Bayesian Statistics with an Example

Bayesian statistics is a framework for updating beliefs or probabilities in light of new evidence. It relies on Bayes' Theorem, which provides a way to revise existing predictions or theories given new data.

Bayes' Theorem Formula

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Where:

- $P(A|B)$: Posterior probability (probability of event A given evidence B).
- $P(B|A)$: Likelihood (probability of evidence B given that event A is true).
- $P(A)$: Prior probability (initial probability of event A before seeing evidence B).
- $P(B)$: Marginal probability (total probability of evidence B).

Inserting an image containing text on climate change to assert that text from the images is actually being extracted

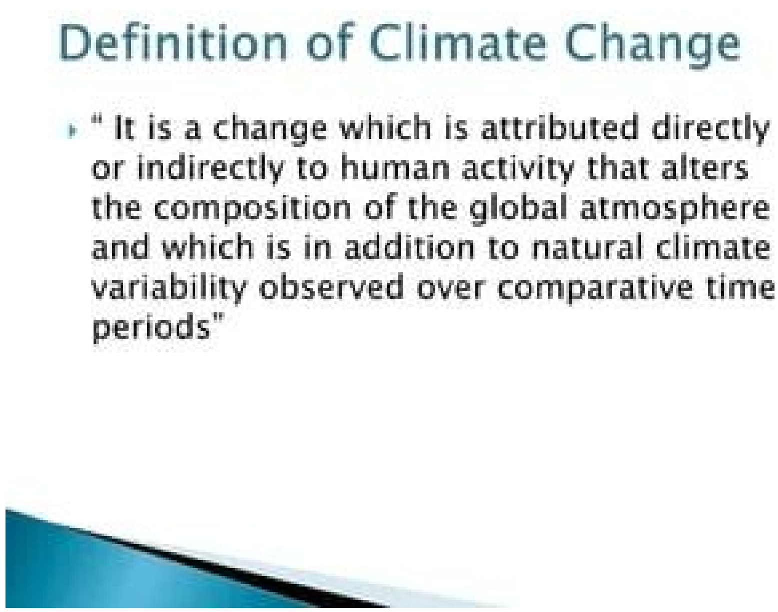

Figure 1: *Definition*

Example: Medical Diagnosis

Imagine a scenario where a patient gets tested for a rare disease.

- **Prior Probability** ($P(\text{Disease})$): The probability that a person has the disease without any test, say 0.01 (1%).
- **Probability of Positive Test Given Disease** ($P(\text{Positive}|\text{Disease})$): If a person has the disease, the test correctly identifies it 95% of the time, so 0.95.

- **Probability of Positive Test Given No Disease** ($P(\text{Positive}|\text{No Disease})$): If a person does not have the disease, the test falsely indicates disease 5% of the time, so 0.05.
- **Probability of Positive Test** ($P(\text{Positive})$): This is a combination of cases where the test is positive due to disease and where it's a false positive.

Applying Bayes' Theorem

We want to find the **posterior probability** that the patient has the disease given a positive test result, $P(\text{Disease}|\text{Positive})$.

- **Prior Probability**: $P(\text{Disease}) = 0.01$
- **Likelihood**: $P(\text{Positive}|\text{Disease}) = 0.95$
- **False Positive Rate**: $P(\text{Positive}|\text{No Disease}) = 0.05$
- **Probability of No Disease**: $P(\text{No Disease}) = 0.99$

Now, compute $P(\text{Positive})$ (Total probability of a positive test):

$$P(\text{Positive}) = P(\text{Positive}|\text{Disease}) \times P(\text{Disease}) + P(\text{Positive}|\text{No Disease}) \times P(\text{No Disease})$$

$$P(\text{Positive}) = 0.95 \times 0.01 + 0.05 \times 0.99 = 0.0095 + 0.0495 = 0.059$$

Finally, compute $P(\text{Disease}|\text{Positive})$:

$$P(\text{Disease}|\text{Positive}) = \frac{P(\text{Positive}|\text{Disease}) \times P(\text{Disease})}{P(\text{Positive})}$$

$$P(\text{Disease}|\text{Positive}) = \frac{0.95 \times 0.01}{0.059} \approx 0.161$$

Interpretation

Even after a positive test result, the probability that the patient actually has the disease is about 16.1%. This shows the importance of considering both the accuracy of the test and the prior probability when interpreting test results.

Bayesian statistics provides a powerful way to update our beliefs with new evidence, giving us a more nuanced understanding of probabilities in light of additional data.