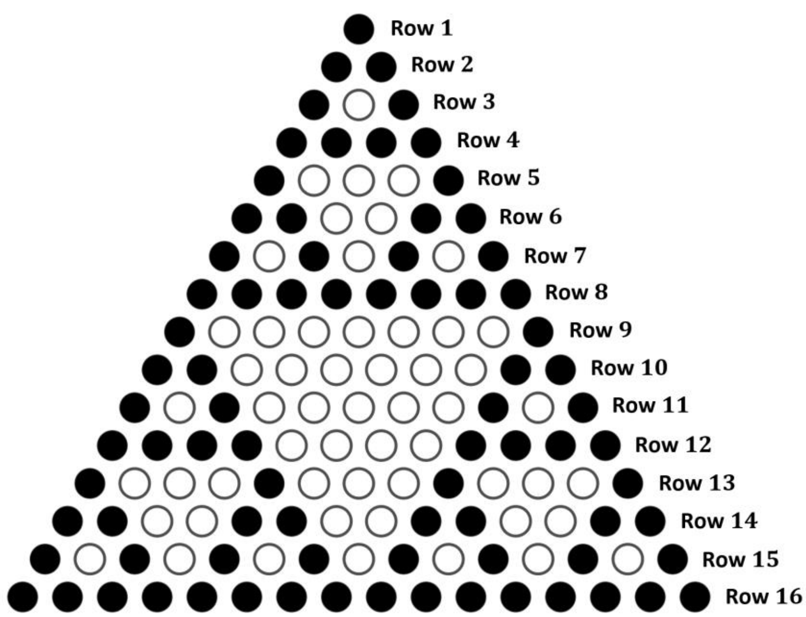
# Probability Theory

## Study Notes

Ken Suman

ksuman@winona.edu




For MN Students Preparing For

High School Mathematics Contests

The triangle on the cover page is constructed by representing the odd numbers in Pascal's triangle as black dots and the even numbers as white dots. The question "What is the probability that a randomly picked number from the first 128 rows of Pascal's triangle is an odd number?" dates back to the nineteenth century French mathematician Édouard Lucas who developed the theory of binomial coefficients modulo a prime power in his text *Théorie Des Nombres*. In Project 1 we present the necessary parts of his theory to answer the above question. But the patterns visible in this diagram are enough to reveal the answer if you look closely.

**Preface**

The intent of this course is to prepare you for solving probability problems on high school level mathematics contests – especially the MSHSML and AMC 10/12 exams. These notes focus on problem solving – except in Chapter 1, which is intended to be the minimum background you will need to navigate through any of the other chapters, in any order that suits your interests. Optimally, you will have the time to read through Chapter 1 prior to attending the SMI lectures and will be prepared for a quick launch into the problem-solving techniques of the remaining chapters. The exercises for Chapter 1 are primarily designed to offer practice in pushing formulas around – a mundane but necessary skill for all mathematicians.

Counting problems (how many ways can you distribute 8 identical black balls into 3 labeled boxes with at least one ball per box) often come in the form of probability problems (what is the probability that each box gets at least one ball if we assume that all possible distributions of 8 identical black balls into 3 labeled boxes are equally likely). Counting techniques was the topic of MSHSML SMI in 2018 and are not covered again in these notes.

These notes stress the areas of probability (apart from counting techniques) that are stressed in MSHSML, AMC 10/12 and other high school mathematics contests. Accordingly, these notes do not cover any aspect of probability that requires calculus or higher levels of mathematical background because calculus is expressly not required for solving any problem on MSHSML or AMC 10/12 exams.

High school level math contests (and hence these notes) stress geometric probability problems in much more detail than is typical in university probability textbooks. Briefly, geometric probability problems involve using geometry (among other subjects) to find lengths, areas and volumes necessary to solve a probability problem. Sometimes the need for geometry is disguised. For example, a classic geometric probability problem asks for the probability that two numbers picked at random from the set of all positive integers will have no common divisor. [By the way, the exact answer (as we will see) is $6/\pi^2$, which is rough $0.61$. Where does this $\pi$ come from?]

Shortcuts are part of game in contest mathematics. It is not infrequent to come across a math contest problem which has, to my mind at least, been specifically designed to allow for both a straightforward (read longer) approach as well as a quick solution *if* you know "the trick". Shortcuts in probability often depend on being about to recognize the symmetry in a problem. I have attempted to bring "shortcuts" to light where I could. In Chapter 5 on *exchangeable random variables* you will see how symmetry in sampling without replacement problems (suppose you draw three times without replacement from an urn containing 5 red and 4 black balls …) can be used to great effectiveness.

I have adopted my own numbering system to identify problems from the old MSHSML tests. In these notes a code such as 5C194 can be parsed by separating the this into three parts. The first two entries in the code identify the test, the next two identify the start date of the school year and the last one (sometimes two) entries represent the question number. So 5C194 identifies question 4 on test 5C from school year 19**94**-1995. For the first two symbols in the code I use, for example, 5A through 5D to identify tests A through D from the 5th meet of that season and I use 5T to identify the team test for that meet. To identify tests from the state tournament the first entry is a T. So TA944 would be question 4 on test A at the state tournament of 1994-1995 and TT944 would be question 4 on the team test at the state tournament of 1994-1995. TI identifies the Invitational test at a state tournament and TM identifies a Math Bowl question at the state tournament.

# Contents

# Chapter 1 The Basics of Probability

The starting point for probability is an ***experiment*** with a given set $S = \{o_1, o_2, \dots\}$ of possible ***outcomes***. The set $S$ of all possible outcomes is called the ***sample space*** of that experiment.

For example, our experiment could be to roll a standard six-sided die. In this case our sample space of all possible outcomes to this experiment is $S = \{1,2,3,4,5,6\}$. In this example, our experiment has a sample space $S$ made up of six outcomes.

Subsets of the sample space $S$ of an experiment are called ***events***. Events are typically denoted by capital letters from the beginning of the English alphabet, such as events $A$ or $B$, etc.

The distinction between outcomes and events is that ***outcomes cannot be further refined*** into smaller pieces.

Continuing with our above example, let $E = \{2,4,6\}$. $E$ is an <u>event</u> of $S$because $E \subset S = \{1,2,3,4,5,6\}$. $\{2\}, \{4\}, \{6\}$ are each <u>outcomes</u> of $S$ because they cannot be broken apart into more refined parts. On the other hand, $E = \{2,4,6\}$ is not an outcome because it can be broken apart into three more refined pieces.

Our intuitive notions of probability stem from chance experiments which we can perform over and over – think again in terms of rolling a die or flipping a coin.

> If we focus on repeatable experiments, we can define the ***probability of an outcome*** $o_j$ as the long run percentage$^\star$ of times this outcome occurs.

$^\star$Long Run Percentage – the theoretical value of the ratio $a_j(m)/m$ as $m$ increases to infinity, where $a_j(m)$ is the number of times outcome $o_j$ occurs in $m$ replications of the experiment.

Events can be written in set notation as the **union** of outcomes.

> This allows us to define $P(A)$, the ***probability of the event $\boldsymbol{A}$***, as the sum of the probabilities of the distinct outcomes making up event $A$.

It should be understood that when we write $P(A)$ or when we say "the probability of event $A$", that is actually a shortcut for the more precise statement "the probability that the outcome of the experiment belongs to the set $A$".

Connecting probability to long run percentages (real or thought experiments) is called the empirical or frequentist approach to probability. Throughout these notes we will adopt the frequentist model.

## The Language of Probability

### "Or" and "And" in Probability

In set theory the ***union*** symbol ∪ translates to the word "**or**" while the **Intersection** symbol ∩ translates to the word "**and**".

This is also true in the language of probability theory. So, for example, we would read $P(A \cup B)$ as the probability that event $A$ occurs "or" the event $B$ occurs and we would read $P(A \cap B)$ as the probability that event $A$ occurs "and" the event $B$ occurs.

### "Or" Can Be Ambiguous Outside of Mathematics

The phrase "$A$ or $B$" in some contexts carries the meaning ***one but not both*** of $A$ and $B$ are true. For example, "This coupon entitles you to a free hamburger or fish sandwich."

In some contexts, "$A$ or $B$" is used to indicate ***not just one but both*** of $A$ and $B$ are true. For example, "I did not see you in the classroom or the hallway."

In some contexts, "$A$ or $B$" is meant to be understood as ***one***, ***both*** or ***neither*** of $A$ and $B$ are true. For example, "Would you like cream or sugar in your coffee?"

Finally, in some contexts "$A$ or $B$" means ***at least one*** of $A$ and $B$ is true. For example, "You will earn an "A" in the class if you have at least a 93 average on all the tests or if you make at least a 98 on the final exam."

In probability, and in mathematics more broadly, the word "***or***" is by definition ***always*** to be interpreted in the "at least one" sense. So $P(A \cup B)$ is necessarily to be interpreted as "the probability that the outcome of the experiment belongs to ***at least one*** of the sets $A$ and $B$".

### Fundamental Properties

**Theorem 1.** If $S$ is the sample space of an experiment, then $P(S) = 1$.

By definition, the sample space $S = \{o_1, o_2, \dots, o_n\}$ of an experiment is the union of all outcomes of that experiment. That is,

$$S = \{o_1\} \cup \{o_2\} \cup \cdots \cup \{o_n\}.$$

Furthermore, $S$ is by definition an event in the sample space $S$.

Therefore, by the definition of the probability of an event as the sum of the probability of the outcomes making up that event, it is necessarily true that

$$P(S) = P(o_1) + P(o_2) + \cdots + P(o_n) = 1.$$

■

Two sets $A$ and $B$ are ***disjoint*** if their intersection (set of shared elements) is the empty set. So the two sets $A = \{2,3\}$ and $B = \{7,11,12\}$ are disjoint because $A \cap B = \emptyset$.

If the set of outcomes making up two events are disjoint, we say those two events are ***mutually exclusive***. Imagine now that our experiment is to roll a pair of dice and to calculate the sum. This is another example of a stochastic repeatable experiment and in this case the sample space would be $S = \{2,3,4,\dots,12\}$. Clearly, $A = \{2,3\}$ and $B = \{7,11,12\}$ are both events of this experiment and clearly these two sets are disjoint. Therefore, we would say that these two events are *mutually exclusive*.

In the case of more than two events, we say the $t$ events $A_1, A_2, \dots, A_t$ are mutually exclusive if the set of outcomes defining these events are ***pairwise disjoint***. That is if $A_i \cap A_j = \emptyset$ for all $i \neq j$.

So, if we define events $A = \{2,3\}$, $B = \{1,3\}$ and $C = \{4,5,6\}$ of the sample space $S = \{2,3,4,\dots,12\}$ then $A \cap B \cap C = \emptyset$ because there is no outcome shared by *all three events*. However, these events are **not** mutually exclusive because these events are not pairwise disjoint. In particular, $A \cap B \neq \emptyset$.

**Theorem 2.** If $A_1, A_2, \dots, A_k$ are mutually exclusive events of the sample space $S$, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_k) = P(A_1) + P(A_2) + \cdots + P(A_k).$$

Consider the special case of events $A = \{2,3\}$, $B = \{4,5,6\}$ and $C = \{7\}$ from the sample space $S = \{2,3,4,\dots,11,12\}$.

In this case, the events $A, B$ and $C$ are mutually exclusive$^\star$ because they are pairwise disjoint. That is, $(A \cap B) = (A \cap C) = (B \cap C) = \emptyset$. Also, we can see that $A \cup B \cup C$ is an event (it is a subset of $S$). To be clear,

$$A \cup B \cup C = \{2,3,4,5,6,7\} \subset \{2,3,4,5,6,7,8,9,10,11,12\}.$$

But (by definition) the *probability of an event* is the sum of the probabilities of the distinct outcomes making up that event. Therefore,

$$P(A \cup B \cup C) = P(\{2\}) + P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\}) + P(\{7\}).$$

But $A$, $B$ and $C$ are also events. Therefore,

$$\begin{aligned} P(A) &= P(\{2\}) + P(\{3\}) \\ P(B) &= P(\{4\}) + P(\{5\}) + P(\{6\}) \\ P(C) &= P(\{7\}). \end{aligned}$$

And in this case, it is clear by inspection that $P(A \cup B \cup C) = P(A) + P(B) + P(C)$.

*Note: The assumption that $A_1, A_2, \dots, A_k$ are mutually exclusive is the **key** to this argument. By definition all elements in a ***set*** are distinct – no repetitions allowed. So even if these events were not mutually exclusive, the probability of shared outcomes would only be counted once in the left-hand side term $P(A_1 \cup A_2 \cup \cdots \cup A_k)$. However, any shared outcomes would be doubly (or multiply) counted in the right-hand side terms $P(A_1) + P(A_2) + \cdots + P(A_k)$. Hence the two sides could ***not*** be equal.

■

The ***complement*** of a set $A$, denoted as $A'$ (alternately denoted as $\overline{A}$ or $A^c$) is defined as the set of all outcomes in $S$ that are <u>not</u> in $S$. Therefore, by definition $A \cap A' = \emptyset$ and $A \cup A' = S$.

Accordingly, if $A$ is an event in the sample space $S$, then by definition $A'$ is also an event in $S$ and furthermore the **events $A$ and $A'$ are mutually exclusive**.

**Theorem 3.** For any event $A$,

$$P(A') = 1 - P(A).$$

Because the events $A$ and $A'$ are mutually exclusive, it follows from Theorem 2 that $P(A \cup A') = P(A) + P(A')$. But $P(A \cup A') = P(S) = 1$ by Theorem 1, therefore,

$$P(A) + P(A') = 1 \text{ and } P(A') = 1 - P(A).$$

■

**Corollary of Theorem 3.** For any experiment, $P(\emptyset) = 0$.

By definition the complement of the sample space of an experiment is the empty set. Therefore, by Theorems 1 and 3, $P(\emptyset) = 1 - P(S) = 1 - 1 = 0$.

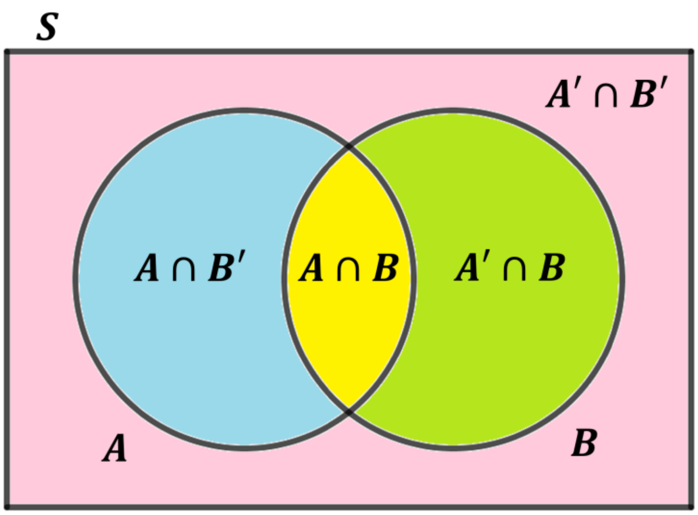
■

**Theorem 4.** If $A$ and $B$ are any* two events in the sample space $S$, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

*Events $A$ and $B$ are not necessarily mutually exclusive in Theorem 4. If $A$ and $B$ are mutually exclusive, then $P(A \cap B) = P(\emptyset) = 0$ and this result is just a special case of Theorem 2.

The essence of Theorem 4 is most clearly seen from a Venn diagram.




Notice that

(i) $(A \cap B')$, $(A \cap B)$ and $(A' \cap B)$ are mutually exclusive events in $S$.

(ii) $A = (A \cap B') \cup (A \cap B)$

(iii) $B = (A \cap B) \cup (A' \cap B)$

(iv) $A \cup B = (A \cap B') \cup (A \cap B) \cup (A' \cap B)$,

Because the blue, yellow and green define mutually exclusive events in $S$, it follows from Theorem 2 that

$$P(A) = P(A \cap B') + P(A \cap B)$$
$$P(B) = P(A \cap B) + P(A' \cap B)$$
$$P(A \cup B) = P(A \cap B') + P(A \cap B) + P(A' \cap B).$$

Hence,

$$\begin{aligned}
P(A \cup B) &= P(A \cap B') + P(A \cap B) + P(A' \cap B) \\
&= P(A \cap B') + P(A \cap B) + P(A' \cap B) + \left(\boldsymbol{P(A \cap B) - P(A \cap B)}\right) \\
&= \left(P(A \cap B') + P(A \cap B)\right) + \left(P(A' \cap B) + P(A \cap B)\right) - P(A \cap B) \\
&= P(A) + P(B) - P(A \cap B).
\end{aligned}$$

■

Having established this formula for $P(A \cup B)$ which applies whether the events $A$ and $B$ are mutually exclusive or not, it is natural to consider if there is a parallel formula for $P(A \cup B \cup C)$ or $P(A \cup B \cup \cdots \cup Z)$ which holds whether the events are mutually exclusive or not. This leads us to the "**method of inclusion-exclusion**". We illustrate the method for the case of five events.

$$\begin{aligned}
P(A \cup B \cup C \cup D \cup E) \quad = \quad & + \quad \left(P(A) + P(B) + P(C) + P(D) + P(E)\right) \\
& - \quad \left(P(A \cap B) + \cdots + P(D \cap E)\right) \\
& + \quad \left(P(A \cap B \cap C) + \cdots + P(C \cap D \cap E)\right) \\
& - \quad \left(P(A \cap B \cap C \cap D) + \cdots + P(B \cap C \cap D \cap E)\right) \\
& + \quad P(A \cap B \cap C \cap D \cap E).
\end{aligned}$$

To be absolutely clear about this notation, the abbreviated notation used in third row of "three-way" intersections would in its entirety be

$$\begin{aligned}
-\Big(P(A \cap B \cap C) &+ P(A \cap B \cap D) + P(A \cap B \cap E) + P(A \cap C \cap D) + P(A \cap C \cap E) \\
&+ P(A \cap D \cap E) + P(B \cap C \cap D) + P(B \cap C \cap E) + P(B \cap D \cap E) \\
&+ P(C \cap D \cap E)\Big).
\end{aligned}$$

The pattern of first adding (including) all "one way" intersections, then subtracting (excluding) all "two way" intersections, then adding (including) all "three way" intersections, etc. continues for cases with any number of events.

Problems requiring the method of inclusion-exclusion come up occasionally on MSHSML tests and we will consider a few of these in our final chapter where we take up "Miscellaneous" problems.

The two results known as **DeMorgan's Laws** are often part of solving inclusion-exclusion problems so we mention them now. Suppose $A_1, A_2, \dots, A_n$ are events in a sample space $S$. Then

$$P\big((A_1 \cup A_2 \cup \cdots \cup A_n)'\big) = P\big(A_1' \cap A_2' \cap \cdots \cap A_n'\big)$$

and

$$P\big((A_1 \cap A_2 \cap \cdots \cap A_n)'\big) = P\big(A_1' \cup A_2' \cup \cdots \cup A_n'\big).$$

The verification of DeMorgan's Laws are an exercise in what is called "set chasing" in contest problem solver blogs. In set theory, we say the sets $A = B$ if for all $x$ it is true that $x \in A \Leftrightarrow x \in B$.

We note that

$$\begin{aligned}
&x \in (A_1 \cup A_2 \cup \cdots \cup A_n)' \\
&\qquad \Leftrightarrow x \notin (A_1 \cup A_2 \cup \cdots \cup A_n) \\
&\qquad \Leftrightarrow (x \notin A_1) \text{ and } (x \notin A_2) \text{ and } \cdots \text{ and } (x \notin A_n) \\
&\qquad \Leftrightarrow \big(x \in (A_1)'\big) \text{ and } \big(x \in (A_2)'\big) \text{ and } \cdots \text{ and } \big(x \in (A_n)'\big) \\
&\qquad \Leftrightarrow x \in \big((A_1)' \cap (A_2)' \cap \cdots \cap (A_n)'\big).
\end{aligned}$$

This establishes that $(A_1 \cup A_2 \cup \cdots \cup A_n)' = A_1' \cap A_2' \cap \cdots \cap A_n'$. From here it follows that

$$P\big((A_1 \cup A_2 \cup \cdots \cup A_n)'\big) = P\big(A_1' \cap A_2' \cap \cdots \cap A_n'\big).$$

The proof of the other DeMorgan's Law follows similarly.

