December 30, 2024 Draft

Physical Principles in Biology

Biology 3550/3551

Spring 2025

Chapter 2: Probability

David P. Goldenberg
University of Utah
goldenberg@biology.utah.edu

Chapter 2

Probability

2.1 An example of a random process: Brownian Motion

I. Some history

Although we tend to think of biological movement, especially that of animals, as being rather directed, at microscopic scales motion in both living and non-living systems can be quite random. The discovery of this kind of motion is credited to a Scottish botanist, Robert Brown (1783–1858). Brown served on a expedition to Australia in 1801–1805 and spent many years afterwards characterizing the plants that he collected on this expedition. Brown was a particularly skilled microscopist, and in 1826 described the motion of tiny particles within pollen grains. He was not the first to observe this kind of motion, but others who had seen it assumed that it reflected some kind of living process. By carefully describing the motions and showing that they could be seen in materials that were clearly not living (like particles of coal dust suspended in water), Brown showed that the motions represented a physical process, rather than a biological one.

A movie of small particles, $\approx$100 nm diameter, undergoing Brownian motion is available online:
https://www.youtube.com/watch?v=cDcprgWiQEY

A theoretical model explaining Brownian motion was presented by Albert Einstein in 1905. This was just one of four major papers that Einstein published in his *annus mirabilis* (miracle year). The others concerned special relativity and the photoelectric effect (for which he won the Nobel Prize in 1921). The paper on Brownian motion would have probably made the reputation of just about any other scientist, but for Einstein it seems almost a footnote.

Einstein's explanation for Brownian motion was that particles in a liquid are constantly being bumped into by molecules, and each collision causes a small motion of the particle. Every once in a while, an imbalance in the number of molecules colliding from one side or the other causes a larger movement in a random direction. This is illustrated in the drawing below:

This drawing comes from an online simulation of Brownian motion.
http://galileoandeinstein.physics.virginia.edu/more_stuff/Applets/Brownian/brownian.html

On the left, a large particle moves randomly because of collisions of more rapidly moving small molecules. On average, the forces on the particle average to zero, but at any instant, there may be an imbalance of collisions from different directions, leading to a small motion. The right-hand panel shows the trajectory of the particle over time, as viewed at lower magnification.

Importantly, Einstein did not just describe this model qualitatively (which others before him had done), but developed a mathematical treatment that made quantitative predictions that could be tested by experiments. Experimental confirmation of this theory provided critical support for the existence of atoms and molecules, an idea that was still contentious at the beginning of the 20th century.

II. A mathematical description - random walks

1. A detailed, exact mathematical description of this process, with explicit descriptions of the behavior each molecule, would be almost impossible.
2. An important aspect of science is deriving an abstract description of a process that captures important elements but is as simple as possible. To quote Einstein (roughly): Theories should be a simple as possible, without being so simple that they fail to account for important observations.
3. The key element of Brownian motion is that in a given time interval, there is an equal probability that a particle will move one way or the opposite way.
4. We can describe the overall behavior of a particle undergoing Brownian motion as a random walk: A process made up of multiple steps separated by random changes in direction.
5. A one-dimensional random walk:
   - Flip a coin.
   - If the coin lands heads-up, take a step to the right. If the coin lands tails-up, take a step to the left.

- Repeat.

6. The Galton probability machine or "Galton board" - aka Plinko: A mechanized demonstration of a one-dimensional random walk, or any process made up of a sequence of binary random events.

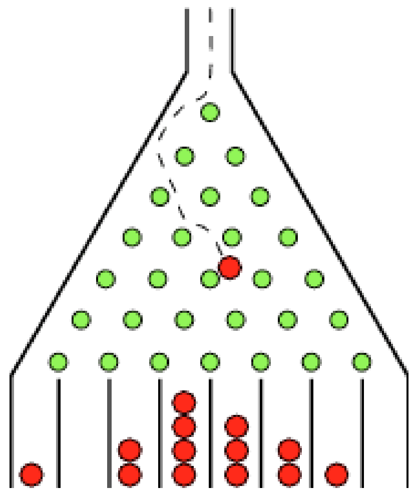


Illustration from http://mathworld.wolfram.com/GaltonBoard.html

- A triangular array of pegs placed on a vertical board. A ball is dropped on to the top peg and bounces to the left or right with equal probability. At each row, the ball hits a peg and moves to the left or right. The balls are collected in bins below the bottom row.
- Devised by Sir Francis Galton, 1822-1911: A cousin of Charles Darwin. Galton played an important role in developing the mathematical description of genetic variation and evolution, but was also a major advocate of the idea that society could be improved by the selective breeding of humans and gave this idea the name eugenics.
- Also known as a probability machine or "Plinko".
- Computer simulation:
  https://phet.colorado.edu/en/simulation/plinko-probability

2.2 Introduction to probability theory

I. Some introductory comments.

In terms of its relevance to science and every day life, probability is arguably one of the most important branches of mathematics. But, it is also has a bit of an odd position within mathematics, and it is, I think, severely under represented in our undergraduate curriculum. It is also one of the most challenging subjects to learn and teach.

1. Why is probability a misfit in the world of mathematics?

If you think about the traditional branches of mathematics, they are generally concerned with the properties of certain kinds of abstract objects:

- Geometry: lines, circles, polygons, planes, spheres, polyhedra, *etc.*
- Number theory: Integers
- Algebra: Polynomials
- Calculus (or analysis for the purists): Functions that change smoothly (usually).

Although all of these have applications in the real world, these branches of mathematics can be discussed completely in the abstract, and that is the way most mathematicians like it!

Probability, on the other hand, deals specifically with the description of real events of a certain type (or models of those events). In particular, probability deals with events about which we are, to some degree, ignorant. We use probability to describe things that have uncertain outcomes. If you think about it, there are lots of things like that!

2. Why is probability so difficult?

   A. One problem is that we constantly use the language of probability in our everyday lives, without necessarily paying attention to exactly what we mean. Some common expressions of a probabilistic nature:

      - It is likely that . . .
      - Chances are . . .
      - I'll bet that . . .

      We are also accustomed to hearing numbers associated with such statements, such as "There will be an 80% chance of rain tomorrow." What do statements like this mean, and where do they come from?

   B. Another problem is that discussions of unpredictable events often have large emotional component. For instance:

      - What are the chances that I will win the lottery?
      - What is the probability that I will get cancer?

      The probabilities of these events might (or might not) be similar to the probability of rain tomorrow, but our emotional responses to them are likely very different.

   C. The calculation of probabilities often involves some rather tricky counting, and the results often contradict our intuition.

   D. The answer to a probability question can depend on exactly how the question is framed. Make sure that you are answering the right question!

II. A coin toss

A typical probabilistic statement: If I toss a coin, the chances it will land heads-up are the same as the chances it will land heads-down.

What is implied by this statement?

- Ignorance: I don't actually know which way the coin will land.
- Knowledge: If I toss the coin a large number of times, the number of times it lands heads-up will be approximately the same as the number of times it lands tails-up.

These answers raise some more questions:

- Why don't I know which way the coin will land? Isn't this just Newtonian mechanics?
- How many times do I have to toss the coin before the number of heads will equal the number of tails? Will they ever be exactly equal?
- Can I say anything more specific about the expected pattern of heads and tail?

For now, we will try to address just one of these questions: Why can't I predict the outcome? The answer is that the final outcome (heads or tails) is extremely sensitive to a large number of small factors that we usually don't have control over. These factors include the exact force applied to the coin, the angle at which the force is applied, any air currents that affect the coin, exactly how the coin hits the surface when it lands. In principle if all of these factors could be controlled and measured, it should be possible to predict the outcome of the toss.

To some degree, the uncertainty of a coin toss is tied to the structure of the coin: The thin edge makes it almost certain to fall one way or the other, and the (near) symmetry makes it equally likely to fall either way. Of course, the coin may be bent or otherwise altered so that the probabilities of heads and tails are not equal.

III. A bit of mathematical formalism.

In order to develop a mathematical theory of probability, we have to make some careful definitions of quantities that we can manipulate. This will seem a bit much for a simple coin toss, but the definitions are important for keeping us straight as we move on to more complicated cases.

1. **Outcomes** For a given experiment, we define a set of distinct *outcomes*. For the coin toss, we define two outcomes, heads (H) and tails (T). Now, we could also consider other outcomes, like dropping the coin, but that makes things more complicated. So, what we usually do is simplify the situation by excluding things like dropping the coin or that it might land on its edge.
2. **Probabilities** For each of the possible outcomes, we define a *probability*, a number (p) constrained such that:
   - p for any given outcome must lie between 0 and 1, inclusive.
   - The sum of the probabilities for all of the possible outcomes is 1.

   For the coin toss, our experience and intuition says that:

   $$p(H) = 1/2$$
   $$p(T) = 1/2$$

What, exactly do we mean by this? This is not quite as obvious as it sounds, and there are actually two major ways of interpreting probabilities, reflecting some rather deep philosophical differences among probabilists. We will use the more intuitive and traditional view, called a "frequentist" interpretation.

A. The frequency interpretation of the statement, $p(H) = 1/2$, is simply that if a "fair" coin is tossed a large number of times, the fraction of times it lands heads-up will be approximately $1/2$, and the fraction will, over time, get closer to $1/2$ as the number of tosses is increased.
   This general trend is called the "law of large numbers".

B. The alternative interpretation of probabilities is called "Bayesian", referring to Thomas Bayes, an 18$^{\text{th}}$ century cleric and mathematician, who devised a very important equation concerning the probabilities of related events. Frequentists do not dispute Bayes' equation, but the Bayesians interpret and apply it more broadly. In brief, the Bayesian approach is used for situations in which we are asking questions for which there are not enough data to make a frequency estimate, such as, "What is the probability that it will rain tomorrow?" Since there has never been even one day exactly like tomorrow, there is no way to know the frequency of rain on such days. But, if we have an initial estimate of the probability, called a *prior* probability, additional information can be used with Bayes' equation to refine the initial estimate to create a *posterior* probability. The Bayesian approach is somewhat controversial, but it has become a very important tool in areas in which exact probabilities are not known, but there is a large amount of data with which to refine the estimates. One common example is filtering e-mail messages for spam.

   For our purposes, the frequency interpretation is most useful. It has a relatively intuitive basis, and most of the problems we will be considering, such Brownian motion and diffusion, involve large numbers of random events.

3. **Sample spaces** We call the set of all possible, distinct outcomes for a given experiment a *sample space*, S. For a single coin toss:

$$S = \{H, T\}$$

We will use curly braces, as above, to enclose the elements of the set. This gets a little more complicated when we consider more complicated experiments, such as multiple coin tosses. For two coin tosses, there are four possible outcomes, and we will define the sample set as:

$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

where the outcomes are defined as ordered pairs, in parentheses, representing the results of the two independent coin tosses.

This is a little bit arbitrary. We could define three outcomes defined in terms of the total number of heads or tails, irrespective of the order:

- Two heads: $2H$
- Two tails: $2T$
- One heads, one tails: $1H1T$

But, the case of $1H1T$ is actually the combination of two outcomes, (H, T) and (T, H), as initially defined.

The major difference between these two ways of defining the outcomes is that the probabilities of the individual outcomes are all equal for the first definition, but not for the second.

In general, we try to define the outcomes and the sample set to make assigning probabilities as simple as possible. This doesn't necessarily mean that the probabilities are all equal, though.

For the plinko, we would define the sample set as the set of all distinct paths through the pegs, not the set of all possible final bins.

Although there might be different ways of defining a sample space for a particular experiment, it must satisfy two requirements:

- The set must be complete, *i.e.,* it must include every possible way that things can end up.
- The items in the set must not overlap.

A consequence of these two requirements is that the sum of the probabilities of the outcomes must be exactly 1.

4. **Events** Formally, an *event* is defined as a subset of the sample set, *i.e.,* a set of zero or more of the possible outcomes.

   For example, with two coin tosses, we could define the events that we considered above:

   - Two heads: $2H = \{(H, H)\}$
   - Two tails: $2T = \{(T, T)\}$
   - One heads, one tails: $1H1T = \{(H, T), (T, H)\}$

   For the plinko, we could define an event as the ball falling into a given bin.

   As noted above, we generally try to define outcomes so that the probabilities can be easily calculated, and then use those probabilities to calculate the probabilities of events, or groups of outcomes.

   The choice of words, "outcomes" and "events", is pretty arbitrary, but the distinction between the two kinds of groupings is important. The outcomes of an experiment are events, but there are usually other events that can be defined as groups of outcomes. The outcomes defined in the sample space must satisfy the requirements specified earlier: They must include all possible outcomes of the experiment, and the sum of their probabilities are one, while there are no general requirements for events.

   We can often define a variety of different events, some of which may overlap. For instance we could define an event such that there is at least one heads.

   $$1^+H = \{(H, T), (H, H), (T, H)\}$$

This event overlaps the events $2H$ and $1H$.

There is no requirement that a set of events be complete or non-overlapping.

