Discrete Probability: Tree Diagrams and Some Common Distributions

Quick Recap

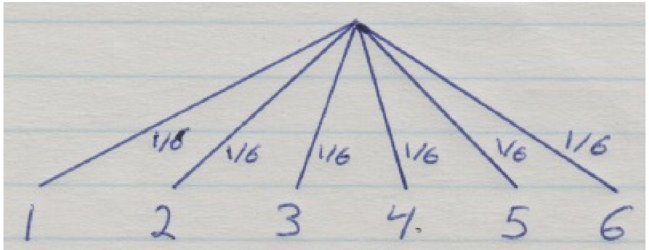
The undetermined *outcome* of an *experiment* involving uncertainty is called a *random variable*, usually represented by the likes of X or Y. The set of all possible outcomes of the experiment is the *sample space* of the experiment, usually denoted by Ω or S. If the sample space is finite or countably infinite, it is said to be *discrete*. A *probability distribution function* for X is a function $m : \Omega \to \mathbb{R}$ such that $m(\omega) \geq 0$ for every outcome $\omega \in \Omega$ and $\displaystyle\sum_{\omega\in\Omega} m(\omega) = m(\omega_1) + m(\omega_2) + m(\omega_3) + \cdots = 1$. $m(\omega)$ is the *probability* of the outcome ω. Note that we're measuring probabilities on a scale of 0 (no chance) to 1 (totally certain), as opposed to using percentages or odds.

An *event* is a collection of possible outcomes, *i.e.* a subset of the sample space, usually denoted by A, B, C, *etc.* The *probability* of an event A, denoted by $P(A)$, is the sum of the probabilities of the outcomes in A, *i.e.* $P(A) = \displaystyle\sum_{\omega\in A} m(\omega)$. The probabilities of events play reasonably nicely with set operations on the events, such as $P\left(\bar{A}\right) = 1 - P(A)$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, and so on.

Tree Diagrams

Tree diagrams are often used to visualize and help analyze experiments in discrete probability, especially experiments involving several stages or components.

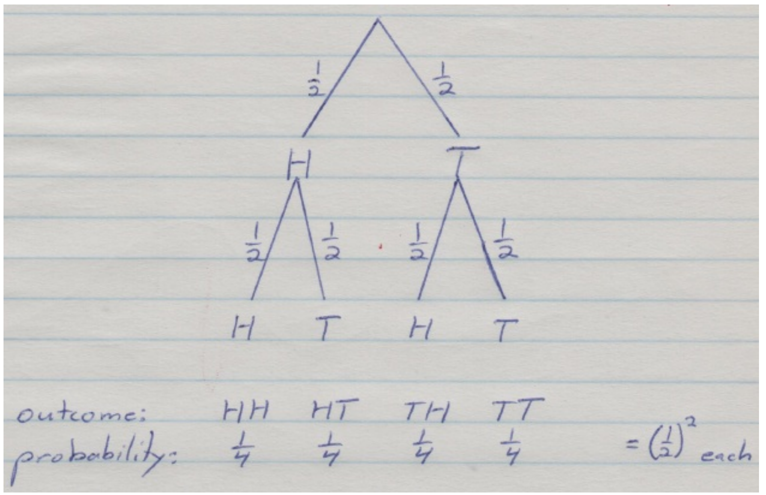
Example 1. For one example, here is tree diagram for the experiment of rolling a fair standard die once:



The tree starts from a single point, representing the start of the experiment, and has a branch leading from the start to each of the six possible outcomes, which are the numbers 1 through 6 that appear on the six faces of the die. Each branch is labelled with the probability that that branch is taken; since the die is fair each of these probabilities is $\frac{1}{6}$.

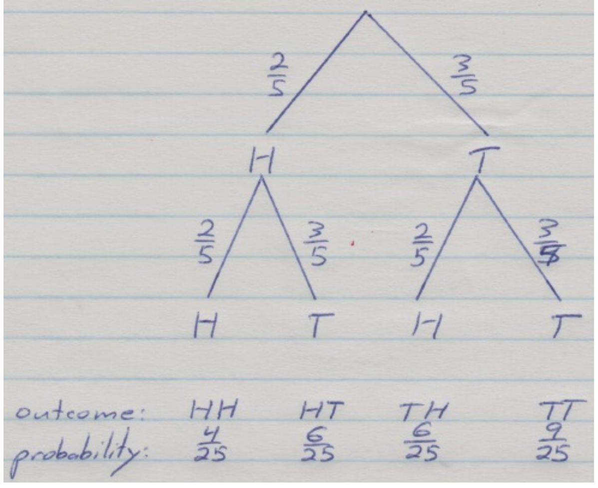
Depending on the application, some features of the full tree diagram could be omitted. If all the branches have equal probability, in particular, it's not uncommon to omit the probabilities attached to each branch. This also often done if you're simply trying to work out what all the outcomes actually are in some multi-stage experiment.

