Statistics 851 (Fall 2013) September 20, 2013
Prof. Michael Kozdron

Lecture #8: Independence and Conditional Probability

Definition. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. The events $A, B \in \mathcal{F}$ are said to be *independent* if

$$\mathbf{P}\{A \cap B\} = \mathbf{P}\{A\} \cdot \mathbf{P}\{B\}.$$

A collection $(A_i)_{i \in I}$ is an *independent collection* if every finite subset J of I satisfies

$$\mathbf{P}\left\{\bigcap_{i \in J} A_i\right\} = \prod_{i \in J} \mathbf{P}\{A_i\}.$$

We often say that (A_i) are *mutually independent*. Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two sub-σ-algebras of $\mathcal{F}$. We say that $\mathcal{A}_1$ and $\mathcal{A}_2$ are *independent* if

$$\mathbf{P}\{A_1 \cap A_2\} = \mathbf{P}\{A_1\} \cdot \mathbf{P}\{A_2\}$$

for every $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$.

Example 8.1. Let $\Omega = \{1, 2, 3, 4\}$ and let $\mathcal{F} = 2^\Omega$. Define the probability $\mathbf{P} : \mathcal{F} \to [0, 1]$ by

$$\mathbf{P}\{A\} = \frac{|A|}{4}, \quad A \in \mathcal{F}.$$

In particular,

$$\mathbf{P}\{1\} = \mathbf{P}\{2\} = \mathbf{P}\{3\} = \mathbf{P}\{4\} = \frac{1}{4}.$$

Let $A = \{1, 2\}$, $B = \{1, 3\}$, and $C = \{2, 3\}$.

- Since
$$\mathbf{P}\{A \cap B\} = \mathbf{P}\{1\} = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbf{P}\{A\} \cdot \mathbf{P}\{B\}$$
we conclude that A and B are independent.

- Since
$$\mathbf{P}\{A \cap C\} = \mathbf{P}\{2\} = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbf{P}\{A\} \cdot \mathbf{P}\{C\}$$
we conclude that A and C are independent.

- Since
$$\mathbf{P}\{B \cap C\} = \mathbf{P}\{3\} = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbf{P}\{B\} \cdot \mathbf{P}\{C\}$$
we conclude that B and C are independent.

However,

$$\mathbf{P}\{A \cap B \cap C\} = \mathbf{P}\{\emptyset\} = 0 \neq \mathbf{P}\{A\} \cdot \mathbf{P}\{B\} \cdot \mathbf{P}\{C\}$$

so that A, B, C are NOT independent. Thus, we see that the events A, B, C are *pairwise independent* but not *mutually independent*.

Notation. We often use *independent* as synonymous with *mutually independent.*

Definition. Let A and B be events with $\mathbf{P}\{B\} > 0$. The *conditional probability* of A given B is defined by
$$\mathbf{P}\{A|B\} = \frac{\mathbf{P}\{A \cap B\}}{\mathbf{P}\{B\}}.$$

Theorem 8.2. *Let $\mathbf{P} : \mathcal{F} \to [0,1]$ be a probability and let A, $B \in \mathcal{F}$ be events. If $\mathbf{P}\{B\} > 0$, then A and B are independent if and only if $\mathbf{P}\{A|B\} = \mathbf{P}\{A\}$.*

Proof. To prove this theorem we must show both implications. Assume first that A and B are independent. Then by definition,
$$\mathbf{P}\{A \cap B\} = \mathbf{P}\{A\} \cdot \mathbf{P}\{B\}.$$
But also by definition we have
$$\mathbf{P}\{A|B\} = \frac{\mathbf{P}\{A \cap B\}}{\mathbf{P}\{B\}}.$$
Thus, substituting the first expression into the second gives
$$\mathbf{P}\{A|B\} = \frac{\mathbf{P}\{A\} \cdot \mathbf{P}\{B\}}{\mathbf{P}\{B\}} = \mathbf{P}\{A\}$$
as required. Conversely, suppose that $\mathbf{P}\{A|B\} = \mathbf{P}\{A\}$. By definition,
$$\mathbf{P}\{A|B\} = \frac{\mathbf{P}\{A \cap B\}}{\mathbf{P}\{B\}}$$
which implies that
$$\mathbf{P}\{A\} = \frac{\mathbf{P}\{A \cap B\}}{\mathbf{P}\{B\}}$$
and so $\mathbf{P}\{A \cap B\} = \mathbf{P}\{A\} \cdot \mathbf{P}\{B\}$. Thus, A and B are independent. $\square$

Theorem 8.3. *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, and suppose that $B \in \mathcal{F}$ is an event with $\mathbf{P}\{B\} > 0$. The function $\mathbf{Q} : \mathcal{F} \to [0,1]$ defined by $\mathbf{Q}\{A\} = \mathbf{P}\{A|B\}$ is a probability on $(\Omega, \mathcal{F})$ called the* conditional probability measure given B.

