MA2202: Probability I

Conditional probability

Spring 2021

Satvik Saha
19MS154

*Indian Institute of Science Education and Research, Kolkata,
Mohanpur, West Bengal, 741246, India.*

Suppose a fair die is rolled and the outcome is even. The probability that the outcome was prime is 1/3. This is because the only even prime is 2, and we know that we have rolled one of $2, 4$ or 6, all of which are equally likely.

Definition 2.1 (Conditional probability)**.** Suppose a random experiment is performed, and we know that an event B occurred. Let $A \in \mathcal{E}$. The probability of occurrence of A given B is called the conditional probability $P(A \,|\, B)$.

Remark. Note that the conditional probability A given B is equivalent to the occurrence of $A \cap B$ computed relative to the probability of the occurrence of B.

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

Lemma 2.1. *If $A_1, \ldots, A_n$ are exhaustive and pairwise mutually exclusive events such that $P(A_i) > 0$, and $B \in \mathcal{E}$, then*

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

Theorem 2.2 (Bayes' Theorem)**.** *If $A, B \in \mathcal{E}$ with $P(B) \neq 0$, we have*

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

Proof. This follows directly from

$$P(A \cap B) = P(A \,|\, B)\, P(B) = P(B \,|\, A)\, P(A).$$ □

Example. Suppose a fair die is rolled and the outcome is even. The probability that the outcome is prime is 1/3. This is because the only even prime is 2, the available even numbers are $2, 4, 6$, and the available primes are $2, 3, 5$. Thus,

$$P(\text{prime} \,|\, \text{even}) = \frac{P(\text{even})}{P(\text{prime})} P(\text{even} \,|\, \text{prime}) = \frac{3/6}{3/6} \cdot \frac{1}{3} = \frac{1}{3}.$$

Corollary 2.2.1. *If $A_1, \dots, A_n$ are exhaustive and mutually exclusive with $P(A_i) > 0$, and $B \in \mathcal{E}$ such that $P(B) > 0$, then*

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

Example. Suppose that a test for a certain disease has a sensitivity of 90% and a specificity of 80%. This means that it correctly identifies a positive 90% of the time, and incorrectly shows a positive for a healthy individual 20% of the time. Also suppose that 5% of the population suffer from this disease.

We note that the probability that a randomly selected person has this disease is $P(\mathrm{D}) = 0.05$. Now, the probability that an infected person gets a positive test result is $P(+ \mid \mathrm{D}) = 0.9$. The probability that an uninfected person gets a positive test result is $P(+ \mid \mathrm{ND}) = 0.2$. Thus, we use Bayes' Theorem to obtain the probability that a person has the disease, given that they test positive.

$$\begin{aligned} P(\mathrm{D} \mid +) &= \frac{P(+ \mid \mathrm{D})\, P(\mathrm{D})}{P(+ \mid \text{Infected})\, P(\mathrm{D}) + P(+ \mid \mathrm{ND})\, P(\mathrm{ND})} \\ &= \frac{0.9 \times 0.05}{0.9 \times 0.05 + 0.2 \times 0.95} \\ &\approx 19\%. \end{aligned}$$

This is surprisingly low. This is because of the high rate of false positives — in a population of 1000 people, there will only be 50 infected people of which $0.9 \times 50 \approx 45$ are identified, but $0.2 \times 950 \approx 200$ healthy people who incorrectly test positive. Thus, of the ≈ 245 positive results, only 45 are genuine, giving an approximate probability $1/5 \approx 20\%$ of having the disease, given a positive test.

The situation changes when we run the test again. The probability that a person has the disease given *two* positive tests is given by

$$\begin{aligned} P(\mathrm{D} \mid ++) &= \frac{P(D)}{P(++)} P(++ \mid D) \\ &= \frac{0.9^2 \times 0.05}{0.9^2 \times 0.05 + 0.2^2 \times 0.95} \\ &\approx 52\%. \end{aligned}$$

Intuitively, when we look at the 245 people who tested positive the first time, around $0.2 \times 200 \approx 40$ people will be false positives the second time around, while around $0.9 \times 45 \approx 41$ people will be true positives. Thus, the probability that a person has the disease given two positive tests is just above 50%.

Note that this gives us a quick way of estimating probabilities; after three tests, we expect around $0.2 \times 40 \approx 8$ false positives and $0.9 \times 41 \approx 37$ true positives, so 37 out of 45 people who test positive three times have the disease. This gives a probability of around 82% that a person who tests positive thrice actually has the disease.

Observe that when testing for a disease with low prevalence, the specificity becomes much more significant than the sensitivity.

Definition 2.2 (Independent events)**.** If $A, B \in \mathcal{E}$ such that $P(A), P(B) \neq 0$, and we have

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

then we say that A and B are independent.

Equivalently, $P(A \cap B) = P(A)P(B)$.

Example. Suppose a die is rolled twice. The probability that the first roll is a 4 is 1/6, and so is the probability that the second roll is a 2. Now, if we were given that the first roll was a 4, it wouldn't affect the likelihood of the second roll being a 2, and vice versa. Thus, the probability of obtaining a 4 and 2 in succession is simply $(1/6)^2 = 1/36$. The two rolls are independent.

Definition 2.3 (Pairwise independent events)**.** A set of events $S \subseteq \mathcal{E}$ is called pairwise independent if the events A and B are independent for all $A, B \in S$.

Definition 2.4 (Mutually independent events)**.** The set of events $A_1, \ldots, A_n$ is called mutually independent if

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$

for all subsets $I \subseteq \{1, \ldots, n\}$.

Remark. A set of mutually independent events is always pairwise independence, since we must have $P(A_i \cap A_j) = P(A_i)\,P(A_j)$ for all $1 \leq i, j \leq n$. The converse is not true.