TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI

Probability Introduction
MATHematics MAGnification™

Dr. Ralph deLaubenfels

TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI

Teacher-Scholar Institute

Columbus, Ohio

2018

PROBABILITY INTRODUCTION MAGNIFICATION

This is one of a series of very short books on math, statistics, and physics called "Math Magnifications." The "magnification" refers to focusing on a particular topic that is pivotal in or emblematic of mathematics.

OUTLINE

Probability is something we use constantly, often unconsciously. This magnification is the first of four intuitive introductory magnifications about probability. See [2] for more topics and details, especially extensive examples and homework.

The first chapter begins by describing and illustrating how essential probability is for decision making. We show how probability is sometimes defined as relative frequency and sometimes as area.

The second chapter discusses an important duality in probability, mathematics, science, pseudo-science, and philosophy: *discrete* as opposed to *continuous.* We also introduce *random variable*, assigning numbers to probability spaces, and make a sharp distinction between discrete and continuous random variables.

The appendix gives a unifying idea for random variables.

The only prerequisites for this magnification (at least prior to the appendix) are arithmetic, including fractions and percents, and area of rectangles and triangles. The appendix needs the idea of graphs of functions.

CHAPTER I: Probability Defined and Drawn.

Example 1.1. Here are some decisions you make every day.

Decision 1. Should you carry an umbrella when you leave the house?

Decision 2. Should you put half your income for the rest of your life into the state lottery?

Decision 3. Should you wear a safety helmet 24 hours a day?

Decision 4. Should I have a particular surgical procedure? The word "procedure" makes it sound less personal and more bureaucratic.

In Decision 1, you might not want to get wet, but an umbrella is cumbersome and dangerous; you might jab someone.

In Decision 2, a marketing campaign might show all the great things you can buy if you win the lottery.

In Decision 3, you might injure your head if you fall. Nighttime sleep is no exception; an exciting dream might cause you to fling yourself out of bed head first on the floor.

Decision 4 will probably have colorful and dramatic pictures for either having or not having the procedure.

All these decisions as stated are missing the same key ingredient for making an intelligent decision: *probability.* In Decision 1, you need to know how likely it is to rain; if it's not very likely, you might leave your umbrella home. In Decision 2, you need to know how likely you are to win. In Decision 3, you need to know how likely you are to bang your head uncomfortably. In Decision 4, you need to compare two probabilities, the probability of suffering from not having the surgery and the probability the surgery will hurt you: "Is the cure worse than the disease?", as the saying goes.

Informally, probability tells you how seriously to take a possible occurrence: the larger the probability, the more we worry that the occurrence might actually happen. It is essential for making decisions. To a small child asking if it will rain, the answers "it will rain" or "it won't rain" are probably the best you can do; for an adult that you respect, a probability is much more informative and useful.

Example 1.2. Suppose a tank of water contains 50 turtles, 2 of which are snapping turtles. You reach into the tank and grab a turtle at random. What is the probability you will get snapped?

If this were asked in an introductory lecture on probability, almost everyone, in our experience, would say

$$\text{probability equals } \frac{2}{50} \text{ or } 4\%.$$

Informally, this is *relative frequency*, of what you want (a snapping turtle in the tank) divided by what you could get (any turtle in the tank). To make this more precise (see 1.5), we need some terminology.

Definitions 1.3. A **sample space** or **universe** is the set of all outcomes of an experiment. An **event** is a subset of the sample space. We will not go into which subsets are considered events; it takes hard work to construct a subset that is *not* an event.

Events are the objects we will take probabilities of.

Examples 1.4. Here are some examples of experiments, sample spaces, and events.

(a) Experiment: flip a coin. The sample space S is $\{H, T\}$, where we have denoted H for getting heads, T for getting tails. $A \equiv \{H\}$ (getting heads) is an event.

(b) Experiment: fire a missile into the air. The sample space S is the set of all places where the missile could land (we're assuming the missile does not have sufficient energy to escape the earth's gravitational field), meaning all points on earth, including the oceans. An example of an event is the state of Ohio; that is, "missile lands in Ohio".

(c) Experiment: flip a coin twice. The sample space S is $\{HH, HT, TH, TT\}$. An example of an event is "getting at least one head," or $\{HH, HT, TH\}$.

Notice that (a) and (c) of Examples 1.4 are qualitatively different than (b) of Examples 1.4. We will discuss this in Chapter II ((a) and (c) are *discrete* while (b) is *continuous*).