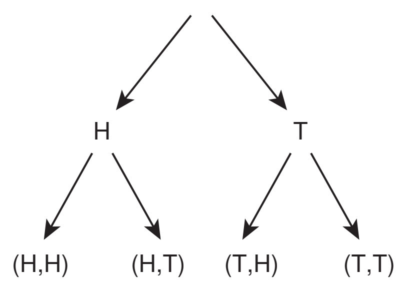
Often, it is the probabilities of events, as defined here, that is most important. For instance, we care about which bin the plinko ball falls in, but not necessarily the specific path it takes there. Thus, we often want to be able to calculate the probabilities of events from the probabilities of outcomes.

IV. Multiplying and adding probabilities.

Provided that we are careful in defining the sample space, the rules for calculating the probabilities of other events are relatively simple. For this discussion, it is useful to introduce another term, *trial*, to indicate a single probabilistic process or experiment. A trial that can have only two outcomes is referred to as a *binary trial* or *Bernoulli trial*, for the Swiss mathematician Jacob Bernoulli (1665–1705). It is also natural to refer to individual trials as events, but this leads to confusion with the definition of events as subsets of the sample set.

1. Sequential independent trials - the product rule.

   We can think of the experiment composed of two coin tosses as two sequential experiments, or trials, each with the sample space, $S = \{H, T\}$. In fact, just about any complicated process can be broken down in this fashion. It is often useful to draw a tree representation of the sequential trials, like the one below:



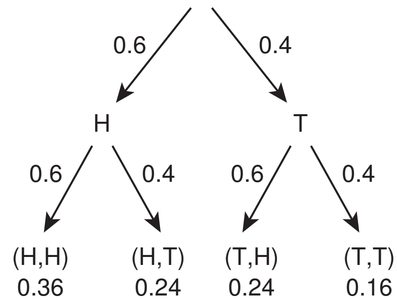
   It's not a coincidence that this looks like the plinko, but there is an important difference: All of the different outcomes are kept separate.

   For the first coin toss, $p(H) = p(T) = 1/2$. Therefore, if we do this experiment many times, we expect heads for the first toss half of the time. Consider just this half, for a moment. For the second toss, we also expect heads half of the time. So, out of all of the two-toss experiments, we expect the outcome (H, H) $1/2 \times 1/2$ of the time. Therefore, $p(H, H) = 1/4$. The same argument can be made for all of the outcomes of this experiment.

   The general statement of this result is that if we have two sequential and independent trials, then we can calculate the probabilities of the final outcomes of the compound experiment as the products of the individual probabilities. We call this the *product rule*. We can extend it to compound experiments of any length.

   Application of the product rule is often associated with the word "and." For instance, the outcome (H, H) can be described as "heads for the first toss *and* heads for the second toss."

In this particular case, the product rule leads to the conclusion that all of the outcomes in the sample space have equal probabilities, 1/4. But this is not always the case. Suppose that we are playing with a coin that has somehow been messed with so that the probability of landing heads-up is 0.6 and the probability of landing tails-up is 0.4. We can still use the same arguments and the product rule:



Notice that the sum of the probabilities is still 1.

2. Groups of non-overlapping events - the addition rule.

   A. We have already used this rule implicitly.
   Consider the event we defined earlier, $1H1T$, *i.e.*, one heads and one tails, irrespective of order. This event is a composite of two outcomes:

   $$1H1T = \{(H,T),(T,H)\}$$

   The probability of $1H1T$ is calculated as the sum of the outcomes:

   $$p(1H1T) = p((H,T)) + p((T,H))$$

   Just as we said that the product rule is associated with the word "and", we can say that the addition rule is associated with "or". The event $1H1T$ can be described as being the result when (H,T) *or* (T,H) is the outcome.
   If $p(H) = p(T) = 1/4$, then $p(1H1T) = 1/2$.

   B. What is $p(1H1T)$ if $p(H) = 0.6$? What can we say in general about $p(1H1T)$ if $p(H)$ is not equal to $p(T)$?
   Consider two extreme cases:

   - $p(H) = 0$ and $p(T) = 1$. Then:

   $$\begin{aligned} p(1H1T) &= p((H,T)) + p((T,H)) \\ &= p(H)P(T) + p(T)P(H) \\ &= 0 \end{aligned}$$

   - $p(H) = 1$ and $p(T) = 0$. It should be apparent that $p(1H1T) = 0$, again.

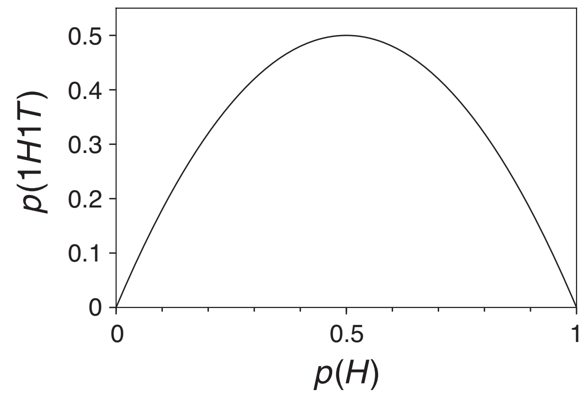
   We can write a general expression for $p(1H1T)$ as a function of $p(H)$, assuming that the two coin tosses are equivalent:

   $$\begin{aligned} P(1H1T) &= p((H,T)) + p((T,H)) \\ &= 2p((H,T)) \\ &= 2p(H)p(T) \end{aligned}$$

We also know that $p(T) = 1 - p(H)$, so:

$$\begin{aligned} p(1H1T) &= 2p(H)(1 - p(H)) \\ &= 2p(H) - 2p(H)^2 \end{aligned}$$

A graph of this function:



With a little bit of calculus, you should be able to confirm that 1/2 is the maximum probability of one heads and one tails. If the coin is biased either way, the probability is less.
Another example: Consider the event we defined earlier, one or more heads.

$$1^+H = \{(H,T), (H,H), (T,H)\}$$

We calculate the probability of this event as the sum of the probabilities of the three outcomes it represents:

$$p(1^+H) = p((H,T)) + p((H,H)) + p((T,H))$$

If the coin is fair, each of the outcomes has equal probability, and $p(1^+H) = 3/4$.
But, there is an even easier way to get this result. The only outcome that is not included in 1^+H is (T,T). Since the sum of the probabilities of all outcomes must be 1:

$$\begin{aligned} p(1^+H) &= 1 - p((T,T)) \\ &= 1 - p(T)p(T) \end{aligned}$$

If the coin is fair, then $p(T) = 1/2$ and $p(1^+H) = 3/4$. Sometimes it is important to consider which probabilities will be the easiest to calculate.

V. A final comment about independent events and the law of large numbers.

Consider the case of a long string of coin tosses. Suppose that 10 straight tosses turn up heads. Someone offers you a bet: If the next toss turns up tails, she will pay you \$1, if it turns up heads, you pay her \$1. Is this a "better-than-even" bet?

The law of large numbers says that eventually the numbers of heads and tails will be close to equal. So, is it time for the coin to show up tails?

No! The coin doesn't know or care about the law of large numbers! Each toss is independent, so the probability of tails for the 11^{th} toss is the same as for the first toss.

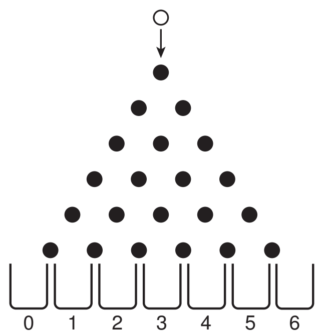
Thinking that "it's time for a tails" is known as the "gambler's fallacy", and has cost many people lots of money over the ages!

But, is there another way of thinking about this situation? What have we assumed about the coin (or its tosser)? If that assumption is called into doubt, how does that change our assessment?

2.3 Plinko probabilities: 6 rows

I. Formulation of the problem

A 6-row plinko:



The white circle represents the ball, and the black circles represent the pegs in the path of the ball.

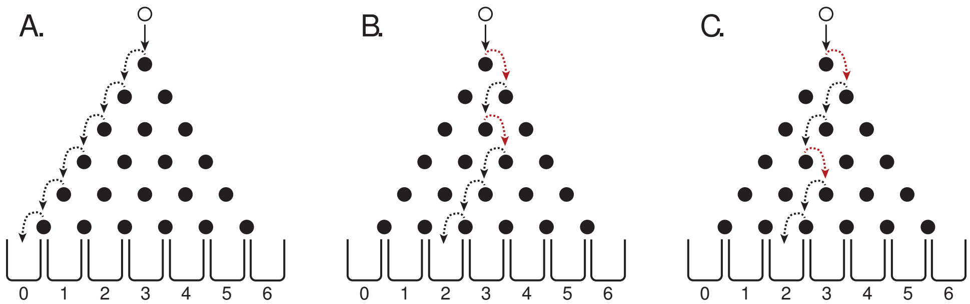
For a general n-row plinko, the bottom row of pegs will contain n pegs. Since a ball can fall to the right or left of each peg, there are $n + 1$ final positions, or buckets, for the balls to fall into. For convenience in what comes later, we label the buckets from 0 to n, or 0 to 6 for the 6-row plinko.

II. Outcomes We have some discretion in defining the outcomes and sample set, so long as we follow the basic rules:

- The outcomes in the sample set must include all possible outcomes.
- None of the outcomes in the sample set can overlap any other outcome.
- The sum of the probabilities of all of the outcomes in the sample set must equal 1.

At first glance, it might make sense to define seven outcomes, corresponding to a ball falling in bucket 0, 1, 2, 3, 4, 5 or 6. We know already, however, that the probabilities of these seven outcomes are not equal, and we will find that calculating them is rather involved. So, instead, we will start by defining the outcomes as all of the possible paths of a ball through the plinko, which all have equal probabilities. Then, we will use the elements in the sample set to calculate probabilities for the events corresponding to a ball landing in each of the buckets.

A few of the outcomes, individual paths, are shown below:



Notice that the paths labeled B and C both lead to bucket 2, but we are treating these as separate outcomes. Notice, also, that both of these paths include two turns to the right, where as the path to bucket 0 includes 0 turns to the right. More generally, any path leading to bucket k must include exactly k turns to the right.

First, we calculate the number of outcomes in the sample set and their probabilities. When the ball hits the single peg in the top row, there are two possible turns, left or right. Similarly, when the ball hits one of the pegs in the second row, there are two possible turns. Each of these turns are independent, just like a series of coin flips. Therefore, the total number of paths is equal to 2^n, where n is the number of rows. For the six-row plinko the number of outcomes is $2^6 = 64$. If the probability of a right or left turn at each peg is equal (0.5), then the probabilities of all of the outcomes are equal, and are equal to one divided by the number of possible outcomes. Thus the probability of each outcome for an $n-$row plinko is 2^{-n}. For the six-row plinko,the probability is 1/64.

III. Events

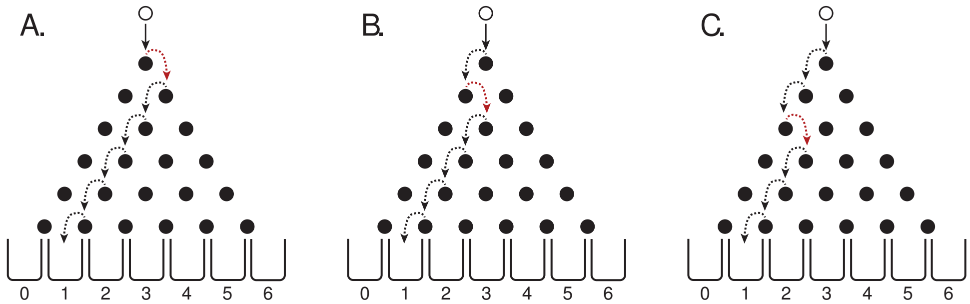
Next, we consider the events corresponding to the ball falling in one of the seven buckets, which we will call $E0$, $E1$, $E2$, $E3$, $E4$, $E5$ and $E6$. One way that we could do this is to write out all of the outcomes (paths) and sort these into those for which the ball lands in bucket 1, 2 and so forth. This would be quite tedious, however, and we would like to be able to do this for much larger numbers of steps. Therefore, we want a more general and efficient way to solve this sort of problem.

1. Paths to buckets 0 and 6

   If we consider first the possible paths to bucket 0, we quickly realize that the ball will reach this bucket only if all of the turns are to the left, as shown in panel A in the figure above. So, the probability of landing in bucket 0, $E0$, is 1/64. Similar reasoning can be applied to conclude that there is only one path to bucket 6, also with a probability of 1/64.

2. Paths to buckets 1 and 5

   In order for the ball to land in bucket 1, the ball must make 1 turn to the right and 5 to the left. Three such paths are shown below:

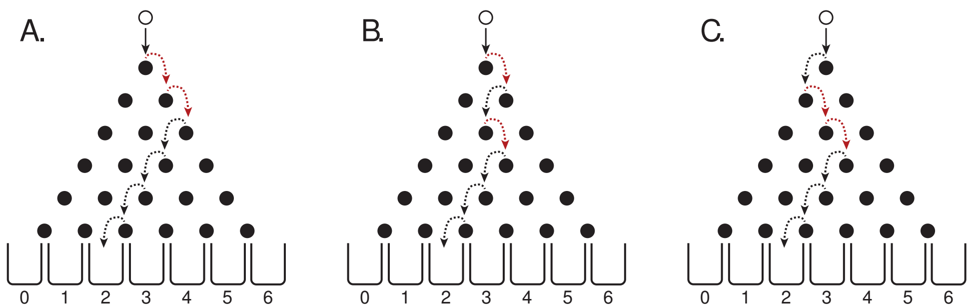


Since there are six rows, at each of which the single turn to the right can occur, there must be six different paths to bucket 1, making up the event $E1$. So, the probability of $E1$ is $6/64 = 3/32$. The same reasoning can be applied to the paths to bucket 5, and the probability of $E5$ is $3/32$.

