Ma 3/103 KC Border
Introduction to Probability and Statistics Winter 2021

Lecture 3: Counting and probability

Relevant textbook passages:

Pitman [10]: Sections 1.5–1.6, 2.1, 2.5; pp. 47–77; Appendix 1, pp. 507–514.

Larsen–Marx [9]: Sections 2.4, 2.5, 2.6, 2.7, pp. 67–101.

3.1 Laplace's model: Uniform probability on finite sets

Recall (Section 1.2) Laplace's [8, pp. 6–7] model of probability as a fraction whose number is the number of favorable cases and whose denominator is the number of all possible cases, where the cases are equally likely. We formalize this as follows. The **uniform probability** on a finite sample space Ω assigns equal probability to each outcome, and every subset of Ω is an event.

3.1.1 Theorem (Uniform probability) *With a* ***uniform probability***[1] *P on a finite set Ω, then for any subset E of Ω,*

$$P(E) = \frac{\# E}{\# \Omega}.$$

> Throughout this course and in daily life, if you come across the phrase **at random** and the sample space is finite, unless otherwise specified, you should assume the probability measure is uniform.

Note that in Laplace's probability model, the only event of probability zero is the empty set, and the only event of probability one is the entire sample space.

3.2 A taxonomy of classic experiments

Many random experiments can be reduced to one of a small number of classic experiments. This characterization is inspired by Ash [2].

- The first kind of random experiment is **sampling** from an urn (see Figure 3.1). In this kind of experiment an urn filled with balls of different colors, or labeled balls. A ball is selected at random (meaning each ball is equally probable).

◦ Note that *coin tossing* is reducible to sampling from an urn with an equal number of balls labeled Heads and Tails.

◦ *Rolling a die* can be viewed as sampling from an urn with balls labeled $1, \dots, 6$ (or of six different colors).

◦ *Dealing a card* is like sampling from an urn with balls labeled $A\clubsuit, A\heartsuit, \dots, K\spadesuit$.

◦ *Roulette* is akin to sampling from an urn with balls labeled $1, \dots, 36, 0, 00$.

For repeated sampling, there are two variations, **sampling with replacement** and **sampling without replacement**.

[1] Some of my colleagues reserve the term "uniform" to refer to a probability space where the sample space is an interval of real numbers, and the probability of a subinterval is proportional to its length. They need to expand their horizons.

Figure 3.1. The archetypal urn.

○ In sampling with replacement, after being drawn, the ball is replaced in the urn, and the contents are remixed. Repeated coin tossing or dice rolling is like sampling with replacement.

○ In sampling with replacement, after being drawn, the ball is discarded, and the composition of the urn has changed. Dealing a poker or bridge game is like sampling without replacement.

• Another kind of random experiment is a **matching** experiment. In this kind of experiment, a randomly selected ball is dropped into a randomly selected labeled bin.

○ A classic textbook question deals with randomly stuffing letters into addressed envelopes.

○ A classic paper [5] treats aerial bombardment like dropping balls into bins. We'll discuss it in Lecture 13.

○ Radioactive decay experiments can be viewed as having bins corresponding time intervals, into which balls symbolizing decays are placed. (This is one way to think about Poisson processes, to be discussed in Lecture 14.)

• Another kind of experiment is **waiting** for something special to occur in another experiment.

○ For instance, we might want to know how long it will take to go broke playing roulette.

○ Or how long between radioactive decays or Southern California earthquakes.

We can also categorize some of the calculations we want to do in connection with these experiments as to whether **order matters** or **order doesn't matter.** But whatever the experiment or type of result we are interested in, remember **Laplace's maxim**.

> To calculate the probability of the event E, when the experimental outcomes are all "equally likely," simply count the number of outcomes that belong to E and divide by the total number of outcomes in the outcome space Ω.

3.2.1 Remark In sampling balls from urns note that all probabilities derived from Laplace's maxim are rational numbers.

3.3 Repeated sampling with replacement

3.3.1 How many different outcomes are there for the experiment of tossing a coin n times?

This is equivalent to repeated sampling with replacement of drawing from an urn with two balls, labeled H and T. Here an *outcome* is the ordered list of T's and H's, so the sample space is $\Omega = \{H, T\}^n$, and the number of possible outcomes is

$$2^n.$$

3.3.2 More general sampling with replacement

For repeated sampling with replacement n times from an urn with m types of balls, the sample space is the set of distinct sequences of draws. Each draw has m possibilities, so there are m choices for the first draw, m for the second draw, etc., so the sample space is $\Omega = \{1, \dots, m\}^n$, so the number of possible outcomes is

$$\underbrace{m \times \cdots \times m}_{n \text{ terms}} = m^n.$$

3.4 Repeated sampling without replacement

Suppose I now have an urn with n distinct objects, (or a deck of cards), and I draw one object at a time, until there are none left. This is **sampling without replacement**. This gives an ordered *list* of length n. How many such lists are there?

3.4.1 Number of lists of length n

Well there are n possibilities for the first draw, and once one has been drawn, there are $n-1$ left, so there are $n-1$ possibilities for the second draw. The with two gone, there are $n-2$ possibilities for the third draw, etc. The total number of lists is gotten by multiplying these, so we have the following.

3.4.1 Proposition *There are*

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$$

distinct lists of n objects.

The number $n!$ is read as "***n* factorial**."

By definition,

$$0! = 1,$$

and we have the following recursion

$$n! = n \cdot (n-1)! \qquad (n > 0).$$

By convention, if $n < 0$, then $n! = 0$.

3.4.2 Example (Number of lists of 4 objects) Suppose the objects are a, b, c, d. The above calculation says that there should be $4 \cdot 3 \cdot 2 \cdot 1 = 24$ distinct lists. Here they are (in alphabetical order):

$$abcd, abdc, acbd, acdb, adbc, adcd,$$
$$bacd, badc, bcad, bcda, bdac, bdca,$$
$$cabd, cadb, cbad, cbda, cdab, csba,$$
$$cabc, dacb, dbac, dbca, dcab, dcba.$$

You can see that each row corresponds to a different first element in the list, etc. All of these lists have the same four letters, but they are different lists because **order matters**. □

3.4.2 Number of lists of length k of n objects

Suppose when sampling without replacement, I do not take all n of the objects out of the urn, but only k of them. How many distinct lists of length k can I make with n objects? As before, there are n choices of the first position on the lists, and then $n-1$ choices for the second position, etc., down to $n-(k-1) = n-k+1$ choices for the k^{th} position on the list. Thus there are

$$\underbrace{n \times (n-1) \times \cdots \times (n-k+1)}_{k \text{ terms}}$$

distinct lists of k items chosen from n items. There is a more compact way to write this. Observe that

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

3.4.3 Proposition *There are*

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

distinct lists of length k chosen from n objects.

We may write this as $(n)_k$, which is read as "***n* order *k***." Note that when $k = n$ this reduces to $n!$ (since $0! = 1$), which agrees with the result in the previous section. When $k = 0$ this reduces to 1, which makes sense since there is exactly one list of 0 objects, namely, the empty list.

3.4.4 Example (Number of lists of length 2 from 4 objects) Suppose the objects are a, b, c, d, but I just want to make lists of length 2. The above calculation says that there should be $4 \cdot 3 = 4!/2! = 12$ distinct lists. Here they are (in alphabetical order):

$$ab, ac, ad,$$
$$ba, bc, bd,$$
$$ca, cb, cd,$$
$$da, db, dc.$$

You can see that each row corresponds to a different first element in the list, etc. The lists ab and ba both appear, because again, **order matters**. □

3.5 Lists versus sets

It very useful to distinguish **lists** and **sets**. Both are collections of n objects, but two lists are different unless each object appears in the same *position* in both lists.

> For lists, **order matters**.
> For sets, **order does not matter**.

For instance,

abcd and *dcba* are distinct lists of four elements, but they comprise the same set.

A list is sometimes referred to as a **permutation** and a set is often referred to as **combination**.

