Lecture 3: Learning to count; Binomial Distribution

Relevant textbook passages:
Pitman [7]: Sections 1.5–1.6, pp. 47–77; Appendix 1, pp. 507–514.
Larsen–Marx [6]: Sections 2.4, 2.5, 2.6, 2.7, pp. 67–101. Sections

The great coin-flipping experiment

There were 184 valid submissions of 128 flips, for a grand total of 23,552! You can soon find the data at http://www.math.caltech.edu/%7E2015-16/2term/ma003/Data/FlipsMaster.txt

Recall that I put predictions into a sealed envelope. Here are the predictions of the average number of runs, by length, compared to the experimental results.

<table>
<thead>
<tr>
<th>Run length</th>
<th>Theoretical average$^a$</th>
<th>Predicted range$^b$</th>
<th>Total runs</th>
<th>Average runs</th>
<th>How well did I do?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32.5</td>
<td>31.615 – 33.385</td>
<td>6206</td>
<td>33.728261</td>
<td>Off by 0.34.</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>7.65 – 8.355</td>
<td>1504</td>
<td>8.173913</td>
<td>Nailed it.</td>
</tr>
<tr>
<td>5</td>
<td>1.96875</td>
<td>1.79 – 2.155</td>
<td>361</td>
<td>1.961957</td>
<td>Nailed it.</td>
</tr>
<tr>
<td>6</td>
<td>0.976563</td>
<td>0.845 – 1.11</td>
<td>184</td>
<td>1.000000</td>
<td>Nailed it.</td>
</tr>
<tr>
<td>7</td>
<td>0.484375</td>
<td>0.39 – 0.58</td>
<td>90</td>
<td>0.489130</td>
<td>Nailed it.</td>
</tr>
<tr>
<td>8</td>
<td>0.240234</td>
<td>0.175 – 0.31</td>
<td>38</td>
<td>0.206522</td>
<td>Nailed it.</td>
</tr>
<tr>
<td>9</td>
<td>0.119141</td>
<td>0.075 – 0.17</td>
<td>14</td>
<td>0.076087</td>
<td>Nailed it.</td>
</tr>
<tr>
<td>10</td>
<td>0.059082</td>
<td>0.03 – 0.095</td>
<td>6</td>
<td>0.032609</td>
<td>Nailed it.</td>
</tr>
<tr>
<td>11</td>
<td>0.0292969</td>
<td>0.01 – 0.055</td>
<td>3</td>
<td>0.016304</td>
<td>Nailed it.</td>
</tr>
<tr>
<td>12</td>
<td>0.0145264</td>
<td>0.0 – 0.035</td>
<td>3</td>
<td>0.016304</td>
<td>Nailed it.</td>
</tr>
<tr>
<td>13</td>
<td>0.00720215</td>
<td>0.0 – 0.02</td>
<td>1</td>
<td>0.005435</td>
<td>Nailed it.</td>
</tr>
<tr>
<td>14</td>
<td>0.00357036</td>
<td>N/A</td>
<td>1</td>
<td>0.005435</td>
<td>Nailed it.</td>
</tr>
<tr>
<td>15</td>
<td>0.00177002</td>
<td>N/A</td>
<td>1</td>
<td>0.005435</td>
<td>Nailed it.</td>
</tr>
</tbody>
</table>

$^a$The formula for the theoretical average is the object of the Optional Exercise.

$^b$This is based on a Monte Carlo simulation of the 95% confidence interval for a sample size of 220, not 184.

Yes! There are Laws of Chance.

How did we do on Heads versus Tails? Out of 23,552 there were:

<table>
<thead>
<tr>
<th>Number</th>
<th>Percent</th>
<th>How close to 50/50 is this? We’ll see in a bit.$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heads</td>
<td>11804</td>
<td>49.881</td>
</tr>
<tr>
<td>Tails</td>
<td>11748</td>
<td>50.119</td>
</tr>
</tbody>
</table>

3.1 Independent events

3.1.1 Definition Events $E$ and $F$ are (stochastically) independent if

$$P(EF) = P(E) \cdot P(F).$$

$^1$It turns out there was one duplicate submission by accident. I have not bothered to recalculate these figures. It won’t have a substantial impact.
3.2 Independent repeated experiments

Our mathematical model of a random experiment is a probability space \((S, \mathcal{E}, P)\). And of the repeated experiment is \((S^2, \mathcal{E}^2, ?)\). The question mark is there because we need to decide the probabilities of events in a repeated experiment. To do this in a simple way we shall consider the case where the *experiments* are independent. That is, the outcome of the first experiment provides no information about the outcome of the second experiment.

Consider a compound event \(E_1 \times E_2\), which means that the outcome of the first experiment was in \(E_1 \in \mathcal{E}\) and the outcome of the second experiment was in \(E_2 \in \mathcal{E}\). The event \(E_1\) in the first experiment is really the event \(E_1 \times S\) in the compound experiment. That is, it is the set of all ordered pairs where the first coordinate belongs to \(E_1\). Similarly the event \(E_2\) in the second experiment corresponds to the event \(S \times E_2\) in the compound experiment. Now observe that

\[(E_1 \times S) \cap (S \times E_2) = E_1 \times E_2.\]

Since the experiments are independent the probability of the intersection \((E_1 \times S)(S \times E_2)\) should be the probability of \((E_1 \times S)\) times the probability of \((S \times E_2)\). But these probabilities are just \(P(E_1)\) and \(P(E_2)\) respectively. Thus for independently repeated experiments and “rectangular events,”

\[\text{Prob}(E_1 \times E_2) = P(E_1) \times P(E_2).\]

This is enough to pin down the probability of all the events in the product algebra \(\mathcal{E}^2\), and the resulting probability measure is called the product probability, and may be denoted by \(P \times P\), or \(P^2\), or by really abusing notation, simply \(P\) again.

