

## Chapter 2: Probability

Curtis Miller

2018-06-13

### Introduction

NEXT WE FOCUS ON probability. **Probability** is the mathematical study of randomness and uncertain outcomes. The subject may be as old as calculus. Modern statistics is based on probability theory, often estimating parameters that arise from a probability model. The subject is big and fascinating and sometimes shockingly counterintuitive. In this chapter we introduce basic ideas in probability theory and the theory of counting and combinatorics.

### Section 1: Sample Spaces and Events

An **experiment** is an activity or process with an uncertain outcome. Example experiments include:

- Flipping a coin
- Flipping a coin until the coin lands heads-up
- Rolling a six-sided die
- Rolling two six-sided dice
- The time in the morning you wake up

When we have an experiment we need to describe the **sample space**,  $\mathcal{S}^1$ , which is the set of all possible outcomes of the experiment. A **set** is loosely defined as a collection of objects.<sup>2</sup> **Events** are subsets of the sample space<sup>3</sup>, defining possible outcomes of an experiment. The **empty set** or **null event**,  $\emptyset$ , is a set with no members; it can be thought of as the event that nothing happens.

---

<sup>1</sup> Another extremely common notation for the sample space is  $\Omega$ .

<sup>2</sup> This definition cannot be rigorous because it leads to paradoxes. Bertrand Russell was able to find sets that, while legally defined this way, cannot logically exist. Examples include “a set of all sets” and “a set of sets that do not have themselves as members.” Axiomatic set theory defines sets in a way that avoids paradoxes but the theory is more complicated than necessary for typical use; the “naive” definition is usually fine.

<sup>3</sup> The sample space is a subset of the sample space and thus is an event, which can be thought of as the event that anything happens.

*Example 1*

Define a sample space for the experiment of flipping a coin. List all possible events for this experiment.

*Example 2*

Define a sample space for the experiment of rolling a six-sided die. List three events based on this sample space.

*Example 3*

Define a sample space describing the experiment of flipping a coin until it lands heads-up. List five events for this sample space.

*Example 4*

Define a sample space describing the experiment of rolling two six-sided die simultaneously. List three events from this sample space.

*Example 5*

Define a sample space describing the experiment of waking up in the morning at a particular time, where the time you wake up at (thought of as a real number) is the outcome of interest. List three events from this sample space.

---

Events can be manipulated in ways to “create” new events. Let  $A$  and  $B$  be events. The **complement** of  $A$ , denoted  $A^4$  is the set of outcomes of  $\mathcal{S}$  not in  $A$ , which in words is the event “not  $A$ ”. The **union** of two sets,  $A \cup B$ , is the set that combines the contents of the sets  $A$  and  $B$ , which in words means “ $A$  or  $B$ ”.<sup>5</sup> The **intersection** of two sets,  $A \cap B$ , is the set that only includes objects that appear in both  $A$  and  $B$ , which in words means “ $A$  and  $B$ ”.

Two sets are **disjoint** if they have no elements in common. In that case,  $A \cap B = \emptyset$ .

An intuitive approach to set theory is the use of **Venn diagrams**, where set-theoretic relations are illustrated by depicting objects as points on a plane and denoting set membership with enclosed regions. Below are Venn diagrams illustrating the relations between two sets just described.

<sup>4</sup> Other common notation includes  $\bar{A}$  and  $A^c$ .

<sup>5</sup> Sets only ever include one copy of each element, so  $\{H, H\} = \{H\}$ . This implies that if there is a copy of  $x$  in both  $A$  and  $B$ , there will not be two copies of  $x$  in  $A \cup B$ ; there is only one copy.

---

*Example 6*

Use a Venn diagram to illustrate  $(A \cup B)' \cup (A \cap B)$ .

*Example 7*

Consider three sets  $A$ ,  $B$ , and  $C$ . Illustrate:

1.  $A \cup B \cup C$

2.  $A \cap B \cap C$

$$3. (A \cap B) \cup (A \cap C) \cup (B \cap C)$$

*Example 8*

Describe the intersection, complement, and union of events described in Examples 1 through 5

---

*Section 2: Axioms, Interpretations, and Properties of Probability*

In probability our objective is to assign numbers to events describing how likely that event is to occur. Thus, a **probability measure**,  $\mathbb{P}$ , is a function taking events as inputs and returning numbers between 0 and 1, and satisfies the following three axioms:

1.  $\mathbb{P}(A) \geq 0$
2.  $\mathbb{P}(\mathcal{S}) = 1$
3. If  $A_1, A_2, \dots$  is a sequence of disjoint events (so that for any  $i \neq j$ ,  $A_i \cap A_j = \emptyset$ ), then  $\mathbb{P}(A_1 \cup A_2 \cup \dots) = \mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ <sup>6</sup>

From these, we get all other intuitive relations in probability.