3.5.1 Number of subsets of size k of n objects

How many distinct subsets of size k can I make with n objects (without repetition, that is sampling without replacement)? (A subset is usually referred to as a **combination** of elements. I find that terminology uninformative at best and misleading at worst.)

Well there are $\frac{n!}{(n-k)!}$ distinct lists of length k chosen from n objects. But when I have a set of k objects, I can write it $k!$ different ways as a list. Thus each set appears $k!$ times in my listing of lists. So I have to take the number above and divide it by $k!$ to get the number of sets.

3.5.1 Proposition *There are*

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

distinct subsets of size k chosen from n objects.

3.5.2 Definition *For natural numbers $0 \leqslant k \leqslant n$*

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

is read as

"n choose k"

For $k > n$ define $\binom{n}{k} = 0$.

It is the number of distinct subsets of size k chosen from a set with n elements. It is also known as the **binomial coefficient**. Note if $k > n$, there are no subsets of size k of a set of size n, so by convention we agree that in this case $\binom{n}{k} = 0$.

Other notations you may encounter include $C(n, k)$, ${}^nC_k$, and ${}_nC_k$. (These notations were easier to typeset in lines of text before the invention of computerized typesetting.)

3.5.3 Example (Number of sets of size 2 from 4 objects) Going back to Example 3.4.4 with four items, a, b, c, d, we had twelve lists of length 2, but if you look at the enumeration, you will find each 2-element set, such as ab at least twice, once as ab and again ba. Thus there are $2 = \binom{2}{1}$ times as many lists as sets. The collection of $\binom{4}{2}$ sets of size 2 is

$$\{ab\}, \{ac\}, \{ad\}, \{bc\}, \{bd\}, \{cd\}.$$

□

3.5.2 Some useful identities

$$\binom{n}{0} = \binom{n}{n} = 1$$

$$\binom{n}{1} = n$$

$$\binom{n}{k} = \binom{n}{n-k}$$

$$\binom{n+1}{k+1} = \binom{n}{k+1} + \binom{n}{k} \tag{1}$$

Here is a simple proof of (1): $\binom{n+1}{k+1}$ is the number of subsets of size $k+1$ of a set A with $n+1$ elements. So fix some element $\bar{a} \in A$ and put $B = A \setminus \{\bar{a}\}$. If E is a subset of A of size $k+1$, then either (i) $E \subset B$, or else (ii) E consists of $\bar{a}$ and k elements of B . (A subset may not satisfy both (i) and (ii).) There are $\binom{n}{k+1}$ subsets E satisfying (i), and $\binom{n}{k}$ subsets satisfying (ii).

Equation (1) gives rise to **Pascal's Triangle**, which gives $\binom{n}{k}$ as the k^{th} entry of the n^{th} row (where the numbering starts with $n = 0$ and $k = 0$). Each entry is the sum of the two (or one) entries diagonally above it:

$$\begin{array}{ccccccccccccc}
&&&&&& \binom{0}{0} &&&&&& \\
&&&&& \binom{1}{0} && \binom{1}{1} &&&&& \\
&&&& \binom{2}{0} && \binom{2}{1} && \binom{2}{2} &&&& \\
&&& \binom{3}{0} && \binom{3}{1} && \binom{3}{2} && \binom{3}{4} &&& \\
&& \binom{4}{0} && \binom{4}{1} && \binom{4}{2} && \binom{4}{3} && \binom{4}{4} && \\
& \binom{5}{0} && \binom{5}{1} && \binom{5}{2} && \binom{5}{3} && \binom{5}{4} && \binom{5}{5} & \\
\binom{6}{0} && \binom{6}{1} && \binom{6}{2} && \binom{6}{3} && \binom{6}{4} && \binom{6}{5} && \binom{6}{6} \\
&&&&&& \text{etc.} &&&&&&
\end{array}
=
\begin{array}{ccccccccccccc}
&&&&&& 1 &&&&&& \\
&&&&& 1 && 1 &&&&& \\
&&&& 1 && 2 && 1 &&&& \\
&&& 1 && 3 && 3 && 1 &&& \\
&& 1 && 4 && 6 && 4 && 1 && \\
& 1 && 5 && 10 && 10 && 5 && 1 & \\
1 && 6 && 15 && 20 && 15 && 6 && 1 \\
&&&&&& \text{etc.} &&&&&&
\end{array}$$

Equation (1) also implies (by the telescoping method) that

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \cdots + (-1)^k \binom{n}{k} = (-1)^k \binom{n-1}{k}.$$

3.5.3 Number of all subsets of a set

You should already know the following,

There are 2^n distinct subsets of a set of n objects.

The set of subsets of a set is known as its **power set**. Let $c(n)$ denote the cardinality of the power set of a set with n elements. Then it is easy to see that $c(0) = 1$, $c(1) = 2$. More generally, $c(n+1) = 2c(n)$: There are two kinds of subsets of $\{x_1, \dots, x_{n+1}\}$, those that are subsets of $\{x_1, \dots, x_n\}$ and those of the form $A \cup \{x_{n+1}\}$ where A is a subset of $\{x_1, \dots, x_n\}$. So $c(n) = 2^n$.

3.5.4 And so ...

If we sum the number of sets of size k from 0 to n, we get the total number of subsets, so

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

This is a special case of the following result, which you may remember from high school. (The special case is $a = b = 1$.)

3.5.4 Binomial Theorem

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

3.5.5 Binomial random variables

The number of Heads in n independent tosses of a fair coin is a simple, yet nontrivial example, of a random variable, called a **binomial random variable**. The distribution of a binomial random variable is called a **binomial distribution**. To find its probability mass function, weed to calculate the probability that there are k Heads in n independent coin tosses.

Let's do this carefully. The sample space Ω is the set of sequences $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)$ of length n where each term ω_i in the sequence is H or T, that is, $\Omega = \{H, T\}^n$. For each point $\boldsymbol{\omega} \in \Omega$, let $A_{\boldsymbol{\omega}} = \{i : \omega_i = H\}$. Since there are only two outcomes, if you know $A_{\boldsymbol{\omega}}$, you know $\boldsymbol{\omega}$ and vice versa.

Now let E be any subset of Ω that has exactly k elements. There is exactly one point $\boldsymbol{\omega} \in \Omega$ such that $A_{\boldsymbol{\omega}} = E$. Thus the number of elements of Ω such that $\# A_{\boldsymbol{\omega}} = k$ is precisely the same as the number of subsets of Ω of size k, namely $\binom{n}{k}$. Thus

$$\text{Prob(exactly } k \text{ Heads)} = \frac{\# \{\boldsymbol{\omega} \in \Omega : \# A_{\boldsymbol{\omega}} = k\}}{\# \Omega} = \frac{\binom{n}{k}}{2^n} = \frac{n!}{k!(n-k)!2^n}.$$

Here is an example with $n = 3$:

$\boldsymbol{\omega}$	$A_{\boldsymbol{\omega}}$
HHH	$\{1,2,3\}$
HHT	$\{1,2\}$
HTH	$\{1,3\}$
HTT	$\{1\}$
THH	$\{2,3\}$
THT	$\{2\}$
TTH	$\{3\}$
TTT	$\varnothing$

For $k = 2$, the set of points $\omega \in \Omega$ with exactly two heads is the set $\{HHT, HTH, THH\}$, which has $3 = \binom{3}{2}$ elements, and probability $3/8$.

We can use Pascal's Triangle to write down these probabilities.

$$
\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & \frac{1}{2} & & \frac{1}{2} & & & \\
 & & \frac{1}{4} & & \frac{2}{4} & & \frac{1}{4} & & \\
 & \frac{1}{8} & & \frac{3}{8} & & \frac{3}{8} & & \frac{1}{8} & \\
\frac{1}{16} & & \frac{4}{16} & & \frac{6}{16} & & \frac{4}{16} & & \frac{1}{16}
\end{array}
$$

(Prob of 0 Heads in 0 tosses)
(Prob of 0, 1 Heads in 1 toss)
(Prob of 0, 1, 2 Heads in 2 tosses)
(Prob of 0, 1, 2, 3 Heads in 3 tosses)
(Prob of 0, ..., 4 Heads in 4 tosses)

etc.