Proof. Define the set function $\mathbf{Q} : \mathcal{F} \to [0,1]$ by setting $\mathbf{Q}\{A\} = \mathbf{P}\{A|B\}$. In order to show that $\mathbf{Q}$ is a probability, we must check both conditions in the definition. Since $\Omega \in \mathcal{F}$, we have
$$\mathbf{Q}\{\Omega\} = \mathbf{P}\{\Omega|B\} = \frac{\mathbf{P}\{\Omega \cap B\}}{\mathbf{P}\{B\}} = \frac{\mathbf{P}\{B\}}{\mathbf{P}\{B\}} = 1.$$
If $A_1, A_2, \ldots \in \mathcal{F}$ are pairwise disjoint, then
$$\mathbf{Q}\left\{\bigcup_{i=1}^{\infty} A_i\right\} = \mathbf{P}\left\{\bigcup_{i=1}^{\infty} A_i \middle| B\right\} = \frac{1}{\mathbf{P}\{B\}}\mathbf{P}\left\{\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right\} = \frac{1}{\mathbf{P}\{B\}}\mathbf{P}\left\{\bigcup_{i=1}^{\infty} (A_i \cap B)\right\}.$$

However, since the (A_i) are pairwise disjoint, so too are the $(A_i \cap B)$. Thus, by countable additivity of the probability $\mathbf{P}$, we see

$$\mathbf{P}\left\{\bigcup_{i=1}^{\infty}(A_i \cap B)\right\} = \sum_{i=1}^{\infty}\mathbf{P}\{A_i \cap B\} = \sum_{i=1}^{\infty}\mathbf{P}\{A_i|B\}\,\mathbf{P}\{B\}$$

which implies that

$$\mathbf{Q}\left\{\bigcup_{i=1}^{\infty}A_i\right\} = \sum_{i=1}^{\infty}\mathbf{P}\{A_i|B\} = \sum_{i=1}^{\infty}\mathbf{Q}\{A_i\}$$

as required. □

Definition. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. A collection of events (E_n) is called a *partition* of Ω if $\mathbf{P}\{E_n\} > 0$ for all n, the events (E_n) are pairwise disjoint, and

$$\bigcup_n E_n = \Omega.$$

Theorem 8.4 (Partition Theorem)**.** *If (E_n) partition Ω and $A \in \mathcal{F}$, then*

$$\mathbf{P}\{A\} = \sum_n \mathbf{P}\{A|E_n\}\,\mathbf{P}\{E_n\}.$$

Proof. Notice that

$$A = A \cap \Omega = A \cap \left(\bigcup_n E_n\right) = \bigcup_n (A \cap E_n)$$

since (E_n) partition Ω. Since the (E_n) are disjoint, so too are the $(A \cap E_n)$. Therefore, by countable additivity of the probability $\mathbf{P}$, we find

$$\mathbf{P}\{A\} = \mathbf{P}\left\{\bigcup_n (A \cap E_n)\right\} = \sum_n \mathbf{P}\{A \cap E_n\}.$$

By the definition of conditional probability, $\mathbf{P}\{A \cap E_n\} = \mathbf{P}\{A|E_n\}\,\mathbf{P}\{E_n\}$ and so

$$\mathbf{P}\{A\} = \sum_n \mathbf{P}\{A|E_n\}\,\mathbf{P}\{E_n\}$$

as required. □

Armed with the partition theorem and the definition of conditional probability, we can now derive Bayes' theorem. Since $A \cap B = B \cap A$ we see that $\mathbf{P}\{A \cap B\} = \mathbf{P}\{B \cap A\}$ and so by the definition of conditional probability

$$\mathbf{P}\{A|B\}\,\mathbf{P}\{B\} = \mathbf{P}\{B|A\}\,\mathbf{P}\{A\}.$$

Assuming that $\mathbf{P}\{A\} > 0$, solving gives

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

If $\mathbf{P}\{B\} \in (0,1)$, then since (B, B^c) partition Ω, we can use the partition theorem to conclude

$$\mathbf{P}\{A\} = \mathbf{P}\{A|B\}\,\mathbf{P}\{B\} + \mathbf{P}\{A|B^c\}\,\mathbf{P}\{B^c\}$$

and so

$$\mathbf{P}\{B|A\} = \frac{\mathbf{P}\{A|B\}\,\mathbf{P}\{B\}}{\mathbf{P}\{A|B\}\,\mathbf{P}\{B\} + \mathbf{P}\{A|B^c\}\,\mathbf{P}\{B^c\}}.$$

More generally, this reasoning leads to the full version of Bayes' theorem.

Theorem 8.5 (Bayes' Theorem)**.** *Let* $(\Omega, \mathcal{F}, \mathbf{P})$ *be a probability space. If* (E_n) *partition* Ω *and* $A \in \mathcal{F}$ *with* $\mathbf{P}\{A\} > 0$*, then*

$$\mathbf{P}\{E_j|A\} = \frac{\mathbf{P}\{A|E_j\}\,\mathbf{P}\{E_j\}}{\sum_n \mathbf{P}\{A|E_n\}\,\mathbf{P}\{E_n\}}.$$

Proof. As above, we have

$$\mathbf{P}\{E_j|A\} = \frac{\mathbf{P}\{A|E_j\}\,\mathbf{P}\{E_j\}}{\mathbf{P}\{A\}}.$$

By the partition theorem, we have

$$\mathbf{P}\{A\} = \sum_n \mathbf{P}\{A|E_n\}\,\mathbf{P}\{E_n\}$$

and so combining these two equations proves the theorem. $\square$