■

**Theorem 5.** Problems with Equally Likely Outcomes

If the sample space $S$ of an experiment consists of $n$ equally likely outcomes $o_1, o_2, \dots, o_n$ and if the event $A$ of $S$ consists of $k$ of those $n$ outcomes, then $\boldsymbol{P(A) = k/n}$.

By Theorem 1, $P(S) = P(o_1) + P(o_2) + \cdots + P(o_n) = 1$. In the case of all equally likely outcomes, it follows that $P(o_j) = 1/n$ for all $1 \leq j \leq n$. Furthermore, by the definition of the probability of an event, if $A = \{o_{i_1}\} \cup \{o_{i_2}\} \cup \cdots \cup \{o_{i_k}\}$ then

$$\begin{aligned} P(A) &= P(\{o_{i_1}\}) + P(\{o_{i_1}\}) + \cdots + P(\{o_{i_k}\}) \\ &= \frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n} = \frac{k}{n}. \end{aligned}$$

■

Theorem 5 is sometimes called the **counting definition** of probability because it can be restated as,

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

for all events $A$ in $S$ whenever all outcomes in $S$ are equally likely.

**Conditional (Rescaled) Probabilities**

A familiar way to think about conditional probability is to make an analogy to test grades. Suppose you have a class where the tests together count for $60\%$ of your overall grade, where the final exam counts for $20\%$ of your grade, your homework counts for $10\%$ and a project counts for $10\%$ of your overall grade.

Now suppose that midway through the semester your teacher announces that he/she is cancelling the required project and will reset the percentages for the tests, final and homework so that the remaining components keep their same relative importance. What would the new percentages have to be?

For grades to keep their "same relative importance" means that if the original plan gave weights $w_1, w_2, w_3, w_4$ to the test, final, homework and project, then the rescaled grades become $kw_1$, $kw_2$ and $kw_3$ after the project has been scrapped. The constant $k$ is that number necessary to make the updated weights again sum of $100\%$. That is,

$$k(60\%) + k(20\%) + k(10\%) = 100\% \Leftrightarrow k = 100/90.$$

So the new grading plan would be to count the tests for

$$(60\%)\cdot\frac{100}{90}=66.\bar{6}\,\%$$

$$(20\%)\cdot\frac{100}{90}=22.\bar{2}\,\%$$

and

$$(10\%)\cdot\frac{100}{90}=11.\bar{1}\,\%.$$

Notice that the <u>relative</u> weights have not changed. The tests are still weighted three times as heavily as the final exam and the final exam is still worth twice as much as the homework.

Now let's keep the same numbers but change the storyline to one where John has agreed to a foot race with his good friends Mary, Max and Min. Max is the track star of the group and his probability of winning the race has been accessed to be $0.60$ (*i.e.* $60\%$). John, Mary and Min come in close together with winning probabilities of $20\%$, $10\%$ and $10\%$ respectively. On the afternoon of the race Mary finds out she has chores and cannot be there for the race after all. What are the new probabilities of winning for Max, John and Min now that Mary has dropped out?

By direct analogy to the question with rescaling test grades they will become $66.\bar{6}\,\%$, $22.\bar{2}\,\%$ and $11.\bar{1}\,\%$ respectively.

Let's generalize this. Suppose we have an experiment with sample space $S=\{o_1,o_2,\dots,o_n\}$. Suppose something causes all the outcomes <u>not</u> in $F=\left\{o_{f_1},\dots,o_{f_r}\right\}\subset S$ to be ruled out as possible outcomes of the experiment. (In analogy with removing the project from the set of graded elements for the class and Mary dropping out of the race.)

What are the conditional (rescaled) probabilities for the remaining outcomes $\left\{o_{f_1},\dots,o_{f_r}\right\}$? These conditional probabilities are denoted by $P\left(o_{f_j}\middle|F\right)$, $j=1,2,\dots,r$.

**Theorem 6.** Conditional Probability of Outcomes

$$P\left(o_{f_j}\middle|F\right)=\frac{P\left(o_{f_j}\right)}{P(F)}$$

for all outcomes in $F=\left\{o_{f_1},\dots,o_{f_r}\right\}\subset S$.

First note that this theorem shows how the conditional probability $P\left(o_{f_j}\middle|F\right)$ can be computed as the ratio of two unconditional probabilities $P\left(o_{f_j}\right)$ and $P(F)$.

The fundamental principle of conditioning on the event $F$ is to declare that all outcomes in $S$ but not in $F$ (*i.e.* the outcomes in $F'$) are no longer possible (*i.e.* have dropped out of the race) and to keep the relative conditional probability of all outcomes in $F$ the same as they were in the original (unconditional) sample space $S$.

*i.e.*

$$\frac{P\left(o_{f_i}|F\right)}{P\left(o_{f_j}|F\right)} = \frac{P\left(o_{f_i}\right)}{P\left(o_{f_j}\right)}.$$

From this principle

$$P\left(o_{f_i}|F\right) = \left(\frac{P\left(o_{f_j}|F\right)}{P\left(o_{f_j}\right)}\right) P\left(o_{f_i}\right) = k \cdot P\left(o_{f_i}\right).$$

The constant $k$ can be determined by the requirement that the sum of these rescaled probabilities equals $1$.

$$1 = \sum_{i=1}^{r} P\left(o_{f_i}|F\right) = \sum_{i=1}^{r} k \cdot P\left(o_{f_i}\right) = k \sum_{i=1}^{r} P\left(o_{f_i}\right) = kP(F).$$

Therefore,

$$k = \frac{1}{P(F)}$$

and

$$P\left(o_{f_i}|F\right) = k \cdot P\left(o_{f_i}\right) = \frac{P\left(o_{f_i}\right)}{P(F)}.$$

■

**Conditional Probability of an Event**

In alignment with how we previously defined (unconditional) probability we can now make the following definition.

> We define $P(A|F)$, the ***probability of the event $A$ conditional on the event $F$***, as the sum of the rescaled probabilities conditional on $F$ of the distinct outcomes making up event $A$.

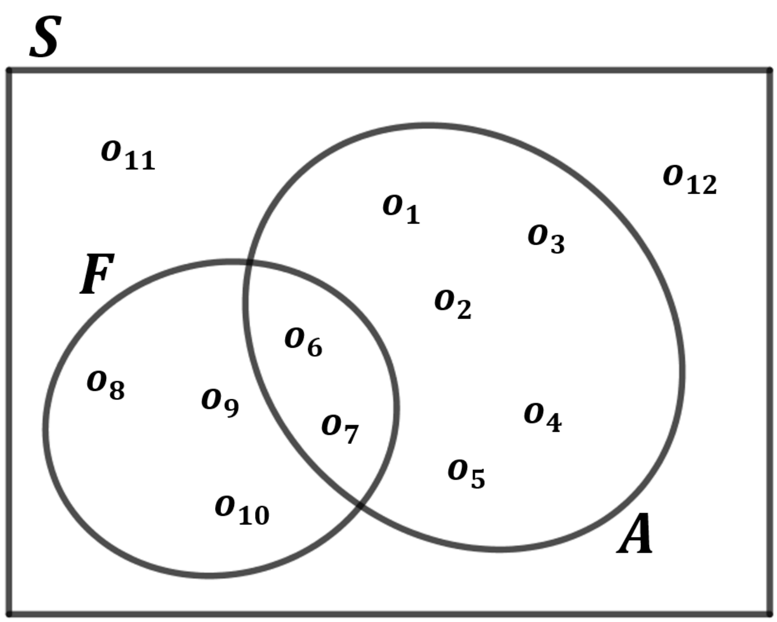
The following result is a consequence of this definition.

**Theorem 7.**

If $A$ and $F$ are events in $S$ then

$$P(A|F) = \frac{P(A \cap F)}{P(F)}.$$

Once again we can use a Venn diagram to illustrate the central idea behind this result.




The two core ideas of conditioning on the event $F$ are

(i) to remove all outcomes not in $F$ from $S$, the sample space of possible outcomes

(ii) to rescale the probabilities of all outcomes in $F$ to make $P(F|F) = 1$.

So, the first step in solving for $P(A|F)$ is set $P(o_j|F) = 0$ for all outcomes $o_j \notin F$. In particular

$$P(o_1|F) = P(o_2|F) = P(o_3|F) = P(o_4|F) = P(o_5|F) = P(o_{11}|F) = P(o_{12}|F) = 0.$$

The second step (as established by Theorem 6) is to set

$$P(o_j|F) = \frac{P(o_j)}{P(F)} \text{ for all outcomes } o_j \in F.$$

We note that in this example we have $A = \{o_1, o_2, o_3, o_4, o_5, o_6, o_7\}$. Hence, by the definition of the condition probability of an event, we have

$$\begin{aligned} P(A|F) &= P(o_1|F) + P(o_2|F) + P(o_3|F) + P(o_4|F) + P(o_5|F) \\ &\qquad + P(o_6|F) + P(o_7|F) \\ &= 0 + 0 + 0 + 0 + 0 + P(o_6|F) + P(o_7|F) \\ &= \frac{P(o_6)}{P(F)} + \frac{P(o_7)}{P(F)} = \frac{P(o_6) + P(o_7)}{P(F)}. \end{aligned}$$

But we also know that in this example $A \cap F = \{o_6, o_7\}$. Therefore,

$$P(A \cap F) = P(o_6) + P(o_7).$$

and

$$P(A|F) = \frac{P(o_6) + P(o_7)}{P(F)} = \frac{P(A \cap F)}{P(F)}.$$

■

**Conditional Probability Versions of Theorems 1 – 5 and DeMorgan's Laws**

Theorems 1 – 5 and DeMorgan's Laws remain valid for conditional probabilities. We can rewrite them as:

> **Theorem 1$'$.** If $C$ is any event in the sample space $S$ of an experiment, then $P(C|C) = 1$.

> **Theorem 2$'$.** If $A_1, A_2, \dots, A_k$ are mutually exclusive events of the sample space $S$ and if $C$ is any event in the sample space $S$, then
> $$P\big((A_1 \cup A_2 \cup \cdots \cup A_k)\big|C\big) = P(A_1|C) + P(A_2|C) + \cdots + P(A_k|C).$$

**Theorem 3′.** For any events $A$ and $C$ in the sample space $S$

$$P(A'|C) = 1 - P(A|C).$$

**Theorem 4′.** If $A, B$ and $C$ are any events in the sample space $S$, then

$$P((A \cup B)|C) = P(A|C) + P(B|C) - P((A \cap B)|C).$$

(**DeMorgan's Laws**)′

If $A_1, A_2, \dots, A_n$ and $C$ are events in the sample space $S$, then

$$P((A_1 \cup A_2 \cup \cdots \cup A_n)'|C) = P\left((A_1' \cap A_2' \cap \cdots \cap A_n')\middle|C\right)$$

and

$$P((A_1 \cap A_2 \cap \cdots \cap A_n)'|C) = P\left((A_1' \cup A_2' \cup \cdots \cup A_n')\middle|C\right).$$

**Theorem 5′.** Problems with Equally Likely Outcomes

If $A$ and $C$ are events in the sample space $S$ such that

(i) $S$ consists of $n$ equally likely outcomes
(ii) $C$ consists of $r$ of the $n$ outcomes in $S$
(iii) $A \cap C$ consists of $k$ of the $n$ outcomes in $S$

then $P(A|C) = k/r$.

**Chain Rule (or General Product Rule) for Probabilities**

**Theorem 8.**

If $A_1, A_2, A_3 \dots, A_n$ are events in $S$ then

$$P(A_1 \cap A_2 \cap \cdots \cap A_n)$$
$$= P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|(A_1 \cap A_2)) \cdot \ \cdots \ \cdot P(A_n|(A_1 \cap A_2 \cap \cdots \cap A_{n-1}))$$

The case of $n = 2$ follows immediately from Theorem 7. Let's step through the logic for the case of $n = 3$. The general proof follows the same pattern with induction.

$$P(A_1) \cdot P(A_2|A_1) \cdot P\big(A_3\big|(A_1 \cap A_2)\big)$$

$$= \Big(P(A_1) \cdot P(A_2|A_1)\Big) \cdot P\big(A_3\big|(A_1 \cap A_2)\big)$$

$$= P(A_1 \cap A_2) \cdot P\big(A_3\big|(A_1 \cap A_2)\big) \qquad \text{by Theorem 7}$$

$$= P(A_1 \cap A_2 \cap A_3) \qquad \text{by Theorem 7.}$$

■

Example

You might well have used Theorem 8 many times without explicit thinking about it as a separate result.

Suppose you draw at random and without replacement from an urn containing 3 white balls and 3 black balls. What is the probability that you get a white ball on the first draw, a white ball on the second draw and a black ball on the third draw?

**Solution**

Define the events $A$:white ball on the first draw, $B$:white ball on the second draw and $C$:black ball on the third draw. Applying Theorem 8,

$$P(A \cap B \cap C) = P(A)P(B|A)P\big(C\big|(A \cap B)\big).$$

There are initially 3 white and 3 black balls so $P(A) = 3/6$. After event $A$ there are 2 white and 3 black balls left in the urn so $P(B|A) = 2/5$. After events $A$ and $B$ there are 1 white and 3 black balls left in the urn so $P\big(C\big|(A \cap B)\big) = 3/4$.

So,

$$P(A \cap B \cap C) = P(A)P(B|A)P\big(C\big|(A \cap B)\big) = \left(\frac{3}{6}\right)\left(\frac{2}{5}\right)\left(\frac{3}{4}\right) = \frac{3}{20}.$$

■

**Independent Events**

We define events $A_1, A_2, \dots, A_k$ of sample space $S$ to be **mutually independent**$^\star$ if

$$P\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r}\right) = P\left(A_{i_1}\right)P\left(A_{i_2}\right) \cdots P\left(A_{i_r}\right)$$

for *all* subsets $\{i_1, i_2, \ldots, i_r\}$ of $\{1,2, \ldots, k\}$ with two or more elements.

$^{\star}$the term "mutually independent" is usually shortened to just "independent" in contest problems

**Example**

To verify that events $A, B$ and $C$ of sample space $S$ are mutually independent requires a separate$^{\star\star}$ proof of each of the $4$ equalities shown below.

<table>
<tr><td>$P(A \cap B) = P(A)P(B)$</td><td>$P(A \cap C) = P(A)P(C)$</td><td>$P(B \cap C) = P(B)P(C)$</td></tr>
<tr><td colspan="3">$P(A \cap B \cap C) = P(A)P(B)P(C)$.</td></tr>
</table>

To verify that events $A, B, C$ and $D$ of sample space $S$ are mutually independent requires a separate$^{\star\star}$ proof of each of the $11$ equalities shown below.

<table>
<tr><td colspan="2">$P(A \cap B) = P(A)P(B)$</td><td colspan="2">$P(A \cap C) = P(A)P(C)$</td><td colspan="2">$P(A \cap D) = P(A)P(D)$</td></tr>
<tr><td colspan="2">$P(B \cap C) = P(B)P(C)$</td><td colspan="2">$P(B \cap D) = P(B)P(D)$</td><td colspan="2">$P(C \cap D) = P(C)P(D)$</td></tr>
<tr><td colspan="3">$P(A \cap B \cap C) = P(A)P(B)P(C)$</td><td colspan="3">$P(A \cap B \cap D) = P(A)P(B)P(D)$</td></tr>
<tr><td colspan="3">$P(A \cap C \cap D) = P(A)P(C)P(D)$</td><td colspan="3">$P(B \cap C \cap D) = P(B)P(C)P(D)$</td></tr>
<tr><td colspan="6">$P(A \cap B \cap C \cap D) = P(A)P(B)P(C)P(D)$.</td></tr>
</table>

$^{\star\star}$No one of these equalities will imply all the others. In particular,

$$P(A \cap B \cap C \cap D) = P(A)P(B)P(C)P(D)$$

does <u>not</u> imply all the other equalities.

The following two theorems are direct consequences of the definition of (mutually) independent events but are useful enough to be highlighted on its own.

**Theorem 9.**

If events $A$ and $F$ of sample space $S$ are mutually independent then

$$P(A|F) = P(A)$$

and

$$P(F|A) = P(F).$$

This result follows immediately from Theorem 7 and the definition of independence.

$$P(A|F) = \frac{P(A \cap F)}{P(F)} = \frac{P(A)P(F)}{P(F)} = P(A)$$

and

$$P(F|A) = \frac{P(F \cap A)}{P(A)} = \frac{P(F)P(A)}{P(A)} = P(F).$$

■

**Theorem 10.**

If the events $A_1, A_2, \dots, A_k$ of sample space $S$ are mutually independent then set functions of pairwise disjoint subsets of the events $A_1, A_2, \dots, A_k$ are mutually independent.

An example of this idea will help to clarify. Suppose $A, B, C, D$ and $E$ are mutually independent events of a sample space $S$. Then by Theorem 10 the three set functions

(i) $(A \cup B)'$
(ii) $C$
(iii) $(D' \cap E)$

are mutually independent because no two of these set functions involve the same events from $A, B, C, D$ and $E$.

**Corollary of Theorem 10**

If events $A$ and $B$ of sample space $S$ are mutually independent then

$\overline{A}$ and $B$ are mutually independent events,

$A$ and $\overline{B}$ are mutually independent events,

$\overline{A}$ and $\overline{B}$ are mutually independent events.

Combining Theorems 9 and 10 we can get a sort of corollary to both which we will simply illustrate through examples.

If events $A, B, C, D$ and $E$ are mutually independent then

$$P\big((A \cup B)'\big|C\big) = P\big((A \cup B)'\big)$$

$$P\big(C\big|(A \cup B)'\big) = P(C)$$

$$P\big((A \cup B)'\big|(D' \cap E)\big) = P\big((A \cup B)'\big)$$

$$P\left((D' \cap E)\Big|(A \cup B)'\right) = P\left((D' \cap E)\right).$$

**Independence and Intuition**

Independence is a term used in probability and in everyday conversation. You might agree that whether it rains tonight is in no way influenced by (*i.e.* is independent of) whether you left your car windows open. But that intuitive understanding of how events influence each other is *for the most part* not sufficient to determine whether events meet the probability definition for independence.

There are a few notable exceptions. In a contest setting, you <u>are</u> expected to assume that the results of successive tosses of a coin are independent events ***whether it is stated or not***. The same goes for the successive rolls of a die and the outcome when drawing balls from an urn if the sampling is done <u>with replacement</u>.

However, apart from such notable exceptions, you should work under the assumption that events are not necessarily independent unless it is explicitly stated in the problem.

Here are a few examples where independence is <u>*implicitly*</u> assumed in some MSHSML problems:

(5C141) … probability that a fair coin will land tails three times in a row

(5D102) … $a$ and $b$ chosen with replacement from $\{1,2,\ldots,8,9\}$

(5C182) … $a$ and $b$ are the results from rolling an $8$-sided die twice

(5C091) … $a$, $b$ and $c$ are the results from rolling a $6$-sided die three times

(5C122) … Kathy flips a fair coin until she get three heads in a row

And here are a few examples where independence is ***explicitly*** stated in some MSHSML problems:

(5C164) … Assuming that Ashley, Ben, and Charlie chose their ice cream flavors independently, what is the probability no flavor is chosen more than once

(5D104) … a frog choses the direction of its next jump independently from previous jumps

(5T104) … a candy cane is broken at two points chosen independently and at random.