3.6 Some examples from card games

3.6.1 How many ways can a standard deck of 52 cards be arranged?

Here the order matters, so we want the number of lists, which is

$$52! \approx 8.06582 \times 10^{67}$$

or more precisely:

$$80,658,175,170,943,878,571,660,636,856,403,766,975,289,505,440,883,277,824,000,000,000,000.$$

This is an astronomically large number. In fact, since the universe is about 10–20 billion years old and there are about 7.28 billion people (according to Siri), if every person on the planet were set to work arranging a deck in a given order, and could do so in one second, it would take about 2×10^{40} lifetimes (to date) of the universe to go through all the possible arrangements of the deck. Nevertheless for this course we make the usual assumption that after shuffling the deck a few times all possible arrangements are equally likely. This is ludicrous. But the assumption may actually give reasonably good results for typical questions we ask about card games, such as those that follow.

For more about the distribution of cards after shuffling, see the papers by Bayer and Diaconis [4] and Assaf, Diaconis, and Soundararajan [3]. A rule of thumb is that it takes at least seven riffle shuffles for the deck to sufficiently mixed up to be able to use the model that all orders of the cards are equally likely—provided what you want to do is to calculate the probabilities of events typically associated with card games.

It is important that when shuffling the deck the deck, the shuffles have some “noise” in them. A **perfect shuffle** is one where the deck is split perfectly in half, and the cards from each half are perfectly alternately interleaved. There are actually two kinds of perfect shuffles—one in which the top of the deck remains the top, and one in which the top card becomes the second card. The problem with perfect shuffles is that the order of cards is known. In fact after eight perfect shuffles fixing the top card, the deck is in the same order as it started. If you can perform perfect shuffles and have an amazing memory (and I have met such people), then you can astonish your friends and family by announcing what the sixteenth card in the deck is.

3.6.2 How many different five-card poker hands are there?

In standard **five-card draw poker**, each player buys into the game by contributing a fixed **ante** to the **pot**. Then each player is dealt five cards before any betting occurs. A round of betting then ensues, after which the players who have not **folded** (dropped out) may discard some of the cards and **draw** fresh ones. They end up with five cards, and more betting ensues. The order in which you receive your cards does not matter for the final outcome, only the set of cards in your hand. But it matters for the betting which cards you get before the draw, and which you get after the draw. In fact, the decision to discard some cards and draw new ones depends on the cards initially dealt, but not their order.

In various forms of **stud poker**, there is no draw, but there is betting before the players receive all their cards, so the order in which you receive cards may influence your bets. Moreover some kinds of stud poker give each player seven cards, from which they select their best five, so the probabilities of your hand being beaten are different from those in five-card poker.

And then there is **Texas hold'em**, which is actually televised, in which each player is dealt two cards and an additional five cards are common to all players. All the computations here are for five-card stud or for the initial hand in five-card draw.

There are

$$\binom{52}{5} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 2,598,960$$

distinct five-card hands.

3.6.3 How many different deals?

How many distinct *deals* of five-card draw poker hands for a seven-player game are there? (The order of hands matters to the betting, but the order of cards within hands does not.)

The number of distinct deals is

$$\underbrace{\binom{52}{5}\binom{47}{5}\binom{42}{5}\binom{37}{5}\binom{32}{5}\binom{27}{5}\binom{22}{5}}_{\text{7 terms}} \approx 6.3 \times 10^{38}.$$

Each succeeding hand has five fewer cards to choose from, the others being used by the earlier hands.

3.6.4 How many five-card poker hands are flushes?

To get a **flush** all five cards must be of the same **suit**. (The suits are clubs, diamonds, hearts, and spades, with sigils inspired by medieval weapons of war.) There are thirteen ranks in each suit, so there are $\binom{13}{5}$ distinct flushes from a given suit. There are four suits, so there are

$$4\binom{13}{5} = 5148 \text{ possible flushes.}$$

(This includes straight flushes.)

A **straight flush** is a flush in which the five cards have consecutive **ranks**. Counting the Ace as a high card, a straight flush may start on any of the nine numbers $2, 3, \ldots, 10$, so there are $4 \times 9 = 36$ possible straight flushes. (I include **royal flushes** as straight flushes. A royal flush is a straight flush with a 10, Jack, Queen, King, and Ace.) Some variants of the rules allow an Ace to be either high or low for the purposes of a straight, which adds another 4 straight flushes for a total of 40.

The probability of a straight flush (without low Aces) is

$$\frac{36}{2,598,000} = \frac{3}{216,580} \approx 0.000014.$$

Thus there are $5148 - 36 = 5112$ flushes that are not straight flushes (allowing for low Aces). So what is the **probability** of a flush that is not a straight flush?

$$\frac{5112}{2,598,960} \approx 0.0020$$

3.6.5 Probability of 4 of a kind

What is the number of (five-card) poker hands that have four of a kind (four cards of the same rank)? Well, there are 13 choices for the rank of the four-of-a-kind, and 48 choices for the fifth card, called the **kicker**, so there are 13×48 distinct hands with four of a kind. There are $\binom{52}{5}$ poker hands, so the probability of four of a kind is

$$\frac{13 \times 48}{\binom{52}{5}} = \frac{1}{4165} \approx 0.00024.$$

3.6.6 Probability of a full house

A **full house** is a poker hand with three cards of one rank and two cards of another rank. How many poker hands are full houses? Well, there are 13 choices for the rank of the three-of-a-kind, and 12 choices for the ranks of the pair, but there are 4 cards of any given rank, so there are $\binom{4}{3} = 4$ sets of three of a kind of a given rank. Like wise there are $\binom{4}{2} = 4$ pairs of a given ranks. So there are $\left(13 \times \binom{4}{3}\right) \times \left(12 \times \binom{4}{2}\right)$ distinct full houses. The probability of a full house is

$$\frac{13 \times \binom{4}{3} \times 12 \times \binom{4}{2}}{\binom{52}{5}} = \frac{6}{4165} \approx 0.00144 \approx 1/700.$$

3.6.7 Probability of a three-of-a-kind hand

A three-of-a-kind poker hand is a hand with three cards of one rank r_1, and two cards of two different ranks, r_2 and r_3, where r_1, r_2, and r_3 are distinct. How many poker hands are a three-of-a-kind?

Here are two ways to approach the problem.

1.) There are 13 ranks to choose from, and for each rank r_1 there are $\binom{4}{3} = 4$ ways to choose the suits for the three-of-a-kind. For each of these $13 \times 4 = 52$ choices, there are 49 cards left, from which we must choose the remaining two cards.

There are twelve remaining ranks and we must choose two distinct ranks—there are $\binom{12}{2}$ ways to do this. Given the choices for the ranks, there are 4 choices of suit each rank, so there are $\binom{12}{2} \times 4 \times 4$ ways to choose the remaining two cards. There are thus

$$\left(13 \times \binom{4}{3}\right) \times \left(\binom{12}{2} \times 4 \times 4\right) = 54,912$$

distinct three-of-a-kind hands, and the probability is

$$\frac{84,912}{\binom{52}{5}} = \frac{88}{4165} \approx 0.0211 \approx 1/48.$$

2.) Another way to reason about the problem is this. As before, there are 13 ranks to choose from, and for each rank r_1 there are $\binom{4}{3} = 4$ ways to choose the three-of-a-kind. For each of these $13 \times 4 = 52$ choices, there are 49 cards left, from which we must choose the remaining dyad (two cards, not necessarily of the same rank). There are $\binom{49}{2}$ ways to do this. But not all of $\binom{49}{2}$ these lead to three-of-a-kind. If one of these two has the same rank r_1 as our first triple, we end up with four-of-a kind, which is a stronger hand. How many ways can this happen? Well there is only one card of rank r_1 left in our remaining 49 and there are 48 not of rank r_1, so there are 48 ways to choose the remaining dyad to get four of a kind. Also, if the two remaining cards have the same rank, then we get a **full house**. There are 12 remaining ranks and for each one there are $\binom{4}{2}$ ways to choose two cards of that rank, and thus end up with a full house. So the number of three-of-a-kind hands is

$$52 \times \Bigg(\underbrace{\binom{49}{2}}_{\text{remaining dyads}} - \underbrace{1 \times 48}_{\text{fours-of-a-kind}} - \underbrace{12 \times \binom{4}{2}}_{\text{full houses}} \Bigg) = 54,912,$$

which agrees with the answer above.