**Proposition 1.**  $\mathbb{P}(\emptyset) = 0$

<sup>6</sup> You may understand this in the more common situation where if  $A \cap B = \emptyset$ ,  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ .

**Proposition 2.**  $\mathbb{P}(A') = 1 - \mathbb{P}(A)$

**Proposition 3.**  $\mathbb{P}(A) \leq 1$  for any event  $A$ .

**Proposition 4.**  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$  for any events  $A$  and  $B$ .

**Proposition 5.** For any events  $A$ ,  $B$ , and  $C$ ,

$$\begin{aligned}\mathbb{P}(A \cup B \cup C) &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \\ &\quad - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) \\ &\quad + \mathbb{P}(A \cap B \cap C)\end{aligned}$$


---

*Example 9*

Reconsider the experiment of flipping a coin, and assume that the coin is equally likely to land with each face facing up. Assign probabilities to all outcomes in the sample space.

*Example 10*

Do the same as Example 9, but when rolling a single dice.

*Example 11*

The dice from Example 10 has been altered with weights. Now, the probability of the dice rolling a 6 is twice as likely as rolling a 1, while all other sides still have the same probability of appearing as before. What is the new probability model?

*Example 12*

Reconsider the experiment of rolling two six-sided die. It's reasonable to assume that each outcome in  $\mathcal{S}$  is equally likely. What, then, is the probability of each outcome in  $\mathcal{S}$ ?

Use this model to find the probability of event  $E$ , where:

1.  $E = \{\text{At least one dice is a 6}\}$
2.  $E = \{\text{The sum of the pips showing on the two die is 5}\}$
3.  $E = \{\text{The maximum of the two numbers showing on the die is greater than 2}\}$

*Example 13*

Reconsider the experiment of flipping a coin until  $H$  is seen. What is one way to assign probabilities to all outcomes of this experiment so that we have a legal probability model? Justify your answer.

With this model, answer the following questions:

1. What is the probability the number of flips needed to see the first  $H$  exceeds 4?
2. What is the probability the number of flips until the experiment ends is between 3 and 20?
3. What is the probability that an even number of flips is seen before the experiment ends?

*Example 14*

In a small town, 20% of the population is considered “wealthy”, 30% of the population identifies as “black”, and 5% of the population is “wealthy” and “black”. Select a random individual from this popula-

tion (everyone equally likely to be selected). What is the probability this individual is “wealthy” and “not black”?

What is the probability this individual is neither wealthy nor black?

*Example 15*

A bag contains balls and blocks. 30% of the bag’s contents are balls. An object is either red or blue, and 40% of the objects are red. An object is made of either wood or plastic, and 65% of the objects are wooden. 10% of the objects are wooden balls, 5% of the objects are red balls, and 20% of the objects are red and plastic. 2% of the objects are red plastic blocks.

Reach into the bag and pick out an object at random, each object equally likely to be selected.

