The notion of “likelihood” is intrinsic to our worldview, so I will not belabor its definition. With this intuition we are going to construct a formal system of probability. This will enable us to formulate odds exactly, to agree on what we mean by odds, and to construct large, complex, and powerful odds machines that can contend with models and decisions which exceed our ability to reason intuitively.

To those who think the ratio of English to equations in my lessons is too high, rejoice, for there are many formulas here to learn; to those who fear walls of equations, simply regard equations as unusual English and read them as such.

## Sample Spaces and Events

What are the possible outcomes of a coin flip? Are they “heads” and “tails”? What if the coin lands on its edge? What if a bird snatches it out of the air? Even in situations where certain outcomes are by far the most likely, the messy universe provides infinitely many other outcomes.

Hence, in formulating a model of an experiment we define the outcomes that we are going to consider. This set of outcomes is called the sample space.

\*\*Sample Space</span>: The set of all possible outcomes of some experiment, often denoted <span class="term">S</span>. Corresponds to the 'universe' in standard set algebra.\*\*

Recall that the elements of a set can be pretty much anything. So the sample space of a coin toss could be defined as S = {H, T}, while the outcome of an experiment to measure height could be the positive real numbers [0, \infty) , an infinite set.

In this course and in standardized tests you will come across questions like “what are the odds of rolling an even number” or “what are the odds of drawing two queens and two hearts when choosing four cards.” We need a systematic way to deal with these complex collections of outcomes.

We call any collection of outcomes an event, and rely on set theory to combine and compose events.

\*\*Event</span>: A subset of S, \(A \subset S\). If the outcome is in A, event A has occurred. Hence we can assign odds to collections of outcomes consistent with our assignment of odds to individual outcomes.\*\*

Comprehension Q’s:

1. Define a reasonable sample space for the following experiments:
   1. Rolling two dice
   2. Measuring stellar masses
   3. Recording someone’s purchases and sales in a market
2. For the sample space S = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}, describe the following events using set theory and English:
   1. Event A = An odd number occurs
   2. Event B = A number greater than 3 occurs
   3. Events A and B occur
   4. Events A and not(B) occur

## Probability Functions

A probability function assigns real numbers to events in the sample space. Put another way, it is a map or function f:B -> [0,1], where B is the collection of all possible events. Note that this is not the same as the sample space. If S = {H, T}, B = {{}, {H}, {T}, {H, T}}. From set theory, this set of all subsets of a set is called the power set, denoted P(S).

If you are surprised to meet a function whose domain consists of elements which are not numbers, consider similar maps in familiar experience. The carnival sign “you must be this tall to ride this ride” is a function from positive real numbers measuring height to the set {“Allowed”, “You shall not pass”}.

A probability function must obey the following axioms:

\*\* defn \*\*

Recall from set theory that disjoint sets are those which share no elements.

\*\* defn disjoint sets := Sets A and B are disjoint if A intersect B = null . A collection of sets {A1, A2, … , An} are mutually disjoint if any given pair of them has intersection null.\*\*

Here are some useful identities for the probabilities of two events. They are all derivable from the Kolmogorov Axioms.

\*\* defns \*\*

## Assigning and Calculating Probabilities

It is important to note that there are many possible probability functions which satisfy the K. axioms for a given experiment. For example, the assignments (H, T) -> (1/2, ½) and (H, T) -> (1/4, ¾) are both valid probability functions according to the axioms; however, only the first one is a good model for a fair coin. The business of whether a probability function models an experiment well is the domain of statistics, where we are headed soon. For now we’re concerned with the formal treatment of probability.

Here’s a standard method to assign valid probability functions to experiments in which the outcomes are equally likely.

Example 1: Consider the experiment of drawing a single card from a standard 52 card deck.

What are the odds of selecting a card that is black (spade or club):

1. and odd numbered?
2. or royal?
3. and either greater than 2 or a spade?
4. Let event A = a black card is drawn and event B = an odd numbered card is drawn. Hence we are interested in finding the odds of the intersection of these events. We assign equal probabilities to every card, hence we need only count the number of cards in this intersection and divide it by 52.

Example 2: Consider an experiment of selecting two word phrases of the form (adjective, noun) from these lists:

(“greedy”, “useless”, “haughty”, “forceful”)

(“hat”, “tirade”, “fish”)

For example, “greedy fish” is one outcome.

1. How many elements are in the sample space?
2. If the outcomes are equally likely, calculate the probability the outcome will make someone chuckle. (To each of the n outcomes, decide whether the outcome will make someone chuckle; count the elements in this event; then find the odds of this event).
3. Let B = The event that the outcome includes “fish”. What are the odds someone will chuckle, given that B occurs?