This result can be generalized to say that for an n-row plinko, there are n paths to bucket 1 and to bucket $n-1$.

3. Paths to buckets 2 and 4

   Things get more complicated when we consider paths to bucket 2, where we must enumerate the possible paths that include exactly two turns to the right. The two turns can occur at any of the six rows, as shown in a few examples:



   In panels A and B, the first turn is to the right, and the second turn to the right is at row 2 (A) or row 3 (B). In panel C, the first turn is to the left, and the two turns to the right occur in rows 2 and 3, followed by 3 turns to the left.

   To determine the number of paths to bucket 2, without drawing them all out, we can calculate the number of paths as follows:

   - Consider the number of positions for the first turn to the right. This can happen at rows 1 through 5. (If the first turn to the right occurs at row 6, there is no chance for a second turn to the right.) If the first turn to the right is at row 1, then the second can occur at rows 2 through 6, corresponding to 5 paths. This is analogous to a 5 row plinko and the number of paths to bucket 1.
   - If the first turn to the right is at row 2, there are only 4 rows left at which the second right turn can occur.
   - Generalizing, the further down the ball moves before the first right turn, the fewer rows there are where the second right turn can occur. Specifically, if the

first turn to the right occurs at row i, then there are $i-1$ possible locations for the second turn.

- For a 6-row plinko the total number of paths to bucket 2 is calculated as:

$$5+4+3+2+1=15$$

By considering the number of ways of placing two *left* turns, we can conclude that there are also 15 paths to bucket 4.

4. Paths to bucket 3

We can now almost fill a table showing the number of baths to each of the buckets

Bucket	Paths
0	1
1	6
2	15
3	
4	15
5	6
6	1

Since we have already concluded that the total number of paths to all of the buckets is 64, and 44 paths are accounted for so far, there must be 20 paths to bucket 3.

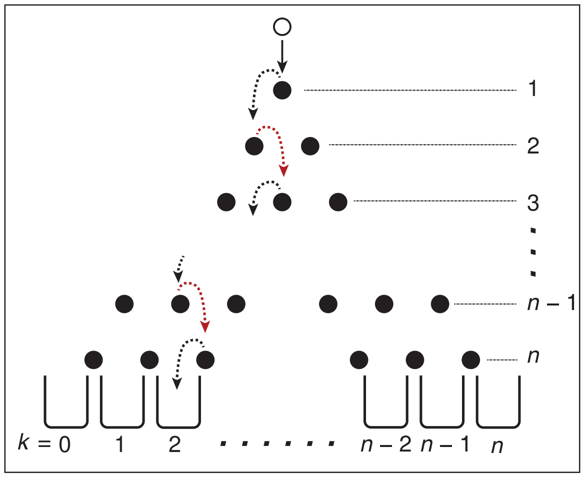
The number of paths and probabilities for all of the buckets can now be listed:

Bucket	Paths	Probability
0	1	1/64
1	6	3/32
2	15	15/64
3	20	5/16
4	15	15/64
5	6	3/32
6	1	1/64

Though we have been able to solve the problem for the 6-row plinko without too much trouble, you will likely guess, correctly, that enumerating all of the paths gets more and more complicated as the number of rows increases. To generalize the solutions to problems of this type of problem, we need to take a different approach.

2.4 Plinko probabilities: The general case for n rows

To keep track of the rows and buckets in the general case of an n-row plinko, we will label them as shown below:



Before trying to solve the general form of this problem, it is useful to step back and look at things a bit differently, and also consider some related probability problems.

I. Another way to count the paths to bucket 2 in a 6-row plinko.

Recall that we concluded that any path to bucket 2 must include 2 turns to the right and 4 turns to the left. A seemingly sensible (but flawed) way of looking at this would be to say that the first turn to the right can occur at any of the 6 rows, and the second turn to the right can occur at any of the 5 rows that are remaining. So, using the product rule, we would calculate the number of paths to bucket 2 as:

$$6 \times 5 = 30$$

Notice that this is twice the number that we calculated earlier! The reason for this is that this calculation has ignored the fact that one of the turns to the right has to come before the other. For instance, for the path that includes right turns at rows 2 and 5, the turn at row 2 has to come first. But, in our second calculation we included both this path and one in which the turn at row 5 comes before the one at turn 2, which is physically impossible! More generally, by simply taking the product of 6 and 5, we have counted twice each of the 15 paths that we counted earlier. But, if we take this into account, and divide 30 by 2, we get the right answer.

So a general strategy might be to calculate the number of all possible placements of the right turns, without worrying at first about the order of the turns, and then correct for the requirement that order does matter.

For the case of bucket 3, we can start by considering (ignoring order) that there are 6 rows where the first turn to right can occur, 5 where the second can occur and 4 where the third can occur. So the total number of paths (with over-counting) is:

$$6 \times 5 \times 4 = 120$$

But, how do we determine how many paths have been over-counted?

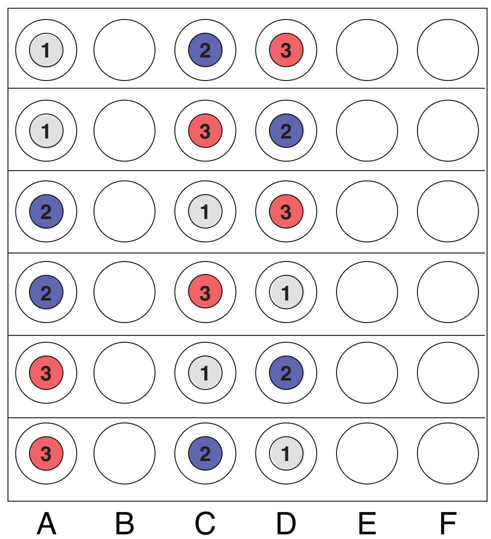
II. Labeled beans in cups

Though the connection may not be apparent just yet, it is useful to consider another type of problem that is popular among probabilists.

Suppose that we have 3 beans, each labeled with a number; 1, 2 or 3, and six cups. How many distinguishable ways are there to place one bean in one of the six cups? This is basically the same as the previous problem: There are 6 possible cups for the first bean, 5 for the second and 4 for the third. So the number of distinguishable different arrangements is:

$$6 \times 5 \times 4 = 120$$

The important point here is that these have *not* been over-counted, because the three beans are distinguishable. For instance, the following 6 arrangements are distinct:



For the general case of k labeled beans in n cups (assuming that $k \leq n$), the number of distinguishable arrangements is:

$$n(n-1)(n-2)\cdots(n-k+1)$$

You should be able to see where the first part of this product comes from, but it may not be so obvious that $(n-k+1)$ is the correct place to end the multiplication. So, you should try out a few examples to convince your self. For instance if $n = 10$ and $k = 6$, $(n-k+1) = (10-6+1) = 5$, and the number of distinct arrangements is:

$$\begin{aligned} n(n-1)(n-2)\cdots(n-k+1) &= 10 \times 9 \times 8 \times 7 \times 6 \times 5 \\ &= 151,200 \end{aligned}$$

There are six terms in the product, corresponding to the 6 labeled beans, and the final term, 5, represents the five empty cups available for the last bean. Notice, also, how quickly the number of possible arrangements has increased with a few more beans and cups!

III. The factorial function, permutations and combinations

Products like the ones used above arise frequently in probability and other areas of mathematics, and there is a function that is particularly useful for working with them. The factorial function is defined only for the integers greater than or equal to 0 and the factorial function of integer k is written as $k!$. The function is defined as:

$$k! = \begin{cases} 1, & \text{if } k = 0 \\ n(n-1)(n-2)\cdots 2\cdot 1, & \text{if } k > 0; \end{cases}$$

Defining 0! as 1 may seem arbitrary (Why isn't it 0?), but this is important in order for the function to behave well when 0! appears.

An immediate application of the factorial function is that $n!$ is the number of ways arranging n labeled beans in n cups. This represents the special case, with $k = n$, of arranging k labeled beans in n cups. From the previous page, the number of distinct arrangements is:

$$\begin{aligned} n(n-1)(n-2)\cdots(n-k+1) &= n(n-1)(n-2)\cdots(n-n+1) \\ &= n(n-1)(n-2)\cdots 2\cdot 1 \\ &= n! \end{aligned}$$

A distinct way of ordering all of the elements in a set is called a *permutation*. The items might be beans with distinct numbers, marbles with different colors or molecules with distinguishable covalent structures or conformations. So, we can say, "There are $k!$ permutations of k labeled beans." This is written as:

$$P(k) = k!$$

Note that we are using the upper-case P here to distinguish permutations from probabilities, written with the lower-case p. Another (mathematically equivalent) example of a set of permutations begins with k labeled marbles in a bag, and we draw all k of them from the bag. There are $k!$ different orders in which the marbles can be drawn.

An extension to this idea is to consider the number of ways of drawing k marbles from a bag starting with $n \geq k$ marbles. Strictly speaking, these are not permutations if $n > k$, because not all n elements are used, but they are often referred to as "k-permutations" of n, written as $P(k, n)$. The number of sequences is calculated as:

$$P(k, n) = n(n-1)(n-2)\cdots(n-k+1)$$

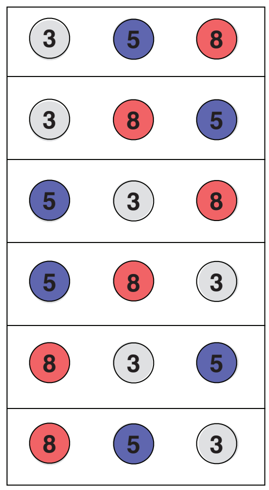
Another way of writing this is:

$$P(k, n) = \frac{n(n-1)\cdots(n-k+1)(n-k)(n-k-1)\cdots 2\cdot 1}{(n-k)(n-k-1)\cdots 2\cdot 1} = \frac{n!}{(n-k)!}$$

This is equivalent to the problem we considered in the previous subsection, the number of distinct ways of distributing k labeled beans into n cups, with only one bean per cup.

Now, we have a nice compact way of writing the result. And, if we have a calculator or computer programmed to calculate the factorial function, it is quite easy to do the calculation.

The term permutation is sometimes confused with *combination*. A combination is a distinct way of selecting a subset of a collection without regard to order. For instance, we might have a bag of 10 marbles, labeled 1 through 10, and without looking, choose 3 of them. From above, we know that there are $P(3,10) = 720$ distinct ways of choosing the three marbles, *if* we treat the different orders of choosing the marbles as distinct from one another. For instance, there are 6 ways of choosing the marbles labeled 3, 5 and 8:



These represent the 6 permutations ($P(3) = 3!$) of the chosen marbles. For any other set of three marbles, there are also 6 permutations. Suppose that, after the 3 marbles have been drawn, the labels were to disappear. The six permutations of each group of 3 marbles would be indistinguishable, and the order in which they were drawn would no longer be discernable. So, to calculate the number of combinations in which 3 marbles can be drawn from a bag of 10, we can do the following:

- Calculate the number of ways in which 3 marbles can be drawn, distinguishing among the different possible orders. This is calculated as:

$$P(3,10) = \frac{10!}{3!} = 720$$

- Calculate the number of ways in which 3 labeled objects can be ordered:

$$P(3) = 3!$$

- Divide the number of ways 3 objects can be drawn from 10 (distinguishing different orders) by the number of ways 3 objects can be ordered:

$$\frac{P(3,10)}{P(3)} = \frac{10!}{(10-3)!} \div 3! = \frac{10!}{8!3!} = \frac{3,628,800}{5040 \cdot 6} = 120$$

To generalize, the number of ways of choosing k objects from a set of n, without distinguishing the order, is calculated as:

$$\frac{n!}{k!(n-k)!}$$

We will come back to say a little more about this function, and its more general applications on page 46.

In the meantime . . .

IV. Back to the plinko

Back on page 39, we suggested that a general way of calculating the number of paths to bucket k in an n-row plinko would be:

- Calculate the number of possible placements of the k turns to the right, without regard to order, realizing that this will lead to over-counting the real number of paths.
  This is is analogous to placing k labeled beans in n cups, with no more than one bean per cup. We have shown above that this is calculated as:

  $$P(k,n) = \frac{n!}{(n-k)!}$$

- Correct the number calculated above by recognizing that, for k turns placed at k specific rows, only one order is physically possible.

  For this part of the calculation, first consider the number of ways of placing k right turns in k specific rows. This is the number of permutations of k objects, $P(k) = k!$. But, we know that only one of these represents a physically possible pathway through the plinko. So, to get the total number of paths with k right turns, we divide the number of possible placements of the k turns to the right, without regard to order, $P(k,n)$, by $P(k)$:

  $$\frac{P(k,n)}{P(k)} = \frac{n!}{k!(n-k)!}$$

Thus, we have the desired result, the number of paths to bucket k in an n-row plinko is calculated as:

$$\frac{n!}{k!(n-k)!}$$

To test this result, you should apply it to the case of the 6-row plinko, for which we previously calculated the number of paths to each bucket (page 38).