3.6.8 Deals in bridge

In Contract Bridge, all fifty-two cards are dealt out to four players, so each player has thirteen. The first player can have any one of $\binom{52}{13}$ hands, so the second may have any of $\binom{39}{13}$ hands, the third may have any of $\binom{26}{13}$ hands, and the last player is stuck with the $\binom{13}{13} = 1$ hand left over.

Thus there are

$$\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13} \approx 5.36447 \times 10^{28}.$$

distinct *deals* in bridge.

After the deal there is a round of bidding, which results in one player becoming *declarer*. The players are divided into teams of two, and arranged around a four-sided table with sides labeled North, East, South, and West. The declarer sits at the South position, and their partner[2] sits at the North position. North's cards are displayed for all the players to see, but the other players are the only ones to see their own hands. South will makes all the plays for North, so North is known as the *dummy*. This gives the declarer and advantage because the declarer sees their cards plus the dummy's cards, so the declare knows which cards their team holds, and by elimination knows which 26 cards the opponents have, but not how they split up. By contrast, West or East does not know the cards held by the other player on their team.

3.6.9 Splits in bridge

Suppose the declarer's opponents have n Clubs between them. What is the probability that they are split k–$(n-k)$ between West and East? This is the probability that West (the player on declarer's left) has k of the n. East will have the remaining $n-k$.

There are $\binom{26}{13} = 10,400,600$ possible hands for West. In order for West's hand to have k Clubs, they must have one of the $\binom{n}{k}$ subsets of size k from the n Clubs. The remaining $13-k$ must be made up from the $26-n$ non-Clubs. There are $\binom{26-n}{13-k}$ possibilities. Thus there are

$$\binom{n}{k}\binom{26-n}{13-k}$$

hands in which West has k clubs, so the probability is

$$\frac{\binom{n}{k}\binom{26-n}{13-k}}{\binom{26}{13}}$$

that West has k clubs.

For the case $n = 3$ this is $11/100$ for $k = 0, 3$, and $39/100$ for $k = 1, 2$.

[2] Some pedants will claim that the use of *they* or *their* as an ungendered singular pronoun is a grammatical error. There is a convincing argument that those pedants are wrong. See, for instance, Huddleston and Pullum [7, pp. 103–105]. Moreover there is a great need for an ungendered singular pronoun, so I will use *they* in that role.

3.6.10 Aside: Some practical advice on gambling

Knowing the probabilities for card or dice games is useful, but probably not the *most* useful thing to know about gambling. For instance, Nelson Algren's three rules for life from his 1956 novel *A Walk on the Wild Side* start with, "Never play cards with a man called Doc." (wikipedia)

But seriously, you should know that it is never a good idea to carry large amounts of cash into a back-alley room full of strangers. Even if they are not just going to rob you at kifepoint, they may try to cheat you, so you have to be very skilled at detecting card manipulation. (I have a friend who, to demonstrate, turned over the top card of a deck of cards and dealt several hands while leaving it in place. Even though everyone knew he was not dealing from the top, no one could see anything amiss.) Even if no one is manipulating the cards, the other players may have signals to share information and coordinate bets. Even if the strangers don't try to cheat you, if you win too much, they may assume that you are cheating them, and unpleasantness may ensue. (This is true even in reputable Las Vegas casinos.)

You are probably better off hustling pool, but that is not perfectly safe either. (E.g., see the 1961 film *The Hustler* directed by Robert Rossen, starring Paul Newman, Jackie Gleason, and Piper Laurie, and nominated for nine Oscars.)

3.7 Bernoulli Trials

Pitman [10]: p. 27

A **Bernoulli trial** is a random experiment with two possible outcomes, traditionally labeled "**success**" and "**failure**." The probability of success is traditionally denoted p. The probability of failure $(1-p)$ is often denoted q. A **Bernoulli random variable** is simply the indicator of success in a Bernoulli trial. That is,

$$X = \begin{cases} 1 & \text{if the trial is a success} \\ 0 & \text{if the trial is a failure.} \end{cases}$$

3.8 The Binomial Distribution

Pitman [10]: § 2.1

If there are n stochastically independent Bernoulli trials with the same probability p of success, the number of successes is a **Binomial random variable**, and its distribution is called the **Binomial distribution** with parameters n and p. This is like *sampling with replacement* from an urn with a fraction p of Successes and $1-p$ of Failures. To get exactly k successes, there must be $n-k$ failures, but **the order** (of the successes and failures) **does not matter** for the count. There are

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

such outcomes, and by independence each has probability $p^k(1-p)^{n-k}$. (Recall that $0! = 1$.) For coin tossing, $p = (1-p) = 1/2$, but in general p need not be $1/2$. The counts $\binom{n}{k}$ are weighted by their probabilities $p^k(1-p)^{n-k}$. Thus

$$P(k \text{ successes in } n \text{ independent Bernoulli trials}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Another way to write this is in terms of the binomial random variable X that counts success in n trials:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Note that the binomial random variable is simply the sum of the Bernoulli random variables for each trial. Compare this to the analysis in Subsection 3.5.5, and note that it agrees because $1/2^n = (1/2)^k (1/2)^{n-k}$. Since $p + (1-p) = 1$ and $1^n = 1$, the Binomial Theorem assures us that the binomial distribution is a probability distribution.

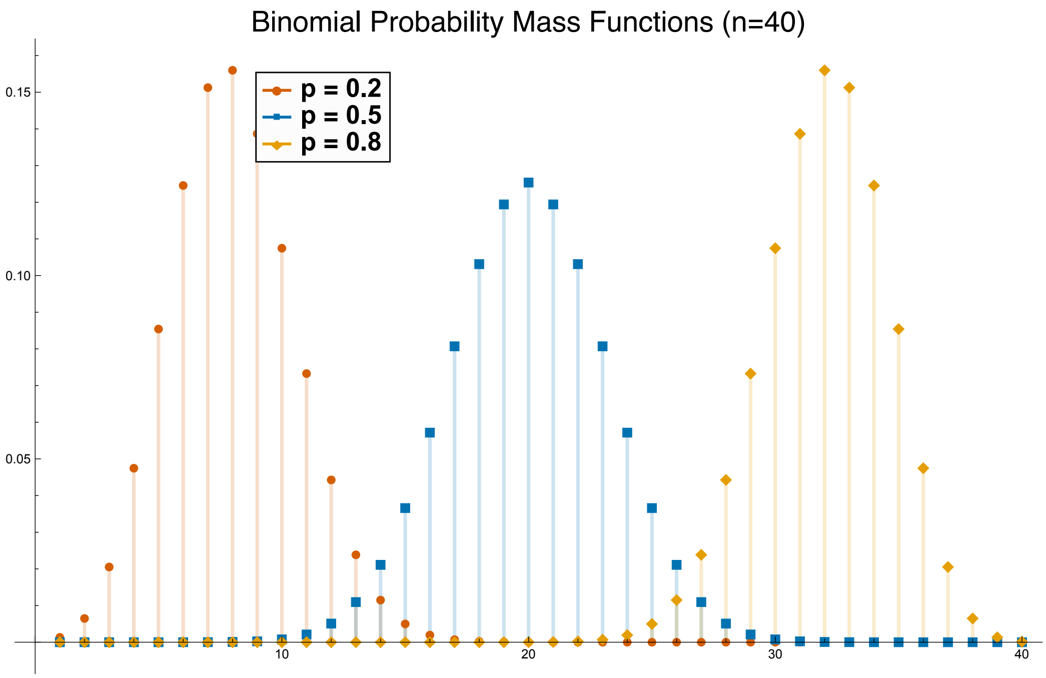

Figure 3.2. Binomial Probabilities

3.8.1 Example (The probability of n heads in $2n$ coin flips) For a fair coin the probability of n heads in $2n$ coin flips is

$$\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}.$$

We can see what happens to this for large n by using Stirling's approximation:

3.8.2 Proposition (Stirling's approximation) *For all $n \geqslant 1$,*

$$n! = \sqrt{2\pi}\; n^{n+(1/2)} e^{-n}\; e^{\varepsilon_n}$$

where $\varepsilon_n \to 0$ as $n \to \infty$. Indeed, $1/(12n+1) < \varepsilon_n < 1/12n$.