Example 2. For a more interesting example, consider the experiment of tossing a fair coin twice. This experiment has the following tree diagram, with some extra commentary below:



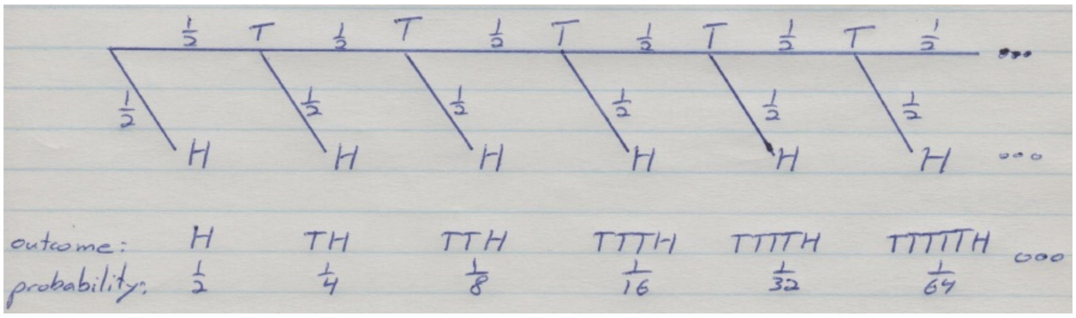
This example illustrates the use of tree diagrams in finding all the outcomes of an experiment and then their probabilities. The possibilities of the first coin toss are given in first row below the start and the possibilities of the second toss are given in the second row. Tracing each possible branch from the start until it ends and writing down the Hs and /or Ts encountered in each branch gives you all the possible outcomes of the experiment. Multiplying the probabilities encountered along each branch together gives you the probability of the corresponding outcome. Once you have the probabilities of the outcomes, you have the probability distribution function for the experiment in hand and can move on to computing the outcomes of events.

Example 3. For a still more interesting example, consider the experiment of tossing an unfair or *biased* coin twice, with the probability of a head being $\frac{2}{5} = 0.4$ and the probability of a tail being $\frac{3}{5} = 0.6$ on each toss. This tree diagram is similar to the previous one, but has different probabilities:



This time the method of finding the probability of each outcome by multiplying together the probabilities encountered along the branch for that outcome is really useful because not all the outcomes have equal probabilities.

Example 4. For yet another example where not all the outcomes have equal probability, consider the experiment of tossing a fair coin until it comes up heads. In this case the tree diagram looks like:



Since the sample space is actually countably infinite instead of finite, we cannot hope to draw the entire tree, so we just draw enough to reveal the patterns that arise. Since the tree is very narrow, it saves space to write this (part of the) tree horizontally instead of vertically ...

Discrete Uniform Distribution

One frequently encountered – and easy to handle – type of discrete probability distribution is the *discrete uniform distribution*, often just called a uniform distribution. (That can be a bit confusing since we will later also encounter the continuous uniform distribution, which are also often abbreviated to just uniform distribution.) This the case when our experiment has a finite sample space and every outcome is equally likely. Examples 1 and 2 above are examples of discrete uniform distributions, whereas Examples 3 and 4 are not. Example 3 may have a finite sample space but the outcomes do not have equal likelihoods, and Example 4 has neither a finite sample space nor has equal likelihood for every outcome.

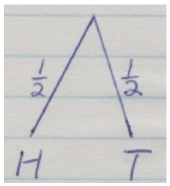
Note that because the probabilities of outcomes have to be at least 0, the sum of the probabilities of all the outcomes has to be 1, and each outcome is as likely as any other, the probability distribution function for a discrete uniform distribution must be $m(\omega) = \frac{1}{n}$ for every outcome ω, where $n \geq 1$ is the number of outcomes in the sample space. Similarly, if A is an event in such a distribution, then $P(A) = \frac{k}{n}$, where k is the number of outcomes in the event A.

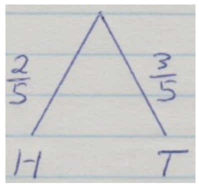
Note also that one cannot have a discrete uniform distribution on a countably infinite sample space. Suppose one assigned some positive number a to be the equal probability that all the outcomes in an infinite sample space have. Then the total probability of all the outcomes would be infinite instead of 1. (What happens if one adds a positive number to iself infinitely many times?) If one tried assigning every outcome a probability of 0 instead, the total probability of all the outcomes would still be 0, which also not equal to 1. (The sum of any number of 0s is still 0.)