The expression that we have derived for the number of paths to bucket k arises in a variety of situations and is commonly written as:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

and spoken as "n choose k". It represents the number of ways of choosing k objects from a set of n, when either:

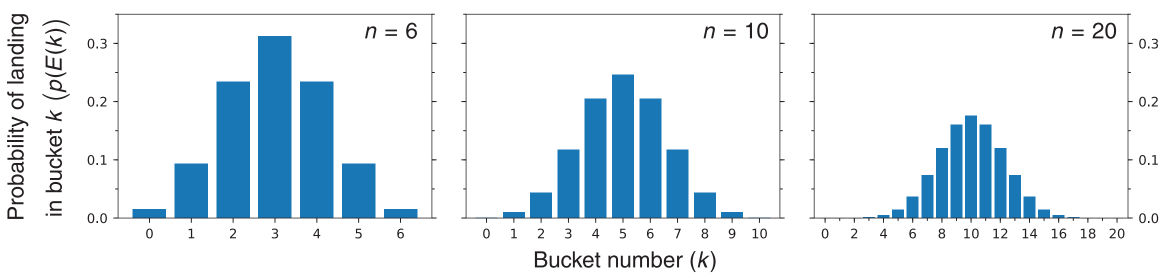
- Only a single order is valid (*i.e.,* the plinko) or
- The order doesn't matter at all (k unlabeled beans placed in n cups, with only one bean per cup allowed).

The total number of paths through an n-row plinko is calculated by multiplying the number of alternatives from the single peg in row 1 (2) by the number of alternatives from a peg in row 2 (2), and then multiplying by the number of alternatives from a peg in row 3 (2), and so on. Thus, the number of paths is 2^n. If, at each peg, the probabilities of turning right or left are equal, then all of the paths will have equal probabilities, equal to 2^{-n}.

The probability of landing in bucket k, that is event $E(k)$, is the sum of the probabilities for all of the paths leading to the bucket. If all of these paths have the same probability, 2^{-n}, then the probability of landing in bucket k is:

$$p(E(k)) = \frac{n!}{k!(n-k)!}2^{-n}$$

The probabilities for plinkos with 6, 10 and 20 rows are shown as bar graphs in the figure below:



Some things to note about these graphs are

- Each graph has the familiar "bell-curve" shape that arises frequently in a variety of contexts.
- As the number of rows, n, increases the maximum probability decreases, as the balls are spread out into more buckets.
- Also as n increases, it becomes increasingly unlikely that a ball will land in one of the buckets near the left or right end. As a fraction of the total number of buckets, the distribution of balls becomes more concentrated towards the center.

We will consider all of these features in more detail as we see the same type of distribution arise in different contexts.

2.5 Biased plinkos

So far, we have assumed that the probability of a ball falling to the left or right at any peg is equal. This assumption leads to the conclusion that all of the paths through the plinko have equal probabilities, and that the different probabilities for landing in the different buckets are only due to the different *numbers* of paths leading to the different buckets. But, things get more interesting when we consider that the probabilities of left and right turns might be unequal for some or all of the pegs.

Suppose that all of the pegs are not quite round, so that the probability of turning to the right, p_R, is 0.6, and the probability of turning left, $p_\mathrm{L} = (1 - p_\mathrm{R})$, is 0.4. For now, we will consider a specific case of a 10-row plinko, starting with the paths leading to bucket 3. The fact that the pegs are biased doesn't change the number of paths leading to a specific bucket, which we calculate as

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{10!}{3!(10-3)!} = 120$$

Now, consider the fact that right and left turns have different probabilities. For each of the 120 paths to bucket 3, there are 3 turns to the right and 7 turns to the left. This represents an "and" situation: The ball must take 3 right turns AND 7 left turns. So, to calculate the probability of each path, we have to multiply the probabilities for 3 right turns and 7 left turns.

$$p_\mathrm{path}(3) = p_\mathrm{R}^3 \cdot p_\mathrm{L}^7$$

Note that the placements of the 3 right turns and 7 left turns does not matter in this context, and this expression applies to all of the paths that lead to bucket 3. Since a ball can land in bucket 3 by any of the paths with equal probability (an "or" situation), the probability of landing in that bucket is the *sum* of all of the probabilities for the individual paths:

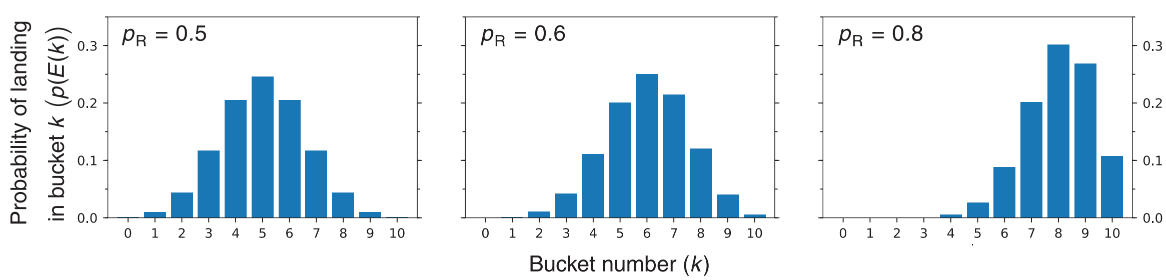
$$p(E(3)) = \frac{10!}{3!(10-3)!} p_\mathrm{R}^3 \cdot p_\mathrm{L}^7$$

For the case where $p_\mathrm{L} = 0.4$ and $p_\mathrm{R} = 0.6$, the probability of a ball landing in bucket 3 is 0.0425, compared to 0.117 for the unbiassed plinko. The bias of each turn towards right has moved the overall distribution of probabilities towards the right, reducing the probabilities of falling on the left-hand side of the plinko.

The more general expression, for bucket k in an n-row plinko is:

$$p(E(k)) = \frac{n!}{k!(n-k)!} p_\mathrm{R}^k \cdot p_\mathrm{L}^{n-k}$$

The bar graphs bellow show the effects of making the right turns progressively more favored, for the case of the 10-row plinko.



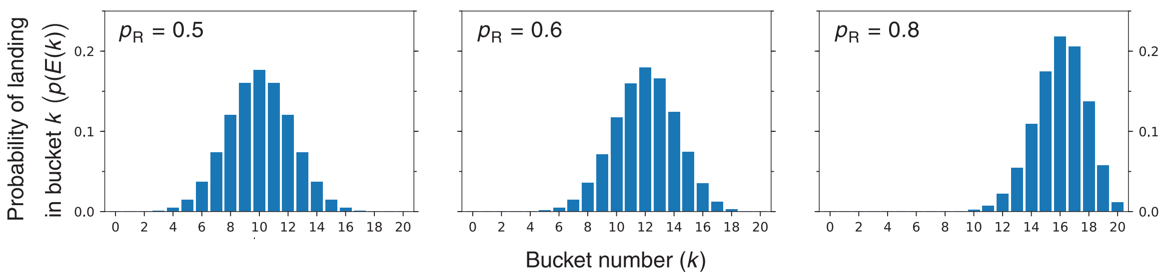
As the graph shows, increasing the probability of a turn to the right at each peg, leads to a progressive shift of the overall distribution to the right. Whereas the probability of a ball landing in bucket 10 is about 0.1% when there is no bias, this probability increases to about 10% if p_{R} is increased to 0.8.

When the probabilities of right and left turns are not equal, there are two competing factors that determine the distribution:

- A statistical factor favoring the central buckets, because there are more paths available toward these buckets than towards the buckets near the left and right hand edges of the plinko.
- A "forcing" factor that causes a systematic tendency towards one side of the plinko or the other.

The forcing factor can be adjusted by changing the relative values of p_{R} and p_{L} as shown in the graphs above. The statistical factor, on the other hand, can be modified by changing the number of rows in the plinko. For instance, with 10 rows, there are 252 paths to the central bucket, as compared to 1 for each of the buckets on the edge and 10 for the buckets one in form the edges. With 20 rows, there are 184,756 paths to the central bucket, as compared to 1 to each of the buckets on the edge and 20 to the buckets one in from the edges. Thus, the statistical bias towards the center is much greater for the 20-row plinko.

The graphs below show the effects of increasing biases to the right for a 20-row plinko.



As expected from the arguments above, the distribution is still shifted towards the right, but the buckets at the far right side are not nearly as favored as they are in the 10-row plinko, because the statistical "resistance" to the bias is greater.

One can imagine other ways in which the plinko could be biased, with only selected pegs with unequal values of p_{R} and p_{L}. Think of some examples of this kind and see if you can calculate the probabilities for these scenarios.

2.6 Binomial coefficients, Pascal's triangle and the binomial distribution function

I. Binomial coefficients in algebra

Recall that the general expression that we found for calculating the number of paths to bucket k in an n-row plinko is:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

This expression arises in a variety of contexts, and the values generated from it are most commonly called *binomial coefficients*. This term reflects their appearance in algebra in the expansion of binomials, which have the general form of:

$$(a+b)^n$$

where n is the order of the binomial. The results for expanding the binomial for $n =$ 0 through 6 are shown below

$$(a+b)^0 = 1$$
$$(a+b)^1 = a+b$$
$$(a+b)^2 = a^2 + 2ab + b^2$$
$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$
$$(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$$
$$(a+b)^5 = a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5$$
$$(a+b)^6 = a^6 + 6a^5b + 15a^4b^2 + 20a^3b^3 + 15a^2b^4 + 6ab^5 + b^6$$

If you examine the coefficients for the 6^{th}-order expanded binomial, you will find that they are exactly the same as the number of paths to the buckets in the 6-row plinko (page 38).

This may seem an odd coincidence, but there is an underlying connection. In the plinko, the number of paths to bucket k reflects the number of ways of combining k turns to the right with $n-k$ turns to the left, and the most paths are found at the center ($k = n/2$ when n is even, or $k = n/2 - 0.5$ and $k = n/2 + 0.5$ when n is odd). In a binomial expansion, the coefficients reflect the number of ways of multiplying together a k times and b $(n-k)$ times, to generate products of the form $a^k b^{n-k}$.

The binomial theorem, in its simplest form, is the equation:

$$(x+a)^n = \sum_{k=0}^{n} \binom{n}{k} x^k a^{n-k}$$

where x and a are real numbers, and n is a positive integer. For a discussion of extension of the theorem to other classes of numbers, see: https://en.wikipedia.org/wiki/Binomial_theorem

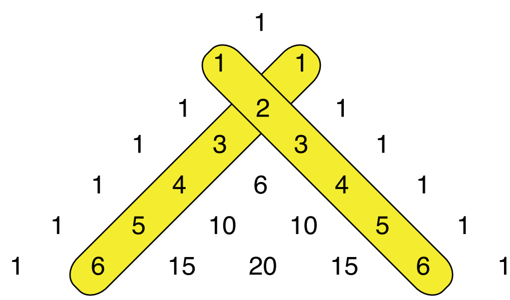
II. Pascal's triangle

The binomial coefficients for increasing values of n can be laid out in a triangle as shown below:

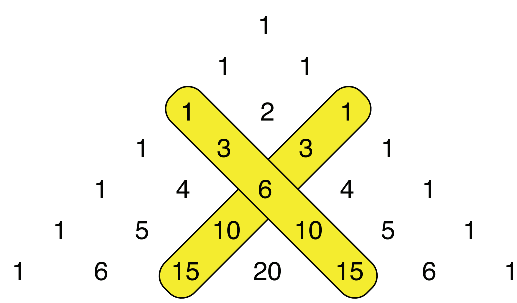
```
                                      Row
                        1· · · · · · · · ·0
                    1       1· · · · · · ·1
                1       2       1· · · · ·2
            1       3       3       1· · ·3
        1       4       6       4       1·4
    1       5       10      10      5    1·5
1       6       15      20      15     6  1·6
```

If the rows in the triangle are labeled from zero for the top row, row n contains the coefficients for the binomial expansion of order n. Although this representation was known before the 2nd century BC, it was brought to prominence in the western world by the French mathematician, Blaise Pascal, in a book published in 1665, two years after his death. It is now most commonly identified as *Pascal's Triangle*.

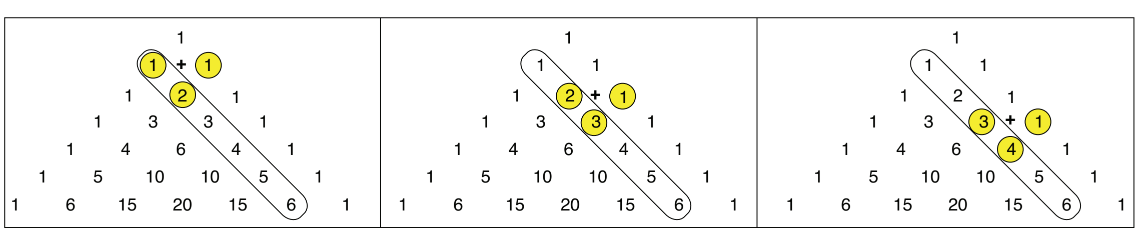
When the coefficients are laid out as a triangle, some interesting patterns become apparent. For instance, the two diagonals starting from the top of the triangle are composed entirely of 1s. The next diagonals down from the top are made up of the sequence of natural numbers: $1, 2, 3, \ldots$.