Apostol [1, Theorem 15.19, pp. 616–618] provides a proof of all but the last sentence, proving only that $0 < \varepsilon_n < 1/(8n)$. See Supplement 6 for a more detailed discussion of Stirling's approximation.

Thus we may write

$$\frac{(2n)!}{n!\,n!} = \frac{\sqrt{2\pi}\,(2n)^{2n+(1/2)}\,e^{-2n}e^{\varepsilon_{2n}}}{2\pi\,n^{2n+1}\,e^{-2n}\,e^{2\varepsilon_n}} = \frac{2^{2n}}{\sqrt{\pi n}}\,e^{\delta_n},$$

where $\delta_n \to 0$ as $n \to \infty$.

So the probability of n heads in $2n$ attempts is

$$\frac{2^{2n}}{\sqrt{\pi n}} 2^{-2n}\,e^{\delta_n} = \frac{1}{\sqrt{\pi n}}\,e^{\delta_n} \longrightarrow 0 \text{ as } n \to \infty.$$

What about the probability of between $n-k$ and $n+k$ heads in $2n$ tosses? Well the probability of getting j heads in $2n$ tosses is $\binom{2n}{j}(1/2)^{2n}$, and this is maximized at $j=n$ (See, e.g., Pitman [10, p. 86].) So we can use this as an upper bound. Thus for $k \geqslant 1$

$$\begin{aligned} P(\text{between } n-k \text{ and } n+k \text{ heads}) &< \frac{2k+1}{\sqrt{\pi n}}\, e^{\delta_n} \\ &\longrightarrow 0 \end{aligned}$$

as $n \to \infty$.

So any reasonable "law of averages" will have to let k grow with n. We will come to this in a few more lectures. □

3.9 The Multinomial Distribution

Larsen–Marx [9]: Section 10.2, pp. 494–499

The binomial distribution arises from *sampling with replacement* from an urn with only two types of balls, Success and Failure. The **multinomial distribution** arises from *sampling with replacement* from an urn with *m types* of balls. A summary of the outcome of a repeated random experiment where each experiment has only two types of outcomes is sometimes represented by a random variable X, which counts the number of "successes." With more than two type of outcomes, the similar summary would be a random vector $\boldsymbol{X}$, where the i^{th} component X_i of $\boldsymbol{X}$ counts the number of occurrences of outcome type i. With m possible outcome types where the i^{th} type has probability p_i, then in n independent trials, if $k_1 + \cdots + k_m = n$,

Pitman [10]: p. 155

$$P(k_i \text{ outcomes of type } i,\ i = 1,\dots,m) = \frac{n!}{k_1! \cdot k_2! \cdots k_m!}\, p_1^{k_1} \cdot p_2^{k_2} \cdots p_m^{k_m}.$$

This is referred to as the Multinomial$(n, \boldsymbol{p})$ distribution where $\boldsymbol{p} = (p_1, \dots, p_m)$. Note that if $m = 2$, then this is just the binomial distribution. In terms of the random vector $\boldsymbol{X} = (X_1, \dots, X_m)$ we write

$$P(\boldsymbol{X} = (k_1, \dots, k_m)) = \frac{n!}{k_1! \cdot k_2! \cdots k_m!}\, p_1^{k_1} \cdot p_2^{k_2} \cdots p_m^{k_m}.$$

Proof: If you find the above claim puzzling, this may help. Recall that in Subsection 3.5.5 we looked at the number of sets of size k and showed that there was a one-to-one correspondence between sets of size k and points in the sample space with exactly k heads. The same sort of reasoning shows that there is a one-to-one correspondence between partitions of the set of trials, $\{1, \dots, n\}$, into m sets $E_1, \dots, E_m$ with $\# E_i = k_i$ for each i and the set of points ω in the sample space where there are k_i outcomes of type i for each $i = 1, \dots, m$. Each such sample point has probability $p_1^{k_1} p_1^{k_2} \cdots p_m^{k^m}$. How many are there?

Well there are $\binom{n}{n-k_1}$ sets of trials of size k_1. But now we have to chose a set of size k_2 from the remaining $n-k_1$ trials, so there are $\binom{n-k_1}{k_2}$ ways to do this for each of the $\binom{n}{n-k_1}$ choices we made earlier. Now we have to choose a set of k_3 trials from the remaining $n-k_1-k_2$ trials, etc. The total number of possible partitions of the set of trials is thus

$$\binom{n}{k_1} \times \binom{n-k_1}{k_2} \times \binom{n-k_1-k_2}{k_3} \times \cdots \times \binom{n-k_1-k_2-\cdots-k_{m-1}}{k_m}.$$

Expanding this gives

$$\frac{n!}{k_1!(n-k_1)!} \times \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!} \times \frac{(n-k_1-k_2)!}{k_3!(n-k_1-k_2-k_3)!} \times \cdots \times \frac{(n-k_1-k_2-\cdots-k_{m-1})!}{k_m!\,\underbrace{(n-k_1-k_2-\cdots-k_{m-1}-k_m)!}_{=0!}}.$$

Now observe that the second term in each denominator cancels the numerator in the next fraction, and (recalling that $0! = 1$) we are left with

$$\frac{n!}{k_1! \cdot k_2! \cdots k_m!}$$

points in the sample space, each of which has probability $p_1^{k_1} \cdot p_2^{k_2} \cdots p_m^{k_m}$. ∎

3.9.1 Example Suppose you roll 9 dice. What is the probability of getting 3 aces (ones) and 6 boxcars (sixes)?

$$\frac{9!}{3!\,0!\,0!\,0!\,0!\,6!}\left(\frac{1}{6}\right)^9 = \frac{84}{10,077,696} \approx 8.3 \times 10^{-6}.$$

(Recall that $0! = 1$.) □

3.10 More on sampling with and without replacement

Move this?

Suppose you have an urn holding N balls, of which B are black and the remaining $W = N - B$ are white. If the urn is sufficiently well churned,[3] the probability of drawing a black ball is simply B/N. Now think of drawing a sample of size $n \leqslant N$ from this underlying population, and ask what the probability distribution of the composition of the sample is.

3.10.1 Sampling without replacement

Sampling without replacement means that a ball is drawn from the urn and set aside. The next time a ball is drawn from the urn, the composition of the balls has changed, so the probabilities have changed as well.

Need a better way to refer to this material.

For $b \leqslant n$, what is the probability that exactly b of the balls are black, and $w = n - b$? are white?

Let's dispose of some obvious cases. In order to have b black and w white balls, we must have

$$b \leqslant \min\{B, n\} \quad \text{and} \quad w \leqslant \min\{W, n\}.$$

There are $\binom{B}{b}$ sets of size b of black balls and $\binom{W}{w}$ sets of size w of white balls. Thus there are $\binom{B}{b}\binom{W}{w}$ possible ways to get exactly b black balls and w white balls in a sample of size $n = w + b$, out of $\binom{N}{n}$ possible samples of size n. Thus

$$P(b \text{ black \& } w \text{ white}) = \frac{\binom{B}{b}\binom{W}{w}}{\binom{N}{n}} = \frac{\binom{B}{b}\binom{W}{w}}{\binom{B+W}{b+w}}.$$

Note that if $b > B$ or $w > W$, by convention $\binom{B}{b} = \binom{W}{w} = 0$ (there are no subsets of size b of a set of size $B < b$), so this formula works even in this case.