## Exercises for Chapter 1

1. (5C183) The probability of event $A$, written as $P(A)$, equals 0.4, and $P(A \text{ and } B) = 0.172$. If $A$ and $B$ are independent events, determine exactly $P(A \text{ or } B)$.
2. If $P(S) = p > 0$, $P(T) = q$ and if $S$ and $T$ are independent events, find $P\big((S \cap T)\big|S\big)$.
3. If $P(S) = p$, $P(T) = q$ and if $0 < p < 1$ and $0 < q < 1$, find $P\big((S \cup T)\big|T\big)$.
4. Two events $A$ and $B$ are such that $P(A) = 0.2$, $P(B) = 0.3$, and $P(A \cup B) = 0.4$. Find the following:

   (a) $P(A \cap B)$

   (b) $P(\overline{A} \cup \overline{B})$

   (c) $P(\overline{A} \cap \overline{B})$

   (d) $P(\overline{A}|B)$.
5. Find $P\big(A \cup (B' \cup C')'\big)$ if $P(A) = 1/2$, $P(B \cap C) = 1/3$ and $P(A \cap C) = 0$.
6. Find $P\left(\left(A \cap \overline{B}\right) \cup \left(\overline{A} \cap B\right)\right)$ if $A$ and $B$ are independent events and if $P(A) = 2/3$ and if $P(B) = 1/4$.
7. Suppose $P(A) = 0.5$ and $P(A \cup B) = 0.6$.

   (a) Find $P(B)$ if $A$ and $B$ are mutually exclusive

   (b) Find $P(B)$ if $A$ and $B$ are independent

   (c) Find $P(B)$ if $P(A|B) = 0.4$.
8. If $P(C) > 0$, $P(A|C) = 0.7$, $P(B|C) = 0.4$ and $P\big((A \cup B)\big|C\big) = 0.8$, find $P\big((A \cap B)\big|C\big)$.
9. If $P(\overline{A}) = 0.3$, $P(B) = 0.4$, and $P(A \cap \overline{B}) = 0.5$, then find $P(A \cap B)$ and $P(A \cup B)$.
10. Find $P\big(A \cup (B^c \cup C^c)^c\big)$ if $A, B$ and $C$ are mutually exclusive events and $P(A) = 3/7$.
11. Find $P\left(A \cup \left(B^c \cup C^c\right)^c\right)$ if $P\big(A^c \cap (B^c \cup C^c)\big) = 0.65$.
12. Suppose that $A$ and $B$ are mutually exclusive events for which $P(A) = 0.3$ and $P(B) = 0.5$ What is the probability that

    (a) either $A$ or $B$ occurs?

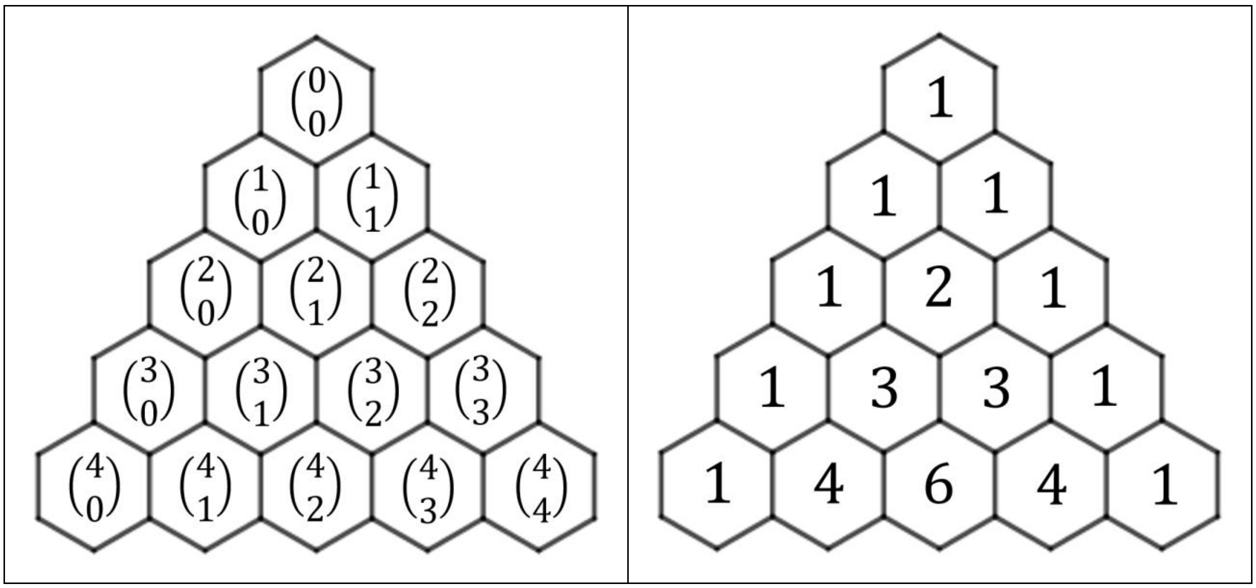
    (b) $A$ occurs but $B$ does not?

## Project 1  Even and Odd Numbers in Pascal's Triangle

Pascal's triangle is defined by taking the $k$ entry in the $n^{th}$ row as the number $\binom{n}{k}$ where

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

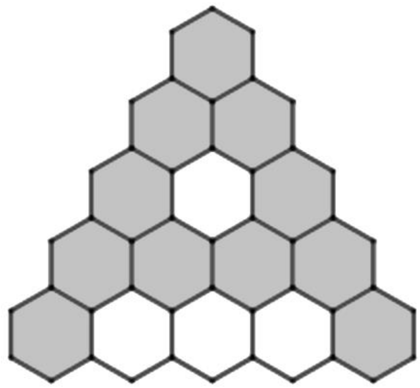
with the definition that $0! = 1$.  Here we show the first five rows of Pascal's triangle and their values.



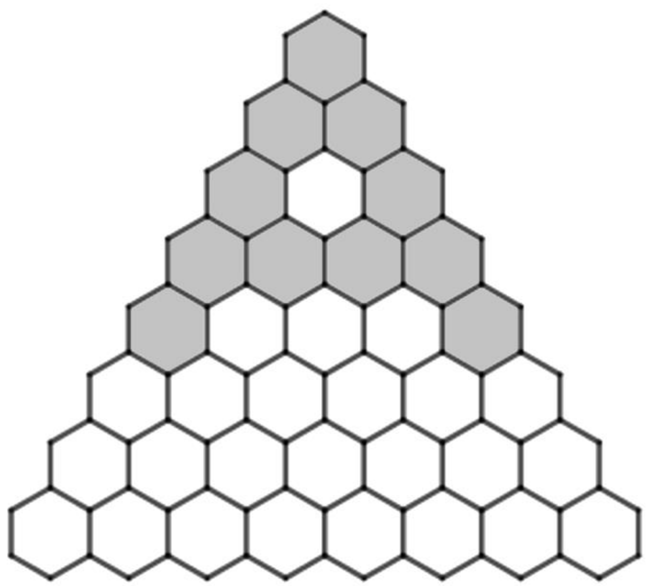

> Project 1:  If you pick a number at random from all values in the first $128$ rows of Pascal's triangle, what is the probability that it will be an odd number?  For example, if you pick a number at random from the first $5$ rows, the probability that it will be an odd number is $11/15$.

If you extend the diagram above to as many rows as you have the patience for, certain patterns start to arise.  It is easier to see the patterns if you don't worry about the actual numbers in each position and just focus on whether the numbers are odd or even.

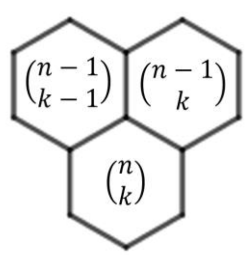
If you shade the hexagons where you have odd numbers (leave the hexagon empty if the number is even), the first five rows would look like



Can you see how to shade row six, seven and eight ***without*** actually computing the binomial coefficients $\binom{n}{k}$, $n = 6{,}7{,}8$ and $k = 0{,}1, \ldots, n$ ?

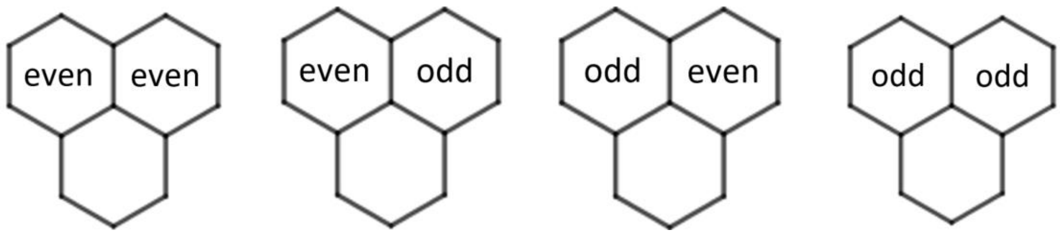


Can you prove the identity $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ ?

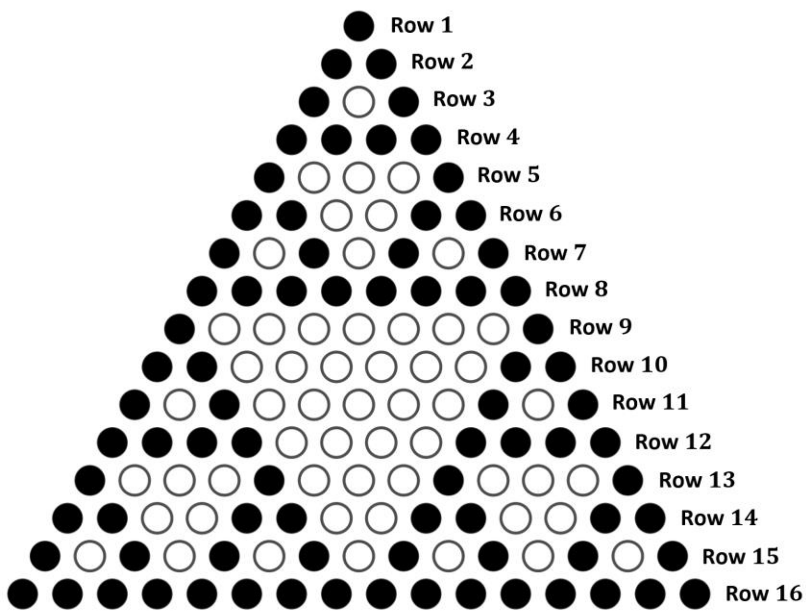



This is the identity that transforms the construction of Pascal's triangle into a series of simple additions.

But how does this identity help in identifying which hexagons will hold even numbers and which will hold odd numbers? Consider all possibilities.

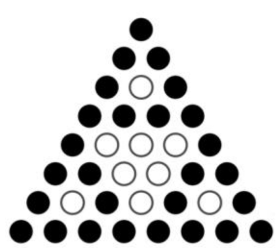



Here are the first sixteen rows of Pascal's triangle with the odd numbers shown as solid black circles.




What patterns can you see? Especially look for patterns in the layout of black dots occurring in the first $2^3$ rows, and then the first $2^4$ rows.

Notice that there are 3 triangles of the form

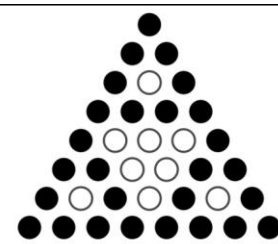


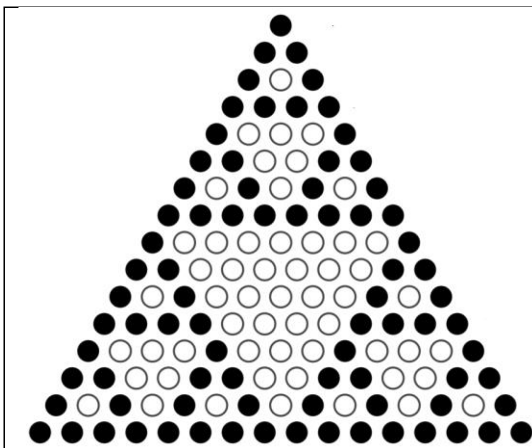
make up Rows 1 to 16 and there are 3 triangles of the form

making up the larger triangle and there are 3 triangles of the form

making up the middle size triangle and there are $3$ solid black dots making up this smallest triangle.

So, the apparent pattern is that there are

|  |  |  |
|---|---|---|
| $3 = 3^1$<br>odd numbers from Row 1 to Row $2^1$ | $3 \times 3 = 3^2$<br>odd numbers from Row 1 to Row $2^2$ | $3 \times 3 \times 3 = 3^3$<br>odd numbers from Row 1 to Row $2^3$ |



$3 \times 3 \times 3 \times 3 = 3^4$
odd numbers from Row 1 to Row $2^4$

The obvious conjecture based on this pattern is that there are $3^7$ odd numbers from Row 1 to Row $2^7 = 128$.

Notice that Row $k$ consists of $k$ dots in total. Hence there are a total of

$$1+2+3+\cdots+127+128=\frac{(128)(129)}{2}=\frac{2^7(2^7+1)}{2}$$

dots in Rows 1 to $128$. So (*assuming this pattern holds and the conjecture is true*) if you pick one of these numbers are random the probability you will get an odd number is the count ratio

$$\frac{3^7}{\frac{2^7(2^7+1)}{2}}=\frac{3^7}{2^6(2^7+1)}\approx 0.265.$$

But how can we prove this? Even though this *really is correct*, at this point it is just a conjecture based on the pattern we hope (but cannot say for sure) will continue beyond the first $16$ rows of Pascal's triangle shown in the diagram above.

To put together an algebraic proof of this conjecture we will start by investigating Pascal's triangle a **row at a time**. To get a concrete sense of the argument we will start by looking at the $11^{th}$ row (not too big, not too small). Then we will generalize the argument to the $n^{th}$ row.

The $11^{th}$ row of Pascal's triangle is made up of the $11$ numbers $\binom{10}{0},\binom{10}{1},\dots,\binom{10}{10}$. We know by the binomial theorem that

$$(1+x)^{10}=\sum_{j=0}^{10}\binom{10}{j}x^j.$$

Suppose, for reasons that will manifest themselves as we get further along in the proof, we express the number 10 in terms of its base 2 expansion. That is,

$$10=1\cdot 2^3+0\cdot 2^2+1\cdot 2^1+0\cdot 2^0=1010_2.$$

Then we can write

$$(1+x)^{10}=(1+x)^{(2^3+2^1)}=(1+x)^{(2^3)}\cdot(1+x)^{(2^1)}.$$

However,

$$\boxed{(1+x)^{(2^n)}\equiv 1+x^{(2^n)}\pmod 2\ \text{ for } n\geq 1.}$$

That is, the coefficient of $x^j$ in $(1+x)^{(2^n)}$ and the coefficient of $x^j$ in $1+x^{(2^n)}$ are equal modulo 2 for all $j=0,1,2,\dots,2^n$.

The gist of the supporting argument for this claim is that (by induction)

$$(1+x)^{(2^n)}=(1+x)^{(2^{n-1})}(1+x)^{(2^{n-1})}$$

$$\equiv \left(1 + x^{(2^{n-1})}\right)\left(1 + x^{(2^{n-1})}\right) \pmod 2$$
$$= \left(1 + 2x^{(2^{n-1})} + x^{(2^n)}\right) \pmod 2$$
$$\equiv \left(1 + x^{(2^n)}\right) \pmod 2.$$

So

$$(1+x)^{10} = (1+x)^{(2^3)} \cdot (1+x)^{(2^1)} \equiv \left(1 + x^{(2^3)}\right)\left(1 + x^{(2^1)}\right) \pmod 2.$$

That is, the coefficients of $x^j$ on the left and right-hand side of this congruency are equal modulo 2. Simplifying, we have

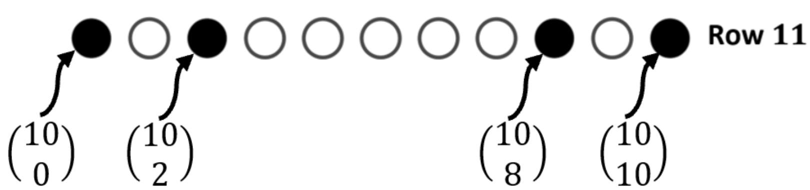
$$(1+x)^{10} = \sum_{j=0}^{10} \binom{10}{j} x^j \equiv 1 + x^{(2^1)} + x^{(2^3)} + x^{(2^1+2^3)} \pmod 2$$

or

$$\binom{10}{0} x^0 + \binom{10}{1} x^1 + \cdots + \binom{10}{10} x^{10} \equiv x^0 + x^2 + x^8 + x^{10} \pmod 2.$$

It follows from this polynomial congruency that $\binom{10}{0}, \binom{10}{2}, \binom{10}{8}, \binom{10}{10}$ are congruent to 1 modulo 2 (*i.e.* odd numbers) and $\binom{10}{1}, \binom{10}{3}, \binom{10}{4}, \binom{10}{5}, \binom{10}{6}, \binom{10}{7}, \binom{10}{9}$ are congruent to 0 modulo 2 (*i.e.* even numbers).

We can see that this is correct if we look back at Row 11 of Pascal's triangle where the solid black dots represent odd numbers.




| |
|---|
| This shows that Row 11 has 4 odd numbers. |

Can this argument be extended to handle counting the number of odd numbers (solid black dots) in any row? Yes.

**Row $n+1$.**

To count the number of odd numbers in Row $n+1$ we would start by finding the base 2 expansion of $n$.

Suppose the base 2 expansion of $n$ is given by $n = 2^{r_1} + 2^{r_2} + \cdots + 2^{r_k}$ where $r_1 > r_2 > \cdots > r_k \geq 0$. Then we have

$$\sum_{j=0}^{n} \binom{n}{j} x^j = (1+x)^n = (1+x)^{2^{r_1}} \cdot (1+x)^{2^{r_2}} \cdots (1+x)^{2^{r_k}}$$

$$\equiv \left(1 + x^{(2^{r_1})}\right)\left(1 + x^{(2^{r_2})}\right) \cdots \left(1 + x^{(2^{r_k})}\right) \pmod 2$$

after again using the identity

$$(1+x)^{(2^n)} \equiv 1 + x^{(2^n)} \pmod 2 \text{ for } n \geq 1.$$

The next step is to verify that when you multiply out

$$\left(1 + x^{(2^{r_1})}\right)\left(1 + x^{(2^{r_2})}\right) \cdots \left(1 + x^{(2^{r_k})}\right)$$

that you get a polynomial with $2 \times 2 \times \cdots \times 2 = 2^k$ distinct terms each with coefficient 1 in the same way as our earlier example where we saw that

$$\left(1 + x^{(2^3)}\right)\left(1 + x^{(2^1)}\right) = x^0 + x^2 + x^8 + x^{10}.$$

It turns that all we need to guarantee that $\left(1 + x^{(2^{r_1})}\right)\left(1 + x^{(2^{r_2})}\right) \cdots \left(1 + x^{(2^{r_k})}\right)$ where $r_1 > r_2 > \cdots > r_k \geq 0$ always simplifies to a polynomial with $2 \times 2 \times \cdots \times 2 = 2^k$ distinct terms each with coefficient 1 is the fact that in every base system the **representation of a number is *unique***.

Together with the polynomial congruency

$$\sum_{j=0}^{n} \binom{n}{j} x^j \equiv \left(1 + x^{(2^{r_1})}\right)\left(1 + x^{(2^{r_2})}\right) \cdots \left(1 + x^{(2^{r_k})}\right) \pmod 2$$

and the fact that $\left(1 + x^{(2^{r_1})}\right)\left(1 + x^{(2^{r_2})}\right) \cdots \left(1 + x^{(2^{r_k})}\right)$ simplifies to a polynomial with $2^k$ terms each with coefficient 1 tells us that exactly $2^k$ of the binomial coefficients from $\binom{n}{0}$, $\binom{n}{1}$, $\binom{n}{2}$, …, $\binom{n}{n}$ will equal 1 modulo 2 (*i.e.* be odd numbers) and the remaining $n - 2^k$ of these binomial coefficients will equal 0 modulo 2 (*i.e.* be even numbers).

Notice that the $k$ in this argument is the number of terms in the base 2 expansion of $n-1$. Another way of saying this is that $k$ equals the number of $1$'s when $n-1$ is written in base 2.