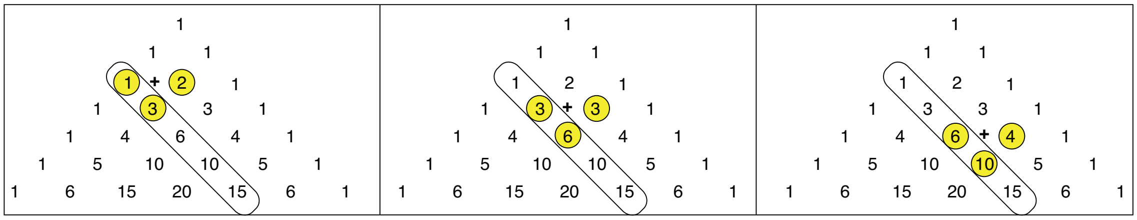
On the other hand, the pattern for the diagonals two down from the top may not seem to show an obvious pattern at first:



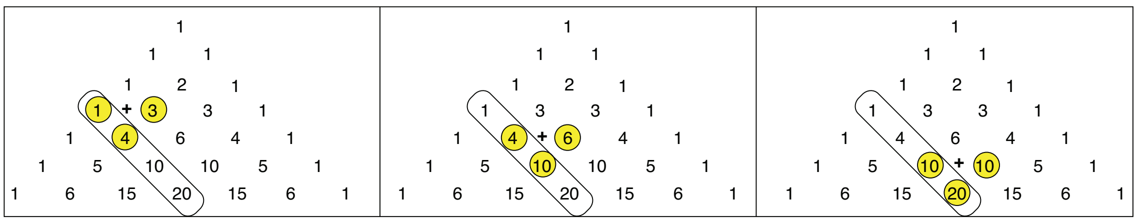
In this case the numbers along the diagonals increase as we move downward, and the increments themseles increase. The series of numbers on each of the diagonals do, in fact, follow a pattern, described by the *figurate numbers*, but that is a rather abstract bit of algebra that we will not pursue. There is, however, an easy way to calculate all of the elements in Pascal's triangle. If we go back to the second diagonal on the right-hand side, we find that each element of the diagonal is the sum of the numbers above and to the sides of it.



The same is true for the elements in the third diagonal on the left:



And, the fourth diagonal:



Indeed, all of the elements of Pascal's triangle can be calculated in this way. So, if we remember only that the outer edges of the triangle are composed entirely of 1s, the rest of the elements in the triangle, to any depth, can be calculated by taking the sum of the two elements above and to the sides of it.

III. The binomial probability distribution function: Trials and successes

The equation that we have derived for for calculating the plinko probabilities has considerably wider applications, and it is also representative of a general class of functions that we call *probability distribution functions*, or *pdf*s (not to be confused with *portable document format*). In particular, this is a class of pdfs described as *discrete probability distribution functions*, which, in general, are used to calculate the probabilities of outcomes or events, for the kinds of process for which the outcomes are discrete, such as tossing coins, rolling dice or the plinko. For many other processes, the outcomes are best described as continuous range of values, for which the probabilities are given by a *continuous probability distribution function*, which we will come to shortly.

Discrete probability distribution functions are commonly identified in the form:

$$p(k; a, b, \dots)$$

where k identifies a specific event, and $a, b \dots$ represent parameters of describing the process and probability function. For the binomial distribution function:

$$p(k; n, p_{\mathrm{s}}) = \frac{n!}{k!(n-k)!} p_{\mathrm{s}}^k (1 - p_{\mathrm{s}})^{n-k}$$

In the general formulation, this distribution represents the number of successes (k) in a series of n successive binary trials, with only two possible outcomes (success and failure), and p_{s} is the probability of success for an individual trial.

Some of the applications of the binomial distribution are:

- The plinko. Here each peg the ball hits represents a trial, with the "success" in this case defined as a turn to the right. In order to land in bucket k, there must be exactly k successes.
- The number of heads (or tails) in n successive coin tosses.
- The number of successes in prescribing a medication to a series of patients with the same condition, where p is the probability of success in any individual case.
- The probability of surviving n potentially deadly events. In this case, $k = n$, since we are only concerned with the case of surviving all n trials.

2.7 Random variables, expected value, variance and standard deviation

I. Playing for money

The origins of probability theory are closely tied to games of chance, and this context still offers one of the most vivid ways to start to think about the subject. To make the plinko more interesting, I might decide to let people drop a ball into a six-row plinko and promise to pay them \$$k$ if the ball lands in bucket k. Unless I only want to hand out money, I will need to charge people to play the game. So, I would like to know how much I need to charge if I want to at least not lose money. In other words how much, on average, should I expect to pay out to each player?

In this case, there is a relatively simple way to solve the problem:

- The probabilities of the ball landing in buckets 0 and 6 are equal. The average payout for these two buckets is $(\$0 + \$6)/2 = \$3$.
- The probabilities of the ball landing in buckets 1 and 5 are equal. The average payout for these two buckets is $(\$1 + \$5)/2 = \$3$.
- The probabilities of the ball landing in buckets 2 and 4 are equal. The average payout for these two buckets is $(\$2 + \$4)/2 = \$3$.
- The payout for bucket 3 is \$3.

So, the overall average payout must be \$3, and I should charge the players at least this much if I don't want to lose money.

The symmetry of the plinko makes the solution to this problem relatively easy to recognize, but in order to deal with more complicated problems, we will introduce some additional concepts, random variables and the expected value. Both of these concepts have been implicit in what we just did, but more formal definitions are called for when extending the ideas.

II. Random variables

A random variable is defined as a variable that is assigned a value for each possible outcome or event from a probabilistic process. Some examples include:

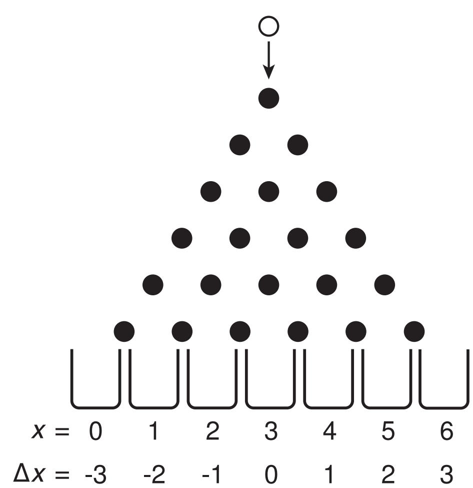
- For a coin toss, we could assign a random variable, x, the value of 1 for heads and 0 for tails. Or, just as arbitrarily, we could define $x = 0$ for heads and $x = 1$ for tails.
- For n successive coin tosses, we could define x to be the number of heads.
- For the plinko, we could define x to be the number associated with the bin that a ball falls in.

For all but the simplest process (*i.e.,* the coin toss), there are likely to be a variety of different random variables that could be defined. For instance, for a series of n coin tosses, we could define the following random variables:

- x_{even}: equal to 1 when the number of heads is even, and 0 when the number of heads is odd.
- x_0: equal to 1 when there are no heads and 0 otherwise.
- x_{run}: equal to the number of successive heads in the longest stretch of heads in the sequence.

Any of these random variables, or many others, could be used for a game of chance, and the trick lies in figuring out how to calculate the probabilities associated with different values for the random variable.

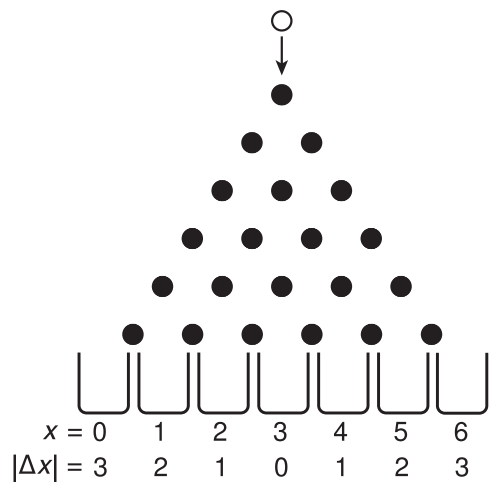
So far for the plinko, we have used the bin number, k as the random variable (without actually calling it that). Now, we will use the variable x to represent the bin number and define some additional random variables in terms of x. For instance, we can define another random variable, Δx, as the position of a bucket from the central bucket. For the six-row plinko the two random variables are related to one another as shown below:



For the six row plinko, the two random variables are related to one another in a simple way:

$$\Delta x = x - 3$$

Another random variable for the plinko is $|\Delta x|$, which represents the distance (in bucket numbers) from the central bucket, as illustrated below:



For the six-row plinko $|\Delta x|$ is related to x according to:

$$|\Delta x| = |x - 3|$$

We could use Δx or $|\Delta x|$ as the basis for gambling. For instance, I could offer to pay players $\$\Delta x$ or $\$|\Delta x|$ when the ball falls in bin x. These games will lead to very different transactions than when I just paid $\$x$. For instance, the negative values of Δx imply that the players will sometimes pay me. This means that I should reconsider how much to charge to play the game, and potential players should reconsider how much they are willing to play.

III. Expected value, or mean of a distribution

The expected value, or *expectation*, for a random variable, x, is the expected average value of x if the process is repeated a large number of times. More formally, consider a process that has m possible outcomes (or a complete set of n non-overlapping events). If the random variable, x, has values of x_i for outcomes $i = 1, 2, 3 \ldots m$, and the probabilities of the outcomes are $p(i)$, then the expected value is defined as:

$$E = \sum_{i=1}^{m} p(i) x_i$$

In essence, E is a weighted sum, in which each of the possible values of x are weighted by the probability of that value of x being the result of a single experiment. If the experiment is repeated a large number of times, we expect the average value of x to approach E. If the experiment represents a game of chance and x is the payout, then E is the expected average payout. This is what we need to calculate how much to charge to play the game.

For the six-row plinko, the possible values of x and the respective probabilities are listed below:

Bucket	x_i	$p(i)$	$p(i)x_i$
0	0	1/64	0
1	1	6/64	6/64
2	2	15/64	30/64
3	3	20/64	60/64
4	4	15/64	60/64
5	5	6/64	30/64
6	6	1/64	6/64
Total		1	192/64 =3

This confirms our earlier deduction that the average payout for the game should be 3.

For the random variable Δx the expected value is quite different, because there are both negative and positive values:

Bucket	Δx_i	$p(i)$	$p(i)\Delta x_i$
0	-3	1/64	-3/64
1	-2	6/64	-12/64
2	-1	15/64	-15/64
3	0	20/64	0
4	1	15/64	15/64
5	2	6/64	12/64
6	3	1/64	3/64
Total		1	0

This result makes sense, since we have, in effect, just moved the labels for the buckets to the right by 3. More generally, if x is a random variable and a is a constant, then:

$$E(x + a) = E(x) + a$$

Also,

$$E(ax) = aE(x)$$

If x and y are two random variables that describe the same set of events, like x and Δx, then

$$E(x+y) = E(x) + E(y)$$

More generally, if x and y are random variables describing the same set of events, and a and b are constants, then

$$E(ax+by) = aE(x) + bE(y)$$

On the other hand, the following relationship is *not* true:

$$E(xy) = E(x)E(y)$$

To help convince yourself of the validity of these relationships (except the last) it may be helpful to test them using x and Δx for the six-row plinko. But, this does not constitute a proof! For that you need to go back to the definition of the expected value.

Finally, consider the case for the random variable $|\Delta x|$

Bucket	$\|\Delta x\|_i$	$p(i)$	$p(i)\,\|\Delta x\|_i$
0	3	1/64	3/64
1	2	6/64	12/64
2	1	15/64	15/64
3	0	20/64	0
4	1	15/64	15/64
5	2	6/64	12/64
6	3	1/64	3/64
Total		1	15/16

This random variable has yet another expected value for the same experiment. Here, it is clear that a relationship that you might think would be true,

$$E(|\Delta X|) = |E(x)|$$

is not.

IV. The variance

Another important parameter that helps define a random variable is called the *variance*. For a discrete random variable, x, with m possible values, the variance, σ^2. is defined as:

$$\sigma^2 = \sum_{i=1}^{m} p(i)(x_i - \mu)^2$$

where μ represents the mean, or expected value, of the random variable. In brief, σ^2 is a measure of the breadth of the distribution. The closer all of the possible values of x are to the mean, the smaller the variance. Of equal importance, if the values of x close to μ are the more probable values, the smaller σ^2 will be. Notice that all of the differences in the sum are squared, which ensures that all of the terms are positive, and values above and below the mean are treated equally.

For the six-row plinko, with the random variable, x, defined as the original bucket number, the variance can be calculated as summarized in the table below. (Recall that the mean value of x is 3.)