Relative frequency probability 1.5. Let's return to scenarios such as the turtle tank in Example 1.2.

If all outcomes in a sample space are equally likely and A is an event, then, denoting by $P(A)$ the **probability of** A,

$$P(A) = \frac{(\text{number of outcomes in } A)}{(\text{number of possible outcomes})}.$$

Here are two ways to recognize "equally likely" outcomes. The phrase "at random," as in Example 1.2, is often a sign of equally likely outcomes, in this case, equally likely turtles. Lack of information, including closing your eyes, is needed for this; if you opened your eyes, you might see and avoid viciously snapping jaws.

A **fair coin** is a coin constructed so that, on each flip, the probability of getting heads equals the probability of getting tails. The adjective "fair" refers to gambling.

Example 1.6. If the coin in Examples 1.4(c) is fair, then the probability of getting at least one head is $\frac{3}{4}$, since our desired event is $\{HH, HT, TH\}$ (three outcomes), while the sample space is $\{HH, HT, TH, TT\}$ (four outcomes).

Notice that, if we had set up our sample space in Examples 1.4(c) as

$$S' = \{\text{no heads, one head, two heads}\},$$

then we would appear to get, with a fair coin, $\frac{2}{3}$, for the probability of getting at least one head. This is FALSE, because the outcomes in S' are not equally likely:

$$P(\text{no heads}) = \frac{1}{4} \neq \frac{1}{2} = P(\text{one head}).$$

Definition 1.7. The probability fraction in 1.5 is not the only possible definition of probability. We will present a completely different picture before the end of this chapter, in 1.10. First let's say as much as possible about probability in general.

Given a sample space S, a **probability** P assigns, to any event A, a number, denoted $P(A)$ (reads "P of A" or "**probability of** A"), and has the following properties:

(1) $0 \leq P(A) \leq 1$, for any event A;

(2) $P(S) = 1$; and

(3) If $A_1, A_2, A_3, A_4, \ldots$ is a sequence of events, such that no pair can occur simultaneously, then

$$P(A_1 \text{ or } A_2 \text{ or } A_3 \text{ or } \ldots) = (P(A_1) + P(A_2) + P(A_3) + \ldots)$$

Very informally, P measures how likely an event is. The reader should check that the probability fraction in 1.5 satisfies (1)–(3) of Definition 1.7.

When $P(A) = 0$, A is called an **impossible event**. It doesn't sound like a compliment; for making decisions, "impossible" translates to "ignore it completely." When $P(A) = 1$, A is called a **sure event**. Property (1) is saying that all events are between the extremes of impossibility and certainty.

Remarks 1.8. Any probability may be thought of as relative frequency over *time*, in the following sense.

Suppose A is an event resulting from an experiment. If we repeat the experiment many times, then

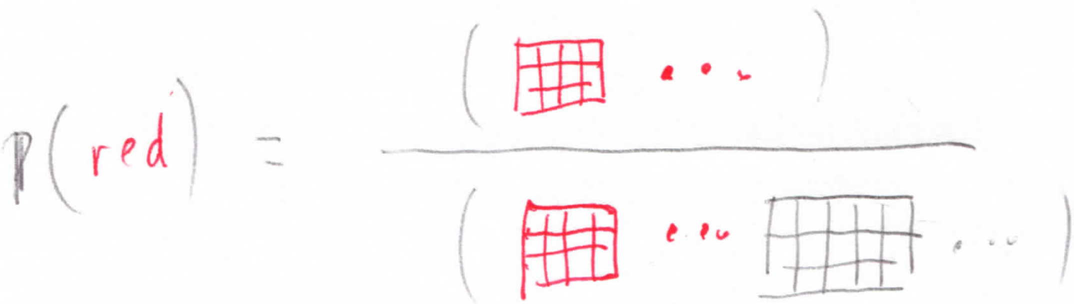
$$P(A) \sim \frac{(\text{the number of times } A \text{ occurred})}{(\text{number of experiments})},$$

with the approximation "$\sim$" getting arbitrarily close as the number of experiments gets large (the precise name for this approximating process is *limit*, the key calculus idea).

Compare the fraction above to the fraction in 1.5.