Bernoulli Trials

An experiment with just two outcomes, not necessarily equally likely, is called a *Bernoulli trial**. It is traditional to call one outcome *success* and the other *failure*, and denote the probability of success by p and the probability of failure by $q = 1 - p$. Note that we must have $0 \leq p \leq 1$, and it's pretty boring if $p = 0$ (no success ever) or $p = 1$ (no failure ever), so we usually assume that $0 < p < 1$.

Example 5. Of course, all that this amounts to tossing a (possibly biased) coin. Here are the tree diagrams for tossing a fair coin and a biased coin (with $m(H) = \frac{2}{5}$ and $m(T) = \frac{3}{5}$), respectively.





Of course, it is up to you define which of heads or tails is "success" and which is "failure". Whichever way you do it, you'll have $p = \frac{1}{2} = q$ for the fair coin. For the biased coin, defining success to be heads and failure to be tails gives $p = \frac{2}{5}$ and $q = \frac{3}{5}$, but defining success to be tails and failure to be tails gives $p = \frac{3}{5}$ and $q = \frac{2}{5}$ instead.

Geometric Distribution

The experiment of repeating a particular Bernoulli trial until you get a success gives rise to a *geometric distribution*. Suppose the underlying Bernoulli trial has probability of success p and probability of failure $q = 1 - p$. In this case, it is traditional to have our random variable X count the number of repetitions required to finally get a success, so the sample space consists of the positive integers: $\Omega = \{1, 2, 3, 4, \dots\}$ and the probability distribution function is given by $m(1) = p$, $m(2) = qp$, $m(3) = q^2p$, $m(4) = q^3p$, and so on. In general, $m(k) = q^{k-1}p$.

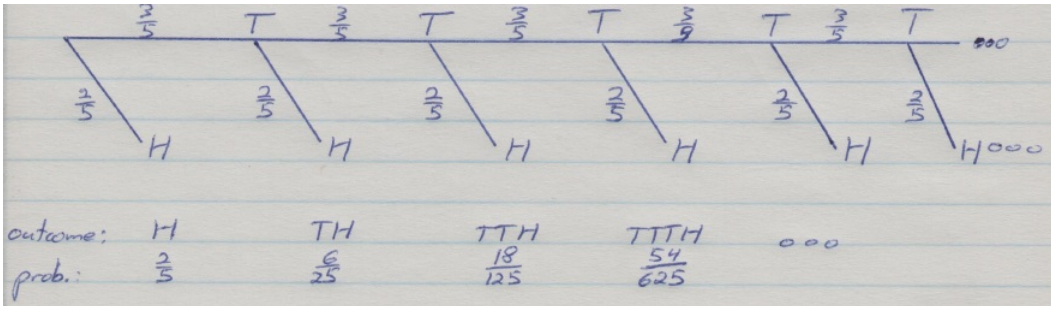
Is this a valid probability distribution function? If $0 < p < 1$ and $q = 1 - p$, then we also have $0 < q < 1$, so $m(k) = q^{k-1}p > 0$ too for every outcome $k \geq 1$. To see that the sum of the probabilities of all the outcomes is 1, consider

$$\sum_{\omega \in \Omega} m(\omega) = \sum_{k=1}^{\infty} m(k) \sum_{k=1}^{\infty} q^{k-1}p = p + qp + q^2p + q^3p + q^4p + \cdots .$$

This is a *geometric series* with initial term $a = p$ and common ratio $r = q$, where $|r| = |q| = q < 1$, so it has a sum of $\dfrac{a}{1-r} = \dfrac{p}{1-q} = \dfrac{p}{1-(1-p)} = \dfrac{p}{p} = 1$, as required.

* Named after Jacob Bernoulli (1654-1705) who made many contributions to calculus and was a pioneer in the study of probability. One of his brothers, Johann, was also a prominent mathematician. Johann and another brother, Nicolaus (an artist), had a number of mathematicians among their descendants, some of whom also worked on probability.

Example 6. For example, suppose we have a biased coin which has a probability of $\frac{2}{5}$ of coming up heads and a probability of $\frac{3}{5}$ of coming up tails on any given toss, just as in Examples 3 and 5. If we toss the coin until it comes up heads, we get the following tree diagram:



If we declare heads to be success and tails failure, this amounts to having a geometric distribution with $p = \frac{2}{5}$ and $q = 1 - p = \frac{3}{5}$.