Bucket	x_i	$p(i)$	$(x_i - \mu)^2$	$p(i)(x_i - \mu)^2$
0	0	1/64	9	9/64
1	1	6/64	4	24/64
2	2	15/64	1	15/64
3	3	20/64	0	0
4	4	15/64	1	15/64
5	5	6/64	4	24/64
6	6	1/64	9	9/64
Total		1	28	1.5

Thus, $\sigma^2 = 1.5$ for this particular random variable. It is rather difficult to compare directly the magnitude of the variance to values of the random variable, because the differences making up the variance have been squared. Although this particular random variable doesn't have units, random variables can have units in other cases. In such cases, the units of the variance are those of the random variable squared. For instance, if the random variable has units of meters, m, the the units of the variance will be m^2. For this reason, another parameter, the standard deviation, σ, is often used and is defined as:

$$\sigma = \sqrt{\sigma^2}$$

Defining things this way, may seem a bit convoluted, but it emphasizes the fact that the variance is defined in terms of a sum of squares, and is positive, and the standard deviation is derived from σ^2, and not the other way around. For the random variable x defined for the 6-row plinko, $\sigma = \sqrt{(1.5)} = 1.225$.

Next, consider the random variable Δx introduced earlier. for this random variable, the mean, μ, is 0.

Bucket	Δx_i	$p(i)$	$(\Delta x_i - \mu)^2$	$p(i)(x_i - \mu)^2$
0	-3	1/64	9	9/64
1	-2	6/64	4	24/64
2	-1	15/64	1	15/64
3	0	20/64	0	0
4	1	15/64	1	15/64
5	2	6/64	4	24/64
6	3	1/64	9	9/64
Total		1	28	1.5

Perhaps surprisingly, the variances for x and Δx are the same. While the means are different, the distribution of values *around* the respective means are the same.

For a binomial distribution defined previously as

$$p(k; n, p_\mathrm{s}) = \frac{n!}{k!(n-k)!} p_\mathrm{s}^k (1 - p_\mathrm{s})^{n-k}$$

where n is the number of trials, k is the number of successes, and p_s is the probability of a single succesful trial, the mean, variance and standard deviation for the random variable, $x = k$, are given by:

$$\mu = np_\mathrm{s}$$

$$\sigma^2 = np_\mathrm{s}(1 - p_\mathrm{s})$$

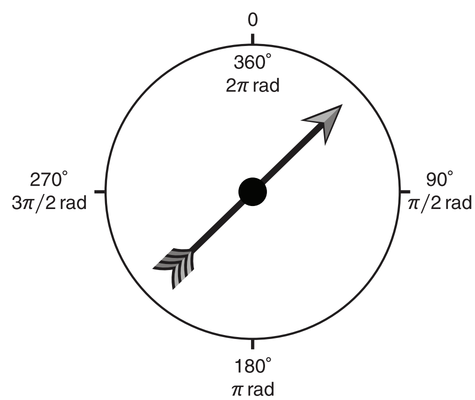
$$\sigma = \sqrt{np_\mathrm{s}(1 - p_\mathrm{s})}$$

Thus, the variance increases in proportion to the number of trials in the binomial experiment. This relationship also indicates that if p_s is closer to 1 or 0, the variance decreases. That is, the more biased the individual trials are, the narrower the distribution. For instance, for the distributions shown on page 45, for a 10-row plinko with $p_\mathrm{s} = 0.5$, 0.6 and 0.8, the corresponding values of the variance are 2.5, 2.4 and 1.6.

2.8 Continuous probability distribution functions

I. The spinner

So far, we have limited our discussion of probability to processes with discrete outcomes, such as coin tosses, dice or the plinko. But, many of the most interesting biological and physical processes give rise to a continuous range of possible outcomes. As a simple example of a process with continuous outcomes, consider a spinner, as used in some board games, consisting of a pointer mounted on a board with a bearing that allows it to spin freely after being given a sharp push, as with a flick of a finger, as illustrated below:



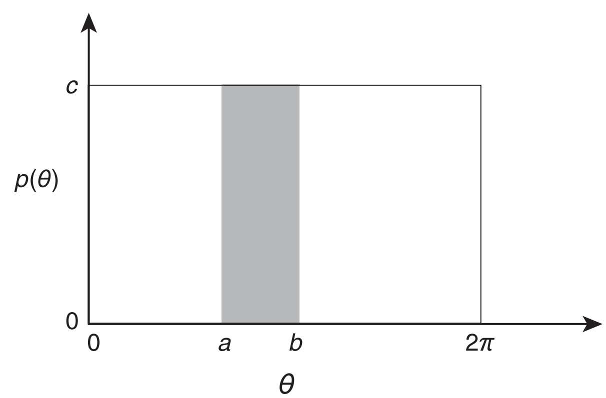
A short time after being pushed, the pointer slows down and stops, pointing in a particular direction. Assuming that the pointer is well balanced and the bearing is very smooth, the pointer should be equally likely to point in any directions. For the purposes of our discussion, we will define the direction of the pointer as the angle between the vertical, as drawn above, and the pointer.

A spinner like this can be used to generate a variety of different random variables. For instance we could divide up the range of angles into two regions; from 0 to π radians and from π to 2π radians. We could then define a random variable so that it is 0 if the pointer lands in the 0–π region and 1 if it lands in the π–2π range. This would be equivalent to a coin toss with a fair coin. We could also divide the range into two non-equal ranges to simulate a biased coin toss. Alternatively, we could divide the range into six regions, to simulate a single six-sided die.

In principle, we can divide the range of angles into smaller and smaller angles, provided that we have a means to measure very small differences in position. This leads to the notion of a continuous random variable that represents the position of the pointer, as shown in the drawing above. We will call this random variable θ and express its value in radians, from 0 to 2π. Like other random variables we have discussed, every possible value θ has associated with it a probability, $p(\theta)$. Since the pointer is equally likely to point in any direction, the value of $p(\theta)$ must be equal for all values of θ. On the other hand, if θ can take on any value in a continuous range, which can be divided into infinitesimally small intervals, then the probability of pointing any single direction must be infinitesimally small, or essentially zero! To resolve this apparent paradox, we interpret continuous probability distribution functions in terms of the probability that the random variable (θ in our case) lies between two defined values. Specifically, the probability that the random variable θ lies between a and b is given by the integral:

$$p(a \leq \theta \leq b) = \int_a^b p(\theta)d\theta$$

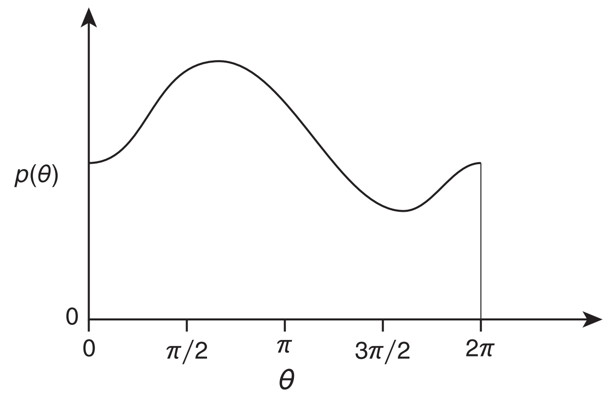
This relationship is illustrated graphically below, for the case of the spinner:



As argued above, the value of $p(\theta)$ must be equal for all values of θ between 0 and 2π. For now, we will call the value of $p(\theta)$ within this range c. Since the pointer must land within the range between 0 and 2π, $p(\theta)$ must be zero elsewhere. The probability that θ lies between a and b is then:

$$\begin{aligned} p(a \leq \theta \leq b) &= \int_a^b p(\theta)d\theta \\ &= \int_a^b c d\theta \end{aligned}$$

This integral corresponds to the area below the horizontal line segment representing $p(\theta) = c$ and bounded by $\theta = a$ and $\theta = b$, as indicated by the shaded box in the drawing above. Assuming that the probability function is, indeed, constant, then the probability is proportional to the difference between a and b, as we would intuitively expect if the spinner is fair. A continuous probability distribution function for which all possible values of the random variable are equal is called a *uniform* probability distribution. A more interesting probability distribution might arise if the pointer was more likely to land in some areas than others, as illustrated in the hypothetical graph below:



In this case, the probability distribution function indicates that the pointer is more likely to land in the region of $\theta \approx 3\pi/4$ than in the region of $3\pi/2$. If this spinner was used in a game of chance, a gambler with this information would be at a distinct advantage over one without it!

Since the spinner must point *somewhere* in the range between $\theta = 0$ and 2π (assuming that it doesn't break), the total probability must be 1. This is equivalent to the requirement that the probabilities of all of the possible outcomes must sum to 1 in a random process with discrete outcomes. For the uniform distribution function for the spinner, we can write this requirement as:

$$\begin{aligned} p(0 \leq \theta \leq 2\pi) = 1 &= \int_0^{2\pi} p(\theta) d\theta \\ &= \int_0^{2\pi} c d\theta \end{aligned}$$

where c is the constant introduced earlier, to which we can now assign a specific value, as follows:

$$\begin{aligned} &1 = \int_0^{2\pi} c d\theta \\ &1 = c\theta\big|_0^{2\pi} = c2\pi - c0 = c2\pi \\ &c = 1/(2\pi) \end{aligned}$$

We can then write the probability distribution function as:

$$p(\theta) = \frac{1}{2\pi}$$

This form of the function is said to be *normalized*, meaning that the integral over all possible values is equal to 1. This term is sometimes confused with the *normal probability function* which refers to a specific continuous probability function that we will discuss bellow and also goes by the name *Gaussian* distribution.

II. Expected value and variance for continuous random variables

Recall that for a discrete random variable, x, we defined the expected value, $E(x)$, as

$$E(x) = \sum_{k=1}^{n} p(x_k)x_k$$

where x_k represents the k^{th} value of x and $p(k)$ is the probability of x_k. For a continuous random variable, the sum above is replaced by an integral:

$$E(x) = \int p(x)x dx$$

where the integral is over all possible values of x. For the spinner random variable, θ, the expected value is calculated as follows (for an unbiased spinner):

$$\begin{aligned}
E(\theta) &= \int_0^{2\pi} p(\theta)\theta d\theta \\
&= \int_0^{2\pi} \frac{1}{2\pi}\theta d\theta \\
&= \frac{1}{4\pi}\theta^2\Big|_0^{2\pi} = \frac{4\pi^2}{4\pi} - 0 \\
&= \pi
\end{aligned}$$

Thus, the average value of θ, over a large number of trials, is expected to be π, that is the mid-point of the range of possible values. Keep in mind, however, that this outcome is no more likely than any other.

For a discrete random variable the variance is defined as the sum:

$$\sigma^2 = \sum_{k=1}^{n} p(k)(k-\mu)^2$$

where μ represents the mean, or expected value, of the random variable.

The equivalent relationship for a continuous random variable is the integral:

$$\sigma^2 = \int p(x)(x-\mu)^2 dx$$

For the random variable θ, the variance is calculated as:

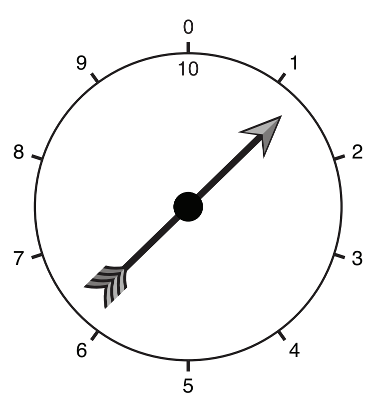
$$
\begin{aligned}
\sigma^2 &= \int_0^{2\pi} p(\theta)(\theta-\pi)^2 d\theta \\
&= \int_0^{2\pi} \frac{1}{2\pi}\left(\theta^2 - 2\pi\theta + \pi^2\right) d\theta \\
&= \frac{1}{2\pi}\left(\frac{1}{3}\theta^3 - \pi\theta^2 + \pi^2\theta\right)\Big|_0^{2\pi} \\
&= \frac{1}{2\pi}\left(\frac{8}{3}\pi^3 - 4\pi^3 + 2\pi^3\right) \\
&= \frac{4}{3}\pi^2 - 2\pi^2 + \pi^2 \\
&= \frac{\pi^2}{3}
\end{aligned}
$$

III. Some other random variables from the spinner

A variety of other random variables might be assigned to the spinner. For instance, we could create a game of chance where the payout ranges from zero to \$10 depending on the position of the pointer, with the payout increasing linearly with θ, from 0 to \$10. We will call this random variable x and define it in terms of θ according to:

$$x(\theta) = \frac{5}{\pi}\theta$$

This amounts to relabeling the spinner, as below:



Since the values of x are evenly distributed over the range of θ, which has a uniform probability distribution function, x should also have a uniform probability distribution. Following the approach used for θ, you should be able to show that $p(x) = 1/10$.

The expected value of x is calculated as

$$
\begin{aligned}
E(x) &= \int_0^{10} p(x) x dx \\
&= \int_0^{10} \frac{1}{10} x dx \\
&= \frac{1}{20} x^2 \Big|_0^{10} = 5
\end{aligned}
$$

This result can derived even more easily by using a relationship introduced on page 53:

$$E(ax) = aE(x)$$

where a is a constant. (Note that x in this equation is not the same as our $x(\theta)$.) Substituting θ for x and $5/\pi$ for a, we can write:

$$
\begin{aligned}
E(\frac{5}{\pi}\theta) &= \frac{5}{\pi} E(\theta) \\
&= \frac{5}{\pi}\pi = 5
\end{aligned}
$$

Note that, just as for θ, the expected value for x is the midpoint in the range of possible values. This is a general property of a uniform probability distribution function, but not of all distribution functions.

