an updated belief about the parameter after observing the data.

**2. Why is Bayesian Statistics Used?**

Bayesian statistics is used for:

* **Incorporating prior knowledge**: Combines previous data, expert opinions, or assumptions with current data.
* **Probabilistic interpretation**: Provides a natural way to quantify uncertainty in parameter estimates.
* **Dynamic updating**: Allows beliefs to be updated as new data becomes available.

**3. Why Use Bayesian Statistics Now?**

In this context, we are estimating the proportion of a population that supports a policy (e.g., mandatory vaccinations). We conducted a survey where:

* **200 individuals were asked**.
* **120 responded "yes"**, and **80 responded "no"**.

We use Bayesian statistics because:

1. **We may have prior knowledge** about . For instance, previous surveys or assumptions might indicate support is around 60%.
2. **We want to quantify uncertainty** in after observing the survey data.
3. **The Beta distribution** provides a convenient framework for modeling the prior and posterior distributions of .

**4. Why Use the Beta Distribution?**

The Beta distribution is a flexible distribution for probabilities (values between 0 and 1). It is defined by two parameters, and :

* : Represents the prior belief about the number of "successes" (e.g., people supporting the policy).
* : Represents the prior belief about the number of "failures" (e.g., people not supporting the policy).

Benefits of using the Beta distribution:

* It is simple to update with data when combined with a Binomial likelihood.
* It can represent a wide range of prior beliefs (uninformative, weakly informative, or strongly informative).

**5. Applying Bayesian Statistics to Our Survey**

**Step 1: Define the Prior**

* If we have no prior knowledge, we use a **uniform prior**: .
  + This prior assumes all values of are equally likely before observing the data.

**Step 2: Collect Data**

* **Survey data**: Out of 200 responses, 120 are "yes" and 80 are "no."

**Step 3: Compute the Posterior**

* With a Beta prior and Binomial likelihood, the posterior is also a Beta distribution:
* Example:
  + Prior: .
  + Data: 120 "yes" and 80 "no."
  + Posterior: .

**Step 4: Visualize the Results**

* **Prior Distribution**: Shows our beliefs about before observing data.
* **Posterior Distribution**: Shows our updated beliefs about after incorporating the data.

**6. Example of Informative Prior**

If prior surveys suggested support was around 30%, we might choose as the prior. The posterior is then updated with the data:

* Prior: .
* Posterior: .

This illustrates how the prior influences the posterior, especially with smaller datasets.

**7. Conclusion**

Bayesian statistics is a powerful tool for estimating unknown parameters while incorporating prior knowledge. In this survey example:

* The posterior reflects a balance between prior beliefs and observed data.
* The more data we have, the less influence the prior has.
* Using a Beta prior simplifies the computations and provides intuitive results.

Bayesian methods are particularly valuable when:

1. Prior knowledge is available and relevant.
2. Uncertainty needs to be quantified probabilistically.
3. Updating beliefs dynamically with new data is important.