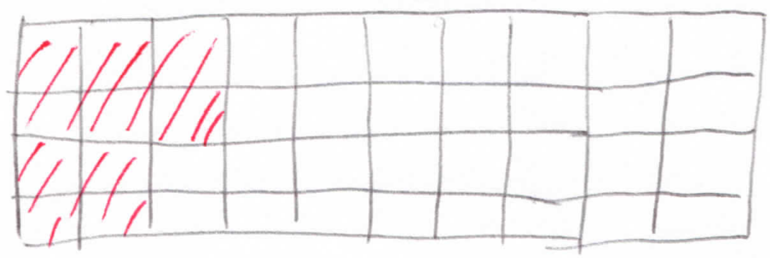
For example, let's say a weather forecast gives a 30% chance of rain. This means that, if we repeat the current observed conditions many times, then rain should result approximately 30% of the time.

Discussion 1.9. In the probability fraction 1.5, represent each outcome as a square; since each outcome is assumed equally likely, each square should be the same size; in the next three drawings below, our desired outcomes are red, all others are black.



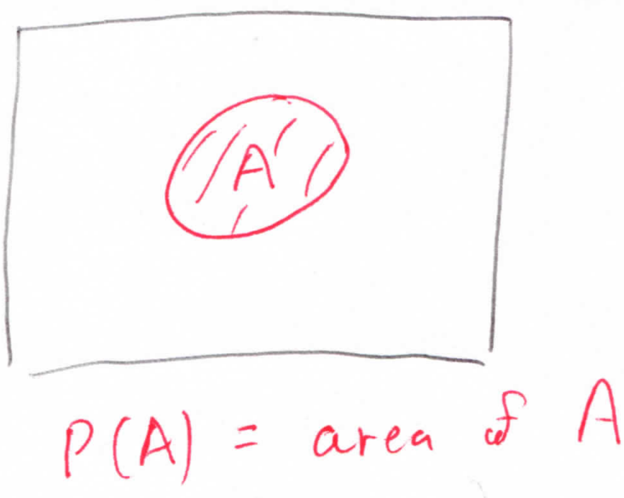
With this representation, the number of outcomes in an event is proportional to the area of the set of outcomes in said event. For example, in the drawing below,

$$P(\text{red}) = \frac{10}{40} = \frac{\text{red area}}{\text{total area}}.$$



The probability of an event is now seen as (relative to the sample space) *area*.

Probability area 1.10. To be more specific, imagine you are firing randomly (eyes closed) at the broad side of a barn. The probability of hitting a target placed on said side is proportional to the area of the target. In the picture below, we have chosen units so that the rectangle representing the sample space has area one, so that relative area of the event A equals just the area of A.



The drawing in 1.10 is a **Venn diagram**, and is of much more interest when there are multiple events.

Definition 1.11. Our definitions and pictures of probability so far deal with somewhat mysterious things, such as sequences of Hs and Ts (see Examples 1.4). What we really want, as scientists or pseudo scientists, is *numbers* to manipulate or make pronouncements about. Our choice of numbers can focus on what we care about; for example, if, in Examples 1.4(c), we get a dollar for every head in our coin flip, we'd like to devote our minds to the number of heads.

A **random variable** assigns a number to each outcome in a sample space. The capital letter X is a popular name for a random variable.

Examples 1.12. (a) In Examples 1.4(c), let X be the number of heads.

(b) In Examples 1.4(b), let X be the distance, in miles, from where we launched the missile to where the missile landed.

See [2] for many famous random variables.

CHAPTER II: Discrete versus Continuous.

A fundamental duality in modeling anything from the universe to a food item is whether it is *discrete*, meaning separate, distinct particles like kernels of corn, or *continuous*, meaning a connected, uninterrupted stream such as gravy. Discrete objects are quantified by counting, continuous objects by continuous measurements such as length, area, or volume.

The classical Greeks, as usual, articulated both sides of this duality thoroughly (see [1, Chapters 6 and 12] and [4, Section 1.1]). Thales, the first real mathematician, in the sense of proving rather than merely asserting, is said to have stated "all is water" (this would be a continuous model of the universe). Democritus believed the universe was made up of indivisible atoms, a discrete model.

It is interesting that information or outlook can make the difference between discrete and continuous. For example, suppose I am waiting for a bus. The bus route goes north until it arrives at the street I am waiting on, then makes a right turn onto my street. A bus on this route is invisible to me until it makes that right turn, then is visible until it reaches me. The bus route is continuous, but the appearance of the bus to my perceptions is discrete: either I see it or I don't, depending on whether it has made that right turn onto my street or not.

