Conditional Probability and Independence

Guy Lebanon

January 6, 2006

In the following we assume that A, B are events in the sample space Ω and P is a probability on Ω.

Definition 1. *Two events A, B are defined to be independent if $P(A \cap B) = P(A)P(B)$.*

The intuitive meaning here is that the fact whether the event A occurred or not does not alter the likelihood of the event B occurring (and vice verse). Recall that if two events are disjoint $A \cap B = \emptyset$, they can't occur simultaneously. Therefore if two events are disjoint (or mutually exclusive) the occurrence of one precludes the occurrence of the other. Hence two events with positive probabilities can't be both disjoint and independent!

Definition 2. *The conditional probability of A given B is defined as (assuming that $P(B) > 0$),*

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

Note that in this case $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$.

The intuitive meaning here is the probability of A occurring assuming that the event B occurred. Note that if A, B are independent, $P(A|B) = P(A)$ – as it should if you put the intuitive meaning of independence and conditional probability together.

Theorem 1. *If A, B are independent, then so are A^c, B and so are A, B^c and so are A^c, B^c.*

Proof. For example,

$$P(A^c \cap B) = P(B \setminus A) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = (1 - P(A))P(B) = P(A^c)P(B)$$

□

Theorem 2 (Bayes)**.**

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

Proof.

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

□

Example: Consider a voting patterns of a group of 100 Americans. They are classified in the following *contingency table* according to their party and whether they live in a city or a small town.

	City	Small Town	
Democrats	30	15	45
Republicans	20	35	55
	50	50	100

We define A as the event that a person selected at random lives in the city, and B as the event that a person selected at random is a democrat. We then have the following quantities (under the classical interpretation) that are called the *joint probabilities* $P(A \cap B) = 0.3, P(A^c \cap B) = 0.15, P(A \cap B^c) = 0.2, P(A^c \cap B^c) = 0.35$. The following probabilities are called the *marginal probabilities*: $P(A) = 0.5, P(B) = 0.45$. We also have the conditional probabilities $P(A|B) = 2/3, P(A|B^c) = 4/11, P(B|A) = 0.6, P(B|A^c) = 0.3$. Clearly A, B are not independent (since $P(A|B) \neq P(A)$). The probability that a randomly drawn person is a democrat is 0.45. If we know that we only consider people that live in the country, we condition on A^c and the probability decreases to 0.3. Using conditional probability we can use additional measurements to get a more accurate estimation of the likelihood of events.

Theorem 3 (General Multiplication Rule)**.** *Generalizing the result $P(A \cap B) = P(A|B)P(B)$ we have that (by induction)*

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n) &= P(A_n|A_1 \cap \cdots \cap A_{n-1})P(A_1 \cap \cdots \cap A_{n-1}) \\ &= \cdots = P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1}). \end{aligned}$$

Theorem 4 (The Law of Total Probability)**.** *Suppose that $\{A_1, \ldots, A_n\}$ are sets such that $A_1 \cup \cdots \cup A_n = \Omega$, $A_i \cap A_j = \emptyset$ for all $i \neq j$ (such sets are called a partition of Ω). Then for every event B we have*

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

Proof. By the third axiom of probability,

$$P(B) = P((B \cap A_1) \cup \cdots \cup (B \cap A_n)) = \sum_{i=1}^{n} P(A_i \cap B) = \sum_{i=1}^{n} P(A_i)P(B|A_i).$$

□

Definition 3. *Let $\{A_i\}_i$ be a finite or infinite sequence of events. If every pair $A_i, A_j, i \neq j$ is independent, we say that the events $\{A_i\}_i$ are pairwise independent. If for every finite subset of events $A_{i_1}, \ldots, A_{i_k}$, we have*

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$$

we say that the events are independent. Pairwise independence is a strictly weaker condition that independence.