

## Lecture 4: Aug 26

### Last time

- Set theory (1.1)

### Today

- Axiomatic Foundations (1.2)

## Basics of Probability Theory

When an experiment is performed, the realization of the experiment is an outcome in the sample space. If the experiment is performed a number of times, then

- different outcomes may occur each time
- some outcomes may repeat
- the “frequency of occurrence” of an outcome can be thought of as a probability

However, we **do not** define probabilities in terms of frequencies but instead take the mathematically simpler axiomatic approach. The axiomatic approach is not concerned with the interpretations of probabilities, but is concerned only that the probabilities are defined by a function satisfying the axioms. Interpretations of the probabilities are quite another matter:

- The “frequency of occurrence” of an event is one example of a particular interpretation of probability.
- Another possible interpretation is a subjective one, where we can think of the probability as a belief in the chance of an event occurring.

### Axiomatic Foundations

For each event  $A$  in the sample space  $S$ , we want to associate with  $A$  a number between zero and one that will be called the probability of  $A$ , denoted by  $\Pr(A)$ . The domain of  $\Pr$  is the set where the arguments of the function  $\Pr(\cdot)$  are defined. It is natural to define the domain of  $\Pr$  as all subsets of  $S$ , that is for each  $A \subset S$ , we define  $\Pr(A)$  as the probability that  $A$  occurs. However, there are some technical difficulties to overcome which requires us to familiarize with the following.

**Definition** A collection of subsets of  $S$  is called a *sigma algebra* (or *Borel field*), denoted by  $\mathcal{B}$ , if it satisfies the following three properties:

1.  $\emptyset \in \mathcal{B}$  (the empty set is an element of  $\mathcal{B}$ ).
2. If  $A \in \mathcal{B}$ , then  $A^c \in \mathcal{B}$  ( $\mathcal{B}$  is closed under complementation).
3. If  $A_1, A_2, \dots \in \mathcal{B}$ , then  $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$  ( $\mathcal{B}$  is closed under countable unions).

From Property (1) and (2), we see that the empty set and its complement  $S$  (since  $S = \emptyset^c$ ) are always in a sigma algebra. In fact, they construct the *trivial* algebra  $\{\emptyset, S\}$  which is the smallest sigma algebra.

By DeMorgan's Law, (3) can be replaced by:

$$3'. \text{ if } A_1, A_2, \dots \in \mathcal{B}, \text{ then } \cap_{i=1}^{\infty} A_i \in \mathcal{B}.$$

This is because:

**Example** If  $S$  is finite or countable (where the elements of  $S$  can be put into one to one correspondence with a subset of the integers), then these technicalities really do not arise, for we define for a given sample space  $S$ ,

$$\mathcal{B} = \{\text{all subsets of } S, \text{ including } S \text{ itself}\}.$$

If  $S$  has  $n$  elements, there are  $2^n$  sets in  $\mathcal{B}$  (why?). [hint: for each element, it is either in or out of a subset, so 2 choices].

**Example** Let  $S = (-\infty, \infty)$ , be the real line. Then  $\mathcal{B}$  is chosen to contain all sets of the form

$$[a, b], (a, b], (a, b), \text{ and } [a, b)$$

for all real numbers  $a$  and  $b$ . Also, from the properties of  $\mathcal{B}$ , it follows that  $\mathcal{B}$  contains all sets that can be formed by taking (possibly countably infinite) unions and intersections of sets of the above varieties.

We now define a probability function.

**Definition** Given a sample space  $S$  and an associated sigma algebra  $\mathcal{B}$ , a *probability function* is a function  $\Pr$  with domain  $\mathcal{B}$  that satisfies

1.  $\Pr(A) \geq 0$  for all  $A \in \mathcal{B}$ .
2.  $\Pr(S) = 1$ .
3. If  $A_1, A_2, \dots \in \mathcal{B}$  are pairwise disjoint, then  $\Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \Pr(A_i)$ .

The above three properties are usually referred to as the Axioms of Probability (or the Kolmogorov Axioms, after A. Kolmogorov, one of the fathers of probability theory). Any function that satisfies the Axioms of Probability is called a probability function.