Another example of continuous becoming discrete is body or personality type. Whatever those words mean, given the infinite variety of homo sapiens, they surely describe something continuous. Certain types of scientists have created a trichotomy of body types or (perhaps going out of vogue now) personality types: endomorph, ectomorph, and mesomorph. Working only with those classifications is now a discrete world, of only three possible outcomes.

Weather forecasts, in particular the probability of rain, are continuous: a probability percentage could be any number from zero to one hundred. These numbers are often discretized, say into "no rain", "rain possible", "rain likely", and "rain certain".

Conversely, one could argue that any seemingly continuous substance is actually discrete, at the atomic or subatomic level.

Then there are objects like grains of sand; one could count grains (discrete), but it's easier to measure volume of sand (continuous).

Mathematics is sometimes broken into discrete math and continuous math: algebra and arithmetic are discrete, while geometry and calculus are continuous. Arguably the most successful theme in the evolution of math has been the intertwining of continuous and discrete math. For example, the interaction between algebra (calculation) and geometry (pictures) gives us the best of both worlds: the precision of algebra and the intuition of geometry; see, for example, [3, Chapter 6] or any of the vector magnifications on this website.

We mention the duality of this chapter because it is fundamental to how we represent and calculate probabilities. We have seen a discrete model of probability in the relative frequency fraction 1.5, where both numerator and denominator come from counting outcomes, and a continuous model in the Venn diagram of probability area 1.10, where probability is area.

Definition 2.1. A random variable is **discrete** if the set of possible values can be written as a sequence $\{x_1, x_2, x_3, \ldots\}$. A random variable X is **continuous** if the set of possible values is a union of intervals, and, for any real number c, $P(X = c)$, the probability that X equals c, is zero.

The random variable in Examples 1.12(a) is discrete, since the set of all possible values is $\{0, 1, 2\}$. The random variable X in Examples 1.12(b) is continuous: the set of all possible values is $\{\text{numbers } x \,|\, 0 \leq x \leq 12,450\}$, since the circumference of the earth is 24,900 miles. It is also true that $P(X = c) = 0$, for any real number c. For example, when we say "the distance is 3.52 miles," we don't really mean that. The alleged distance 3.52 is a *rounded* number, meaning only that

$$3.515 \leq \text{distance} < 3.525.$$

Definition 2.2. If X is a discrete random variable, then the **probability mass function** for X lists, for $k = 1, 2, 3, \ldots$, $p(x_k) \equiv P(X = x_k)$, for all possible values $x_1, x_2, x_3, \ldots$ for X.

Example 2.3. The probability mass function for X as in Examples 1.12(a) is, if the coin being flipped is fair,

x	0	1	2
$p(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

This is saying that

$$\frac{1}{4} = p(0) \equiv P(X = 0), \frac{1}{2} = p(1) \equiv P(X = 1), \frac{1}{4} = p(2) \equiv P(X = 2).$$

Notice that the Hs and Ts from the definition of X are gone. The probability mass function focuses on everything we want to know about X: possible values and how likely each possible value is. For any event A, the probability that X is in A is determined by adding up values of $p(x)$, for x in A. For example, the probability that X is less than 1.6 equals $p(0) + p(1) = \frac{3}{4}$.

Discussion 2.4. If X is a continuous random variable, then a probability mass function is not possible because there are too many possible values of X, and $P(X = c)$ is already guaranteed to be zero, for any real number c.

The relevant definition for continuous random variables, analogous to the probability mass function for discrete random variables, will be put in an appendix (Definition APP.1), to make our pre-appendix exposition accessible to readers who have not seen functions and graphs.

