Bayes' Theorem

Daniel Rostovtsev
Date: May 13, 2017

Given two events A and B, Bayes' Theorem relates the likelihood of A given B with the posterior likelihood of B given A. It allows for old probabilistic predictions to be altered given new information, and its proof relies on conditional probability. Thomas Bayes had an intuition for conditional probability as early as the 1700s, but he considered it to be a fundamental axiom of his theory.

It wasn't until the mid twentieth century that mathematicians like Kolmogorov and Cox realized that the concept of probability is best understood as extensions of a more basic set of axioms. These axioms, coined the Kolmogorov axioms, provide a deeper insight into how our minds think about probability, and it is worthwhile to examine complex ideas, like conditional probability, as results of those axioms, rather than as axioms of their own.

Sample Space

Let Ω be the sample space of an experiment, then Ω is the set of all possible outcomes, or results, of that experiment.

Probability Space

Let Ω be the sample space of some experiment, then the probability space $\mathcal{F} = \mathcal{P}(\Omega)$, where $\mathcal{P}(\Omega)$ is the power set of Ω.

For example, the sample space of flipping a coin two times would be $\Omega = \{HH, HT, TH, TT\}$, and the probability space would be $\mathcal{F} = \{\emptyset, \{HH\}, \{HT\}, \cdots, \{HH, HT, TH\}, \cdots, \Omega\}$ (or all the possible subsets of Ω). Today, probability is understood to be a function mapping a probability space to the real numbers lying between zero and one.

Kolmogorov's Probability Axioms

A probability measure, or a probability P, is a function $P : \mathcal{F} \to \mathbb{R}$ such that the following is true:

i. $P(E) \geq 0 \;\forall\; E \in \mathcal{F}$

ii. $P(\emptyset) = 0$ and $P(\Omega) = 1$

iii. Let $E_1, \cdots, E_n$ be disjoint elements of $\mathcal{F}$, then: $P(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i)$

Corollary 1

Let $A, B \in \mathcal{F}$, then $A \subset B \implies P(A) \leq P(B)$.

Proof. Let A' be the complement of A in B, then all elements of B not in A are in A'. By the third probability axiom, $P(B) = P(A \cup A') = P(A) + P(A')$, so $P(B) = P(A) + P(A')$. Since $P(A') \geq 0$, it follows that $P(A) \leq P(B)$.

□

Corollary 2

The $A \in \mathcal{F}$, then $P(A) \leq 1$

Proof. Since $A \in \mathcal{F}$, it follows that $A \subset \Omega$. By the first corollary, $P(A) \leq P(\Omega)$, and by the second axiom $P(\Omega) = 1$, therefore $P(A) \leq 1$.

□

It is usually best to think about probabilities of events as the sums of the probabilities of specific outcomes, since any event is the union of specific outcomes. Outcomes in probability theory have a special name: "elementary elements." They are the building blocks of probability.

Elementary Element

An element, $E \in \mathcal{F}$ is an elementary element if E has no proper subset.

Proposition

E is an elementary element of $\mathcal{F}$ if and only if it contains only one element $\omega \in \Omega$.

Proof. E is an elementary element if and only if E has only one element. Since $E \subset \Omega$, $E = \{\omega\}$ where $\omega \in \Omega$.

$\square$

Now that there is a solid foundation for the meaning of a probability, the idea of a conditional probability can be accurately described. A conditional probability $P(A|B)$ (or the probability of A "given" B) is the likelihood of an event A when B is certain. More precisely, a conditional probability $P(A|B)$ is a special kind of probability measure over Ω for which all outcomes not contained in B have a probability of 0. Formally, it is defined as follows:

Conditional Probability

Given a probability P, and a seed event B, the conditional probability is a probability measure over Ω such that, for an outcome $\omega \in \Omega$ and a scalar $\alpha \in \mathbb{R}$, the following is true:

$$P(\omega|B) = \begin{cases} 0 & \text{if } \omega \notin B \\ \alpha P(\omega) & \text{if } \omega \in B \end{cases}$$

Corollary 1

$\sum_{\omega \in \Omega} P(\omega|B) = 1$

Proof. By the third axiom, $\sum_{\omega \in \Omega} P(\omega|B) = P(\bigcup_{\omega \in \Omega} \omega) = P(\Omega)$. By the second axiom $P(\Omega) = 1$, since P is a probability measure over Ω. Therefore $\sum_{\omega \in \Omega} P(\omega|B) = 1$.

$\square$

Corollary 2

$\alpha = \dfrac{1}{P(B)}$

Proof. $\sum_{\omega \in \Omega} P(\omega|B) = \sum_{\omega \in B} P(\omega|B)$ since $P(\omega|B) = 0$ if $\omega \notin B$. Since, if $\omega \in B$, $P(\omega|B) = \alpha P(\omega)$, it follows that:

$\sum_{\omega \in B} P(\omega|B) = \sum_{\omega \in B} \alpha P(\omega) = \alpha \sum \omega \in \Omega P(\omega) = \alpha P(B) \therefore \sum_{\omega \in \Omega} P(\omega|B) = \alpha P(B)$

By the first proposition $\sum_{\omega \in \Omega} P(\omega|B) = 1$, so

$\alpha P(B) = 1 \therefore \alpha = \dfrac{1}{P(B)}$

$\square$

Proposition

$$P(\omega|B) = \begin{cases} 0 & \text{if } \omega \notin B \\ \dfrac{P(\omega)}{P(B)} & \text{if } \omega \in B \end{cases}$$

Proof. Follows directly from the definition and the second corollary. □

Theorem of Conditional Probability

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

Proof. $P(A|B) = \sum_{\omega \in A} P(\omega|B) = \sum_{\omega \in A \cap B} P(\omega|B) = \frac{\sum_{\omega \in A \cap B} P(\omega)}{P(B)} = \frac{P(A \cap B)}{P(B)}$ □

Now, having a clear definition of conditional probability, the proof of Bayes' Theorem is simple.

Bayes' Theorem

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

Proof. From the theorem of conditional probability:

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

However, $A \cap B = B \cap A$, so $P(A \cap B) = P(B \cap A)$, and thus:

$$P(A|B)P(B) = P(B|A)P(A)$$

Therefore $P(A|B) = \dfrac{P(B|A)P(A)}{P(B)}$ □