Chapter 2. Probability.

2.1 Probability measures.

Given an experiment and an event A we need to associate a probability, $\Pr(A)$, with the event. The formal (axiomatic) definition of a probability measure below indicates the restrictions we will impose on any such assignment of probabilities to events.

Definition. *A probability measure Pr is a function which assigns probabilities to events (subsets of Ω) and satisfies the following axioms.*
Axiom 1: For every event A, $Pr(A) \geq 0$.
Axiom 2: $Pr(\Omega) = 1$.
Axiom 3: For every sequence $\{A_i\} = \{A_1, A_2, \ldots\}$ of disjoint events ($A_iA_j = \emptyset$ for all $i \neq j$) $Pr(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} Pr(A_i)$.

Axiom 1 states that the probability of an event cannot be negative. Axiom 2 states that when the experiment is conducted something must happen. Axiom 3 states that if an event can be decomposed (partitioned) into a sequence of disjoint subevents (the A_i), then the probability of the event (the union) must be equal to the sum of the probabilities of the disjoint subevents in the partition since there is no overlap among these subevents.

Aside: The event space. *A rigorous definition of a probability measure requires the specification of an event space $\mathcal{A}$ containing all of the events for which Pr is defined. If Ω is finite or countably infinite, then $\mathcal{A}$ can be taken to be the collection of all subsets of Ω. However, if Ω is uncountably infinite, then technicalities arise which require restrictions on $\mathcal{A}$. We will not dwell on the details but will mention these technicalities when they arise.*

2.2. Properties of probability measures.

Theorem 2.1. *$Pr(\emptyset) = 0$.*

Proof. For $i = 1, 2, \ldots$, let $A_i = \emptyset$. Then the sequence $\{A_i\}$ is a sequence of disjoint events and by Axiom 3 $\Pr(\emptyset) = \Pr(\bigcup_{i=1}^{\infty} \emptyset) = \sum_{i=1}^{\infty} \Pr(\emptyset)$. The only value of $\Pr(\emptyset)$ for which this is possible is $\Pr(\emptyset) = 0$. $\square$

Theorem 2.2. *For $n \geq 2$, if $A_1, \ldots, A_n$ are disjoint events, then*

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

Proof. For $i = n+1, n+2, \ldots$, let $A_i = \emptyset$. Then the sequence $\{A_i\}$ is a sequence of disjoint events and by Axiom 3 and Theorem 2.1 $\Pr(\bigcup_{i=1}^{n} A_i) = \Pr(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{n} \Pr(A_i)$. $\square$

Theorem 2.3. *If A and B are disjoint events, then*

$$Pr(A \cup B) = Pr(A) + Pr(B).$$

Proof. Follows from Theorem 2.2 with $n = 2$, $A_1 = A$, and $A_2 = B$. □

Theorem 2.4. *For any event A,*

$$Pr(A^c) = 1 - Pr(A).$$

Proof. Follows from Theorem 2.3 and the fact that $A \cup A^c = \Omega$. □

Theorem 2.5. *For any events A and B,*

$$Pr(A) = Pr(AB) + Pr(AB^c).$$

Proof. Follows from Theorem 2.3 and the fact that $AB \cup AB^c = A$. □

Theorem 2.6. *If $A \subset B$, then*

$$Pr(A) \leq Pr(B).$$

Proof. Follows from Theorem 2.5 and the fact that $A \subset B$ implies that $BA = A$. □

Theorem 2.7. *For any event A,*

$$0 \leq Pr(A) \leq 1.$$

Proof. Follows from Theorem 2.6 and the fact that $\emptyset \subset A \subset \Omega$. □

Theorem 2.8. *For any events A and B,*

$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(AB).$$

Proof. Let A and B be given. The result follows from Theorem 2.2 and the representations $A = AB \cup AB^c$, $B = AB \cup A^cB$, and $A \cup B = AB \cup AB^c \cup A^cB$ of A, B, and $A \cup B$ as unions of disjoint events. □

Theorem 2.9. *For any events A, B, C,*

$$Pr(A \cup B \cup C) = Pr(A) + Pr(B) + Pr(C) - Pr(AB) - Pr(AC) - Pr(BC) + Pr(ABC).$$

Proof. Let A, B, and C be given. Noting that $(A \cup B)C = AC \cup BC$, application of Theorem 2.8 yields $\Pr(A \cup B \cup C) = \Pr(A \cup B) + \Pr(C) - \Pr(AC \cup BC)$. By Theorem 2.8 we also have

$\Pr(A\cup B) = \Pr(A)+\Pr(B)-\Pr(AB)$ and $\Pr(AC\cup BC) = \Pr(AC)+\Pr(BC)-\Pr(ABC)$. Combining these three expressions yields the result. □

Theorem 2.10. *For any events $A_1, \ldots, A_n$,*

$$Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i Pr(A_i) - \sum_{i<j} Pr(A_iA_j) + \sum_{i<j<k} Pr(A_iA_jA_k) + \cdots + (-1)^{n+1} Pr(A_1 \cdots A_n)$$

Proof. This can be proved using a straightforward but tedious inductive argument along the lines of the proof of Theorem 2.9. □

2.3. Discrete sample spaces

A sample space Ω is said to be discrete if it contains a finite or countably infinite number of elementary outcomes. That is, either $\Omega = \{\omega_1, \ldots, \omega_N\}$ for some positive integer N or the elements of Ω can be arranged in a sequence $\Omega = \{\omega_1, \omega_2, \ldots\}$.

