Conditional Probability and Bayes' Theorem

Mathematics for Artificial Intelligence

1 Conditional Probability

1.1 Definition

The **conditional probability** of event A given that event B has occurred is defined as:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) > 0.$$

This measures the probability that A occurs under the condition that B is known to have occurred.

1.2 Interpretation

- $P(A)$ is the probability of A in the entire sample space.
- $P(A \mid B)$ is the probability of A *within the restricted space where B has already happened.*

1.3 Example

A box contains 3 red balls and 2 blue balls. One ball is drawn and not replaced. What is the probability that the second ball is red given that the first ball was red?

$$P(\text{Red}_2 \mid \text{Red}_1) = \frac{2}{4} = \frac{1}{2}$$

This is because removing one red ball leaves 2 red out of 4 remaining balls.

1.4 Multiplication Rule

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

More generally,

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_1 \cap A_2) \cdots$$

This is heavily used in probabilistic models such as Hidden Markov Models and Bayesian Networks.

2 Law of Total Probability

Suppose $B_1, B_2, \dots, B_n$ form a partition of the sample space S where:

$$B_i \cap B_j = \emptyset \quad (i \neq j), \qquad \bigcup_{i=1}^{n} B_i = S.$$

Then for any event A:

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i)\, P(B_i).$$

This law is foundational for computing marginal probabilities in Bayesian models.

3 Bayes' Theorem

3.1 Formula

Bayes' theorem relates conditional probabilities:

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

Using the Law of Total Probability:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{\sum_{i=1}^{n} P(B \mid A_i)\, P(A_i)}.$$

3.2 Interpretation

- $P(A)$ is the **prior probability**.
- $P(B \mid A)$ is the **likelihood**.
- $P(A \mid B)$ is the **posterior probability**.

Bayes' theorem is fundamental in machine learning methods such as:

- Naive Bayes Classifier
- Bayesian Inference
- Probabilistic Graphical Models

3.3 Example: Medical Diagnosis

A disease affects 1% of the population. A diagnostic test has the following properties:

- True positive rate: $P(+ \mid D) = 0.95$
- False positive rate: $P(+ \mid \neg D) = 0.10$

What is the probability that a person actually has the disease given that they tested positive?

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

First compute $P(+)$ using the Law of Total Probability:

$$P(+) = P(+ \mid D)\, P(D) + P(+ \mid \neg D)\, P(\neg D)$$

$$P(+) = (0.95)(0.01) + (0.10)(0.99) = 0.0095 + 0.099 = 0.1085$$

Now the posterior:

$$P(D \mid +) = \frac{0.95 \times 0.01}{0.1085} \approx 0.0876$$

Even with a positive test, the chance of having the disease is only about **8.76%** — because the disease is rare.

4 Bayes' Theorem in Machine Learning

4.1 Naive Bayes Classifier

Assumes features $X_1, X_2, \ldots, X_n$ are conditionally independent given class C:

$$P(C \mid X_1, X_2, \ldots, X_n) \propto P(C) \prod_{i=1}^{n} P(X_i \mid C)$$

Used widely in:

- Spam detection
- Text classification
- Sentiment analysis

4.2 Bayesian Updating

Posterior becomes the new prior:

$$\text{Posterior} \rightarrow \text{Updated Prior (for next round of inference)}$$

This is essential in:

- Bayesian Optimization
- Kalman Filters
- Reinforcement Learning (belief updates)