1. What is the probability the object selected is a ball, red, or wooden?

2. What is the probability the object is a red wooden ball?

3. What is the probability that the object is a blue plastic block?

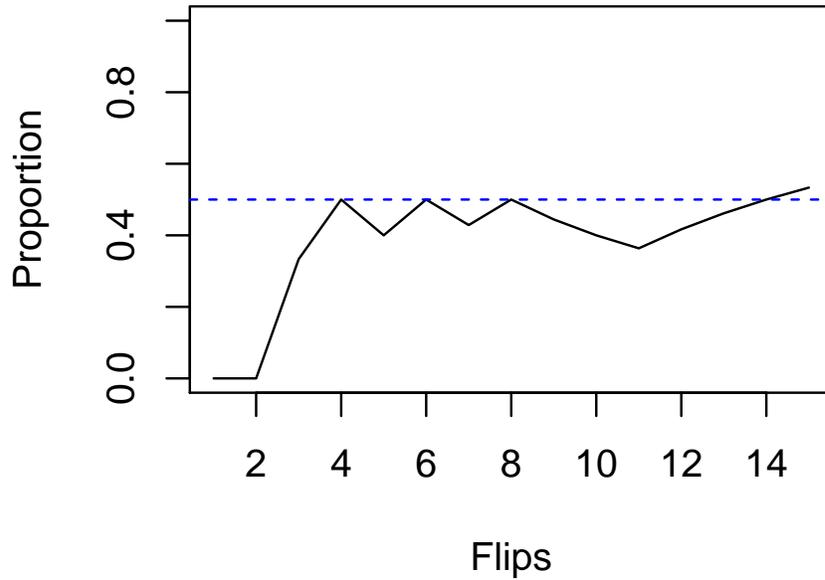
---

How do we interpret probabilities? The **frequentist interpretation of probability**<sup>7</sup> interprets probabilities as the long-run relative frequency as we repeat an experiment many times. For example, if we were to flip a fair coin many times, the proportion of times the coin lands heads up would approach  $\frac{1}{2}$ .

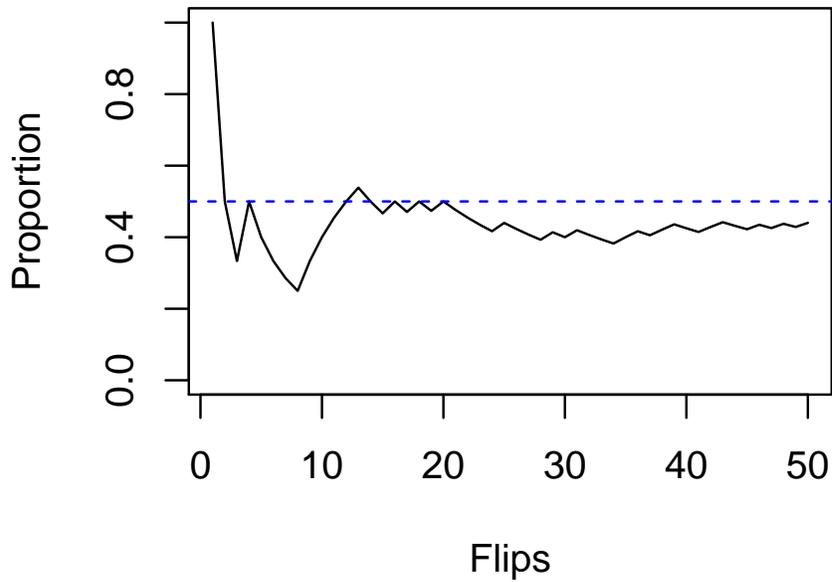
The chart below illustrates this idea.

<sup>7</sup> This interpretation isn't the only one. Any interpretation limits the kind of questions you can obtain probabilities for. In this case, the frequentist interpretation suggests that probabilities can be assigned only to repeatable experiments. While the frequentist interpretation is simple it can lead to convoluted language as we avoid referencing probabilities for nonrepeatable circumstances. The convoluted interpretation of confidence intervals, for example, is due to this interpretation of what a probability means. It turns out though that the rigorous mathematical theory of probability, which is based on measure theory and real analysis, does not care about the "interpretation" of a probability, so all the mathematics remain the same no matter what interpretation we choose.