These probabilities are known as the **hypergeometric distribution**.

Aside: According to Larsen and Marx Larsen–Marx [9, p. 111], a **hypergeometric series** is an infinite sum of the form

$$1 + \frac{\alpha\beta}{\gamma}x + \frac{\alpha(\alpha+1)\,\beta(\beta+1)}{2!\gamma(\gamma+1)}x^2 + \cdots + \left(\prod_{j=0}^{n-1}\frac{(\alpha+j)(\beta+j)}{(1+j)(\gamma+j)}\right)x^n + \cdots$$

that were first studied by Euler in 1769.

[3] To churn an urn has a nice poetic ring. In fact, in James Taylor's song *Steamroller* he refers to himself as "a churning urn of burning funk."

A special case is where $\alpha = 1$, and $\beta = \gamma$, in which case it reduces to the **geometric series** $1 + x + x^2 + \cdots$. For its relation to the hypergeometric distribution, take a sample of size n from an urn with W white balls and B black balls ($N = B + W$). Set $\alpha = -n$, $\beta = -B$, and $\gamma = W - n + 1$. Then (multiplying each α and β term by -1)), the coefficient of x^k in the hypergeometric series is

$$\begin{aligned}
\frac{n(n-1)\cdots(n-k+1)\,B(B-1)\cdots(B-k+1)}{k!(W-n+1)(W-n+2)\cdots(W-n+k)} &= \frac{\dfrac{n!}{(n-k)!}\,\dfrac{B!}{(B-k)!}}{k!\dfrac{\big(W-(n-k)\big)!}{(W-n)!}} \\
&= \frac{B!}{k!(B-k)!}\,\frac{(W-n)!}{\big(W-(n-k)\big)!}\,\frac{n!}{(n-k)!} \\
&= \binom{B}{k}\binom{W}{n-k}\frac{n!(W-n)!}{W!} \\
&= \underbrace{\frac{\binom{B}{k}\binom{W}{n-k}}{\binom{N}{n}}}\cdot\frac{\binom{N}{n}}{\binom{W}{n}},
\end{aligned}$$

where the underbraced term is the hypergeometric probability of getting k black balls in the sample.

3.10.2 Sampling without replacement with more than two types

Suppose an urn has three types of balls, Good, Bad, and Ugly. If the urn has $N = B + G + U$ balls, the probability of getting g good, b bad, and u ugly, is just

$$P(g \text{ good } \& \ b \text{ bad } \& \ u \text{ ugly}) = \frac{\binom{G}{g}\binom{B}{b}\binom{U}{u}}{\binom{G+B+U}{g+b+u}}.$$

And you can generalize from here.

3.10.3 Sampling with replacement

Sampling with replacement means that after a ball is drawn from the urn, it is returned, and the balls are mixed well enough so that each is equally likely. Thus repeated draws are independent and the probabilities are the same for each draw.

What is the probability that sample consists of b black and w white balls? This is just the binomial probability

$$P(b \text{ black } \& \ w \text{ white}) = \binom{n}{b}\left(\frac{B}{N}\right)^b\left(\frac{W}{N}\right)^w.$$

3.10.4 Comparing the two sampling methods

Intuition here can be confusing, since without replacement every black ball drawn reduces the pool of black balls making it less likely to get another black ball relative sampling with replacement, but every white ball drawn makes more likely to get a black ball. On balance you might think that sampling without replacement favors a sample more like the underlying population.

To compare the probabilities of sampling without replacement to those with replacement, we can rewrite the hypergeometric probabilities to make them look more like the binomial probabilities as follows.

$$P(\text{exactly } b \text{ balls out of } n \text{ are black}) = \frac{\binom{B}{b}\binom{W}{w}}{\binom{N}{n}} = \frac{\frac{B!}{b!(B-b)!}\,\frac{W!}{w!(W-w)!}}{\frac{N!}{n!(N-n)!}} = \frac{\frac{B!}{(B-b)!}\,\frac{W!}{(W-w)!}}{\frac{N!}{(N-n)!}}\,\frac{n!}{b!w!},$$

or in terms of the "order notation" (Subsection 3.4.2) we have

$$P(b \text{ black } \& \ w \text{ white}) = \binom{n}{b} \frac{(B)_b \, (W)_w}{(N_n)}$$
$$= \binom{n}{b} \frac{\overbrace{B(B-1)\cdots(B-b+1)}^{b \text{ terms}} \cdot \overbrace{W(W-1)\cdots(W-w+1)}^{w \text{ terms}}}{\underbrace{N(N-1)\cdots(N-n+1)}_{n \text{ terms}}}$$

for sampling without replacement versus

$$P(b \text{ black } \& \ w \text{ white}) = \binom{n}{b} \left(\frac{B}{N}\right)^b \left(\frac{W}{N}\right)^w = \binom{n}{b} \frac{\overbrace{B \times \cdots \times B}^{b \text{ terms}} \times \overbrace{W \times \cdots W}^{w \text{ terms}}}{\underbrace{N \times \cdots \times N}_{n = b+w \text{ terms}}}.$$

for sampling with replacement.

The ratio of the probability without replacement to the probability with replacement can be written as

$$\frac{B}{B} \cdot \frac{B-1}{B} \cdots \frac{B-b+1}{B} \cdot \frac{W}{W} \cdot \frac{W-1}{W} \cdots \frac{W-w+1}{W} \cdot \frac{N}{N} \cdot \frac{N}{N-1} \cdots \frac{N}{N-n+1}.$$

If $b = 0$, the terms involving B do not appear, and similarly for $w = 0$. Whether this ratio is greater or less than one is not obvious. But if we increase N keeping B/N (and hence W/N) constant, then holding the sample size n fixed, each term in this ratio converges to 1 for each b. Therefore the ratio converges to one.

That is, the difference between sampling with and without replacement holding the sample size constant becomes insignificant as N gets large, holding B/N and W/N fixed.

But how big is big enough? The only time sampling with replacement makes a difference is when the same ball is chosen more than once. The probability that all n balls are distinct is $\frac{N}{N} \cdot \frac{N-1}{N} \cdot \cdots \cdot \frac{N-n+1}{N}$, so the complementary probability (of a duplicate) is $1 - \prod_{k=0}^{n} \big(1 - (1/N)\big)$. Now us the Taylor series approximation that $\ln(1-x) \approx -x$, to get that $\log P(\text{duplicate}) \approx -\sum_{k=0}^{n} k/N = -n(n+1)/2N$. (The probability is less than one, so its logarithm is negative.) So as Pitman [10, p. 125] asserts, if $n \ll \sqrt{N}$, this probability is very small. With modern software, you can see for yourself how the two sampling methods compare. See Table 3.1 for a modest example of results calculated by Mathematica 12.

Aside: David Freedman [6] notes the following inequality regarding the difference between the two methods. Consider an urn with n distinct balls, and let $N = \{1, \dots, n\}$. A sample of size k can be thought of as an element of N^k. The probability measure P on N^k induced by sampling with replacement is given by $P(i_1, \dots, i_k) = 1/n^k$. There are $(n)_k = n!/(n-k)!$ distinct samples of size k when sampling without replacement, so the probability measure Q on N^k satisfies $Q(i_1, \dots, i_k) = 1/(n)_k$ if $i_1, \dots, i_k$ are all distinct, and $= 0$ otherwise.