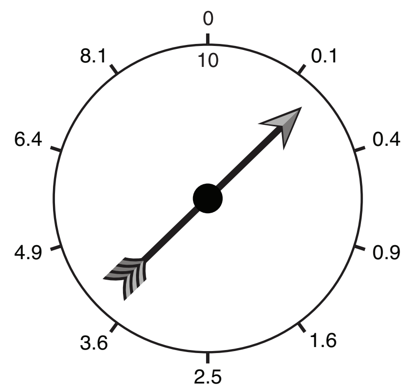
The variance of the new random variable is calculated as:

$$
\begin{aligned}
\sigma^2 &= \int_0^{10} p(x)(x-5)^2 dx \\
&= \int_0^{10} \frac{1}{10} \left(x^2 - 10x + 25\right) dx \\
&= \frac{1}{10} \left(\frac{1}{3}x^3 - 5x^2 + 25x\right) \Bigg|_0^{10} \\
&= \frac{1}{10} \left(\frac{1}{3} 1000 - 500 + 250\right) \\
&= \frac{25}{3} \approx 8.333
\end{aligned}
$$

As an example of a random variable that does not have a uniform distribution function, but is still based on the spinner, we can define y as:

$$y(\theta) = \frac{10}{4\pi^2}\theta^2$$

As in the previous example, a constant of multiplication has been introduced to make the range of possible values lie between 0 and 10. We can use this definition to relabel the spinner dial:



Now, we find that the values are not evenly distributed around the dial. For instance the range of y-values from 0 to 2.5 represents half of the dial, meaning that values in this range are expected to occur half the time, and values from 2.5–10 are expected the other half.

To derive the probability distribution function for y, we can use its relationship to θ, for which we do know the distribution function. For y, the expression $p(y)dy$ represents the probability that y lies within a small region, dy, of a specific value of y. Similarly, the expression $p(\theta)d\theta$ represents the probability that θ lies within $d\theta$ of θ. If y is $y(\theta)$ for a specific value of θ and dy is the small region of y corresponding to the small region of θ, $d\theta$, then the two probabilities must be equal, and we can write:

$$p(y)dy = p(\theta)d\theta$$

Taking some mathematical liberties, this can be rewritten in terms of the derivative, $d\theta/dy$:

$$p(y) = \frac{d\theta}{dy}p(\theta)$$

To find the derivative, $d\theta/dy$, we need the function $\theta(y)$, which can be obtained by rearranging the definition of y as a function of θ:

$$y = \frac{10}{4\pi^2}\theta^2$$

$$\theta^2 = \frac{4\pi^2}{10}y$$

$$\theta = \frac{2\pi}{\sqrt{10}}y^{1/2}$$

The derivative of θ with respect to y is:

$$\frac{d\theta}{dy} = \frac{2\pi}{\sqrt{10}}\frac{1}{2}y^{-1/2} = \frac{\pi}{\sqrt{10}}y^{-1/2}$$

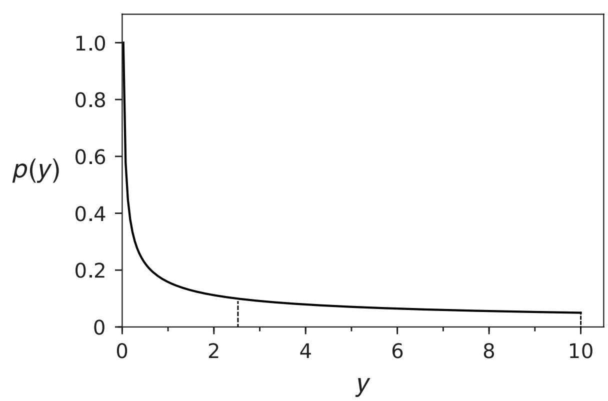
Recalling that $p(\theta) = 1/(2\pi)$, the desired probability function, $p(y)$, can now be written as:

$$
\begin{aligned}
p(y) &= \frac{d\theta}{dy} p(\theta) \\
&= \frac{1}{2\pi} \frac{\pi}{\sqrt{10}} y^{-1/2} = \frac{1}{2\sqrt{10}} y^{-1/2}
\end{aligned}
$$

Since $p(y)$ was derived from a normalized probability density function, $p(\theta)$, $p(y)$ should be normalized as well. To be sure, though, we can calculate the integral of $p(y)$ from 0 to 10.

$$
\begin{aligned}
\int_0^{10} p(y) dy &= \int_0^{10} \frac{1}{2\sqrt{10}} y^{-1/2} dy \\
&= \frac{1}{\sqrt{10}} y^{1/2} \Big|_0^{10} \\
&= \frac{1}{\sqrt{10}} \left(10^{1/2} - 0^{1/2}\right) = 1
\end{aligned}
$$

Thus, $p(y)$ is, indeed normalized. A plot of $p(y)$ is shown below:



Note that $p(y)$ is not defined for $y = 0$, but it can be evaluated for any value of y arbitrarily close to 0. As expected from the relabeled spinner dial shown on page 63, the distribution favors smaller values of y. For instance, one half of the area under the curve lies between 0 and 2.5, and the other half lies between 2.5 and 10, as indicated by the vertical dashed lines.

Calculating the expected value of y is a bit more involved than in the previous examples, because the probability function is not a simple constant:

$$
\begin{aligned}
E(y) &= \int_0^{10} p(y) y dy \\
&= \int_0^{10} \frac{1}{2\sqrt{10}} y^{-1/2} y dy = \int_0^{10} \frac{1}{2\sqrt{10}} y^{1/2} dy \\
&= \frac{1}{2\sqrt{10}} \frac{2}{3} y^{3/2} \Big|_0^{10} \\
&= \frac{1}{3\sqrt{10}} \left(10^{3/2} - 0^{3/2}\right) \\
&= \frac{10}{3} \approx 3.333
\end{aligned}
$$

Recall that the expected value of x, which also covers the range from 0 to 10, is 5. The lower value of $E(y)$ reflects the non-uniform probability distribution function for this variable, which more heavily favors lower values.

The variance is calculated as:

$$
\begin{aligned}
\sigma^2 &= \int_0^{10} p(y)(y - 10/3)^2 dy \\
&= \int_0^{10} \frac{1}{2\sqrt{10}} y^{-1/2} \left(y^2 - \frac{20}{3} y + \frac{100}{9}\right) dx \\
&= \frac{1}{2\sqrt{10}} \int_0^{10} \left(y^{3/2} - \frac{20}{3} y^{1/2} + \frac{100}{9} y^{-1/2}\right) dy \\
&= \frac{1}{2\sqrt{10}} \left(\frac{2}{5} y^{5/2} - \frac{40}{9} y^{3/2} + \frac{200}{9} y^{1/2}\right) \Big|_0^{10} \\
&= \frac{1}{2\sqrt{10}} \left(\frac{2}{5} 10^{5/2} - \frac{40}{9} 10^{3/2} + \frac{200}{9} 10^{1/2}\right) \\
&= \frac{80}{9} \approx 8.888
\end{aligned}
$$

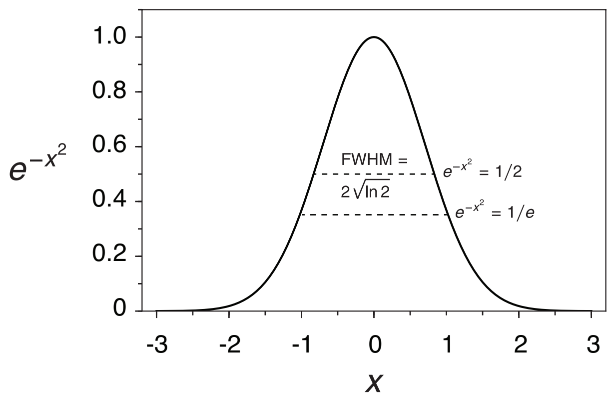
Notice that the variance is just a bit larger than for the random variable x, which has a uniform probability distribution function.

2.9 The Gaussian, or normal, probability distribution function

One of the most important continuous probability distribution functions is commonly referred to as a *Gaussian* or *normal* distribution function. The Gaussian function, in various forms, also arises in areas outstide of probability and statistics. At it's simplest, a Gaussian function has the form:

$$f(x) = e^{-x^2}$$

where $e \approx 2.71828$ is the base of the natural logarithms. A graph of the function has the familiar bell shape shown below:



The function has its maximum value, 1, when the exponent is 0 and decreases as x becomes either positive or negative. When $x = 1$ or -1, the function equals $1/e \approx 0.3679$. Another useful parameter for describing functions that describe peaks is the *full width at half maximum*, FWHM. For the simple Gaussian function, it is easy to show that FWHM $= 2\sqrt{\ln 2}$.

The simple Gaussian function, and forms derived from it, have the rather inconvenient property that its antiderivative cannot be written in terms of a finite number of simple functions. As a consequence, integrals over finite ranges of x cannot be evaluated exactly. However, the indefinite integral over all values of x can be shown to be:

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

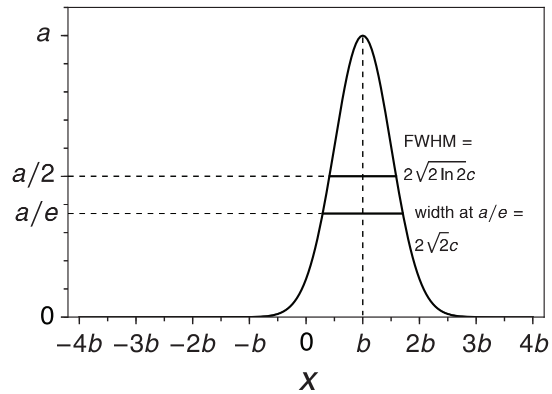
Thus, this form is not a properly normalized probability distribution function. It is also striking that integration of a function defined in terms of an important irrational number, e, is related to a second fundamental irrational number, π.

I. The general form of the Gaussian function

A more general form of the Gaussian function can be written as:

$$f(x) = ae^{-\frac{(x-b)^2}{2c^2}}$$

This form introduces three parameters, a, b and c, which affect the shape of the curve in different ways, as illustrated in the figure below.



The parameter a determines the value of the function at its maximum, where the exponent of e equals zero, which occurs when the value of x is equal to b. The width of the peak is determined by c^2: The larger the value of c^2, the more slowly the exponent decreases as x increases or decreases away from b, and the wider the peak is. As shown in the figure, both the full width at half maximum ($FWHM$) and the width at the maximum value divided by e (a/e) are proportional to c (which is assumed here to be positive).

The integral of the general form of the Gaussian function is:

$$\int_{-\infty}^{\infty} ae^{-\frac{(x-b)^2}{2c^2}}dx = a\sqrt{2c^2\pi}$$

If a is set so that it is equal to $1/\sqrt{2c^2\pi}$, then the value of the integral is equal to one:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2c^2\pi}}e^{-\frac{(x-b)^2}{2c^2}}dx = 1$$

The function thus has the required property of a normalized probability distribution function. However, the form that is usually used in probability and statistics uses the symbol μ in place of b and σ^2 in place of c^2, to give:

$$p(x) = \frac{1}{\sqrt{\sigma^2 2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

As with the other continuous probability distribution functions we have looked at, the (normalized) Gaussian distribution gives the probability that the variable x lies between two points, x_1 and x_2, when the function is integrated between the two points.

$$p(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\sigma^2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

In this context, μ is the mean value (or expectation value) of the distribution and σ^2 is the variance, as defined earlier.