So, how many odd numbers are there in Row 42 (just to pull a number out of the air) of Pascal's triangle?

$$42-1=41=32+8+1=2^5+2^3+2^0=101001_2.$$

Now notice that $41=101001_2$ has 3 $1$'s. (Alternatively, we can say that that the base 2 expansion of $41=2^5+2^3+2^0$ has 3 terms.) Either way, we can conclude that Row 42 has $2^3=8$ odd numbers. More precisely, $8$ of the $42$ binomial coefficients $\binom{41}{0}, \binom{41}{1}, \binom{41}{2}, \binom{41}{3}, \ldots, \binom{41}{41}$ which make up Row 42 of Pascal's triangle are odd numbers.

How many odd numbers are there in Row 64 of Pascal's triangle?

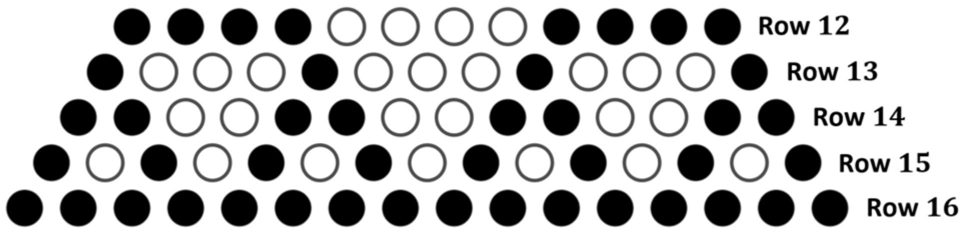
$$64-1=63=32+16+8+4+2+1=111111_2.$$

We see that the base 2 representation of $64-1=63$ consists of six $1$'s. Hence, Row 64 of Pascal's triangle contains $2^6$ odd numbers. More preciously, $2^6=64$ of the $64$ binomial coefficients $\binom{63}{0}, \binom{63}{2}, \binom{63}{3}, \ldots, \binom{63}{62}, \binom{63}{63}$ which make up Row 64 of Pascal's triangle are odd numbers.

The general conclusion we have arrived at is …

If $n-1$ written in base 2 contains $k$ $1$'s, then Row $n$ of Pascal's triangle will contain $2^k$ odd numbers.

For clarification with all this new notation, let's check this result against what we know about Row 12 through Row 16 in Pascal's triangle.




| | | | |
|---|---|---|---|
| Row 12 | $12-1=8+2+1$ $=1011_2$ | $3$ 1's in $1011_2$ | $2^3=8$ odd numbers in Row 12 |
| Row 13 | $13-1=8+4$ $=1100_2$ | $2$ 1's in $1100_2$ | $2^2=4$ odd numbers in Row 13 |
| Row 14 | $14-1=8+4+1$ $=1101_2$ | $3$ 1's in $1101_2$ | $2^3=8$ odd numbers in Row 14 |
| Row 15 | $15-1=8+4+2$ $=1110_2$ | $3$ 1's in $1110_2$ | $2^3=8$ odd numbers in Row 15 |
| Row 16 | $16-1=8+4+2+1$ $=1111_2$ | $4$ 1's in $1111_2$ | $2^4=16$ odd numbers in Row 16 |

Checking against the diagram above showing Rows 12 through 16 we can see that the results are all correct.

**Counting the Number of Odd Numbers in Row 1 through Row 128.**

Constructing the list of the base 2 form of the $2^7=128$ numbers from $0$ to $127$ is equivalent to filling the seven boxes

$$\underline{\quad}\ \underline{\quad}\ \underline{\quad}\ \underline{\quad}\ \underline{\quad}\ \underline{\quad}\ \underline{\quad}$$

in all possible ways where each box has to filled with a $0$ or a $1$. Remember that

$$0=\left(\underline{0}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}\right)_2 \quad \text{and} \quad 127=\left(\underline{1}\ \underline{1}\ \underline{1}\ \underline{1}\ \underline{1}\ \underline{1}\ \underline{1}\right)_2.$$

By the combinatorial formula for combinations, there are $\binom{7}{k}$ different ways to fill these boxes with $k$ $1$'s (and $7-k$ $0$'s), $k=0,1,2,\dots,7$.

Thus, $\binom{7}{k}$ of the rows from $1$ to $128$ will contain $2^k$ odd numbers. Hence the total number of odd numbers in Row $1$ to Row $128$ of Pascal's triangle equals

$$\binom{7}{0}2^0+\binom{7}{1}2^1+\binom{7}{2}2^2+\binom{7}{3}2^3+\binom{7}{4}2^4+\binom{7}{5}2^5+\binom{7}{6}2^6+\binom{7}{7}2^7.$$

But, by the binomial theorem we also know that

$$(1+x)^7 = \sum_{j=0}^{7} \binom{7}{j} x^j.$$

Thus, the total number of odd numbers in Row 1 to Row 128 of Pascal's triangle equals simplifies to $(1+2)^7 = 3^7$.

So, our conjecture that the probability of getting an odd number if you pick a number at random from Row 1 to Row 128 of Pascal's triangle equals

$$\frac{3^7}{\frac{2^7(2^7+1)}{2}} = \frac{3^7}{2^6(2^7+1)} \approx 0.265$$

is now formally verified.

■

**Probability that a Binomial Coefficient is Not Divisible by the Prime $p$**

The above result can be extended to find the probability that a randomly selected binomial coefficient is not divisible by the prime number $p$.

> The probability of selecting a number which is not divisible by the prime number $p$ if you randomly pick a binomial coefficient from Rows 1 to $p^m$ of Pascal's triangle (inclusive) equals
>
> $$\frac{(p+1)^m}{2^{m-1}(p^m+1)}$$
>
> for all positive integers $m$.

We will list the important steps that are needed in the proof of this result but will leave the details (which are somewhat more involved than for the special case $p = 2$ given above) to the solution manual. Note that the solution derived in the previous section where $p = 2$ and $m = 7$ is consistent with the above general result.

**Step 1.**

$$(1+x)^{(p^n)} \equiv 1 + x^{(p^n)} \pmod{p} \text{ for } p \text{ prime and integer } n \geq 1.$$

**Step 2. (Lucas's Theorem)**

Let $p$ be a prime number and let

$$n = n_0p^0 + n_1p^1 + n_2p^2 + \cdots + n_kp^k \qquad (0 \le n_i < p,\ \ i = 0,1,\dots,k)$$

$$r = r_0p^0 + r_1p^1 + r_2p^2 + \cdots + r_kp^k \qquad (0 \le r_i < p,\ \ i = 0,1,\dots,k)$$

then

$$\binom{n}{r} \equiv \binom{n_0}{r_0}\binom{n_1}{r_1}\cdots\binom{n_k}{r_k} \quad \text{mod}(p)$$

with the understanding that $\binom{n_j}{r_j} = 0$ for any $r_j > n_j$.

**Step 3.**

Let $p$ be a prime number and let

$$n = n_0p^0 + n_1p^1 + n_2p^2 + \cdots + n_kp^k \qquad (0 \le n_i < p,\ \ i = 0,1,\dots,k).$$

Then $(n_0 + 1)(n_1 + 1)\cdots(n_k + 1)$ of the $n + 1$ binomial coefficients $\binom{n}{0}, \binom{n}{1}, \dots, \binom{n}{n}$ are <u>not</u> divisible by $p$.

**Step 4.**

There are

$$\left(\frac{p(p+1)}{2}\right)^m$$

binomial coefficients in Rows 1 to $p^m$ of Pascal's triangle that are not divisible by the prime number $p$.

**Step 5.**

There are

$$\frac{p^m(p^m + 1)}{2}$$

binomial coefficients in Rows 1 to $p^m$ of Pascal's triangle.

**Step 6.**

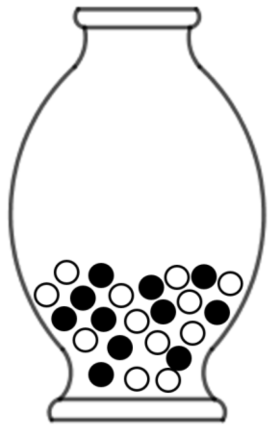
The probability that a binomial coefficient picked at random from Rows 1 to $p^m$ of Pascal's triangle taken will <u>not</u> be divisible by $p$ equals the count ratio

$$\frac{\text{\# of binomial coefficients in Rows 1 to } p^m \text{ not divisible by } p}{\text{total \# of binomial coefficients in Rows 1 to } p^m} = \frac{(p+1)^m}{2^{m-1}(p^m + 1)}.$$

■

## Chapter 2 Binomial Random Variables

I. Suppose an urn contains $w$ white balls and $b$ black balls. You reach into this urn and randomly select a ball, note its color and ***then return the ball to the urn***. You repeat this process for a total of $n$ draws. Let $X$ equal the number of times you select a white ball and let $P(X = x)$ represent the probability that you get exactly $x$ white balls in these $n$ draws.



Then

$$P(X = x) = \begin{cases} \binom{n}{x}\left(\frac{w}{w+b}\right)^x\left(\frac{b}{w+b}\right)^{n-x} & x \in \{0,1,\dots,n\} \\ 0 & x \notin \{0,1,\dots,n\}. \end{cases}$$

II. Suppose you roll a fair die a total of $n$ times. Let $X$ equal the number of times you roll either a $3$ or $6$. Then

$$P(X = x) = \begin{cases} \binom{n}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{n-x} & x \in \{0,1,\dots,n\} \\ 0 & x \notin \{0,1,\dots,n\}. \end{cases}$$

III. You have a penny that has been altered in such a way that it lands heads up with probability $5/8$ when flipped. Suppose you flip this penny $n$ times and let $X$ equal the number of times it lands heads in these $n$ flips. Then

$$P(X = x) = \begin{cases} \binom{n}{x}\left(\frac{5}{8}\right)^x\left(\frac{3}{8}\right)^{n-x} & x \in \{0,1,\dots,n\} \\ 0 & x \notin \{0,1,\dots,n\}. \end{cases}$$

Now consider inventing a generic experiment that incorporates the common aspects of these three problems above.

Suppose our generic experiment consists of $n$ repeated trials where these trials meet the following conditions:

(1) the trials are independent, *i.e*. the outcome in one trial will have no impact on the outcome of a different trial

(2) there are only two possible outcomes for a trial. We will refer to these two possible outcomes as "success" or "failure"

(3) $P(\text{success})$ is the same for every trial. We will use the letter $p$ to represent $P(\text{success})$ for notational simplicity

(4) the total number of trials performed is fixed (not random)

(5) our interest is in the total number of successes that occur in these fixed number of trials

Trials that meet these conditions are called **Bernoulli trials** in honor of Jacob Bernoulli (1654-1705) who wrote about such trials in his seminal paper *Ara Conjectandi* (The Art of Conjecturing).

Let $X =$ number of successes that occur in $n$ Bernoulli trials. Then

$$P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x \in \{0,1,\dots,n\} \\ 0 & x \notin \{0,1,\dots,n\} \end{cases}$$

where $p = P(\text{success})$.

In the language of probability and statistics, whenever $P(X = x)$ is determined by the above formula we say that $X$ is a **binomial random variable** or that $X$ follows the **binomial distribution**.

To understand where this result comes from, let's examine the case where $X = 2$ heads in $n = 3$ flips of a penny with $p = P(\text{heads}) = 5/8$. Let $\underline{H}$ represent heads and $\underline{T}$ represent tails. Then

$$P(X = 2 \text{ heads}) = P\left(\left(\underline{H} \text{ and } \underline{H} \text{ and } \underline{T}\right) \text{ or } \left(\underline{H} \text{ and } \underline{T} \text{ and } \underline{H}\right) \text{ or } \left(\underline{T} \text{ and } \underline{H} \text{ and } \underline{H}\right)\right)$$

$$
\begin{aligned}
&= P(\underline{H} \text{ and } \underline{H} \text{ and } \underline{T}) + P(\underline{H} \text{ and } \underline{H} \text{ and } \underline{T}) + P(\underline{H} \text{ and } \underline{H} \text{ and } \underline{T}) \\
&= \left(P(\underline{H}) \cdot P(\underline{H}) \cdot P(\underline{T})\right) + \left(P(\underline{H}) \cdot P(\underline{T}) \cdot P(\underline{H})\right) + \left(P(\underline{T}) \cdot P(\underline{H}) \cdot P(\underline{H})\right) \\
&= \left(\frac{5}{8}\right)^2 \left(\frac{3}{8}\right) + \left(\frac{5}{8}\right)^2 \left(\frac{3}{8}\right) + \left(\frac{5}{8}\right)^2 \left(\frac{3}{8}\right) \\
&= 3 \cdot \left(\frac{5}{8}\right)^2 \left(\frac{3}{8}\right).
\end{aligned}
$$

As a reminder, $P(A \text{ or } B) = P(A) + P(B)$ when events $A$ and $B$ are mutually exclusive (*i.e.* they cannot both be true at the same time) and $P(A \text{ and } B) = P(A) \cdot P(B)$ when events $A$ and $B$ are independent (*i.e.* one event occurring or not does not influence whether the other event occurs or not).

It is straightforward to extend this to the general case where we want to find the probability of $x$ heads in $n$ flips with $p = P(\text{heads})$.

Look back at each of the parts in the above answer

$$P(X = 2) = 3 \cdot \left(\frac{5}{8}\right)^2 \left(\frac{3}{8}\right).$$

The leading coefficient 3 was just the number of possible ways to arrange two $\underline{H}$'s and one $\underline{T}$. In general, there are $\binom{n}{x}$ possible ways to arrange $x$ $\underline{H}$'s and $(n - x)$ $\underline{T}$'s.

The factor $p = P(\text{heads}) = 5/8$ was raised to the second power and the factor $(1 - p) = P(\text{tails}) = 3/8$ was raised to the first power because there two $\underline{H}$'s and one $\underline{T}$. In the general case with $x$ $\underline{H}$'s and $(n - x)$ $\underline{T}$'s these factors become $p^x(1 - p)^{n-x}$.

This establishes the general result. If we let $X$ equal the number of successes that occur in $n$ Bernoulli trials then

$$P(X = x) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x} & x \in \{0,1, \dots, n\} \\ 0 & x \notin \{0,1, \dots, n\} \end{cases}$$

where $p = P(\text{success})$.

**Negative Binomial Model**

The starting point for the *negative* binomial model is nearly identical to our starting point the binomial model.

We consider a generic experiment that consists of repeated trials where these trials meet the following conditions:

(1) the trials are independent, *i.e*. the outcome in one trial will have no impact on the outcome of a different trial

(2) there are only two possible outcomes for a trial. We will refer to these two possible outcomes as "success" or "failure"

(3) $P(\text{success})$ is the same for every trial. We will use the letter $p$ to represent $P(\text{success})$ for notational simplicity

These are the same first three assumptions about the repeated trials in the binomial model. It is in the fourth and fifth assumptions where the difference between the binomial and the negative binomial shows up.

In the binomial model we took

(4) the total number of trials performed is $n$ ($n$ is fixed, not random)

But in the negative binomial model the

(4) trials continue until we get observe the $r^{th}$ success ($r$ is fixed, not random)

In the binomial model

(5) our interest is in the total number of successes that occur in these $n$ trials

But in the negative binomial model

(5) our interest is in the total number of trials that required to observe $r$ successes

The following chart makes clear the essential distinction between the "Binomial" and "Negative Binomial" models.

| | **Total Number of Trials Performed** | **Total Number of Successes Observed** |
|---|---|---|
| **Binomial** | Fixed (predetermined) | Random |
| **Negative Binomial** | Random | Fixed (predetermined) |

**Theorem**

Let $X$ equal the number of (Bernoulli) trials required in order to get a total of $r$ successes. Then

$$P(X=k)=\begin{cases}\binom{k-1}{r-1}p^r(1-p)^{k-r} & k\in\{r,r+1,r+2,\dots\}\\ 0 & k\notin\{r,r+1,r+2,\dots\}.\end{cases}$$

The above formula is called the ***negative binomial distribution***. Note that the range of $k$ (the possible values of $k$) start at $r$ because it certainly takes at least $r$ trials to reach the $r^{th}$ success. But there is no upper limit on the number of trials it might take you to reach the $r$ success. It is unlikely, but you might be there forever waiting for the $r^{th}$ success to occur.

To justify this formula we start by defining the events:

$A$: get the $r^{th}$ success on the $k^{th}$ Bernoulli trial

$B$: get exactly $(r-1)$ successes in the first $(k-1)$ Bernoulli trials

$C$: get a success on the $k^{th}$ Bernoulli trial.

Then

$$P(\text{get the } r^{th} \text{ success on the } k^{th} \text{ Bernoulli trial})=P(A).$$

Remember that the trials stop right after observing the $r^{th}$ success. This means that the **last trial had to be a success** (in fact, the $r^{th}$ success). Hence saying that we got the $r^{th}$ success on the $k^{th}$ trial is equivalent to saying that we got $r-1$ successes in the first $k-1$ trials and then got a success on the $k^{th}$ trial.

That is,

$$P(\text{get the } r^{th} \text{ success on the } k^{th} \text{ Bernoulli trial})$$

$$=P(A)=P(B\cap C).$$

But from Chapter 1,

$$P(B\cap C)=P(C|B)P(B).$$

Now look again at the definition for the event $B$: getting exactly $(r-1)$ successes in the first $(k-1)$ Bernoulli trials. We can write this as

$$P(B) = P(X = r - 1 \text{ successes in } n = k - 1 \text{ trials})$$

to help bring home the point. We are asking for the probability of a given number of successes $(r-1)$ in a given number of trials $(k-1)$. But that is what the Binomial probability gives us. That is, $B$ follows the *binomial* distribution. So,

$$\begin{aligned} P(B) &= \binom{k-1}{r-1} p^{r-1}(1-p)^{(k-1)-(r-1)} \\ &= \binom{k-1}{r-1} p^{r-1}(1-p)^{k-r}. \end{aligned}$$

Now look at the factor $P(C|B)$, which is asking for the probability of getting a success on the $k^{th}$ trial conditional on the event that we got exactly $(r-1)$ successes in the first $(k-1)$ trials. But remember that all trials are independent. The probability of getting a success on ***any*** trial is $p$. That is, $P(C|B) = p$.

So, we have established that

$$\begin{aligned} P(X = k) &= P(\text{get the } r^{th} \text{ success on the } k^{th} \text{ Bernoulli trial}) \\ &= P(C|B) \cdot P(B) \\ &= p \cdot \binom{k-1}{r-1} p^{r-1}(1-p)^{k-r} \\ &= \binom{k-1}{r-1} p^{r}(1-p)^{k-r}. \end{aligned}$$

■

**Multinomial Probability Distribution**

The ***multinomial distribution*** is an extension of the binomial distribution to the case where there are ***more than two possible outcomes per trial***.

Suppose an experiment consists of repeated trials where these trials meet the following conditions:

(1) the trials are independent, *i.e.* the outcome in one trial will have no impact on the outcome of a different trial

(2) there are $k$ possible outcomes on any trial

(3) $p_j = P(\text{type } j \text{ outcome})$ is the same for every trial, for each $j = 1,2,\ldots,k$

Let $X_j$ equal the number of Type $j$ outcomes that occur in $n$ repeated trials. Then

$$P(X_1 = x_1, \ldots, X_k = x_k) = \begin{cases} \dfrac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} & \begin{array}{c} x_j \in \{0,1,2,\ldots,n\}, j = 1,\ldots,k \\ x_1 + x_2 + \cdots + x_k = n \end{array} \\ 0 & \text{else.} \end{cases}$$

## Exercises for Chapter 2

Exercises 1 to 24 were all taken from old MSHSML tests. Before trying to solve them, read them several times. As you are reading them, consider how each of them depends in various ways on a binomial, negative binomial or multinomial random variable.

