Last time, we talked about events. Now, we introduce some basic definitions and theorems about probability.

1 More Examples

1.1 Birthday Paradox

Let us suppose that we have a set of $k$ people, and each of them were born on one of 365 days in the year (assume no one was born on Feb. 29).

Then $\Omega = [365]^k$, under the uniform distribution. Let $E$ be the event where no two people have the same birthday. What is $P(E)$?

There are $\frac{365!}{(365-k)!}$ ways of listing $k$ different birthdays. So the probability is

$$\frac{365 \cdot 364 \cdots (365-k+1)}{365^k}.$$  

If $k$ is even 23 or higher, this means that there is a probability of less than $1/2$ that two people have the same birthday!

1.2 Bernoulli Trials

Suppose we wish to flip a loaded coin (heads with probability $p$) until we get heads. Then $\Omega = \{1, 2, \ldots \}$, where each outcome is the number of flips required. What is $P(\omega)$? We need $\omega$ flips iff the first $\omega - 1$ are all tails, and the $\omega$th is heads – this has probability $(1 - p)^{\omega - 1}p$.

2 Boole’s Inequality

Boole’s inequality in some ways is so obvious that it makes one wonder how someone got name recognition for proving it. Yet, it is a very powerful result because the hypotheses required are so weak. Later, we will discuss independence, and show many nice results when certain events are independent. Boole’s inequality does not require independence – that’s why it’s so powerful.

**Definition 2.1.** Let $A_1, A_2$ be events. We define $P(A_1 \lor A_2) = P(A_1 \cup A_2)$ and $P(A_1 \land A_2) = P(A_1 \cap A_2)$.

**Theorem 2.2 (Boole’s Inequality).** If $A_1, \ldots, A_m$ are events in a probability space $\Omega$, then :

$$P(A_1 \lor A_2 \lor \cdots \lor A_m) \leq \sum_{i=1}^{m} P(A_i)$$

**Proof.** The LHS of the formula is:

$$\sum_{\omega \in \bigcup_{i=1}^{m} A_i} P(\omega)$$

And the RHS is
\[ \sum_{i=1}^{m} \sum_{\omega \in A_i} P(\omega) \quad (2) \]

Each \( \omega \) in \( \bigcup_{i=1}^{m} A_i \) appears \textit{exactly} once in (1), and appears \textit{at least} once in (2). Since \( P(\omega) \) is always nonnegative, it follows that (2) is at least as large as (1).

3 Conditional Probability

\textbf{Definition 3.1 (Probability of \( A \) given \( B \)).} Let \( A \) and \( B \) be events in a probability space. We define \( P(\!A|B\!) \), the probability of \( A \) given \( B \), to be \( P(A \land B)/P(B) \).

This definition makes sense, since it tells us how often \( A \) occurs, if we know that \( B \) occurs – if \( B \) occurs, and \( A \) occurs also, then it is the case that \( A \land B \) occurs.

\textbf{Fact 3.2.} Suppose that \( P(A_1|A_2) > P(A_1) \). Then \( P(A_2|A_1) > P(A_2) \). The same holds if we use “\(<\)” or “\(\leq\)” instead of “\(>\)”.

\textbf{Proof.} Here, only the “\(>\)” case is proved. Suppose that \( P(A_1|A_2) > P(A_1) \). Then it follows that \( \frac{P(A_1 \land A_2)}{P(A_2)} > P(A_1) \). Thus, \( \frac{P(A_2 \land A_1)}{P(A_1)} > P(A_2) \), and we are done.

This fact allows us to make the following definition without worrying about the order of \( A_1 \) and \( A_2 \):

\textbf{Definition 3.3 (Independent Events).} Let \( A_1 \) and \( A_2 \) be events in \( \Omega \):

1. \( A_1 \) and \( A_2 \) are independent if \( P(A_1|A_2) = P(A_1) \).
2. \( A_1 \) and \( A_2 \) are positively correlated if \( P(A_1|A_2) > P(A_1) \)
3. \( A_1 \) and \( A_2 \) are negatively correlated if \( P(A_1|A_2) < P(A_1) \).
4. \( A_1 \) and \( A_2 \) are dependent if they are not independent.