**Example** Consider the simple experiment of tossing a fair coin (just once), so  $S = \{H, T\}$ . A reasonable probability function is the one that assigns equal probabilities to heads and tails, that is,

$$\Pr(\{H\}) = \Pr(\{T\}).$$

Since  $S = \{H\} \cup \{T\}$ , we have, from Axiom 2,  $\Pr(\{H\} \cup \{T\}) = 1$ . Also,  $\{H\}$  and  $\{T\}$  are disjoint, so  $\Pr(\{H\} \cup \{T\}) = \Pr(\{H\}) + \Pr(\{T\})$ . Collectively, we have

$$\begin{aligned}\Pr(\{H\}) &= \Pr(\{T\}) \\ \Pr(\{H\} \cup \{T\}) &= 1 \\ \Pr(\{H\} \cup \{T\}) &= \Pr(\{H\}) + \Pr(\{T\})\end{aligned}$$

Therefore,  $\Pr(\{H\}) = \Pr(\{T\}) = \frac{1}{2}$ .

**Example** If  $S$  is finite or countable (where the elements of  $S$  can be put into 1 – 1 correspondence with a subset of the integers), then these technicalities really do not arise, for we define for a given sample space  $S$ ,

$$\mathcal{B} = \{\text{all subsets of } S, \text{ including } S \text{ itself}\}.$$

If  $S$  has  $n$  elements, there are  $2^n$  sets in  $\mathcal{B}$  (why?). [hint: for each element, it is either in or out of a subset, so 2 choices].

**Example** Let  $S = (-\infty, \infty)$ , the real line. Then  $\mathcal{B}$  is chosen to contain all sets of the form

$$[a, b], (a, b], (a, b), \text{ and } [a, b)$$

for all real numbers  $a$  and  $b$ . Also, from the properties of  $\mathcal{B}$ , it follows that  $\mathcal{B}$  contains all sets that can be formed by taking (possibly countably infinite) unions and intersections of sets of the above varieties.

We now define a probability function.

**Definition** Given a sample space  $S$  and an associated sigma algebra  $\mathcal{B}$ , a *probability function* is a function  $\Pr$  with domain  $\mathcal{B}$  that satisfies

1.  $\Pr(A) \geq 0$  for all  $A \in \mathcal{B}$ .
2.  $\Pr(S) = 1$ .
3. If  $A_1, A_2, \dots \in \mathcal{B}$  are pairwise disjoint, then  $\Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \Pr(A_i)$ .

The above three properties are usually referred to as the Axioms of Probability (or the Kolmogorov Axioms, after A. Kolmogorov, one of the fathers of probability theory). Any function that satisfies the Axioms of Probability is called a probability function.

**Example** Consider the simple experiment of tossing a fair coin (just once), so  $S = \{H, T\}$ . A reasonable probability function is the one that assigns equal probabilities to heads and tails, that is,

$$\Pr(\{H\}) = \Pr(\{T\}).$$

Since  $S = \{H\} \cup \{T\}$ , we have, from Axiom 1,  $\Pr(\{H\} \cup \{T\}) = 1$ . Also,  $\{H\}$  and  $\{T\}$  are disjoint, so  $\Pr(\{H\} \cup \{T\}) = \Pr(\{H\}) + \Pr(\{T\})$ . Collectively, we have

$$\begin{aligned}\Pr(\{H\}) &= \Pr(\{T\}) \\ \Pr(\{H\} \cup \{T\}) &= 1 \\ \Pr(\{H\} \cup \{T\}) &= \Pr(\{H\}) + \Pr(\{T\})\end{aligned}$$

Therefore,  $\Pr(\{H\}) = \Pr(\{T\}) = \frac{1}{2}$ .

## Caculus of Probabilities

We start with some fairly self-evident properties of the probability function when applied to a single event.

**Theorem** If  $\Pr$  is a probability function and  $A$  is any set in  $\mathcal{B}$ , then

1.  $\Pr(\emptyset) = 0$ , where  $\emptyset$  is the empty set;
2.  $\Pr(A) \leq 1$ ;
3.  $\Pr(A^c) = 1 - \Pr(A)$ .

*proof:*