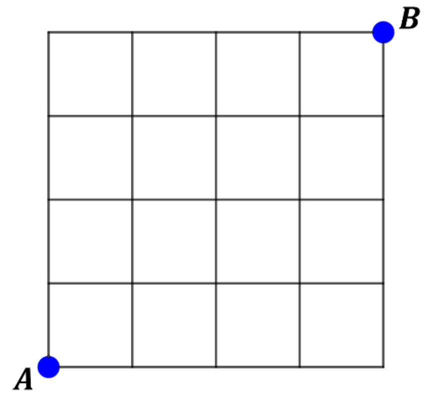
1. (5C191) Jamie flips four fair coins. Determine exactly the probability that she gets more heads than tails.
2. (5C194) Amy and Ben each have biased coins but they are unfair in different ways: Amy's coin comes up <u>heads</u> only $1/3$ of the time and Ben's coin comes up <u>tails</u> only $1/3$ of the times. They both flip their respective coins twice. Determine exactly the probability that they get the same number of heads.
3. (5C174) A server noted that when diners have cherry pie for dessert, $3$ out of $5$ will leave a big tip. If there are $6$ diners who have cherry pie, what is the probability that a big tip is left by exactly $4$ of them?
4. (5C153) Three quarters and three dimes are tossed in the air. Determine exactly the probability that the same number of quarters and dimes turn up heads.
5. (TT156) An unfair coin lands tails with a probability of $1/5$. When tossed $n$ times, the probability of exactly three tails is the same as the probability of exactly $4$ heads. What is the value of $n$?
6. (5C141) What is the probability that a fair coin, flipped three times, will land all tails?
7. (5C142) Determine exactly the probability that a student, by randomly guessing, achieves a score of $3$ out of $5$ on a pop quiz whose questions are multiple choice with four choices each.
8. (MB0610) A die is rolled six times. What is the probability of getting either a $1$ or a $6$ on at least three rolls?
9. (5A054) (a) Amy, Beth, and Christine toss a coin $15$,$16$ and $17$ times respectively. Which girl is least likely to get more heads than tails?
10. (5A054) (b) Amy, Beth and Christine toss a coin $18$,$19$ and $20$ times respectively. Which girl is least likely to get more heads than tails?
11. (5C054) Sarah was sent to get $8$ cans of soda to have on hand for the study session. When she got to the machine, she found that she had six choices. Remembering that she had a die in her purse for some homework for her probability class, she decided to roll it $8$ times, choosing the first flavor if a $1$ came up, etc. for all six possibilities. Using this scheme, what is the probability that she gets exactly four diet cokes?

12. (TI0512) John flips a fair coin $12$ times. What is the probability that he gets more heads than tails?

13. (TI008) If the probability of $A$ beating $B$ is $3/5$ (so the probability of losing to $B$ is $2/5$), find the probability of $A$ winning exactly $8$ of $12$ games the teams play during a season.

14. (5C974) In the game of Zonk, one throws six dice on each turn. What is the probability that on a random throw, exactly three dice will be $2$'s?

15. (TT974) John starts at $A$ and walks toward $B$ (eight blocks), moving only to the right or up (east or north). Jane starts at $B$ and walks toward $A$ (eight blocks), moving only to the left or down (west or south). If they start at the same time and walk at the same rate, and they each choose their direction at intersections (when they have a choice) in a random manner, what is the probability they will meet along the way?




16. (TC942) A fair coin is flipped $8$ times. What is the probability of obtaining exactly $3$ heads?

17. (TI9213) A pitcher throws a sequence of pitches and a blind umpire calls each one a ball or strike at random, meaning that independent of previous calls, the probability that a particular pitch will be called a strike is $1/2$. The following questions assume that the batter does not swing at any pitch (*i.e.* is taking).

    (a) What is the probability that the first three pitches will be called strikes?

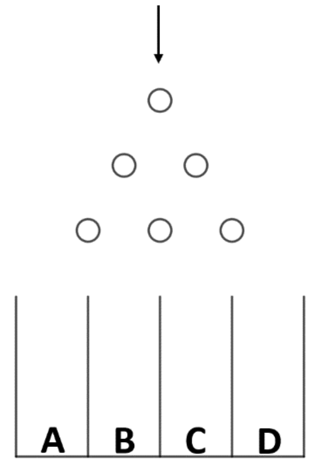
    (b) What is the probability that after two pitches, the count will be $1$ ball, $1$ strike?

    (c) What is the probability that the batter will be called out (have $3$ strikes called) before a fifth pitch has been thrown? (No more pitches will be thrown to a batter once three strikes have been called.)

    (d) What is the probability that the umpire will call $3$ strikes before calling $4$ balls.

18. (TI8912) It is rumored that Eric Vee, Math Leaguer from Moose Lake, beats his father Stan, Math League coach at Barnum, about $2/3$ of the time in chess. Assuming this to be true, what is the probability that Eric will beat his father in a best of three match (which, in the usual fashion, is over as soon as someone wins two games)?

19. (5D852) A ball in a certain pinball machine bounces to your left with probability $1/3$ and to your right with probability $2/3$ whenever it comes to a bumper. It falls through the configuration shown, hitting a bumper at each level, finally coming to rest in one of the slots $A, B, C$ or $D$. What is the probability that the ball will come to rest in slot $C$?




20. (TI8414) Oddsmakers believe the probability of the Tigers beating the Cubs in a single game is $0.6$. The Tigers and the Cubs are to play a series of up to seven games, the first team to win four games being declared the champion.

    (i) What probability should they assign to the outcome of the Cubs winning the series in exactly $6$ games?

    (ii) What probability should they assign to the outcome that the Tigers will win the series in less than $7$ games?

21. (3T836) Rodney Roller is down to his last turn in a Yahtzee game. He has two $6$'s on the table, but he needs one more. He gets to roll three dice. What is the probability that Rodney will get <u>at</u> <u>least</u> one $6$ when he rolls the three dice?

22. (5C182) When tossing two $8$-sided dice, the sides of each die being numbered $1$ through $8$, determine the probability of rolling two numbers $a$ and $b$, such that $|a - b| < 3$.

23. (TC894) A fair die is rolled three times. What is the probability that two of the three rolls, but not all three will be equal?

24. Every hospital has backup generators for critical systems should the electricity go out. Independent but identical backup generators are installed so that the probability that at least one system will operate correctly when called upon is no less than $0.99$. Let $n$ denote the number of backup generators in a hospital. How large must $n$ be to achieve the specified probability of at least one generator operating, if the probability that any backup generator will work correctly is $0.95$?

25. You flipped a coin $10$ times and got $8$ heads and this made you wonder if this was a "fair" coin (i.e. $50/50$ chance of heads/tails). To find out you decide to run an experiment

consisting of $5$ replications of flipping this coin $10$ times (in each replication) in a controlled manner. If this really is a fair coin, what is the probability of observing at least one replication where you observe at least $8$ heads?

26. A tire maker knows from past experience that $20\%$ of their top-of-the-line brand tire will not satisfy the conditions of their warranty. The tire maker also knows from experience that only $10\%$ of their customers of this top-of-the-line brand tire will bother to make a claim on their warranty when they could. What is the probability that a particular tire store selling this top-of the-line brand tire will not see any appropriate claims against the warranty from their next $30$ customers buying this brand of tire?
27. Consider a multiple-choice examination with $10$ questions, each of which has $4$ possible answers. If a student knows the correct answer with probability $0.8$ and guesses with probability $0.2$, what is the probability that this student will score at least an $80$? Assume each question is worth $10$ points and no partial credit is given. Also assume that the questions are answered independently, that is, whether this student answers question #3 correct or incorrect will not influence whether they answer question #7 correct or incorrect?
28. A fair die is rolled $n$ times. What is the smallest value of $n$ such that the probability of getting at least one six in these $n$ rolls is $0.95$ or higher?
29. Suppose you know from experience that $1\%$ of the parts coming off an assembly line at a local manufacturing plant are defective.

    (i) What is the probability that a lot of 500 will have less than $3$ defectives in it?

    (ii) Suppose you cannot tell by just looking whether a part is defective but rather have to subject the part to a test in order to tell. What is the probability that you will have to test more than $150$ parts before you find a defective one?

30. An airline knows from experience that $10\%$ of the people holding reservations on a given flight will not appear. The plane holds $90$ people. If $95$ reservations have been sold, what is the probability that the airline will be able to accommodate everyone appearing for the flight?

31. Suppose that a four-engine plan can fly if at least two engines work and suppose that a two-engine plan can fly if at least one engine works. Would you rather fly on a four-engine or two-engine plan?

32. During the 1978 baseball season, Pete Rose of the Cincinnati Reds set a National League record by hitting safely in 44 consecutive games. Assume that Rose is a 300 hitter, (i.e. the probability he hits safely on any given time at bat is .300) and assume that he comes to bat four times each game. If each at bat is assumed to be an independent event, what is the probability of hitting safely in 44 consecutive games?

33. A coin is altered so that the probability that it lands on heads is less than $1/2$ and then the coin is flipped four times, the probability of an equal number of heads and tails is $1/6$. What is the probability that the coin lands on heads? (2010 AMC 12 A Problem 15)

34. Coin $A$ is flipped three times and coin $B$ is flipped four times. What is the probability that the number of heads obtained from flipping the two fair coins is the same? (2004 AMC 10a Problem 10)

35. A coin with an unknown probability of landing heads is tossed ten times and lands on heads exactly three times. Find the conditional probability that the first toss landed on heads. (Source: https://math.la.asu.edu/~jtaylor/teaching/Spring2017/STP421/problems/ps2-solutions.pdf)

**Solution**

Let $p$ equal the unknown probability that this coin will on heads on any toss. Let $E$ be the event of getting exactly 3 heads in 10 tosses. By the binomial distribution we have

$$P(E) = \binom{10}{3} p^3(1-p)^7.$$

Let $H_1$ be the event that the first toss landed on heads. Then

$$P(H_1|E) = \frac{P(E|H_1)P(H_1)}{P(E)}.$$

The distribution of $E$ conditional on the information that the first toss landed on heads is the same as the unconditional probability of getting 2 heads in 9 tosses. But this is again modeled by the binomial distribution. That is,

$$P(E|H_1) = \binom{9}{2} p^2(1-p)^7.$$

Also, we know that $P(H_1) = p$. Therefore,

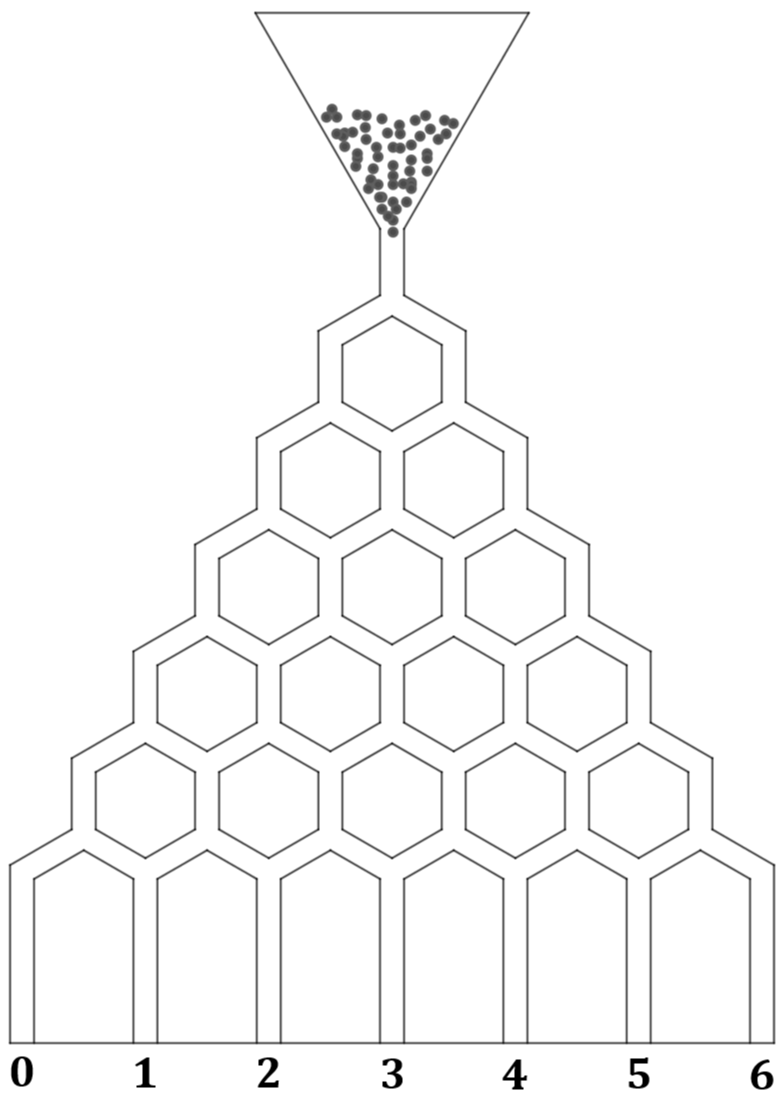
$$P(H_1|E) = \frac{P(E|H_1)P(H_1)}{P(E)} = \frac{\binom{9}{2} p^2(1-p)^7 \cdot p}{\binom{10}{3} p^3(1-p)^7}$$

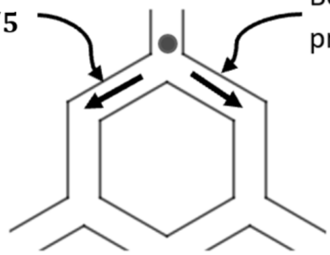
$$= \frac{\binom{9}{2}}{\binom{10}{3}} = \frac{9!}{10!} \cdot \frac{3!}{2!} = \frac{3}{10}.$$

■

## Project 2 The Quincunx - A Discrete Optimization Problem

Small metal balls are poured into the top funnel of the device pictured below. This device is alternately called a *quincunx* (by the way, that's a big 26 point word in Scabble) or sometimes a Galton's board. Depending on whether the quincunx is perfectly level or not, a ball could have different probabilities of bouncing left and right. Assuming that each ball bounces to the left with probability $3/5$ and to the right with probability $2/5$ at each intersection and assuming the balls bounce independently of each other throughout the process, **what is the most likely column** (from Column $0$ on the far left to Column $6$ on the far right) **that a ball will land in**?






No matter which path a ball takes there are exactly six intersections where that ball has to either bounce to the left or bounce to the right. How could a ball land in Column $0$? The only way is for that ball to bounce to the right at none of the intersections. Where will a ball land if it bounces to the right at exactly one of the intersections? It is instructive to trace the paths with exactly your finger that bounce to the right exactly once (and hence to the left five times). You can see that all such paths a ball will land in Column 1.

Each intersection is actually a Bernoulli trial.

> Are the direction of the bounces made independently of other? Yes. The probability that a ball will bounce left (or right) at the third intersection, for example, is in no way influenced by which way the ball bounced at the first two intersections.
>
> Are there only two outcomes at each intersection? Yes. The ball can only bounce left or right.
>
> Is the probability of bouncing left (or right) constant at each intersection? Yes. A fix probability of $3/5$ of bouncing left at each intersection.
>
> Are there a fixed number of trials (intersections)? Yes. Exactly 6.

If we label bouncing to the right as our "success" and if we define $X$ as the number of successes in the $n = 6$ trials (intersections), then $X$ follows a binomial distribution with $p = 1 - 3/5 = 2/5$.

That is, the probability that a ball will land in Column $k$ equals the probability that $X = k$ for a binomial distribution.

$$P(X = k) = \binom{6}{k}\left(\frac{2}{5}\right)^k\left(\frac{3}{5}\right)^{6-k}, \quad k = 0,1,\ldots,6.$$

Hence, the method to find the most likely column has come down to determining which value of $k$ maximizes the above formula for $P(X = k)$.

With a calculator we can quickly determine that

| $k$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $P(X=k)$ | 0.047 | 0.187 | 0.311 | 0.276 | 0.138 | 0.037 | 0.004 |

So, the most likely column is Column 2 with $P(X=2)=0.311$.

Can we solve this problem as a function of a generic $n$ and $p$? That is, can we determine, as a function of $n$ and $p$, the value of $x$ where

$$a(x)=\binom{n}{x}p^x(1-p)^{n-x},\ \ x=0,1,\dots,n.$$

is maximized?

For those of you who have already taken calculus, your mind might already be jumping to "take a derivative, set it equal to $0$, solve for $x$".

The problem is that $a(x)$ is only defined at the integers $x=0,1,\dots,n$. The derivative is not defined in this case. Is there an alternative method that will work when our function is only defined at a finite number of integers? Consider the ratio $r(x)=a(x)/a(x-1)$. If we can find where $r(x)$ switches from greater than 1 to less than 1 (or vice versa) then we can find where $a(x)$ is maximized (or minimized) because $r(x)>1$ if and only if $a(x)>a(x-1)$.

$$r(x)=\frac{a(x)}{a(x-1)}=\frac{\binom{n}{x}p^x(1-p)^{n-x}}{\binom{n}{x-1}p^{x-1}(1-p)^{n-x+1}}$$

$$=\frac{\frac{n!}{x!\,(n-x)!}}{\frac{n!}{(x-1)!\,(n-x+1)!}}\frac{p^x}{p^{x-1}}\frac{(1-p)^{n-x}}{(1-p)^{n-x+1}}$$

$$=\frac{n-x+1}{x}\left(\frac{p}{1-p}\right).$$

Now note that

$$r(x)=\frac{n-x+1}{x}\left(\frac{p}{1-p}\right)>1$$

$$\Leftrightarrow (n - x + 1)p > x(1 - p)$$

$$\Leftrightarrow x < np + p.$$

This shows that if $np + p$ is not an integer, then

$$a(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

is maximized at the unique $x$ between $np + p - 1$ and $np + p$. And if $np + p$ is an integer, then $a(x)$ is maximized at $x = np + p - 1$.

But we also might notice that in the case $np + p$ is an integer then

$$\begin{aligned} \frac{a(np + p)}{a(np + p - 1)} &= \frac{n - np - p + 1}{np + p} \left( \frac{p}{1 - p} \right) \\ &= \frac{n - np - p + 1}{np + p} \left( \frac{p}{1 - p} \right) = \frac{np - np^2 - p^2 + p}{np + p - np^2 - p^2} = 1. \end{aligned}$$

So, in the case $x = np + p$ is an integer $a(np + p - 1) = a(np + p)$ and as we already have shown that $a(x)$ is maximized at $x = np + p - 1$, it follows that $a(x)$ is also maximized at $x = np + p$.

Does this give us the correct answer to the problem we previous solved when $n = 6$ and $p = 2/5$? In this case

$$np + p = 6 \left( \frac{2}{5} \right) + \frac{2}{5} = \frac{14}{5}$$

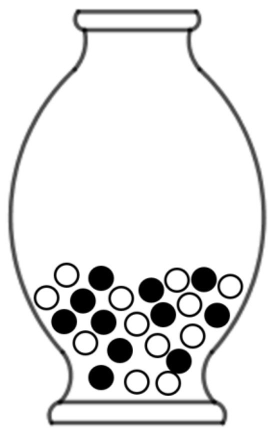
which is not an integer and according to the above general solution when $np + p$ is not an integer then the unique maximum occurs at the unique $x$ between $np + p - 1 = 7/5$ and $np + p = 12/5$. We can see that the unique integer in this interval is $10/5 = 2$ which agrees with our previous answer.

■

# Chapter 3 Hypergeometric Random Variables

Suppose an urn contains $w$ white balls and $b$ black balls. You reach into this urn and randomly select a ball, note its color but ***then do not return the ball to the urn***. You repeat this process for a total of $n$ draws. Let $X$ equal the number of times you select a white ball and let $P(X = x)$ represent the probability that you get exactly $x$ white balls in these $n$ draws.

By the way, in some hypergeometric problems the phrasing might be to select $n$ balls "one-by-one without replacement" and in others the phrasing might be to select $n$ balls "all at once". From a probability point of view, these are *equivalent* if your only objective is to count the number of white balls among the $n$ balls drawn.