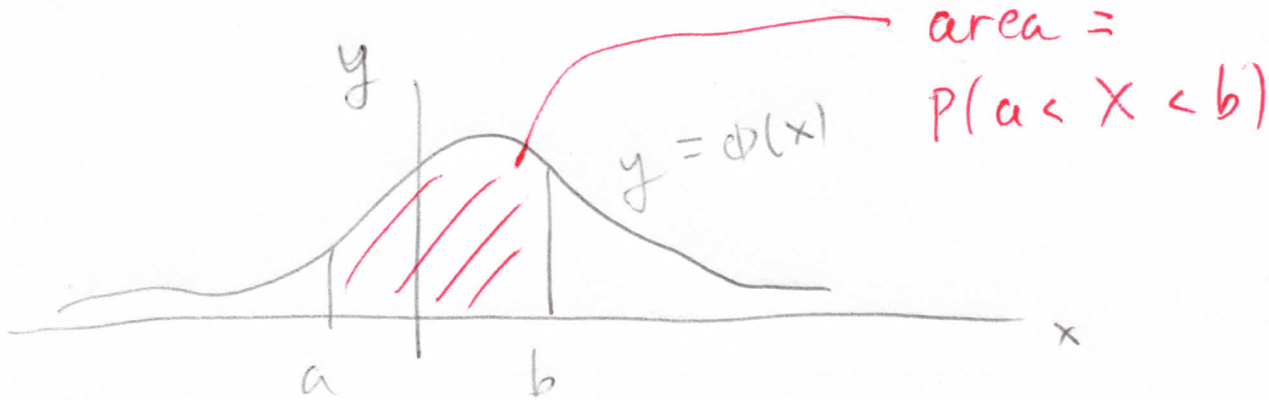
Suffice it to say here that calculating probabilities for a continuous random variable means calculating *area*, as with the Venn diagram 1.10. See Example APP.2 in the appendix.

See [2] for examples of discrete versus continuous variables.

APPENDIX

For this appendix only, the reader needs to know about the graph of a function.

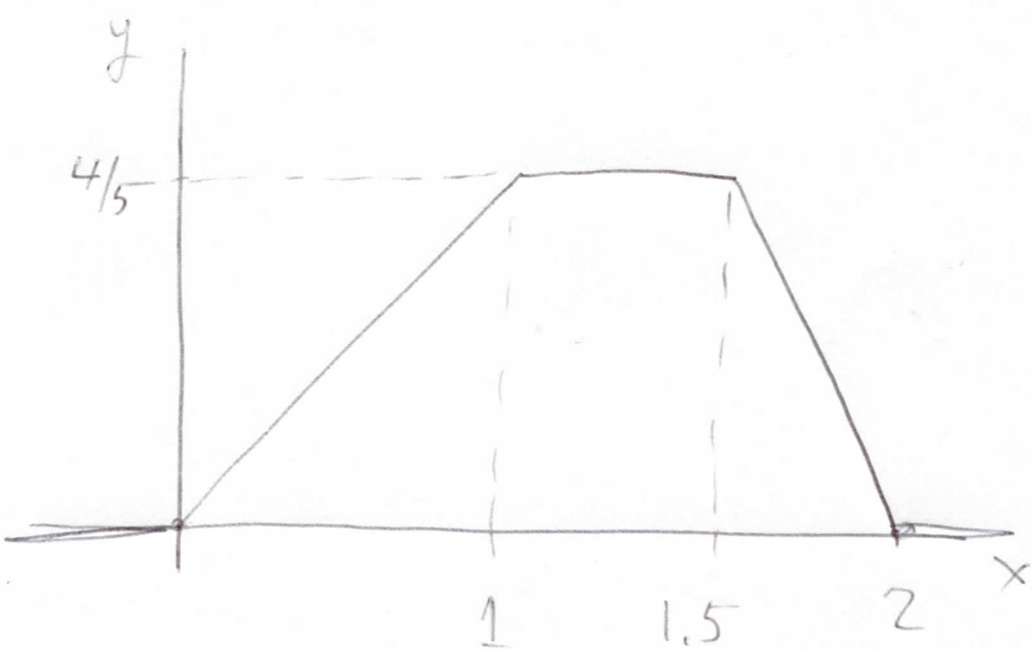
Definition APP.1. If X is a continuous random variable, then a **probability density function** for X is a function ϕ such that, for any real numbers $a < b$, $P(a < X < b) \equiv$ the probability that X is between a and b, is the area between $y = \phi(x), x = a, x = b$, and the y axis.



Example APP.2. Suppose the function ϕ whose graph is drawn below is a probability density function for the random variable X. Find

(a) $P(X > 1)$, the probability that X is greater than 1.