The **total variation norm** $\|P - Q\|$ is defined to be $\sup\{|P(E) - Q(E)| : E \subset N^k\}$ and it is achieved for the event $G = \{(i_1, \dots, i_k) \in N^k : \text{all } i_j \text{ are distinct}\}$ and

$$\|P - Q\| = Q(G) - P(G) = 1 - P(G) = 1 - (n)_k/n^k.$$

He derives as a corollary, the following inequality:

$$1 - e^{-\frac{1}{2}\frac{k(k-1)}{n}} < \|P - Q\| < \frac{1}{2}\frac{k(k-1)}{n}.$$

b	Without Replacement	With Replacement	Ratio
0	0.33048	0.34868	0.94780
1	0.40800	0.38742	1.0531
2	0.20151	0.19371	1.0403
3	0.051794	0.057396	0.90240
4	0.0075532	0.011160	0.67680
5	0.00063980	0.0014880	0.42997
6	0.000030998	0.00013778	0.22498
7	8.1440×10^{-7}	8.7480×10^{-6}	0.093096
8	1.0411×10^{-8}	3.6450×10^{-7}	0.028564
9	5.1992×10^{-11}	9.0000×10^{-9}	0.0057769
10	5.7769×10^{-14}	1.0000×10^{-10}	0.00057769

Probability of b black balls in a sample of size $n = 10$ for a population of size $N = 100$, $B = 10$, $W = 90$.

b	Without Replacement	With Replacement	Ratio
0	0.34850	0.34868	0.99950
1	0.38761	0.38742	1.0005
2	0.19379	0.19371	1.0004
3	0.057348	0.057396	0.99917
4	0.011125	0.011160	0.99683
5	0.0014782	0.0014880	0.99340
6	0.00013625	0.00013778	0.98887
7	8.6016×10^{-6}	8.7480×10^{-6}	0.98326
8	3.5597×10^{-7}	3.6450×10^{-7}	0.97660
9	8.7200×10^{-9}	9.0000×10^{-9}	0.96889
10	9.6017×10^{-11}	1.0000×10^{-10}	0.96017

Probability of b black balls in a sample of size $n = 10$ for a population of size $N = 10,000$, $B = 1000$, $W = 9000$.

Table 3.1. Sampling without replacement vs. sampling with replacement.

Freedman then uses this to illustrate the difference between a sample of size $k = 1000$ versus a smale of size $k = 5000$ for an un urn with $n = 100,000,000$ balls. (Think a poll of U.S. voters.) But we don't need the inequality if we have Mathematica. After about 73 seconds for each calculation we find that for $k = 1000$ that $\|P - Q\| \approx 0.00498$, but for $k = 5000$, $\|P - Q\| \approx 0.117$. To three decimal places, these are Freedman's lower bounds. The point is, that for the worst case event, sampling 5000 balls with or without replacement can lead to a significant difference in the probabilities of particular events. But as Freeman pointed out, the worst case is not usually what we care about.

So where does the inequality come from? Note that

$$\frac{(n)_k}{n^k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} = \prod_{j=0}^{k-1} \frac{n-j}{n} = \prod_{j=1}^{k-1} \left(1 - \frac{j}{n}\right).$$

Now use the inequality that[4]

$$1 - \sum_{j=1}^{k} x_j < \prod_{j=1}^{k} (1 - x_j) < e^{-\sum_{j=1}^{k} x_j}$$

and applied to the case $x_j = j/n$, so that $\sum_{j=1}^{k-1} j/n = \frac{1}{2}\frac{k(k-1)}{n}$. (The first inequality is strict only if the sample size $k > 1$.)

3.11 Symmetry

When all the points in the sample space are equally likely (the uniform probability case) the probability of an event depends only on its **cardinality**, that is, the number of points of the sample space that it contains. This is the whole basis for the counting approach to calculating probabilities. Sometimes though it is perhaps not obvious that two events have the same number of points, and hence the same probabilities.

One way to show that two events E and F have the same number of points is to find a one-to-one function from the sample space Ω onto itself that maps E onto F. (Such a function is known as a **permutation** of Ω, or a **bijection** of Ω onto itself, or sometimes a **symmetry** of Ω.)

3.11.1 Symmetry Lemma *Let P be the uniform (counting) probability on the finite set Ω, and let $\varphi \colon \Omega \to \Omega$ be a one-to-one function from Ω onto itself. If*

$$\varphi(E) = \{\varphi(\omega) : \omega \in \Omega\} = F,$$

then $\# E = \# F$, so

$$P(E) = P(F).$$

This nothing deep, but it has some useful consequences.

- Let

$$U = \{1, \dots, m\}.$$

Think of U as an urn with m distinct balls. We assume that each ball is equally likely to be drawn so that $P(k) = 1/m$ for each $k \in U$. If we draw n balls independently from U *with replacement*, the sample space

$$\Omega = U^m.$$

[4] The first inequality is proven by induction on k, and is strict only if $k > 1$. Note that for $0 < x_j < 1$, $j = 1, \dots, k$, we have

$$(1 - x_1)(1 - x_2) = 1 - x_1 - x_2 + x_1 x_2 > 1 - (x_1 + x_2).$$

Now assume

$$(1 - x_1) \cdots (1 - x_{k-1}) > 1 - (x_1 + \cdots + x_{k-1}).$$

Multiplying both sides by $1 - x_k > 0$ we get

$$\begin{aligned}(1 - x_1) \cdots (1 - x_k) &> \big(1 - (x_1 + \cdots + x_{k-1})\big)(1 - x_k) \\ &= 1 - (x_1 + \cdots + x_k) + x_k(x_1 + \cdots + x_{k-1}) \\ &> 1 - (x_1 + \cdots + x_k).\end{aligned}$$

To get the second inequality above, the **subgradient inequality** days that for any strictly convex function f we have

$$f(x) > f(x_0) + f'(x_0)(x - x_0), \quad \text{for } x \neq x_0,$$

evaluated at $x_0 = 0$ and $x = -x_j$ to get

$$e^{-x_j} > 1 - x_j \implies e^{-\sum_j x_j} > \prod_j (1 - x_j), \quad (0 < x_j < 1,\ j = 1, \dots, k).$$

A typical element of Ω will be denoted $\omega = (u_1, \dots, u_n)$, and by independence each has probability $1/m^n$.

Pick some subset $A \subset U$, and fix i and j and consider the two events,

$$E_i = \{\omega = (u_1, \dots, u_n) \in U^m : u_i \in A\} \quad \text{and} \quad E_j = \{\omega = (u_1, \dots, u_n) \in U^m : u_j \in A\}.$$

These are the events that i^{th} and the j^{th} balls drawn belong to A.

Consider now the permutation of $\Omega = U^n$ that switches the i^{th} and the j^{th} terms in the sequence:

$$\varphi(u_1, \dots, u_{i-1}, u_i, u_{i+1}, \dots, u_{j-1}, u_j, u_{j+1}, \dots, u_n) \\ = (u_1, \dots, u_{i-1}, u_j, u_{i+1}, \dots, u_{j-1}, u_i, u_{j+1}, \dots, u_n),$$

assuming $i < j$, which entails no loss of generality.

Then

$$\varphi(E_i) = E_j, \text{ so } P(E_i) = P(E_j).$$

In fact, since we are making independent draws with replacement, $P(E_i) = P(E_j)$ is just the probability of drawing an element of A out of the urn, namely $\# A/m$.

• The point just made is fairly obvious, but now consider what happens if we draw $n \leqslant m$ balls *without replacement*. Now the sample space Ω is the set of lists of length n of elements of U, which has cardinality $m!/n!$.

Now with events E_i and E_j as above and the mapping φ as above, we still have

$$\varphi(E_i) = E_j, \text{ so } P(E_i) = P(E_j).$$

In fact, by considering E_1, we see that its probability is just the same as drawing an element of A out of the urn, namely $\# A/m$.

• **3.11.2 Proposition** *So the probability that the i^{th} ball drawn belongs to A is the same as the probability that the first ball belongs to A, namely $\# A/m$, regardless of whether we are sampling with or without replacement!*

• So for instance, in dealing cards from a deck (which is just like sampling without replacement) the probability that the first card is an Ace is $1/13 = 4/52$, which is the same as the probability that the second card is an Ace, which is the same as the probability that the fifty-second card is an Ace.

