Bayesian Analysis

Bayes' Rule

Bayes' rule (or Bayes' theorem) was named after Thomas Bayes (1701-1761), letting A and B be two events,

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

- Math to English translation:
  - $P(A)$: probability that event A happens
  - $P(A \mid B)$: given that event B happens, the probability that event A happens
- The proof of Bayes' rule is straight-forward (ignoring technical assumptions)

$$P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

  Here $P(A \cap B)$ is the probability that events A and ("∩") event B both happen

Example

Scenario: A student takes a rapid COVID test and it comes back **positive**.

Assume the following:

- The test is 95% for infected individuals: $P(\text{Positive} \mid \text{COVID}) = 0.95$
- The test gives a false positive 3% of the time: $P(\text{Positive} \mid \text{No COVID}) = 0.03$
- Only 1% of people in the school currently have COVID: $P(\text{COVID}) = 0.01$

But how likely is it that the student is actually infected ($P(\text{COVID} \mid \text{Positive})$?

Solution

Step 1: Bayes’ Rule

We want to find $P(\text{COVID} \mid \text{Positive})$ using

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)},$$

where $A \equiv \text{has COVID}$ and $B \equiv \text{tests positive}$

Step 2: Compute total probability of a positive test

$$P(B) = P(B \mid A) \times P(A) + P(B \mid \text{not } A) \times P(\text{not } A) = 0.95 \times 0.01 + 0.03 \times 0.99 =$$

Step 3: Apply Bayes' Rule

$$P(A \mid B) = \frac{0.95 \cdot 0.01}{0.0392} = \frac{0.0095}{0.0392} \approx 0.242$$

Conclusion: Even after a positive test, there's not even 50% chance that the student actually has COVID.

Bayes' Rule in Statistical Modeling

Recall Bayes' rule

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

Let $A \equiv \{\text{True parameter equals } \theta\}$ and $B \equiv \{\text{Observe dataset } \mathcal{D}\}$. We can write

$$P(\theta \mid \mathcal{D}) = \frac{P(\mathcal{D} \mid \theta)\, P(\theta)}{P(\mathcal{D})}$$

- $P(\theta)$ **Prior** belief/distribution about parameter θ
- $P(\mathcal{D} \mid \theta)$: **Likelihood** of seeing data D if θ were true
- $P(\theta \mid \mathcal{D})$: **Posterior** belief *after* seeing data
- $P(\mathcal{D})$: Evidence/normalizing constant

Bayesian vs. Frequentist Thinking

Feature	Frequentist	Bayesian
Parameters (θ)	Fixed but unknown	Random variables with distributions
Data (D)	Random draw from population	Observed *fixed* evidence
Estimand	Point estimate $\hat{\theta}$	Full posterior $P(\theta \mid D)$
Inference output	Confidence intervals	Full posterior $P(\theta \mid D)$, credible interval

Both a powerful tools for data analysis:

- Use Frequentist methods for its convenience and "objectivity"
- Use Bayesian methods to incoporate prior knowledge, combine multi-view data, handle network data

Recipe for Bayesian Analysis

1. Choose a prior $P(\theta)$
   - Encodes what we believe *before* seeing data
   - Should be chosen *before* looking at the data to avoid double-counting
2. Write down the likelihood $P(D \mid \theta)$
   - Comes from the data-generating model (e.g., Binomial for coin flips)
3. Obtain the posterior $P(\theta \mid D) \propto P(\theta)P(D \mid \theta)$
   A. Analytic solution for *conjugate* pairs of prior and likelihood
   B. Sampling from the posterior
   C. Approximation (variational inference)
4. Summarize and inference
   - Posterior mean, median, credible interval
   - Posterior predictive distribution for new data

Examples

1. One sample test of means (freq v.s. bayes)
2. Conjugate priors v.s. sampling
3. Variational inference