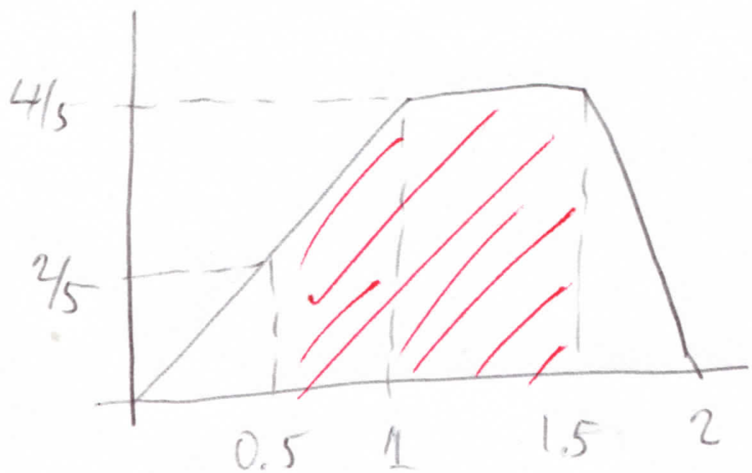
(b) $P(0.5 < X < 1.5)$, the probability that X is between 0.5 and 1.5.



Solutions. (a)

$$\text{area} = \frac{4}{5}(0.5) + \frac{1}{2}\left(\frac{4}{5}\right)(0.5) = \frac{2}{5} + \frac{1}{5} = \frac{3}{5}$$

(b)



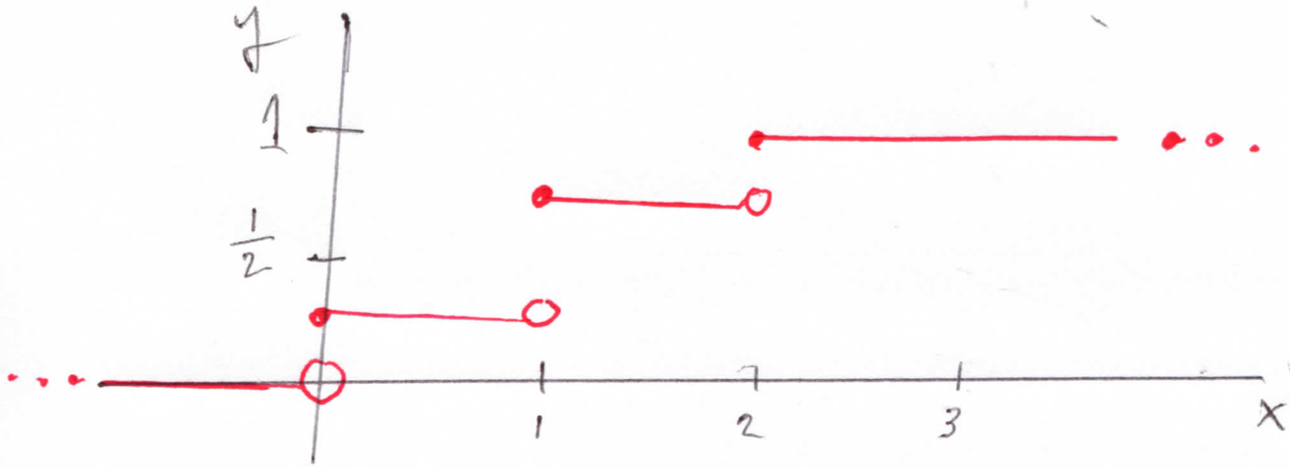
$$\text{area} = \frac{1}{2}\left(\frac{4}{5}\right)(1) - \frac{1}{2}\left(\frac{2}{5}\right)(0.5) + \frac{4}{5}(0.5) = \frac{2}{5} - \frac{1}{10} + \frac{2}{5} = \frac{7}{10}$$

Definition APP.3. Since our strategies for discrete and continuous random variables are so different, it is reassuring that there is a unifying idea, that works for any random variable.