\textbf{Theorem 3.4.} If \( A_1 \) and \( A_2 \) are independent events, then \( P(A_1 \land A_2) = P(A_1)P(A_2) \).

\textbf{Proof.} By independence, we have:

\[ P(A_1) = \frac{P(A_1 \land A_2)}{P(A_2)} \]

Therefore,

\[ P(A_1)P(A_2) = P(A_1 \land A_2) \]
Examples:

1. Rolling two dice, $x_1$ and $x_2$. Let $A$ be the event $x_1 = 3$ and $B$ be the event $x_2 = 4$. Then $A$ and $B$ are independent.

2. However, if $A$ is $x_1 \geq 3$ and $B$ is $x_1 \geq x_2$, then $P(B) = \frac{7}{12}$, but $P(B|A) = \frac{1}{2}$, so $A$ and $B$ are not independent.

4 Examples

4.1 Balls in Boxes

Suppose we have $n$ distinguishable balls and $n$ distinguishable boxes, and we throw each ball into a box randomly. A box may contain any number (including 0) of balls.

Then $\Omega = [n]^m = \{(b_1, b_2, \ldots, b_m)\}$, where $b_i$ denotes the box containing ball $i$. $P$ will be the uniform distribution.

Let $E$ be the event “Box 1 is empty”. That is, $b_i > 1$ for all $i$. Let $A_i$ be the event “ball $i$ is not in box 1”. Then $A_i$ and $A_j$ are independent for all $i \neq j$.

Furthermore, $P(A_i) = \frac{n-1}{n}$. Note also that $E = \bigcap_{i=1}^m A_i$, so $P(E) = \prod_{i=1}^m A_i$.

Therefore, $P(E) = \left(\frac{n-1}{n}\right)^m$. If $m = cn$, and $P_n(E)$ be the probability that box 1 is empty when there are $n$ boxes. Then $\lim_{n \to \infty} P_n(E) = e^{-c}$.

4.2 Random Walk

A particle is at point 0 on the real line, and at each second may go either one unit to the left, or one unit to the right. Suppose $n$ seconds pass – so the particle makes $n$ moves total. Suppose that $n = 2m$. What is the probability that after $n$ moves, the particle is back at position zero?

Here, $\Omega$ is the set of all possible sequences consisting of $L$ and $R$ of length $n$, and $E$ is the set of all of these that have precisely $m$ $L$’s. Therefore, we get

$$P(E) = \frac{\binom{n}{m}}{2^n}$$

4.3 Coloring Problem

Let $A_1, A_2, \ldots, A_n$ be subsets of $A$ and $|A_i| = k$ for $1 \leq i \leq n$. If $n < 2^{k-1}$ then we may color the elements of $A$ red or blue so that each $A_i$ contains a red and a blue element.

To see this, let us randomly color $A$. Then $\Omega = \{R, B\}^A$, with the uniform distribution. Let $BAD$ be the set of colorings that fail the desired property. Then we want to show $P(BAD) < 1$. Let $BAD(i)$ be the event $A_i$ is all blue or all red. Then $P(BAD(i)) = 2^{1-k}$. Then:
\[ P(BAD) = P\left( \bigcup_{i=1}^{n} BAD(i) \right) \]
\[ \leq nP(BAD(i)) = n2^{1-k} < 2k - 12^{1-k} = 1 \]

So \( P(BAD) < 1 \) and we are done.

The above proof is an example of the probabilistic method. We show that it is possible to do something by attempting to do this thing randomly, and showing that the probability of success is nonzero. Usually, such proofs give no insight into how to actually do something successfully.

5 Law of Total Probability

**Theorem 5.1.** Let \( B_1, B_2, \ldots, B_n \) be pairwise disjoint events which partition \( \Omega \), for any other event \( A \),

\[ P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i) \]

**Proof.** Exercise. \( \square \)

**Example:**
Suppose we have two crooked dice, so that if the outcome of the first is \( X \) then the outcome of the second \( Y \) satisfies \( Y \) is equally likely to be \( X-1, X, X+1 \) (If \( X = 1 \) or \( X = 6 \), then instead we have two equally likely values)

What is the probability that \( X = Y \), if \( P(X = i) = \frac{1}{6} \) for each \( i \)?