Then

$$P(X = x) = \begin{cases} \dfrac{\binom{w}{x}\binom{b}{n-x}}{\binom{w+b}{n}} & x \in \{\max\{0, n-b\}, \dots, \min\{w, n\}\} \\ 0 & x \neq \{\max\{0, n-b\}, \dots, \min\{w, n\}\}. \end{cases}$$

In probability and statistics, whenever $P(X = x)$ is determined by the above formula we say that $X$ is a **hypergeometric random variable** or that $X$ follows the **hypergeometric distribution**.

First off, let's plug in some numbers to guide us through the messy looking range of $X$.

$$x \in \{\max\{0, n-b\}, \dots, \min\{w, n\}\}.$$

For now, let $\alpha$ (alpha) represent the smallest number and let $\omega$ (omega) represent the largest number of white balls we can get when drawing $n$ times without replacement from an urn containing $w$ white balls and $b$ black balls.

What we want to verify is that in all situations $\alpha = \max\{0, n-b\}$, the <u>max</u>imum of the two numbers $0$ and $n-b$ and that $\omega = \min\{w, n\}$, the <u>min</u>imum of the two numbers $w$ and $n$.

Let's consider various cases:

<u>Case 1a</u>. $(n = 3, w = 4, b = 6)$

Suppose we draw $n = 3$ times without replacement from an urn containing $w = 4$ white balls, $b = 6$ black balls. In this scenario we could draw as few as $0$ white balls.

<u>Case 1b</u>. $(n = 3, w = 4, b = 1)$

Suppose we draw $n = 3$ times without replacement from an urn containing $w = 4$ white balls, $b = 1$ black ball. Obviously, it is always true that we cannot draw less than $0$ white balls but in this case, we can also see that it would be *impossible* to draw exactly $0$ white balls. While we might get the $b = 1$ black ball on one of the $n = 3$ draws, we would have then run out of black balls and the other $n - b = 3 - 1 = 2$ draws would necessarily result in white balls.

What Cases 1a, 1b show us is that the number of white balls drawn cannot be less than <u>*either*</u> $0$ or $n - b$. But this is logically equivalent to the single statement that the number of white balls drawn cannot be less than the <u>*maximum*</u> of $0$ and $n - b$.

That is, $\alpha = \max\{0, n-b\}$.

<u>Case 2a</u>. $(n = 3, w = 4, b = 6)$

Here again we will draw $n = 3$ times without replacement from an urn containing $w = 4$ white balls, $b = 6$ black balls. By drawing a white ball each time draw we could draw as many as $n = 3$ white balls.

<u>Case 2b</u>. $(n = 3, w = 2, b = 6)$

In this case it would be *impossible* to get a white on each of the $n = 3$ draws because we are sampling without replacement and there are only $w = 2$ white balls in the urn.

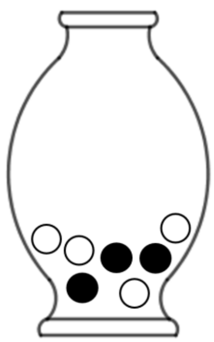
What Cases 2a, 2b show us is that the number of white balls drawn cannot be greater than <u>*either*</u> $n$ or $w$. But this is logically equivalent to the single statement that the number of white balls drawn cannot be greater than the <u>*minimum*</u> of $0$ and $w$.

That is, $\omega = \min\{w, n\}$. Hence, we have established that $x \in \{\max\{0, n-b\}, \dots, \min\{w, n\}\}$.

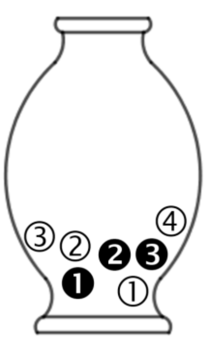
**Derivation of the Hypergeometric Probability Distribution**

$$P(X=x)=\begin{cases}\dfrac{\binom{w}{x}\binom{b}{n-x}}{\binom{w+b}{n}} & x\in\{\max\{0,n-b\},\dots,\min\{w,n\}\}\\[2ex] 0 & x\neq\{\max\{0,n-b\},\dots,\min\{w,n\}\}.\end{cases}$$

We will start our derivation by focusing on the special case of an urn with $w=4$ identical white balls and $b=3$ identical black balls. If we randomly take out $n=3$ balls from this urn, what is the probability of getting $x=2$ white balls (and 1 black ball)?

| | |
|---|---|
|  | What is the probability of getting 2 white balls and 1 black ball if we randomly select 3 balls from this urn with unlabeled balls? |

But before we even being on solving this, let's introduce the related problem when the balls are labeled. What is the probability of getting $x=2$ white balls (and 1 black ball) if we randomly take out $n=3$ balls from the urn below with $w=4$ labeled white balls and $b=3$ labeled black balls?

| | |
|---|---|
|   | What is the probability of getting 2 white balls and 1 black ball if we randomly select 3 balls from this urn with labeled balls? |

Practical experience tells us that the labels will <u>not</u> change how likely we are to get 2 white and 1 black if we randomly draw out 3 balls from this urn.

However, once we attach labels, we can view this as an urn where all the $w + b$ balls are *distinguishable*. How does it help us to how distinguishable balls in the urn?

The basic formula for *combinations* tells us there are $\binom{7}{3}$ distinct samples of size $n = 3$ from an urn of $w + b = 7$ **distinct** objects. The combinations formula does not apply unless all the objects are distinct.

Furthermore, because we are selecting balls at random, all $\binom{7}{3}$ samples are equally likely.

How many distinct ways can we select $x = 2$ white balls from this urn with $w = 4$ white balls? The same formula for combinations tells us there are $\binom{4}{2}$ distinct ways to select the $x = 2$ white balls from this urn. By the same reasoning there are $\binom{3}{1}$ ways to select $n - x = 3 - 2 = 1$ black balls from this urn.

Because we want to select $x = 2$ white balls **and** $n - x = 1$ black ball, the product rule for counting tells us there $\binom{4}{2} \cdot \binom{3}{1}$ ways to do both.

That is, there are a total of $\binom{4}{2}\binom{3}{1}$ ways to select 2 white balls and 1 black ball from this urn with 4 distinct white balls and 3 distinct black balls.

And because the $\binom{7}{3}$ total number of ways to select $n = 3$ balls (without any restriction on colors) are all equally likely, the ratio

$$\frac{\binom{4}{2}\binom{3}{1}}{\binom{7}{3}}$$

of counts gives the probability of getting 2 white and 1 black when we select 3 balls from the above urn (whether the labels are attached or not).

It is straightforward to extend this to the general case where we draw $n$ times (without replacement) from an urn containing $w$ white balls and $b$ black balls and we want to find the probability of getting $x$ white (and hence $n - x$ black) balls.

$$\frac{\binom{4}{2}\binom{3}{1}}{\binom{7}{3}} \longrightarrow \frac{\binom{w}{x}\binom{b}{n-x}}{\binom{w+b}{n}}.$$

**Multivariate Hypergeometric Model**

Suppose an experiment consists of drawing marbles at random and without replacement from an urn containing initially containing $r_j$ Type $j$ balls, $j = 1,2, \dots, k$.

Let $X_j$ equal the number of Type $j$ balls selected in $n$ draws. Then

$$P(X_1 = x_1, \dots, X_k = x_k) = \begin{cases} \dfrac{\binom{r_1}{x_1}\binom{r_2}{x_2}\cdots\binom{r_k}{x_k}}{\binom{r_1 + r_2 + \cdots + r_k}{x_1 + x_2 + \cdots + x_k}} & \begin{array}{c} x_j \in \{0,1,2, \dots, n\}, j = 1, \dots, n \\ x_1 + x_2 + \cdots + x_k = n \end{array} \\ 0 & \text{else} \end{cases}$$

**Negative Hypergeometric Distribution**

**Polya's Method for Estimating How Many Typo's Are Still in a Document**

## Exercises for Chapter 3

**Read** each of the following problems which were all taken from old MSHSML tests and consider how each of them depends in various ways on a hypergeometric random variable. (*i.e.* Identify what is playing the role of the white balls and the black walls, identify how is it made clear we are sampling *without* replacement and identify how many balls we are sampling in total).

1. (5T194) Johnny has a drawer containing two green socks, two red socks, two black socks, two white socks and two blue socks. Johnny likes to wear matching socks, but he is color blind and cannot distinguish red from green. Johnny randomly pulls our four socks from his drawer. Determine exactly the probability that Johnny does not get a pair of socks of the same color, but *thinks* he does.
2. (TI199) I have a bag with $21$ beads of various colors, the majority being green. If I remove two beads (without replacement), the probability I will end up with one green and one non-green is $3/7$. How many green beads were in the bag originally?
3. (5C171) Sal's drawer contains $4$ black socks and $2$ red socks. If he chooses two socks at random without replacement, what is the probability he chooses two socks of the same color? Express your answer as a quotient of two relatively prime integers.
4. (5C102) An ordinary deck of $52$ playing cards is shuffled and $2$ cards are dealt face up. Calculate the probability that at least one of these is a spade. *[NYCC, Fall 1984]*
5. (MB097) Ten playing cards are laying face down on a table. Exactly three of them are aces. You get to turn over two cards. What is the probability that at least one of them is an ace?
6. (TI086) A hand of $5$ cards is drawn from a standard deck of $52$ cards. Find the probability of getting exactly one ace.
7. (TT082) A hand of $5$ cards is drawn from a standard deck of $52$ cards. Successful participants in this morning's Invitational found that the probability of getting exactly one ace is $0.299474$. You get one point for finding the probabilities (again to six decimal places) of each of the other possibilities.
8. (TC063) A cooler contains $4$ bottles of carbonated mineral water and $6$ bottles of non-carbonated spring water. Not realizing there is a choice, Alicia and Beth each plunge their hands into the ice and grab a bottle. What is the probability that both girls get the same kind of water?
9. (5C052) If you draw three cards from an ordinary $52$ card deck, what is the probability of drawing at least one of the twelve face cards (*i.e.* a Jack, Queen, or King)?

10. (5C012) Nine playing cards are lying face down on a table. Exactly two of them are Aces. You pick up two cards. What is the probability that at least one of them will be an Ace?

11. (5T003) A box contains eleven balls numbered $1,2,3,\ldots,11$. If six balls are drawn simultaneously at random, what is the probability that the sum of the numbers on the balls drawn is odd? [1984 AHSME, prob. 19]

12. (TI0010) Alicia grabs two bagels from a bag containing $10$ whole wheat and $6$ pumpernickel bagels. What is the probability that she gets $1$ of each?

13. (5C991) Cards are turned over one at a time (without replacement) from an ordinary well-shuffled decks ($52$ cards). What is the probability that the first two cards are aces? (There are four aces in a deck.)

14. (TC994) A piggy bank contains $1$ quarter, $3$ dimes, $2$ nickels, and $1$ cent. Needing $46$ cents, Aaron shakes out four coins. Assuming the coins are equally likely to be shaken out, what is the probability that he has shaken out exactly $46$ cents?

15. (TI9911) A bag contains white marbles and black marbles. If you were to reach in without looking and pull out one marble, the probability that it would be white is $1/4$. However if you were to pull out two marbles, the probability that both would be white is $1/18$. How many white marbles are in the bag?

16. (5T985) Three cards are drawn at random from a standard deck of $52$ playing cards. Calculate the probability (rounded to four decimal places) that at least one of these three cards is a face card. (Here, a face cards is a jack, queen, or king. The motivation for this problem is the question of whether or not one would give even odds that at least one of the three cards is a face card.)

17. (TT983) A bag contains $16$ billiard balls, some black and the remainder white. Two balls are drawn at the same time. It is equally likely that the two balls will be the same color as different colors. How are the balls divided within the bag?

18. (5C972) In your drawer there is a pair of green gloves and a pair of brown gloves. You reach in with your eyes closed and pull out two gloves at random. What is the probability that you have a matched pair?

19. (5T913) I have $n$ nuts in my hand, $m$ of which will fit a bolt on my bicycle. If I select $2$ of the nuts at random, there is an even chance (*i.e.* the probability is $1/2$) that both will fit the bolt. What is the smallest value of $n$?

20. (5C902) There are $4$ black socks and $6$ brown socks all mixed together in a drawer in a dark room. If Pat grabs two socks from the drawer and takes them into the light, what is the probability that they will match?

21. (5C903) There are $4$ black socks and $6$ brown socks all mixed together in a drawer in a dark room. If Pat grabs three socks from the drawer and takes them into the light,

    a) what is the probability of a match?

    b) what is the probability of at least two black socks?

22. (5T881) Amy, who puts only nickels and dimes in her bank, currently has an even number of nickels. When two coins are drawn at random from this bank, the probability that both are dimes is $1/2$. What is the smallest amount of money that Amy might have?

23. (TD883) A box contains nine balls numbered $1$ through $9$. If five balls are drawn at random and without replacement, what is the probability that the sum of the numbers on the balls will be odd?

24. Suppose that $100$ cards marked $1,2, \dots ,100$ are randomly arranged in a line. What is the probability that exactly $8$ of the first $20$ cards in this line are marked with an even number?

25. Suppose you play a lottery where you choose five different numbers from the integers $1, 2, \dots, 45$. Then you choose one number from $1, 2, \dots, 45$. This last number is often called the *powerball* and can (but does not have to) repeat one of the first five numbers you chose. Lottery officials choose five numbers and the powerball in the same way. What is the probability that you match $3$ of the first $5$ but do not match the powerball?

26. 15 women and 9 men were eligible for promotion at a university and from this group 2 women and 4 men were selected for promotion. Is there reason to suspect gender bias?

27. An urn contains $3$ red, $5$ blue and $7$ yellow balls. You pull out $4$ balls (without replacement). What is the probability you get exactly $2$ yellow balls?

28. A Hypergeometric Problem Nested Inside a Binomial Problem

    Imagine a game that consists of two stages. In the first stage you have to draw $3$ balls at random and without replacement from an urn containing $4$ black and $8$ white balls. According to the rules, in order to be able to proceed to the second stage of this game you have to get at least one black ball in this sample of size three during the first phase.

    (a) Out of ten players (all starting with the same urn of $4$ black and $8$ white balls), what is the probability that five or more will make it to the second stage?

    (b) How many out of these ten players do you expect to make it to the second stage?

29. Suppose there are two identical urns and that each urn contains $N$ balls numbered from $1$ to $N$. One of these urns is given to Mary and the other is given to Bob. Mary reaches into her urn and randomly selects $n$ balls without replacement and makes a list of the numbers

she got. Bob reaches into his urn and randomly selects $k$ balls without replacement and makes a list of the numbers he got.

Let $X$ equal how many numbers are on both Mary's and Bob's list. Find $P(X = x)$.

30. In the game of Texas Hold'em, players are each dealt two private cards, and five community cards are dealt face-up on the table. Each player makes the best 5-card hand they can with their two private cards and the five community cards. What is the probability that a particular player can make a flush of spades (i.e. 5 spades)?

    https://brilliant.org/wiki/hypergeometric-distribution/

## Project 3 The Capture-Recapture Problem

In order to estimate $N$, the total number of bluegills in Goose Lake in Waseca, $250$ bluegills are captured, tagged and released by the DNR. A few days later a sample of $150$ bluegills were caught. It was found that $16$ bluegills in this sample were tagged. Assume that the $150$ bluegills in this sample were held in a holding tank for a few days to ensure that no bluegill was sampled more than once.

Find the value of $N$ that maximizes the probability that the DNR would have found $16$ tagged bluegills in the sample of $150$. In short, we want to find the value of $N$ that is the most consistent with the observed data. This is called the ***capture-recapture*** method for estimating $N$.

### Solution

Let $Y$ represent how many tagged bluegill will be caught in a sample of size $150$ bluegills.

We need to find the value of $N$ that maximizes $L(N) = P(Y = 16)$.

We can think of Goose Lake as an urn with $250$ white balls (tagged blue gills) and $N - 250$ black balls (untagged blue gills). From this urn (lake) we take out a random sample of size $150$ without replacement and find that we got $16$ white balls (tagged blue gills).

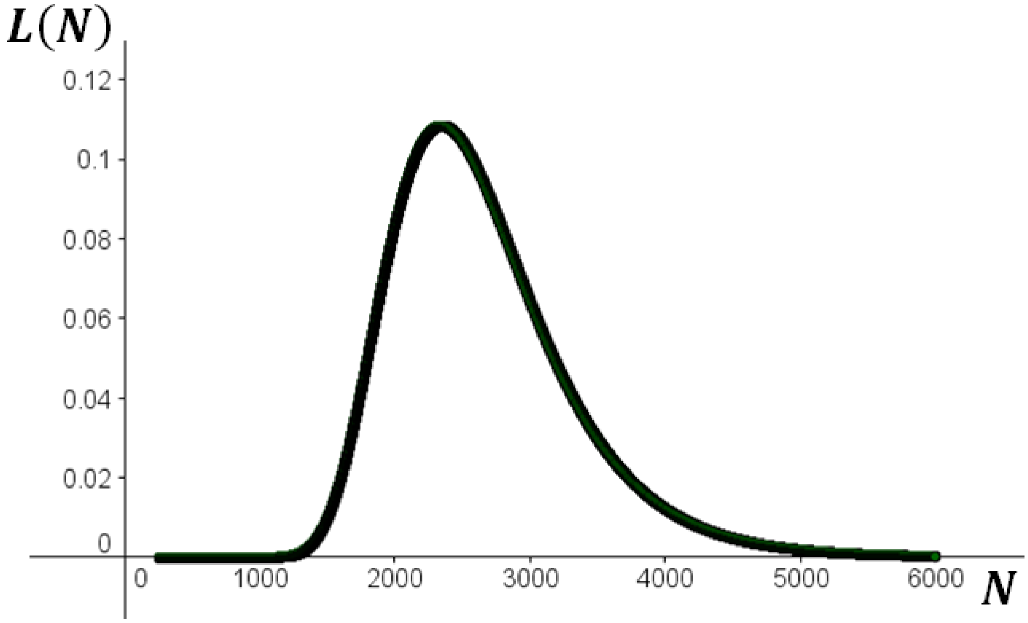
So, $Y$, the number of tagged bluegill caught in the sample of size $150$, follows a **hypergeometric distribution**. More precisely,

$$P(Y = y) = \frac{\binom{250}{y}\binom{N-250}{150-y}}{\binom{N}{150}}.$$

So,

$$L(N) = P(Y = 16) = \frac{\binom{250}{16}\binom{N-250}{150-16}}{\binom{N}{150}}.$$

Our goal is to the find the positive integer $N$ that maximizes $L(N)$. Here is a plot of $L(N)$ for every integer $N$ from $250$ to $6000$.




From the plot it appears that the value of $N$ that maximizes $L(N)$ is in the neighborhood of 2300 but it's kind of hard to tell.