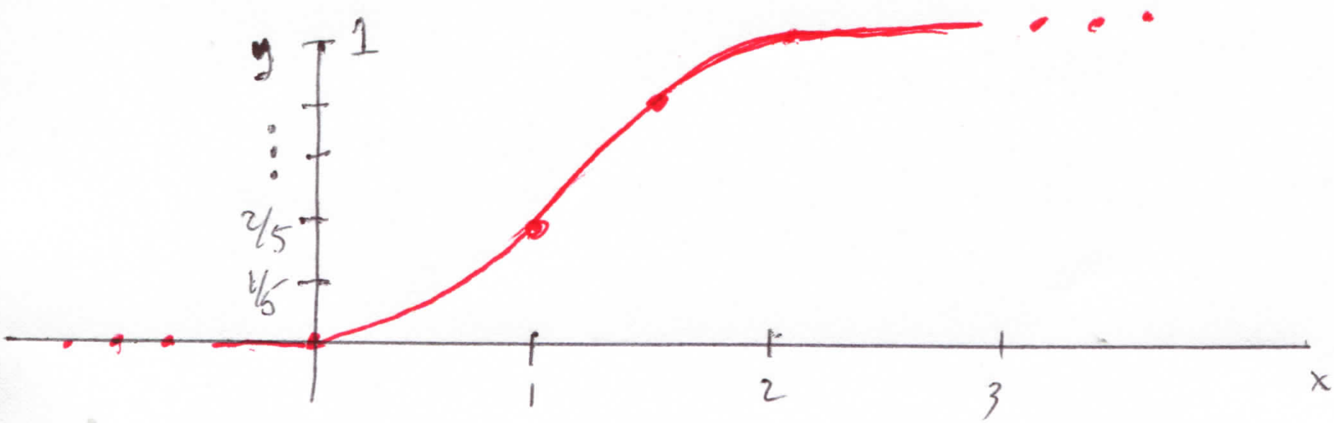
If X is a random variable, the **cumulative distribution function** for X is the following function.

$$F_X(x) \equiv P(X \leq x) \quad (x \text{ real}).$$

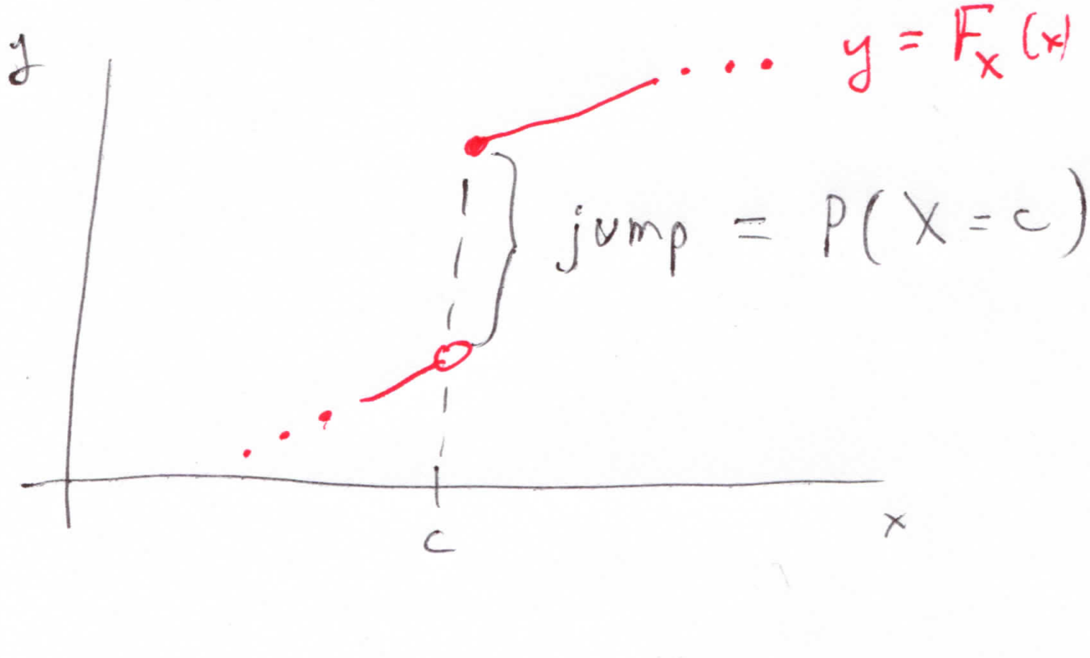
Examples APP.4. (a) toss fair coin twice, X equal to number of heads (see Example 2.3).



(b) Example APP.2.



A number c such that $P(X = c) > 0$ corresponds to a jump in the graph of F_X at $x = c$, as drawn below. In the language of calculus, F_X then has a *discontinuity* at $x = c$. See Examples APP.4(a).



REFERENCES

1. W.S. Anglin and J. Lambek, "The Heritage of Thales," Springer, 1995.

2. Ralph deLaubenfels, "Fun with Introductory Probability," www.teacherscholarinstitute/Books/Probability.pdf

3. Ralph deLaubenfels, "Linear Algebra," or E Pluribus Unum, http://teacherscholarinstitute.com/FreeMathBooksHighschool.html (2017).

4. J. Stillwell, "Mathematics and Its History," Springer-Verlag, 1989.