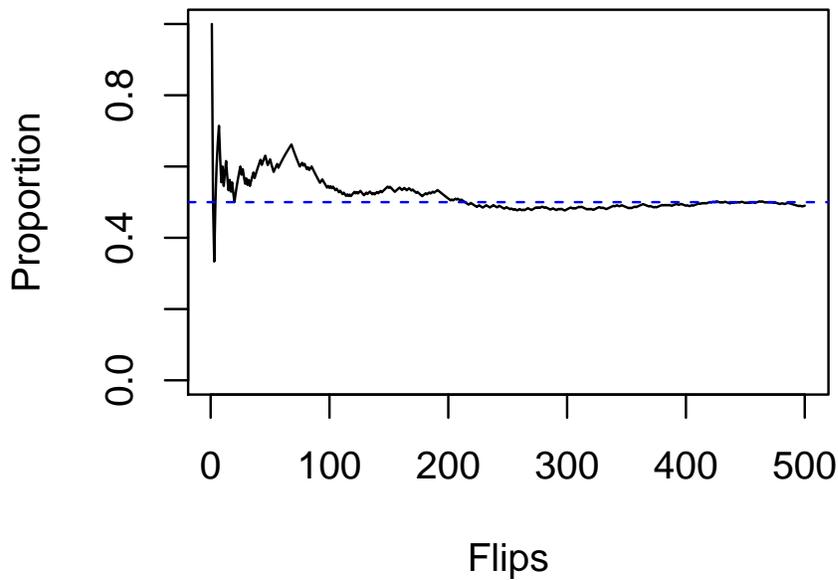
```
set.seed(11618) # Choosing a number to set the seed, for replicability
n <- 15
flips <- rbinom(n, 1, 0.5)
heads <- cumsum(flips == 1)
plot(1:n, heads/(1:n), type = "l", ylim = c(0, 1), xlab = "Flips",
     ylab = "Proportion")
abline(h = 0.5, col = "blue", lty = "dashed")
```



```
n <- 50
flips <- rbinom(n, 1, 0.5)
heads <- cumsum(flips == 1)
plot(1:n, heads/(1:n), type = "l", ylim = c(0, 1), xlab = "Flips",
     ylab = "Proportion")
abline(h = 0.5, col = "blue", lty = "dashed")
```



```
n <- 500
flips <- rbinom(n, 1, 0.5)
heads <- cumsum(flips == 1)
plot(1:n, heads/(1:n), type = "l", ylim = c(0, 1), xlab = "Flips",
     ylab = "Proportion")
abline(h = 0.5, col = "blue", lty = "dashed")
```



### Section 3: Counting Techniques

Consider a burger shop, Bob's Burgers, that offers three types of bread: white, rye, and sourdough. A burger can come with or without cheese. How many burgers are possible?

We first answer this question using a **tree diagram**:

Or we can answer using the **product rule**:

**Proposition 6.** *If there are  $n_1$  possibilities for choice 1,  $n_2$  possibilities for choice 2, ...,  $n_k$  possibilities for choice  $k$ , then there are  $n_1 n_2 \dots n_k = \prod_{i=1}^k n_i$  total possible combinations.*

Using the product rule:

---

*Example 16*

The sandwich shop Deluxe Deli offers four bread options (white, sourdough, whole wheat and rye), five meat options (turkey, ham, beef, chicken, no meat), six cheese options (cheddar, white cheddar, swiss, American, pepperjack, no cheese), with or without lettuce, with or without tomatoes, with or without bacon, with or without mayonaise, and with or without mustard. How many sandwiches are possible?

---

Suppose that out of  $n$  possibilities we will be choosing  $k$ . We have two essential questions to answer:

1. Do we choose with or without replacement?
2. Does order matter?

Depending on our answer our question has different solutions, summarized below:

|             | With replacement     | Without replacement                  |
|-------------|----------------------|--------------------------------------|
| Ordered     | $n^k$                | $P_{n,k} = \frac{n!}{(n-k)!}$        |
| Not ordered | $\binom{k+n-1}{n-1}$ | $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ |

### Justifications



---

*Example 17*

When we roll two six-sided die, we assume each outcome is equally likely (if the dice are different colors). How many possible outcomes are there? What about for three six-sided die?

*Example 18*

A high school has 27 boys playing men's basketball. In basketball, there are five positions: point guard (PG), shooting guard (SG), small forward (SF), power forward (PF), and center (C). Each assignment of player to position is unique. How many teams can be formed?

*Example 19*

When playing poker, players draw five cards from a 52-card deck. Every card is distinct, but the order of the draw does not matter. How many hands are possible?

```
# Example 16
```

```
6^2
```

```
## [1] 36
```

```
6^3
```

```
## [1] 216
```

```
# Example 17
```

```
factorial(27)/factorial(27 - 5)
```

```
## [1] 9687600
```

```
# Example 18
```

```
choose(52, 5)
```

```
## [1] 2598960
```

*Example 20*

You want to choose a dozen donuts from a donut shop. There are eight different kinds of donuts. How many boxes of a dozen donuts are possible?

`choose(12 + 8 - 1, 8 - 1)`

## [1] 50388

For the next few examples, we will be using a standard<sup>8</sup> 52-card deck of playing cards. In this deck, each card belongs to one of four suits: spades (♠), hearts (♥), clubs (♣), and diamonds (♦). Each card has a face value, which is either Ace (A), King (K), Queen (Q), Jack (J), or a number between 2 and 10; there are 13 possible face values. Hearts and diamonds are colored red, while spades and clubs are colored black. The notation  $8♦$  means “eight of diamonds”,  $K♠$  means “king of spades”, and so on.