• Which answers a question my physical therapist asked. He and his cousin play a card game where one player turns over cards until a red ace appears. Then the second player turns over cards until the second red ace appears. The winner is the player who turns over the fewest cards. A recent game did not end until the last card. He was disappointed to find out that this was not as rare as he had thought. It is just the probability that the last card is a red ace (1/26).

• Now letting E_i be the event that the i^{th} card is an Ace, doesn't the above imply that the probability that the first card is an Ace and the second card is an Ace and etc. and the fifty-second card is an Ace must be equal to $(1/13)^{52}$? NO! These events are clearly *not* independent events. The probability that every card is an Ace is zero, at least for a standard deck of cards.

• The last point suggests the following ugly calculations.

The number of sequences of deck of cards that put an Ace on top is $4 \times 51!$, as there are four aces and once one is on top the remaining 51 can be in any order.

To get an Ace in the second position, you can have either an Ace on top or a non-Ace on top. There are $4 \times 3 \times 50!$ to have an Ace on top and in second place. There are $48 \times 4 \times 50!$ ways to put one of 48 non-Aces on top, and an Ace second, and the remaining 50 cards can be in any order. Adding these two (the events are disjoint) gives a total of $4 \times (48 + 3) \times 50! = 4 \times 51!$, same as an Ace on top.

You could continue on to calculate the probability that the third cards is an Ace, etc., but the calculations just get uglier. The symmetry argument is much simpler.

3.12 Matching

There are n consecutively numbered balls and n consecutively numbered bins. The balls are arranged in the bins (one ball per bin) at random (all arrangements are equally likely). What is the probability that at least one ball matches its bin? (See Exercise 28 on page 135 of Pitman [10].)

Intuition is not a lot of help here for understanding what happens for large n. When n is large, there is only a small chance that any given ball matches, but there are a lot of them, so one could imagine that the probability could converge to zero, or to one, or perhaps something in between.

Let A_i denote the event that Ball i is placed in Bin i. We want to compute the probability of $\bigcup_{i=1}^{n} A_i$. This looks like it might be a job for the Inclusion–Exclusion Principle, since these events are not disjoint. Recall that it asserts that

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) = & \sum_i p(A_i) \\ & - \sum_{i<j} P(A_i A_j) \\ & + \sum_{i<j<k} P(A_i A_j A_k) \\ & \vdots \\ & + (-1)^k \sum_{i_1<i_2<\cdots<i_k} P(A_{i_1} A_{i_2} \cdots A_{i_k}) \\ & \vdots \\ & + (-1)^{n+1} P(A_1 A_2 \cdots A_n). \end{aligned}$$

Consider the intersection $A_{i_1} A_{i_2} \cdots A_{i_k}$, where $i_1 < i_2 < \cdots < i_k$. In order for this event to occur, ball i_j must be in bin i_j for $j = 1, \ldots, k$. This leaves $n - k$ balls unrestricted, so there are $(n - k)!$ arrangements in this event. And there are $n!$ total arrangements. Thus

$$P(A_{i_1} A_{i_2} \cdots A_{i_k}) = \frac{(n-k)!}{n!}.$$

Note that this probability depends only on k (and n), not on the particular set $\{i_1, \ldots, i_k\}$. Now there are $\binom{n}{k}$ size-k sets of balls. Thus the k-term in the formula above is

$$\sum_{i_1<i_2<\cdots<i_k} P(A_{i_1} A_{i_2} \cdots A_{i_k}) = \binom{n}{k} \frac{(n-k)!}{n!}.$$

Therefore the Inclusion–Exclusion Principle reduces to

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!}.$$

Here are the values for $n = 1, \dots, 10$:

n:	Prob(match)
1:	1
2:	$\frac{1}{2} = 0.5$
3:	$\frac{2}{3} \approx 0.666667$
4:	$\frac{5}{8} = 0.625$
5:	$\frac{19}{30} \approx 0.633333$
6:	$\frac{91}{144} \approx 0.631944$
7:	$\frac{177}{280} \approx 0.632143$
8:	$\frac{3641}{5760} \approx 0.632118$
9:	$\frac{28673}{45360} \approx 0.632121$
10:	$\frac{28319}{44800} \approx 0.632121$

Notice that the results converge fairly rapidly, but to what? The answer is $\sum_{k=1}^{\infty}(-1)^{k+1}/k!$, which you may recognize as $1 - e^{-1}$. (See Supplement 1.)

3.13 Waiting: The Negative Binomial Distribution

Pitman [10]: p. 213
Larsen–Marx [9]: § 4.5

The **Negative Binomial Distribution** is the probability distribution of the number of independent trials need for a given number of heads. What is the probability that the r^{th} success occurs on trial t, for $t \geqslant r$?

For this to happen, there must be $t - r$ failures and $r - 1$ successes in the first $t - 1$ trials, with a success on trial t. By independence, this happens with the binomial probability for $r - 1$ successes on $t - 1$ trials times the probability p of success on trial t:

$$\text{NB}(t; r, p) = \binom{t-1}{r-1} p^{r-1}(1-p)^{(t-1)-(r-1)} \times p = \binom{t-1}{r-1} p^{r}(1-p)^{t-r} \qquad (t \geqslant r).$$

Of course, the probability is 0 for $t < r$. The special case $r = 1$ (number of trials to the first success) is called the **Geometric Distribution**:

$$\text{NB}(t; 1, p) = \binom{t-1}{0} p^{0}(1-p)^{t-1} \times p = p(1-p)^{t-1} \qquad (t \geqslant 1).$$

Warning: The definition of the negative binomial distribution here is the same as the one in Pitman [10, p. 213] and Larsen–Marx [9, p. 262]. Both Mathematica and R use a different definition. They define it to be the distribution of the number of failures that occurs before the r^{th} success. That is, Mathematica's `PDF[NegativeBinomialDistribution[r, p], t]` is our $\text{NB}(t + r; r, p)$. Mathematica and R's definition assigns positive probability to 0, ours does not.

Bibliography

[1] T. M. Apostol. 1969. *Calculus, Volume II: Mulit-variable calculus and linear algebra, with applications to differential euations and probability*, 2d. ed. Waltham, Massachusetts: Blaisdell.

[2] R. B. Ash. 2008. *Basic probability theory*. Mineola, New York: Dover. Reprint of the 1970 edition published by John Wiley and Sons.

[3] S. Assaf, P. Diaconis, and K. Soundararajan. 2011. A rule of thumb for riffle shuffling. *Annals of Applied Probability* 21(3):843–875. http://www.jstor.org/stable/23033357

[4] D. Bayer and P. Diaconis. 1992. Trailing the dovetail shuffle to its lair. *Annals of Applied Probability* 2(2):294–313. http://www.jstor.org/stable/2959752

[5] R. D. Clarke. 1946. An application of the Poisson distribution. *Journal of the Institute of Actuaries* 72:481.

[6] D. Freedman. 1977. A remark on the difference between sampling with and without replacement. *Journal of the American Statistical Association* 72(359):681. http://www.jstor.org/stable/2286241

[7] R. Huddleston and G. K. Pullum. 2005. *A student's introduction to English grammar*. Cambridge: Cambridge University Press.

[8] P. S. Laplace. 1995. *A philosophical essay on probability*. New York: Dover Publications. The Dover edition, first published in 1995, is an unaltered and unabridged republication of the work originally published by John Wiley and Sons in 1902, and previously reprinted by Dover in 1952. The English translation by F. W. Truscott and F. L. Emory, is from the sixth French edition of the work titled *Essai philosophique sur les probabilités*, published by Gauthier–Villars (Paris) as part of the 15-volume series of Laplace's collected works. The original French edition was published in 1814 (Paris).

[9] R. J. Larsen and M. L. Marx. 2012. *An introduction to mathematical statistics and its applications*, fifth ed. Boston: Prentice Hall.

[10] J. Pitman. 1993. *Probability*. Springer Texts in Statistics. New York, Berlin, and Heidelberg: Springer.