What we want to find is that integer $N$ where

$$L(N) > L(N-1) \quad \text{and} \quad L(N) > L(N+1).$$

But this is logically equivalent to finding the integer $N$ where

$$\frac{L(N)}{L(N-1)} > 1 \quad \text{and} \quad \frac{L(N+1)}{L(N)} < 1.$$

Let's see if we can find the value of $N$ where this happens.

$$\frac{L(N)}{L(N-1)}$$

$$= \frac{\dfrac{\binom{250}{16}\binom{N-250}{150-16}}{\binom{N}{150}}}{\dfrac{\binom{250}{16}\binom{(N-1)-250}{150-16}}{\binom{N-1}{150}}}$$

$$= \frac{(N-250)!}{(N-251)!}\ \frac{(N-150)!}{(N-151)!}\ \frac{(N-1)!}{N!}\ \frac{(N-385)!}{(N-384)!}$$

$$= \frac{(N-250)(N-150)}{N(N-384)}.$$

Therefore,

$$\frac{L(N)}{L(N-1)} > 1$$

$$\Leftrightarrow \frac{(N-250)(N-150)}{N(N-384)} > 1$$

$$\Leftrightarrow (N-250)(N-150) > N(N-384)$$

$$\Leftrightarrow N < \frac{(250)(150)}{16} = 2343.75.$$

Similarly,

$$\frac{L(N+1)}{L(N)} < 1$$

$$\Leftrightarrow (N-249)(N-149) < (N+1)(N-383)$$

$$\Leftrightarrow N > \frac{383 + (249)(149)}{16} = 2342.75.$$

Therefore, the integer value of $N$ that maximizes $L(N)$ is bounded by

$$2342.75 < N < 2343.75$$

which means that the capture-recapture estimate of $N$ is

$$\hat{N} = 2343.$$

**General Capture Recapture Problem**

Now suppose that we do this for the generic problem where we initially catch, tag and release $t$ bluegills. We will again let $N$ represent the total number of bluegills in the lake. Later we catch $s$ bluegills and find that $y$ of these fish are tagged. Then our hypergeometric likelihood function for $Y$, the random variable representing how many tagged fish we will catch in the sample is given below.

$$L(N) = P(Y = y) = \frac{\binom{t}{y}\binom{N-t}{s-y}}{\binom{N}{s}}.$$

If we go through the algebra and see how all the ratios of factorials simplify for this generic version of the problem we will find that

$$\frac{L(N)}{L(N-1)} > 1 \Leftrightarrow N < \frac{t \cdot s}{y}$$

and

$$\frac{L(N+1)}{L(N)} < 1 \Leftrightarrow N > \frac{t \cdot s}{y} - 1.$$

That is, $\hat{N}$ is that integer between $ts/y - 1$ and $ts/y$. We can also write this in terms of the floor function.

$$\hat{N} = \left\lfloor \frac{t \cdot s}{y} \right\rfloor$$

where $\lfloor x \rfloor$ equals the largest integer less than or equal to $x$.

*Could we have anticipated this result?* Yes! The percent of tagged fish in the second sample should be approximately the same as the percent of (all) tagged fish in the lake.

That is,

$$\frac{y}{s} \cdot 100\% \cong \frac{t}{N} \cdot 100\%.$$

Solving for $N$ gives us

$$N \cong \frac{t \cdot s}{y}.$$

■

# Chapter 4 Conditional Probability

### The Phrase "Given that" is Ambiguous in Probability

Consider the following standard conditional probability problem:

> You flip a fair penny and a fair nickel. What is the probability that both coins landed heads ***given that*** at least one coin landed heads?

The phrase "given that" is a key identifier of conditional probability problems. Precisely what do we mean by the phrase "given that"?

Let's consider three related storylines. In each case you are given the information that at least one coin landed heads. As you read these storylines think about whether and/or how the difference in the storylines impacts the probability that both coins landed heads

**Storyline 1.** Your teacher asks you to close your eyes and to flip a fair penny and a fair nickel. Your teacher sees the results but before you or anyone else gets the chance to see how the coins landed, your teacher covers each coin with a playing a card that completely hides the results. Your teacher tells you that the penny is on the left and the nickel is on the right.

Now you leave the room and your friend comes in. **Your friend looks under *both* cards and then covers them back up.**

Finally, you come back into the room and ask your friend "Did I get at least one heads?" Your friend truthfully answers "Yes" and also tells you the exact process they used to learn this information.

Given what you now know, what is the probability that both coins landed heads?

**Storyline 2.** Your teacher asks you to close your eyes and to flip a fair penny and a fair nickel. Your teacher sees the results but before you or anyone else gets the chance to see how the coins landed, your teacher covers each coin with a playing a card that completely hides the results. Your teacher tells you that the penny is on the left and the nickel is on the right.

Now you leave the room and your friend comes in. **Your friend has decided ahead of time that they will only look to see the outcome of the coin under the card on the left.**

Finally, you come back into the room and ask your friend "Did I get at least one heads?" Your friend truthfully answers "Yes" and also tells you the exact process they used to learn this information.

Given what you now know, what is the probability that both coins landed heads?

**Storyline 3.** Your teacher asks you to close your eyes and to flip a fair penny and a fair nickel. Your teacher sees the results but before you or anyone else gets the chance to see how the coins landed, your teacher covers each coin with a playing a card that completely hides the results. Your teacher tells you that the penny is on the left and the nickel is on the right.

Now you leave the room and your friend comes in. **Your friend looks under a randomly selected card (assume $50$-$50$ chance they looked under the card on the left or the card on the right) and then puts the card back over the coin.**

Finally, you come back into the room and ask your friend "Did I get at least one heads?" Your friend truthfully answers "Yes" and also tells you the exact process they used to learn this information.

Given what you now know, what is the probability that both coins landed heads?

In each storyline you flipped two fair coins and you were ***given that*** at least one coin landed heads.

- Does it matter ***how*** your friend came to get this information they gave you? Ans: Yes.
- Can it change the probability that both coins landed heads? Ans: Yes.
- Assuming that it does, does it change ***how*** we go about solving each problem? Ans: Yes.

The key to sorting this all out depends on **establishing the appropriate sample space and the appropriate conditional sample space** for each storyline.

| |
|---|
| Best Advice: Even in a contest setting where time is critical, carefully enumerate and account for *all* random aspects of the story when building your sample space. |

The sample space for the initial flips of the penny and nickel would just be

$$S = \left\{(\overset{\text{penny}}{H},\overset{\text{nickel}}{T}),(T,H),(H,H),(T,T)\right\}.$$

Note: We included $(H,T)$ as well as $(T,H)$ in the sample space because they are distinguishable.

In Storylines 1 and 2 there are no additional experiments performed with random outcomes. So, the above sample space $S$ applies to both storyline 1 and storyline 2.

Because both the penny and the nickel are assumed to be fair coins, all four of these outcomes are equally likely and hence each outcome has probability $1/4$.

However, in Storyline 3, your friend *randomly* selects whether they will look under the left or right card. We will designate these as $L$ and $R$ respectively. This additional random aspect of the story needs to be integrated into your sample space. So, for Storyline 3, our new sample space $S_3$ becomes:

$$S_3 = \Big\{(\overset{\text{penny}}{H}, \overset{\text{nickel}}{T}, \overset{\text{card lifted}}{L}), (H,T,R), (T,H,L), (T,H,R), (H,H,L), (H,H,R), (T,T,L), (T,T,R)\Big\}.$$

Because both the penny and the nickel are assumed to be fair coins and because your friend randomly selected whether to look under the left or right card, all eight of these outcomes are equally likely and hence each outcome has probability $1/8$.

In Storyline 1 you were given that at least one coin was a heads. So, $C_1$, the reduced (conditional) sample space for Storyline 1 is therefore,

$$C_1 = \Big\{(\overset{\text{penny}}{H}, \overset{\text{nickel}}{T}), (T,H), (H,H)\Big\}.$$

As all outcomes in the unconditional sample space for Storyline 1 were equally likely, all outcomes in its conditional sample space $C_1$ must remain equally likely.

Only one of these three equally likely conditional outcomes has both coins land heads. Thus, for **Storyline 1, the probability that both coins landed heads equals $\mathbf{1/3}$**.

In Storyline 2 you were given that the penny landed heads (the coin under the left card). So, $C_2$, the reduced (conditional) sample space for Storyline 2 is therefore,

$$C_2 = \Big\{(\overset{\text{penny}}{H}, \overset{\text{nickel}}{T}), (H,H)\Big\}.$$

As all outcomes in the unconditional sample space for Storyline 2 were equally likely, all outcomes in its conditional sample space $C_2$ must remain equally likely.

In just one of these two equally likely outcomes will both coins be heads. That is, for **Storyline 2, the probability that both coins landed heads equals $\mathbf{1/2}$**.

In Storyline 3 you were given that the coin under a randomly picked card was a heads. The outcomes $(H,T,R), (T,H,L), (T,T,L)$ an $(T,T,R)$ because in each of these cases your friend the coin your friend would have seen was a $T$ (tails).

So, $C_3$, the reduced (conditional) sample space for storyline three is therefore, penny nickel card lifted

$$C_3 = \Big\{(H, T, L), (T, H, R), (H, H, L), (H, H, R)\Big\}.$$

As all outcomes in the unconditional sample space for Storyline 3 were equally likely, all four outcomes in its conditional sample space $C_3$ must remain equally likely.

In two of these four remaining equally likely outcomes both coins are heads. That is, for **Storyline 3, the probability that both coins landed heads equals $\mathbf{2/4 = 1/2}$**.

Summarizing, we have shown that in Storylines 2 and 3 the probability that both coins landed heads equaled $1/2$. However, in Storyline 1 the probability that both coins landed heads equaled $1/3$.

Is this what you expected?

**Two Children Problem**

9. (5C981) A woman goes to visit the house of some friends whom she has not seen in many years. She knows that, besides the two married adults in the household, there are two children of different ages. But she does not know their genders. When she knocks on the door of the house, a boy answers. What is the probability that the other child is a boy?

**Solution**

This problem is known as the "Two Children Problem" and shows up in many forms on puzzle blogs. This problem is notorious for tripping up problem solvers intuition and is also notorious for tripping up problem writers for creating ambiguity in the wording.

To make the solution provided here complete the underlying assumptions being used must be stated explicitly.

Assumptions:

(1) the gender determination of all children in a family are independent events

(2) $P(\text{boy}) = P(\text{girl}) = 1/2$ for all children in a family

(3) $P(\text{younger child answers the door}) = P(\text{older child answers the door}) = 1/2$.

The unconditional sample space must account for the gender of the younger child, the gender of the older child and which child (younger or older) opens the door.

| **Younger** | **Older** | **Opens Door** | **Probability** |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Male | Male | Younger | $1/8$ |
| Male | Male | Older | $1/8$ |
| Male | Female | Younger | $1/8$ |
| Male | Female | Older | $1/8$ |
| Female | Male | Younger | $1/8$ |
| Female | Male | Older | $1/8$ |
| Female | Female | Younger | $1/8$ |
| Female | Female | Older | $1/8$ |

Which of these eight outcomes are **not** possible once we see that a boy answers the door? Answer: (Male, Female, Older), (Female, Male, Younger), (Female, Female, Younger), (Female, Female, Older)

Removing these four outcomes and rescaling the remaining possibilities leaves us with our conditional sample space and their probabilities.

| **Younger** | **Older** | **Opens Door** | **Conditional Probability** |
|---|---|---|---|
| Male | Male | Younger | $1/4$ |
| Male | Male | Older | $1/4$ |
| Male | Female | Younger | $1/4$ |
| Female | Male | Older | $1/4$ |

Now we can easily find $P(\text{other child is boy}|\text{boy opens the door})$. We just need to sum the conditional probabilities where both children are boys.

$$P(\text{other child is boy}|\text{boy opens the door}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

■

The solution provided by MSHSML for this problem was $1/3$. This answer is correct if we amend the storyline as follows: "Suppose the visitor knows there are two children in the family they are visiting and when the visitor knocks the mother answers the door. The visitor then asks the mother if she has at least one boy and the mother informs her that she does have at least one boy but does not elaborate further. Based on this information what is the probability that the mother has two boys?"

**Solution**

Let $A$ be the event that the younger child is a boy and let $B$ be the event that the older child is a boy. Then the question in this amended form becomes

$$P(\text{two boys}|\text{at least one boy})$$

$$= P\big((A \cap B)\big|(A \cup B)\big)$$

$$= \frac{P\big((A \cap B) \cap (A \cup B)\big)}{P(A \cup B)}$$

$$= \frac{P(A \cap B)}{P(A \cup B)}$$

$$= \frac{P(A \cap B)}{P(A) + P(B) - P(A \cap B)}$$

$$= \frac{P(A)P(B)}{P(A) + P(B) - P(A)P(B)}$$

$$= \frac{\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}{\left(\frac{1}{2}\right) + \left(\frac{1}{2}\right) - \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}$$

$$= \frac{\frac{1}{4}}{1 - \frac{1}{4}} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

■

Notice that in both the original and the amended storylines the visitor learns that the family they are visiting has at least one son. The difference is in ***how*** this information is learned. In the original version this information is learned as the outcome of a random event (namely, which child answers the door). In contrast, in the amended version this information is learned as "given information" and not through any random event.

**Multiplication Rule**

Let $A_1, \dots, A_n$ be events such that $P(A_1 \cap \cdots \cap A_n) > 0$. Then

$$P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P\big(A_3\big|(A_1 \cap A_2)\big) \cdots P\big(A_n\big|(A_1 \cap \cdots \cap A_{n-1})\big)$$

**Rule of Total Probability**

Let $\{H_n\}, n = 1,2, \dots$ be a sequence of events such that

$$H_i \cap H_j = \emptyset,\ \ i \neq j \ \text{ and } \ H_1 \cup \cdots \cup H_n = \Omega$$

Suppose that $P(H_n) > 0$ for all $n \geq 1$. Then for any event $A$

$$P(A) = \sum_{j=1}^{\infty} P(H_j)P\big(A\big|H_j\big)$$

**Bayes's Rule**

Let $\{H_n\}, n = 1,2, \dots$ be a sequence of mutually exclusive events such that $P(H_j) > 0$, $j = 1,2, \dots$ and $H_1 \cup \cdots \cup H_n = \Omega$. Let $A \subset \Omega$ be an event with $P(A) > 0$. Then for $j = 1,2, \dots$

$$P\big(H_j\big|A\big) = \frac{P(H_j)P\big(A\big|H_j\big)}{\sum_{k=1}^{\infty} P\,(H_k)P(A|H_k)}$$

**Extended Bayes' Rule**

Let $\{H_k\}$ be a sequence of mutually exclusive events such that $(H_k) > 0$, $k = 1,2, \dots$, and $\bigcup_{k=1}^{\infty} H_k = \Omega$. Let $A$ be an event with $P(A \cap B) > 0$. Then for $j = 1,2, \dots$

$$P(H_j|\,(A \cap B)) = \frac{P\big(H_j|B\big)P\big(A|\big(H_j \cap B\big)\big)}{\sum_{k=1}^{\infty} P(H_k|B)\,P(A|(H_k \cap B))}.$$

**Conditioning on what you wish you knew**

**First Step Analysis**

Hence the following general result about mutually independent events applies:

> If $E_1, \dots E_k, E_{k+1}, \dots, E_n$ are mutually independent events then for general set functions $f$ and $g$, $f(E_1, \dots, E_k)$ is independent of $g(E_{k+1}, \dots, E_n)$.

For certain problems, it may be more convenient to use the Bayes' ratio to evaluate comparative odds of two events than Bayes' theorem itself. For three events $A, B, C$

$$\frac{p(A|C)}{p(B|C)} = \frac{p(C|A)}{p(C|B)} \cdot \frac{p(A)}{p(B)}$$

**What is the conditioning event (Pfeiffer)**

Independent Events in a Discrete Uniform Probability Space

## Exercises for Chapter 4

1. (5T184) Two teams, $A$ and $B$, are playing in a tournament. They will play until one team wins four games (no ties allowed). The probability of either team winning the first game is $50\%$. For both teams the probability of winning the very next game after winning one is $60\%$, winning a third game after winning two in a row $70\%$, and winning a fourth game in a row is $75\%$. Determine the probability team $A$ wins in exactly $5$ games.

2. (5A164) Two people play a game, taking turns drawing a ball from an urn randomly, without replacing them. The urn starts out with $3$ black balls and $4$ white balls. If a player draws a black ball, their turn is complete. If a player draws a white ball, they must also (nonrandomly) take two black balls from the urn before they complete their turn. A player loses when they cannot complete their turn. If you draw first, determine the exact probability you will win the game.

3. (5C122) Kathy flips a fair coin until she gets three heads in a row. If it is known that she stopped after seven flips, what is the probability her <u>first</u> flip was heads?

4. (5C102) An ordinary deck of $52$ playing cards is shuffled and $2$ cards are dealt face up. Calculate the probability that at least one of these is a spade. ***[NYCC, FALL 1984]***

5. (5C083) At a carnival game, a black velvet pouch contains either a gold or silver coin, with equal probability of each. You win if you can correctly guess the color of the coin. While you are contemplating your choice, the game operator says, "Let me help you." He drops in two gold coins, shakes the pouch, then reaches in and draws out two gold coins. What is the probability the remaining coin is gold? (Variant of one of Lewis' Carroll's famous "*Pillow Problems*".)

Historical Note: Problem (5C083) given here is a variant of problem #5 in Charles L. Dodgson's book *Pillow Problems Thought Out During Wakeful Hours*. His book is a collection of $72$ problems and was first published in 1895 as Part 2 of his book *Curiosa Mathematica*. Later editions were published under his more familiar pen name Lewis Carroll.

The book is still available as a Dover edition but you can also get a free copy in many places on the web. Problems 5,10,16,19,23,27,38,41,45,50, 58 and 66 all involve probability. The last problem in the book, No. 72, is a tongue-in-check situation where Dodgson "proves" that if two balls are in a bag and if each is equally likely to be black or white and if their colors are determined independently then it will always be the case that one is black and the other is white. It is a problem of the same type you can find in problem books where you are given a proof that $1 = 2$ and the question is to spot the mistake in the proof. His Problem 72 is famous because the mistake in the proof is very well hidden.

6. (5C063) A bag contains $4$ red and $8$ white marbles, well mixed. One marble is removed and replaced by two marbles of the other color. Again, after thorough mixing, a marble is draw. What is the probability that this last marble to be removed is red? [original source AHSME 1995 #20]

7. (MB062) A bag contains ten balls numbered $1,2, \dots ,10$, that are thoroughly mixed. Alice draws a ball at random, looks at the number, and replaces the ball in the bag, which is then mixed again. Beth then draws a ball at random. What is the probability that the girls will have drawn the same number?

8. (5C991) Cards are turned over one at a time (without replacement) from an ordinary well-shuffled deck (52 cards). What is the probability that the first two cards are aces? (There are four aces in a deck.)

9. (5D942) A bag of popping corn contains $2/3$ white kernels and $1/3$ yellow kernels. Only $1/2$ of the white will pop, whereas $2/3$ of the yellow will pop. After the corn is popped, you select sight unseen a single popped piece from a well-mixed bag. What is the probability that it will be yellow?

10. (5C902) There are $4$ black socks and $6$ brown socks all mixed together in a drawer in a dark room. If Pat grabs two socks from the drawer and takes them into the light, what is the probability that they will match?

11. (5T894) Assume that a certain test for cancer is $98\%$ accurate, by which we mean that if the test is administered to $100$ people with cancer, it will detect the cancer in $98$ of them; and if it is administered to $100$ people without cancer, it will show $98$ of them to be free of cancer. If this test is given to $10{,}000$ people $(1/2)\%$ of whom actually have cancer, and if Mr. Casetest is one of the people who is told that he has cancer, what is the probability that Mr. Casetest really does have cancer? [Adapted from <u>Innumeracy</u> by John Paulas.]

12. Dodgson's *Pillow Problem* #16.

    Imagine two urns where the first urn is known to contain jone ball and that ball is equally likely to be either black or white. The second urn is known to contain three balls, two black and one white. A white ball is added to Urn 1 and afterwards a ball is randomly selected and removed from Urn 1. This ball is seen to be white. This ball is not replaced back into Urn 1. After this, an urn is randomly selected from these two urns and a ball is randomly selected from that urn. What is the probability that this last selected ball is white?

13. (Freund) Consider a deck of $4$ cards consisting of the Ace of Hearts, Ace of Diamonds, King of Hearts, and the Queen of Hearts. Suppose these $4$ cards are shuffled and then you draw two cards from this deck and set them apart. Consider the four different scenarios.

    (a) Your friend looks at one of the two cards you selected and then tells you that your selection contains the Ace of Hearts.

(b) Your friend looks at one of the two cards you selected and then tells you that your selection contains an Ace.
(c) Your friend looks at both of the cards you selected and then tells you that your selection contains the Ace of Hearts.
(d) Your friend looks at both of the cards you selected and then tells you that your selection contains an Ace.

In all four scenarios, find the conditional probability that you selected both Aces.

14. A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the information that $85\%$ of the cabs in the city are Green and $15\%$ are Blue. A witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors $80\%$ of the time and failed $20\%$ of the time.