A probability distribution on a finite sample space $\Omega = \{\omega_1, \ldots, \omega_N\}$ is an assignment of probabilities to the elementary outcomes (elements) of Ω. More formally, given a finite sample space $\Omega = \{\omega_1, \ldots, \omega_N\}$, a collection of probabilities $p_1, \ldots, p_N$, with $0 \leq p_i \leq 1$ and $p_1 + \cdots + p_N = 1$, determines a probability distribution on Ω with $\Pr(\omega_i) = p_i$ for $i = 1, \ldots, N$. Note that in most situations we can remove any elements with zero probability and there are at least two elements with positive probability, thus, we can assume that $0 < p_i < 1$.

More generally, letting $\Omega = \{\omega_1, \omega_2, \ldots\}$ denote a discrete sample space (finite or countably infinite) a probability distribution on Ω is a sequence $p_1, p_2, \ldots$ of probabilities ($\Pr(\omega_i) = p_i$) with $0 \leq p_i \leq 1$ for all i and $\sum_{i=1}^{\infty} p_i = 1$.

With a discrete sample space $\Omega = \{\omega_1, \omega_2, \ldots\}$ we can define probabilities for every subset of Ω. Given an event A, *i.e.*, given $A \subset \Omega$, the probability of the event A is the sum of the probabilities of the elementary outcomes which belong to A, *i.e.*,

$$\Pr(A) = \sum_{\omega_i \in A} \Pr(\omega_i) = \sum_{\{i:\omega_i \in A\}} p_i.$$

The simplest way to assign probabilities to the elements of a finite sample space $\Omega = \{\omega_1, \ldots, \omega_N\}$ is to assume that the N elementary outcomes are equally probable so that $\Pr(\omega_i) = \frac{1}{N}$ for $i = 1, \ldots, N$. When the N elementary outcomes are assumed equally probable, the probability of an event A is $\Pr(A) = \frac{N_A}{N}$, where N_A is the number of elementary outcomes which belong to A. In other words, with equally probable outcomes, the probability of event A is the ratio of the number of outcomes "favorable" for A to the number of "possible" outcomes. This simple situation is convenient for demonstrating concepts; but, the usefulness of the assumption of a finite sample space with equally probable outcomes as a model for an idealized version of reality is restricted to games of chance and combinatorial problems.

For completeness we will now describe a simple example involving an uncountably infinite sample space and an analog of equally probable outcomes for such a sample space. Consider a spinner atop a disk and an experiment consisting of spinning the spinner and noting the location of its pointer on the circumference of the disk. If the circumference is divided into N arcs, then we can represent an elementary outcome by the appropriate integer in the finite sample space $\Omega = \{1, \ldots, N\}$. On the other hand, if we think of the circumference as a continuum, then, assuming that the disk has radius one, we can represent an elementary outcome as a real number in the uncountably infinite sample space $\Omega = [0, 2\pi)$. With $\Omega = \{1, \ldots, N\}$ we can assign positive probabilities to each of the elementary outcomes and use these to find probabilities of events. However, with $\Omega = [0, 2\pi)$ it is clear that we cannot assign positive probabilities to the elementary outcomes. In this case if we assume that all of the events of interest can be represented as unions of arcs, then we can define probabilities for all possible arcs and use these to find probabilities of events. Now suppose that we have an fair spinner with the property the likelihood of stopping at any particular point on the circumference of the disk is the same for each point. In the finite case, if each arc is of the same length, then, for this fair spinner, we can assume that each of the N elementary outcomes has probability $\frac{1}{N}$. In the case $\Omega = [0, 2\pi)$, with this fair spinner we can assume that for any specified arc of length α the probability of the event that the pointer lands in the arc is equal to $\frac{\alpha}{2\pi}$. With this uniform distribution on the circumference of the disk the probability that the pointer lands at a specified point is zero but the probability that it lands in a specified arc, of length α, which contains the point is $\frac{\alpha}{2\pi} > 0$. Note that if the arc containing the point is made shorter and shorter, then the arc degenerates to the point and the probability $\frac{\alpha}{2\pi}$ approaches zero; thus, the notion of a point having probability zero is consistent with this assignment of probability to an arc.