<sup>8</sup>The “standard” deck is the French deck, the most common deck in the English-speaking world. Other European countries have their own traditional decks.

### *Example 21*

A poker hand is “four of a kind” if four cards have the same face value. How many four-of-a-kind hands exist?

### *Example 22*

A poker hand is “full house” if two cards have the same face value and three different cards have another common face value. How many “full house” hands exist?

```
# Example 20
(a1 <- 13 * 48)
```

```
## [1] 624
```

```
# Example 21
(a2 <- 13 * choose(4, 3) * 12 * choose(4, 2))
```

```
## [1] 3744
```

### Example 23

A “flush” is a poker hand where all cards belong to the same suit. How many “flush” hands exist (including straight flush hands)?

### Example 24

A “straight” is poker hand where the cards can be arranged in sequence: for example,  $5\spadesuit 6\clubsuit 7\clubsuit 8\heartsuit 9\heartsuit$  is a straight (suit does not matter). A “straight flush” is both a straight and a flush, so it is a flush with all cards belonging to the same suit (and the best possible hand). How many straight flush hands exist? How many straight hands exist (that are *not* straight flushes)?

```
# Example 22
(a3 <- 4 * choose(13, 5))

## [1] 5148

# Example 23
(a4 <- 10 * 4^5 - 4 * 10)

## [1] 10200
```

---

For finite sample spaces, there is a natural probability measure, defined below for a set  $A$ .

---

### *Example 25*

Use the natural probability measure to compute the probability of each poker hand mentioned in Examples 21 to 24.

```

s <- choose(52, 5)
a1/s # Exc. 20

## [1] 0.000240096

a2/s # Exc. 21

## [1] 0.001440576

a3/s # Exc. 22

## [1] 0.001980792

a4/s # Exc. 23

## [1] 0.003924647

```

---

#### Section 4: Conditional Probability

Consider flipping a fair coin three times. What is the probability the same face will appear three times?

Now suppose I told you that the first two flips were  $HH$ . What is the probability of this event now?

This demonstrates the need for **conditional probability**, which is a probability of an event given the fact that another event has occurred. The probability of  $A$  given  $B$  has occurred, denoted  $\mathbb{P}(A|B)$ , is:

There is an illustration for making this definition intuitive:

Given a conditional probability we can also compute  $\mathbb{P}(A \cap B)$ :

Given  $\mathbb{P}(A|B)$ , what is  $\mathbb{P}(A'|B)$ ?

---

*Example 26*

Use the definition of conditional probability to compute the probability of the event that all three coins have the same face up when flipped given the first two flips were heads.

*Example 27*

Suppose that you were dealt two cards of a five-card poker hand, which are  $K\heartsuit 8\heartsuit$ . Given this information, what is the probability your complete hand will be a full house?

```
# Hands with KH 8H
(den <- choose(50, 3))

## [1] 19600

# Full house hands with KH 8H
(num <- 2 * choose(3, 1) * choose(3, 2))

## [1] 18

num / den

## [1] 0.0009183673

(num/s) / (den/s)

## [1] 0.0009183673
```

### *Example 28*

Suppose your five-card poker hand is a flush. What is the probability it is a straight flush?

### *Example 29*

Suppose your five-card poker hand is a straight. What is the probability it is a straight flush?

*Example 30*

In a certain village 20% of individuals are considered “wealthy” and 35% are considered “black”. Among blacks, 60% are not considered “wealthy”. If you chose a random individual from this village, what is the probability this individual is “black” and “wealthy”?