    What is the probability that the cab involved in the accident was Blue rather than Green? Surprisingly, in spite of the witness testimony, the hit-and-run cab is more likely to be Green than Blue. (Source: A. Tversky, D. Kahneman, <u>Evidential impact of base rates</u>, in <u>*Judgement under uncertainty: Heuristics and biases*</u>, D. Kahneman, P. Slovic, A. Tversky (editors), Cambridge University Press, 1982.)

15. The probability that $A$ can solve a certain problem is $2/5$ and that $B$ can solve it is $1/3$. If they both try it, independently, what is the probability that it is solved? (Source: *Probability*, James R. Gray, 1967, Page 33, Problem 22)

16. A man draws a card at random from a pack of $52$ playing cards. He then draws as many cards from the remainder of the pack as the number on the card already drawn, ace counting one and king, queen and jack each counting ten. What is the probability that the ace of spades is among the cards drawn? (Source: *Probability*, James R. Gray, 1967, Page 37, Problem 38)

17. A man chooses a painting from a group containing eight originals and two copies. He consults an expert whose chance of judging either an original or a copy correctly is $5/6$.

    (a) If the expert considers that the chosen painting is an original, what is the probability that this is so?

    (b) If the expert considers the painting is a copy and the man returns it and choses another painting at random from the other nine, what is the probability that this second painting is an original? (Source: *Probability*, James R. Gray, 1967, Page 39, Problem 45)

18. One hundred bags each contain two balls. In $99$ bags one ball is white and one ball is black; the hundredth bag contains two white balls. One bag is selected at random a ball is drawn and a ball is drawn at random from it.

    (a) What is the probability that this ball is white?

    (b) If the ball selected was found to be white, what is the probability that the other ball in that bag is also white?

    (c) If the selected white ball was replaced and a second ball drawn at random from the same bag, what is the probability that this second ball is white?

    (d) If the procedure is repeated and it is found that $n$ balls drawn at random and with replacement from the same bag were all white, what is the probability that both balls in the selected bag are white?

    (e) Find the smallest value of $n$ for which this exceeds $0.95$. (Source: *Probability*, James R. Gray, 1967, Page 40, Problem 48)

19. $A$ speaks truth $3$ times out of $4$, and $B$ $7$ times out of $10$; they both assert that a white ball has been drawn from a bag containing $6$ balls all of different colours: find the probability of the truth of the assertion. (Source: *Higher Algebra*, Hall and Knight, Fourth Edition, Macmillan and Co. (1891), Section 478, example, pages 397-398.)

20. **Bertrand's Box Problem** There are three boxes. One box contains two gold coins, one box contains one gold and one silver coin, and one box contains two silver coins. You picked a box at random and then selected a coin at random from that box which turns out to be gold. Find the probability that the second coin in the box selected is also gold.
    Source: https://en.wikipedia.org/wiki/Bertrand%27s_box_paradox

21. The Miracle Marble Manufacturing Company manufactures orange marbles and purple marbles. A bag of their marbles may contain any combination of orange and purple marbles (including all orange or all purple) and all combinations are equally likely.

    Henry bought a bag of their marbles and pulled one out at random. It was purple. What is the probability that if he pulled out a second marble at random it would also be purple?

    (Source: *Jim Totten's Problems of the Week*, Problem 318, Editors John McLoughlin, Joseph Khoury, Bruce Shawyer, pages 270-271.)

# Project 4  Painted Cubes Problem

Suppose you have 27 unpainted cubes stacked as shown to make one big $3 \times 3 \times 3$ cube (Figure A). The exterior of Figure A is painted but none of the interior sides are painted (Figure B).

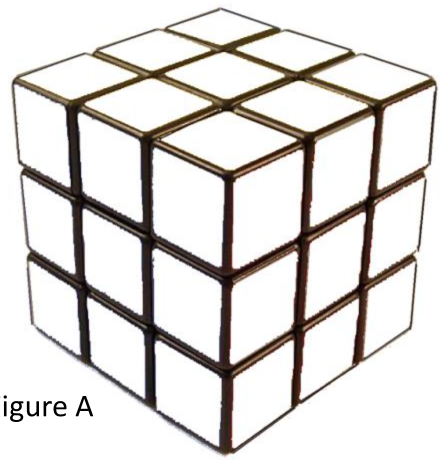


Figure A

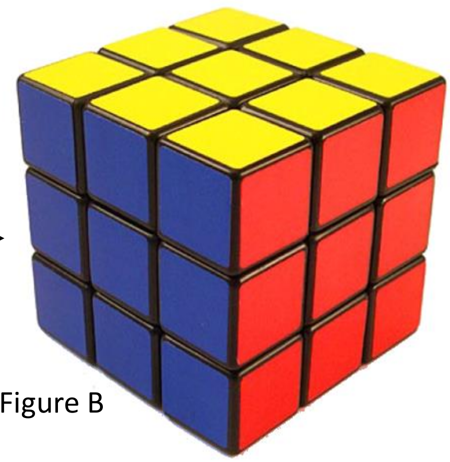


Figure B

Now suppose you separate these 27 small cubes, put them in a bag and randomly select one. Furthermore, suppose you don't actually get to see your selected cube until after you have placed it on the table in front of you.

If none of the five sides of this small cube that you can see without picking it up off the table have paint on them,

*what is the probability that the bottom side of this cube which is face own on the table is painted?*

The earliest source found for this problem is a 2013 paper by Peter Zoogman ([https://www.cfa.harvard.edu/~pzoogman/files/Bayes2.pdf](https://www.cfa.harvard.edu/~pzoogman/files/Bayes2.pdf)). It has appeared on numerous puzzle blogs since then.

The puzzle arises when solvers forget to **account for every random aspect of the problem in the sample space**.

There are <u>two</u> random parts to this problem. (It is the second part that is sometimes overlooked.)

(1) which cube you pick out of the bag

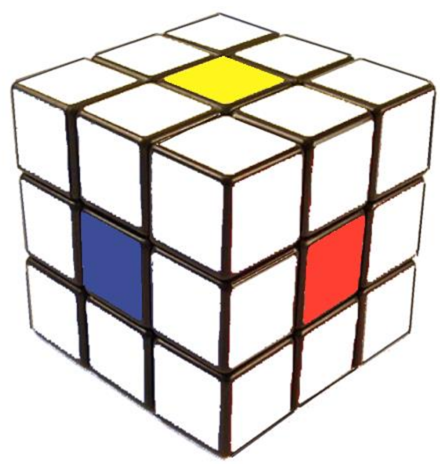
(2) which face of the cube you picked in Step (1) goes face down on the table.

There are $27$ cubes and $6$ faces per cube. All $27 \times 6$ (cube, face down-side) pairs are given to be equally likely.

From the information given we know that the cube selected from the bag either

has 5 unpainted sides and the single painted side was placed face down on the table
or
has 6 unpainted sides and any of its six sides was placed face down on the table.

Of the $27$ cubes, exactly six have only one painted face. Namely the six center exterior cubes, which includes the blue, yellow and red cubes shown below.



Only one of the $27$ cubes has no painted faces (the centermost cube which has no exposed faces).

From the information given we know that the cube selected from the bag has to be one of these $6 + 1 = 7$ cubes.

How many of the $27 \times 6$ (cube, face down-side) pairs are still possible in our conditional sample space (where the unexposed bottom face is the only possible painted face)?

For the six cubes with one painted face, the painted face would have to be on the table. So, each of these six cubes contributes just one (cube, face down-side) pair which is still possible.

For the one cube with no painted faces, any of its six sides could be the side face-down on the table. So, this cube contributes six (cube, face down-side) pairs which are still possible.

Of the original $27 \times 6$ equally like (cube, face down-side) pairs, only $6 + 6 = 12$ are still possible with the given information. Because the original $27 \times 6$ (unconditional) outcomes were equally likely, the remaining $12$ outcomes must still be equally likely. Hence each of these $12$ outcomes has probability $1/12$.

Of these $12$ outcomes, $6$ have a painted bottom side. Hence the conditional probability that the unseen bottom side is painted is $6/12 = 1/2$.

∎

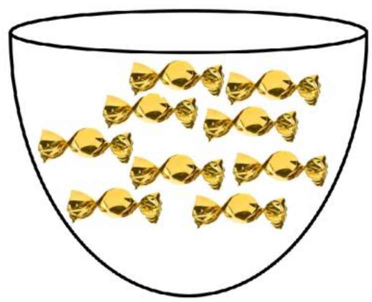
## Chapter 5  Exchangeable Random Variables

A box contains $3$ red and $5$ white balls. You reach into the box and randomly draw out a ball. What is the probability that it is red? We can all agree it is just $3/8$.

Now suppose we draw out a second ball, without replacing the first one. What is the probability that the second ball is red *taking into account that the first ball was red*. Obviously, there are now $7$ balls left and $2$ of them are red. So, the probability must be $2/7$.

But can we determine the probability that the second ball is red *without knowing the color of the first ball*? Is there a single answer?

Suppose that all eight balls are wrapped like a piece of candy so that the color of a ball cannot be determined just by inspection.



Now suppose we randomly draw out a wrapped ball from the initial distribution of $3$ red and $5$ balls and set it aside without unwrapping it. Then suppose we randomly draw out a second wrapped ball. Can we find the probability this second ball is red without knowing the color of the first ball?

To be certain, there is a unique answer to this question. If we repeated this experiment many, many times the percentage of times the second ball is red would begin to converge to a single number and that number is the probability the second ball is red *without taking into account the color of the first ball*.

How can we find this number using the rules of probability? Recall the formula

$$P(B) = P(B|A)P(A) + P(B|A')P(A')$$

which is valid for all generic events $A$ and $B$. Let $B$ represent the event that the second ball drawn is red and let $A$ represent the event that the first ball drawn is red. In this case, $A'$ represents the event that the first ball drawn is <u>not</u> red or equivalently that the first ball drawn is white.

Now we are in position to find $P(B)$, the probability that the second ball is red unconditionally. That is *without any knowledge of the color of the first ball*. Plugging into the above formula we have,

$$P(\text{second ball drawn out is red})$$

$$= P(\text{second ball red}|\text{first ball red})P(\text{first ball red})$$

$$+ P(\text{second ball red}|\text{first ball white})P(\text{first ball white})$$

$$= \left(\frac{2}{7}\right)\left(\frac{3}{8}\right) + \left(\frac{3}{7}\right)\left(\frac{5}{8}\right) = \frac{3(2+5)}{7 \cdot 8} = \frac{3}{8}.$$

Notice what we have just shown. When we do not take into account the color of the first ball,

$$P(\text{first ball drawn out is red}) = P(\text{second ball drawn out is red}).$$

Was this what you expected? Is it intuitive to you? Would this equality hold on the next draw as well?

In fact, it does. $P(i^{th}$ ball drawn out is red$) = 3/8$ for all $i = 1,2, \dots ,8$. Let's verify this mathematically for just one more step.

Let $A_r$ $(A_w)$ be the event that the first ball is red (white). Let $B_r$ $(B_w)$ be the event that the second ball is red (white) .

Let $D$ be the event that the third ball drawn out is red and define the following four events.

$$\begin{aligned} C_1 &= A_r \cap B_r \\ C_2 &= A_r \cap B_w \\ C_3 &= A_w \cap B_r \\ C_4 &= A_w \cap B_w \end{aligned}$$

Notice that $P(C_1 \cup C_2 \cup C_3 \cup C_4) = 1$ and $P\left(C_i \cap C_j\right) = 0$ for all $i \neq j$. Hence, it follows from the rule of total probability that

$$P(D) = P(D \cap C_1) + P(D \cap C_2) + P(D \cap C_3) + P(D \cap C_4)$$

and by substitution

$$P(D) = P(D \cap A_r \cap B_r) + P(D \cap A_r \cap B_w) + P(D \cap A_w \cap B_r) + P(D \cap A_w \cap B_w).$$

Now using the multiplication rule,

$$\begin{aligned} P(D) = &P\left(D\middle|(B_r \cap A_r)\right)P(B_r|A_r)P(A_r) \\ &+ P\left(D\middle|(B_w \cap A_r)\right)P(B_w|A_r)P(A_r) \\ &+ P\left(D\middle|(B_r \cap A_w)\right)P(B_r|A_w)P(A_w) \end{aligned}$$

$$+ P\big(D\big|(B_w \cap A_w)\big)P(B_w|A_w)P(A_w).$$

From here we get

$$P(D) = \left(\frac{1}{6}\right)\left(\frac{2}{7}\right)\left(\frac{3}{8}\right) + \left(\frac{2}{6}\right)\left(\frac{5}{7}\right)\left(\frac{3}{8}\right) + \left(\frac{2}{6}\right)\left(\frac{3}{7}\right)\left(\frac{5}{8}\right) + \left(\frac{3}{6}\right)\left(\frac{4}{7}\right)\left(\frac{5}{8}\right)$$
$$= \frac{6 + 30 + 30 + 60}{6(7)(8)} = \frac{126}{6(7)(8)} = \frac{21}{7(8)} = \frac{3}{8}.$$

As predicted, we have found that the unconditional probability that the third ball is red equals the unconditional probability that the second ball is red which also equals the unconditional probability that the first ball is red.

But what is the intuition behind this and why is this important? First, the intuition behind this result. Imagine we shook the balls around in the box, dumped them onto a table and lined them up.

Without any loss of generality, we can associate the first ball drawn with the ball on the far left. The second ball drawn is then associated with the ball immediately to its right. And so on. Is there anything in this process that would make a red ball more likely to land in one position in this line over another? No. But this means that for all $i = 1,2, \dots ,8$

$$P(i^{th} \text{ ball drawn out is red}) = P(\text{a red ball lands in the } i^{th} \text{ position}) = 3/8.$$

We can extend this to unconditional problems involving multiple balls. For example, suppose we draw out all eight balls from this box of 3 red and 5 white balls one after the other and without replacement.

What is the probability that the first ball drawn is red and the last ball drawn is white?

In terms of the equivalent row model described above, this probability about balls in a box is equal to the probability that the first ball in the row is red and the last ball in the row is white.

But this must be the same as the probability the first ball in the row is red and the second ball in the row is white which in turn must be the same as the probability that the first ball drawn out of the urn is red and the second ball drawn out of the urn is white. In terms of our earlier notation this is just

$$P(A_r \cap B_w) = P(B_w|A_r)P(A_r) = \left(\frac{5}{7}\right)\left(\frac{3}{8}\right).$$

The formal word for this type of symmetry among events is *exchangeability*.

**It is important to not get confused between *exchangeable* events and *independent* events.**

Recall that events $A$ and $B$ are independent if $P(B|A) = P(A)$. But while the events $A$: the first ball is white and $B$: the second is red are exchangeable, they are not independent. More specifically,

$$P(\text{first ball is white and the second ball is red})$$

$$= P(A_w \cap B_r) = P(B_r|A_w)P(A_w) = \left(\frac{3}{7}\right)\left(\frac{5}{8}\right) = \frac{15}{56}$$

and

$$P(\text{first ball is red and the second ball is white})$$

$$= P(A_r \cap B_w) = P(B_w|A_r)P(A_r) = \left(\frac{5}{7}\right)\left(\frac{3}{8}\right) = \frac{15}{56}.$$

These probabilities agree so these events **are exchangeable**.

Now check for independence. We see that

$$P(\text{second ball is red}|\text{first ball is white}) = \frac{5}{7}$$

but as we determined above

$$P(\text{second ball is red}) = \frac{3}{8}.$$

These probabilities do not agree so these events **are <u>not</u> independent**.

**Drawing Straws**. Siblings Adam, Bob and Carrie all want to have the first chance to play a new video game they bought together. Their mom cuts a straw into three pieces of slightly different lengths and holds them together in such a way that all three pieces appear to be the same length. They agree that whoever draws the longest straw gets to go first. Adam, Bob and Carrie, in that order, each take a straw from their mom's hand. Is this a fair method for all three siblings?

It is. We could model this as a box containing one red and two white balls. Drawing the red ball would correspond to drawing the longest straw. But we know from exchangeability that for each of $i = 1,2,3$, the $P(i^{th}$ ball drawn is red$) = 1/3$.

That is, Adam, Bob and Carrie have an equal probability of drawing the longest straw.

## Exercises for Chapter 5

1. (5C013) A box contains exactly seven chips, four red and three white. Chips are randomly drawn one at a time without replacement until all the red chips are drawn or all the white chips are drawn. What is the probability that the last chip drawn is white?

   [Variation, AMC, 2001, No. 11]

2. (TC013) A bag contains $5$ red balls and $4$ white ones. They are to be drawn one at a time without replacement until all the red balls are drawn, or all the white ones are drawn. What is the probability that the last ball drawn is white?

3. A box contains $6$ red balls, $7$ green balls, and $9$ yellow balls. Eleven balls are chosen at random one after the other without replacement. What is the probability that the 3rd ball chosen was yellow given that the 9th ball chosen was green?

4. A box contains $a$ white balls and $b$ black balls. Balls are drawn from the box at random and without replacement. What is the probability that all black balls are drawn before the last white ball?

5. You play a game with a friend where you each start with a shuffled deck of $52$ cards turned face down. At each turn you both flip over the top card in your deck. If the cards are the same color, you win that turn. If the cards are not the same color, your friend wins that turn. On the next turn you repeat the process but with just the cards remaining in your decks. What is the probability that you win the fifth turn (the cards are both red or both black on the fifth turn) of this game?

6. Four balls are drawn at random and without replacement from a box containing $10$ red balls and $8$ white balls and then discarded without noting the color of any of these four balls. At this point four more balls are drawn out at random and without replacement. What is the probability that exactly two of this second set of four balls are red?

## Project 5  Shout Stop Whenever You're Ready

**(Part I of Shouting Stop)**

Back in the mid 1980's there was a popular TV game show called *Press Your Luck*.  Three contestants faced a large electronic monitor made up of eighteen squares.  One square at a time would be lit up (as the $6000 + One Spin in the figure below).  A square would only stay lit for less than half a second and then a different square would light up, seemingly in some random order.  Additionally, where the prizes and the three "whammies" (the grinning devils as seen below) were located on the monitor would change just as quickly and again in what seemed to be in some random order.


The three contestant rotated turns and on a contestant's turn they would shout "Stop" while simultaneously hitting a big button to freeze the monitor when they felt it to be a lucky moment.  At that point, if the lit square the monitor stopped on was a prize you won that prize.  Then you had the option to "press your luck" and take another turn or to pass the turn to the next contestant.  You could keep pressing your luck and continue to take another turn as long as your wanted.  However, if the monitor stopped on one "whammies" you would lose your turn and your prizes. There were other rules to add to the excitement but that was the basic setup.

If the game was truly as random as it appeared then a contestant had a 1 in 6 chance of getting a "whammy" on any turn. Now it really happened that a contestant managed to avoid getting a "whammy" on 45 consecutive turns. How unlikely was that? (You can google "Press Your Luck" to get the details on this contestant.)

**(Part II of Shouting Stop)**

Now consider a different game altogether that starts with a deck of 52 cards placed face down on a table. You can assume the cards are well shuffled (*i.e.* in a random order). In this game cards are taken off the top of the deck one at a time and turned over so you can see what that card is. This revealed card is <u>not</u> returned to the deck. Then the next card from the deck is taken off the top and turned over so you can see what that card is. The game continues like this with the revealed cards <u>not</u> being returned to the deck.

Now, at any point in this game you can yell "Stop". At that point, one more card is taken off the top of the remaining deck and is turned over. If that card is a red card (a diamond or heart), you are a winner*!*

*What would be your strategy for deciding when to yell "Stop" and what would be your probability of winning with this strategy?*

To be clear, you <u>are</u> allowed in this game to modify your strategy at any point in the game based on the cards you see turning up before you yell "stop".

Consider doing a small-scale simulation of this problem by starting with just $3$ cards of each suit instead of the full $13$. Go through $n = 50$ replications of a fixed strategy and see what percentage of the time you win. Of course, the larger you make $n$ the more accurate your estimate will be of the true probability of winning with that strategy.