Note: Sample Space and Probability

Author: Hao Chung *last revised: September 23, 2019*

1 Probability Model

A fundamental notion in probability theory is *random experiment*; that is, an experiment whose result cannot be determined in advance. For example, tossing a coin is a random experiment. The result of a random experiment is called *outcome* . For example, the possible outcomes of tossing two coins may be HH, HT, TH or TT. After an experiment, only one of the outcomes will occur (the outcomes are mutually exclusive).

The set of all possible outcomes of an experiment is called the *sample space* of the experiment, which is denoted by Ω. An *event* is a subset of the sample space; that is, a set of outcomes. An event is said to occur if the outcome of the experiment is in the event. Let $\mathcal{F}$ be a set of events that satisfies

1. If $A \in \mathcal{F}$, then $A^{\complement} \in \mathcal{F}$.
2. If $A_1, A_2, \cdots$ is a countable sequence of sets in $\mathcal{F}$, then $\bigcup_i A_i$ is also in $\mathcal{F}$.

The set that satisfies the two conditions above is called *σ-algebra*, so $\mathcal{F}$ is also a σ-algebra.

Definition 1 (probability measure)**.** A *probability measure* $\Pr : \mathcal{F} \to \mathbb{R}$ is a function that assigns a number $\Pr(A)$ to each event $A \in \mathcal{F}$ satisfying the three axioms

1. **Non-negativity.** $\Pr(A) \geq 0$, for any event A.
2. **Normalization.** The probability of the entire sample space Ω is equal to 1, that is, $\Pr(\Omega) = 1$.
3. **Additivity.** If A and B are two disjoint events, then the probability of their union satisfies
$$\Pr(A \cup B) = \Pr(A) + \Pr(B).$$
In general, for any sequence of events $A_1, A_2, \cdots$, that are mutually exclusive, the probability of their union satisfies
$$\Pr\left(\bigcup_i A_i\right) = \sum_i \Pr(A_i).$$

We refer to $\Pr(A)$ as the probability of the event A.

Proposition 2 (union bound[1])**.**
$$\Pr(A \cup B) \leq \Pr(A) + \Pr(B).$$

A *probability space* is a triple $(\Omega, \mathcal{F}, P)$ where Ω is a sample space, $\mathcal{F}$ is a σ-algebra representing a set of events and P is a probability measure.

2 Conditional Probability

Definition 3 (conditional probability)**.** Let A and B be two events with $\Pr(B) \neq 0$. The *conditional*

[1]The union bound is also known as *Boole's inequality.*

probability of A given B is defined by

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

The idea is that if we given the event B is occured, the relative sample space becomes B rather than Ω and conditional probability is now a probability measure on B.

Theorem 4 (Law of Total Probability)**.** *Let $B_1, \cdots, B_n$ be disjoint events that form a partition of the sample space with $\Pr[B_i] > 0$ for all i. Then, for any event A, we have*

$$\Pr(A) = \sum_i \Pr(A \cap B_i) = \sum_i \Pr(A|B_i)\Pr(B_i).$$

Proof. Because $B_1, \cdots, B_n$ form a partition of the sample space Ω, we have

$$\Pr(A) = \Pr(A \cap \Omega) = \Pr\left(A \cap \bigcup_{i=1}^n B_i\right) = \Pr\left(\bigcup_{i=1}^n (A \cap B_i)\right).$$

Because $B_1, \cdots, B_n$ are disjoint events, by additivity axiom, it follows that

$$\Pr\left(\bigcup_{i=1}^n (A \cap B_i)\right) = \sum_{i=1}^n \Pr(A \cap B_i).$$

Finally, by the definition of conditional probability, we have

$$\sum_{i=1}^n \Pr(A \cap B_i) = \sum_{i=1}^n \sum_i \Pr(A|B_i)\Pr(B_i),$$

so we complete the proof. □

Remark 5. Note that we can adopt the law of total probability only when $B_1, \cdots, B_n$ form a partition of the sample space. The equation $\Pr(B|A) \cdot \Pr(A) = \Pr(B)$ is not generally true.