---

A **partition** is a division of  $\mathcal{S}$  into sets  $A_1, \dots, A_n$  such that for  $i \neq j$ ,  $A_i \cap A_j = \emptyset$ , and  $\bigcup_{i=1}^n A_i = \mathcal{S}$ . Below is an illustration:

**Theorem 1** (Law of Total Probability). *Let  $A_1, \dots, A_n$  be a partition of  $\mathcal{S}$  and  $B$  be an event. Then:*

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B|A_i) \mathbb{P}(A_i)$$

We can then state Bayes’ Theorem<sup>9</sup>:

**Theorem 2** (Bayes’ Theorem). *Let  $A_1, \dots, A_n$  be a partition of  $\mathcal{S}$  and  $B$  be an event. Then:*

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i) \mathbb{P}(A_i)}{\sum_{j=1}^n \mathbb{P}(B|A_j) \mathbb{P}(A_j)}$$


---

<sup>9</sup> Bayes’ Theorem is also seen in the simpler form where the partition is  $A$  and  $A'$ , in which case the statement becomes

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A) \mathbb{P}(A)}{\mathbb{P}(B|A) \mathbb{P}(A) + \mathbb{P}(B|A') \mathbb{P}(A')}$$

*Example 31*

Roll a fair six-sided dice. Then, after observing the number of pips, roll another dice until more than that number of pips appears. What is the probability that the second die roll will show four pips?

*Example 32*

In a city Uber and Lyft transport passengers. 95% of drivers work for Uber, and 5% work for Lyft. One day there is a hit-and-run accident and a witness claims that she noticed the driver worked for Lyft. Lyft's defense attorneys subject her to testing, and in testing determine that she correctly identifies a car as belonging to Lyft 90% of the time but will claim a vehicle belongs to Lyft incorrectly 20% of the time. Based on this evidence, how likely is it that the driver who hit the pedestrian worked for Lyft?

---

### Section 5: Independence

Two events  $A$  and  $B$  are **independent** if  $\mathbb{P}(A|B) = \mathbb{P}(A)$ . In some sense, information about the event  $B$  gives no information about whether  $A$  happened.

Use this to compute  $\mathbb{P}(B|A)$ .

Use this to compute  $\mathbb{P}(A'|B)$ .

A consequence of this definition of independence:<sup>10</sup>

<sup>10</sup> In fact, this may be a more common definition of independence.

Below is a graphical representation of independence:<sup>11</sup>

<sup>11</sup> Notice that independence is *not* the same as being disjoint. In fact, two disjoint events are not independent except in the most trivial cases. (That is,  $S$  and  $\emptyset$  are technically independent.)

---

*Example 33*

Consider rolling a 6-sided dice. Show that the events  $A = \{\text{Number does not exceed 4}\}$  and  $B = \{\text{Number is even}\}$  are independent.

---

Suppose we have events  $A_1, \dots, A_n$ . These events are **mutually independent** if, for  $k \leq n$ :

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}\left(\bigcap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k \mathbb{P}(A_{i_j}) = \mathbb{P}(A_{i_1}) \mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_k})$$

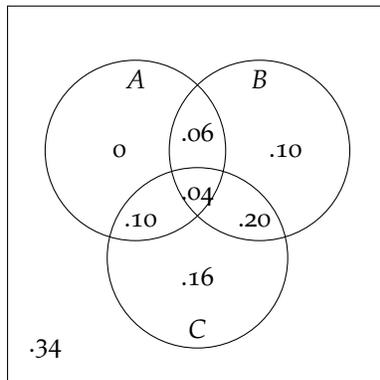
This definition cannot be simplified to  $\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \dots \mathbb{P}(A_n)$ , as demonstrated below.<sup>12</sup>

<sup>12</sup> This example was written by George [2004] and is available here: <http://www.engr.mun.ca/~ggeorge/MathGaz04.pdf>

---

*Example 34*

Using the diagram below for finding probabilities, compute  $\mathbb{P}(A \cap B \cap C)$  and  $\mathbb{P}(A) \mathbb{P}(B) \mathbb{P}(C)$ . Are  $A$ ,  $B$ , and  $C$  mutually independent?



*Example 35*

We flip eight fair coins. What is the probability of  $H$ ?  $TH$ ?  $TTH$ ?  $TTTH$ ? In general, what is the probability for a sequence of  $n$  flips to have  $n - 1$   $T$  and a  $H$  at the end?

*Example 36*

Below is a system of components. A signal will be sent from one end of the system, and will be successfully transmitted to the other end if no intermediate components fail. Each component functions independently of the others. What is the probability a transmission is sent successfully?

*Example 36*

Below is another system of components. A signal will be sent from one end of the system, and will be successfully transmitted to the other end if no intermediate components fail. Each component functions independently of the others. What is the probability a transmission is sent successfully?

*References*

Glyn George. Testing for the independence of three events. *Mathematical Gazette*, 88, nov 2004.