II. Approximation of the binomial distribution by the Gaussian distribution

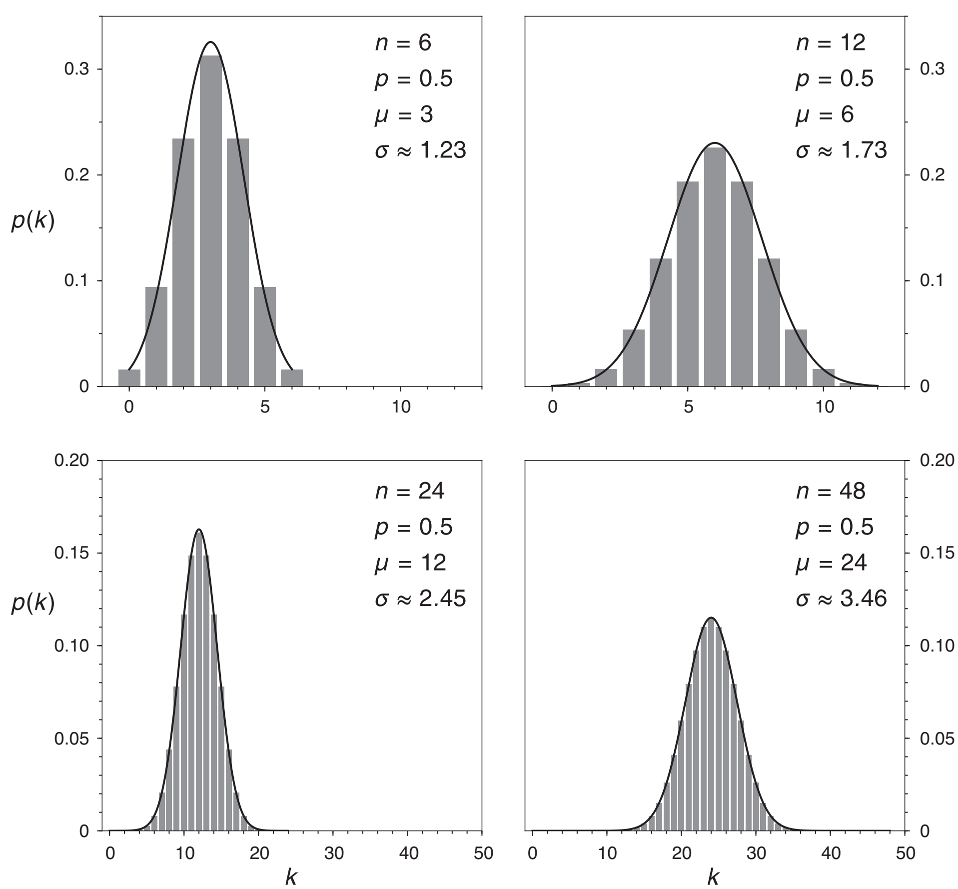
One important feature of the Gaussian (or normal) probability distribution function is that it represents the limiting case of the discrete binomial distribution when the number of trials (rows in the plinko, for instance) becomes large. Unfortunately, rigorously demonstrating this relationship is not simple. Instead, we will simply demonstrate how closely the two relationships match one another as the number of trials increases. Recall that the binomial distribution is given by

$$p(k; n, p_\mathrm{s}) = \frac{n!}{k!(n-k)!} p_\mathrm{s}^k (1-p_\mathrm{s})^{n-k}$$

where n is the number of trials, k is the number of successes, and p_s is the probability of a single succesful trial. Recall also that the expected value, or mean, of the binomial distribution is np_s, and the variance is $np_\mathrm{s}(1-p_\mathrm{s})$. For given values of n and p_s, the Gaussian distribution with the same mean and variance can be written as:

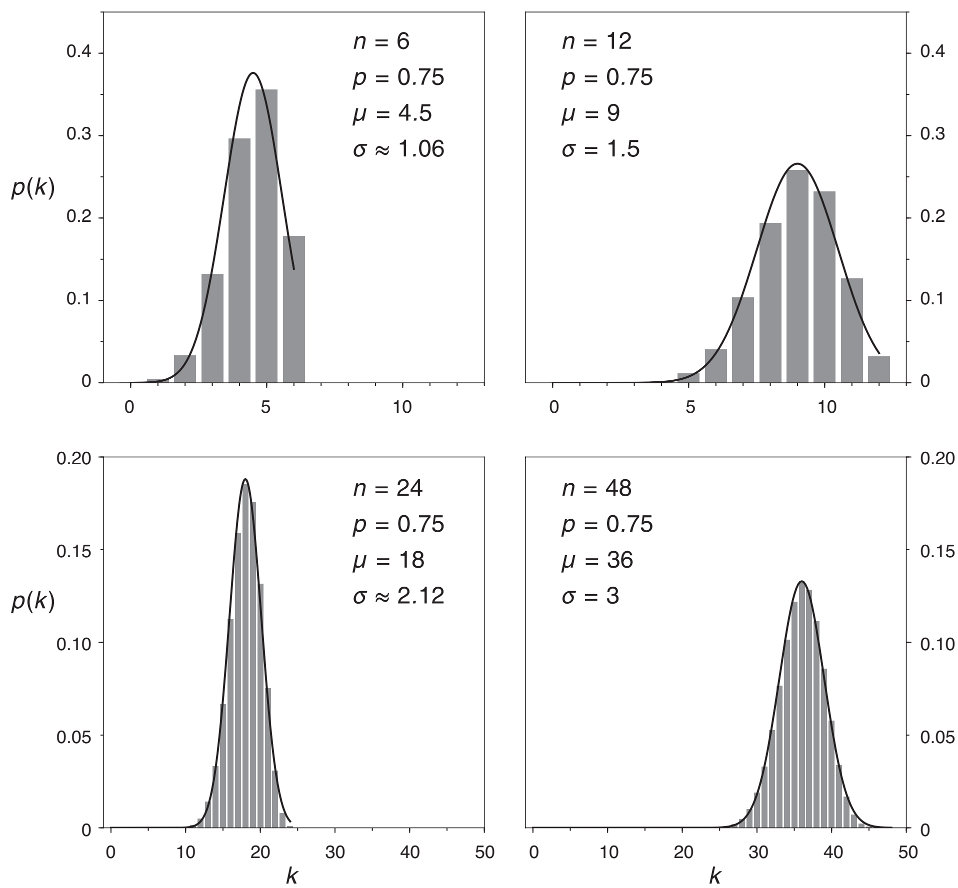
$$\begin{aligned} p(k) &= \frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{(k-\mu)^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{np_\mathrm{s}(1-p_\mathrm{s})2\pi}} e^{-\frac{(k-np_\mathrm{s})^2}{2np_\mathrm{s}(1-p_\mathrm{s})}} \end{aligned}$$

The figure below shows direct comparisons between the binomial and matched Gaussian distribution for $p_\mathrm{s} = 0.5$ and $n = 6$, 12, 24, and 48.



As you can see, the Gaussian distribution is a close match to the binomial distribution, even for n as small as 6. For even moderately large values of n, it is much easier to calculate values for the Gaussian distribution than for the binomial distribution, where the values of the coefficients quickly become very large.

The figure below shows the same comparisons between the binomial and Gaussian distributions, but now with $p_s = 0.75$, so that the distributions are shifted to the right. With the biased distributions, the match between the binomial and Gaussian distributions is not quite as close as when $p_s = 0.5$ The reason for this is that the Gaussian distribution is always symmetrical about the mean, even if the mean is shifted from the unbiased value. On the other hand, the binomial distribution becomes skewed when p_s is not equal to 0.5, especially for relatively small values of n. As n increases, the binomial distribution becomes more symmetrical, for a given value of p_s, and is better matched by the Gaussian distribution.



A general rule of thumb[1] states that the Gaussian distribution, with μ set to np_s and σ^2 set to $np_s(1 - ps)$ is a "good" approximation to the binomial if the following two

[1] https://en.wikipedia.org/wiki/Binomial_distribution#Normal_approximation

conditions are satisfied:

$$n > 9\frac{1 - p_{\mathrm{s}}}{p_{\mathrm{s}}}$$

and

$$n > 9\frac{p_{\mathrm{s}}}{1 - p_{\mathrm{s}}}$$

Do the examples shown above appear to confirm this rule of thumb?

2.10 Simulating randomness with a computer: (Pseudo) random numbers

One of the most interesting, and useful, applications of computers in science is the simulation of processes that have an underlying random, or unpredictable, character. Such processes include the diffusion of particles, mutations of genes and quantum-mechanical phenomena. The basic idea of such simulations is to figuratively flip a coin, throw dice or spin a roulette wheel to decide the outcome of specific events in a simulation. The simulation is usually repeated many times in order to describe the distribution of possible outcomes. The technique is usually attributed to two mathematicians, John Von Neumann and Stansilaw Ulam, who used it to study nuclear physics problems near the end of World War II and attached the name "Monte Carlo" to this kind of calculation. (Presumably, they thought that Monte Carlo sounded more glamorous than Wendover.)

Although they may not always seem it, computers are, by design, extremely predictable machines. So, the problem arises, how do we simulate a random event, like a coin flip? The answer is to generate a sequence of numbers using a completely predictable algorithm, that *appears* to have come from a random physical process. These numbers are properly called "pseudo-random numbers", but the shortened term "random number" is often used. For instance, a simulation of a series of coin tosses might be represented as a series of 1s and 0s. Over a long period, the number of 1s and 0s should be roughly, but not exactly, equal. But, the sequence should not be a simple alteration of 1s and 0s, since we know that "runs" of "heads" and "tails" are common. More generally, each number should be unpredictable from the previous one, *unless* one knows the algorithm. Very demanding tests for random number generators have been devised, and the development of improved generators is, itself, an ongoing endeavor.

A fairly simple approach to generating random numbers involves taking a "seed" value, applying some arithmetic operation to it, dividing the result by some other number and returning the remainder. The result is then used as the seed for calculating the next number in the series. One widely used algorithm uses the following equation to calculate number X_{n+1} from X_n:

$$X_{n+1} = (aX_n + b) \mod c$$

where a, b and c are integers, and the operator $\mod c$ indicates the remainder of dividing the quantity $(aX_n + b)$ by c. For instance:

$$5 \mod 4 = 1$$

The remainder has to be less than the value chosen for `c`, so this establishes a maximum number of unique numbers that can be generated. Eventually, a number will be returned a second time, and from that point on the series repeats exactly. How well this algorithm works depends critically on the choice of constants.

Many computer languages include built in functions for generating random numbers. For instance, the Python language includes a module, called `random`, that provides a much better random number generator than the one described above, as well as several nifty variations for special purposes. The function `random.random()` returns a number, x, such that $0 \leq x < 1$, ss illustrated below:

```
>>> random.random()
0.63876690995825403
>>> random.random()
0.98645481541390223
```

In general, the initial seed for a random number generator can be either set to a specific value derived from information provided by the computer or some outside source. A common way of setting the seed is to derive it from the time when the program is started, using the computer's clock.

The Python random module includes a function that allows the user to specify the seed. The listing below shows what happens if the same seed is used a second time:

```
>>> random.seed(12345)
>>> random.random()
0.41661987254534116
>>> random.random()
0.010169169457068361
>>> random.random()
0.82520650925374317
>>> random.seed(12345)
>>> random.random()
0.41661987254534116
>>> random.random()
0.010169169457068361
```

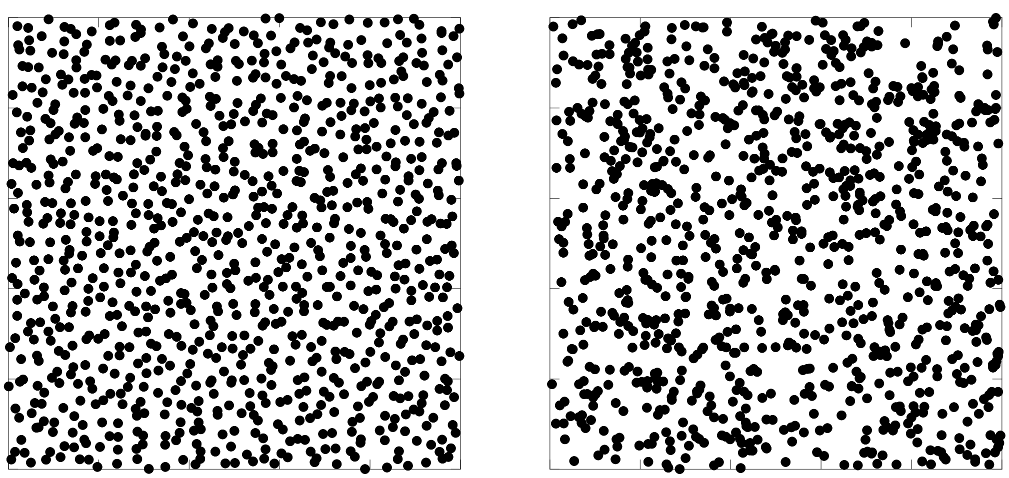
For a given seed, the same sequence of "random" numbers will be generated again. This is a clear demonstration that the numbers generated from most random number generators aren't truly "random." Sometimes, though, it is useful to be able to use the same set of random numbers multiple times, for instance in testing a computer program or algorithm.

In addition to pseudo-random number generators, there are ways to generate "true" random numbers from physical processes. One of these involves monitoring the decay of a radioactive element. Although the average number of decay events over a period of time can be well known, the intervals between successive events is random. A website that provides

"Hot bits" derived in this fashion is:
https://www.fourmilab.ch/hotbits/
The hardware used at this site can only generate numbers at a modest rate, but they can be used as the seeds for a pseudo-random number generator. At one time, there was a website that generated random numbers from images of a lava lamp, as described on Wikipedia:
https://en.wikipedia.org/wiki/Lavarand
Again, the idea was to generate true random numbers that could be used as seeds for pseudo-random number generators.

More recently, a variety of hardware devices have been developed that generate true random numbers from various forms of electronic noise or quantum mechanical phenomena. Some of these are relatively inexpensive USB dongles and some are built into newer computers, including those using newer Intel microprocessor chips. It may seem surprising that there would be so much demand for something random! But the reason for this demand is that random numbers play a central role in cryptography, including securing data that is transmitted over the internet. As the security problems of the modern age grow, random number generators are becoming increasingly critical and are coming under ever more careful scrutiny.

One of the useful things that we can do with a random number generator is to simulate some process and look at the distribution, just to get a sense of what a truly random process looks like. The figure below shows two distributions of points on a square:



For one of these figures, I choose random x and y values for 1,000 points and plotted them. For the other, I placed a similar number of points using another procedure.

Which one is the true random distribution? How could you decide?