Theorem 6 (Bayes' Theorem)**.** *Let A and B are two events with $\Pr(B) \neq 0$. Then*

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

In general, let $A_1, \cdots, A_n$ are mutually disjoint events, $\bigcup_i A_i = \Omega$ and $\Pr(A_i) > 0$ for all i ($A_1, \cdots, A_n$ is a partition of the sample space). Then, for all i, it holds that

$$\Pr(A_i|B) = \frac{\Pr(B|A_i)\Pr(A_i)}{\sum_{j=1}^n \Pr(B|A_j)\Pr(A_j)}.$$

Proof. A direct calculation shows that

$$\Pr(A_i|B) = \frac{\Pr(A_i \cup B)}{\Pr(B)} = \frac{\Pr(B \cup A_i)}{\Pr(B)} = \frac{\Pr(B|A_i) \cdot \Pr(A_i)}{\Pr(B)} = \frac{\Pr(B|A_i) \cdot \Pr(A_i)}{\sum_{j=1}^n \Pr(B|A_j)\Pr(A_j)}.$$

□

Remark 7. The law of total probability and the Bayes' theorem are the consequence of the definition of conditional probability.

3 Independence

Definition 8 (independent)**.** Two events A and B are called *independent* if

$$\Pr(A \cap B) = \Pr(A)\Pr(B).$$

Proposition 9 (characteristic of independence)**.** *Two events A and B are independent if and only if*

$$\Pr(A|B) = \Pr(A).$$

Proof. By the definition of conditional probability, we have

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

Since A and B are independent, we have $\Pr(A \cap B) = \Pr(A)\Pr(B)$. Thus,

$$\frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A)\Pr(B)}{\Pr(B)} = \Pr(A),$$

so we complete the proof. □

In particular, a collection of events $A_1, \cdots, A_n$ is called *pairwisely indenpendent* if any of two events are indenpendent, that is,

$$\Pr(A_i \cap A_j) = \Pr(A_i)\Pr(A_j), \text{ for all } i \neq j \in 1, \cdots, n.$$

Generally, a collection of events $A_1, \cdots, A_n$ is called *mutually indenpendent* if any subcollection $A_{i_1}, \cdots, A_{i_m}$, it holds that

$$\Pr\left(\bigcap_{j=1}^{m} A_{i_j}\right) = \prod_{j=1}^{m} A_{i_j} \Pr(A_{i_j}).$$

Example 10. Let A_1, A_2 and A_3 be three events. Suppose we have

1. $\Pr(A_1 \cap A_2) = \Pr(A_1) + \Pr(A_2)$
2. $\Pr(A_1 \cap A_3) = \Pr(A_1) + \Pr(A_3)$
3. $\Pr(A_2 \cap A_3) = \Pr(A_2) + \Pr(A_3)$
4. $\Pr(A_1 \cap A_2 \cap A_3) \neq \Pr(A_1) + \Pr(A_2) + \Pr(A_3)$.

Then, A_1, A_2 and A_3 are pairwisely indenpendent but not mutually indenpendent. If we want that they are mutually indenpendent, we need $\Pr(A_1 \cap A_2 \cap A_3) = \Pr(A_1) + \Pr(A_2) + \Pr(A_3)$.

Remark 11*.* The following sentences are quoted from Terence Tao's Wordpress[2]:

> In order to have the freedom to perform extensions every time we need to introduce a new source of randomness, we will try to adhere to the following important dogma: probability theory is only "allowed" to study concepts and perform operations which are preserved with respect to extension of the underlying sample space. (This is analogous to how differential geometry is only "allowed" to study concepts and perform operations that are preserved with respect to coordinate change, or how graph theory is only "allowed" to study concepts and perform operations that are preserved with respect to relabeling of the vertices, etc...)

[2] https://terrytao.wordpress.com/2010/01/01/254a-notes-0-a-review-of-probability-theory/