The point to remember is that independent experiments give rise to products of probabilities.

How do we know when two experiments are independent? We rely on our knowledge of physics or biology or whatever to tell us that the outcome of one experiment yields no information on the outcome of the other. It’s built into our modeling decision. I am no expert, but my understanding is that quantum entanglement implies that experiments that our intuition tells are independent are not really independent. But that is an exceptional case. For coin tossing, die, rolling, roulette spinning, etc., independence is probably a good modeling choice.

3.3 Generally accepted counting principles

The Uniform Probability (or counting) model was the earliest and hence one of the most pervasive probability models. For that reason it is important to learn to count. This is the reason that probability and combinatorics are closely related.

Lists versus sets

If you’ve had CS 1, you’ve studied Python which has both *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 instance,

\[123\text{ and }213\]

are distinct lists of three elements, but the same set.

A list is sometimes referred to as a *permutation*.

Number of lists of length \(n\)

If I have \(n\) distinct objects, how many distinct ways can I arrange them into a list (without repetition)? Think of the objects being numbered and starting out in a bag and having to be distributed among \(n\) numbered boxes.
There are \( n \) choices for box 1, and for each such choice, there are \( n - 1 \) for position 2, etc., so all together

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 \textbf{\( n \) \textit{factorial}},

By definition, \( 0! = 1 \), and we have the following recursion

\[ n! = n \cdot (n - 1)! \quad (n > 0). \]

**Number of lists of length \( k \) of \( n \) objects**

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

\[
n \times (n - 1) \times \cdots \times (n - k + 1) \\
= n \times (n - 1) \times \cdots \times (n - k + 1) \times (n - k) \times (n - k - 1) \times \cdots \times 2 \times 1 \\
= \frac{n!}{(n - k)!}
\]

Thus

there are \( \frac{n!}{(n - k)!} \) distinct lists of length \( k \) chosen from \( n \) objects.

Note that when \( k = n \) this reduces to \( n! \) (since \( 0! = 1 \)), which agrees with the result in the previous section.

**Number of subsets of size \( k \) of \( n \) objects**

How many distinct subsets of size \( k \) can I make with \( n \) objects? (A subset is sometimes referred to as a \textbf{\textit{combination}} of elements.) 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. Thus

there are \[ \frac{n!}{(n-k)! \cdot k!} \] distinct subsets of size $k$ chosen from $n$ objects.

3.3.1 Definition For natural numbers $0 \leq k \leq n$

\[ \binom{n}{k} = \frac{n!}{(n-k)! \cdot k!} \]

is read as

"$n$ choose $k$"

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.

Other notations you may encounter include $C(n,k)$, $^nC_k$, and $nC_k$. (These notations are easier to typeset in lines of text.)

Some useful identities

\[ \binom{n}{n} = 1 \]
\[ \binom{n}{1} = n \]
\[ \binom{n}{k} = \binom{n}{n-k} \]
\[ \binom{n+1}{k+1} = \binom{n}{k+1} + \binom{n}{k} \] (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) $E \subseteq B$, or else (ii) $E$ consists of $\bar{a}$ and $k$ elements of $B$. 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 number is the sum of the two above it:

```
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
etc.
```
Equation (1) also implies (by the telescoping method) that
\[
\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \cdots + (-1)^k \binom{n}{k} = (-1)^k \binom{n-1}{k}.
\]

**Number of all subsets of a set**

Given a subset \( A \) of a set \( X \), its **indicator function** is defined by

\[
1_A(x) = \begin{cases} 
1 & x \in A, \\
0 & x \notin A.
\end{cases}
\]

There is a one-to-one correspondence between sets and indicator functions. How many different indicator functions are there? For each element the value can be either 0 or 1, and there are \( n \) elements so there are \( 2^n \) distinct subsets of a set of \( n \) objects.

**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 or Ma 1a.

**3.3.2 Binomial Theorem**

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

**3.4 Examples of counting and probability**

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 \( S \).

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

\( 2^n \)
Binomial probabilities

What is the probability of getting $k$ heads in $n$ independent tosses of a fair coin?

Let’s do this carefully. The sample space $S$ is the set of sequences of length $n$ where each term $s_i$ in the sequence is $H$ or $T$. For each point $s \in S$, let $A_s = \{i : s_i = H\}$. Since there are only two outcomes, if you know $A_s$, you know $s$ and vice versa. Now let $E$ be any subset of $S$ that has exactly $k$ elements. There is exactly one point $s \in S$ such that $A_s = E$. Thus the number of elements of $S$ such that $|A_s| = k$ is precisely the same as the number of subsets of $S$ of size $k$, namely $\binom{n}{k}$. Thus

$$\text{Prob(exactly } k \text{ Heads}) = \frac{|\{s \in S : |A_s| = k\}|}{|S|} = \frac{\binom{n}{k}}{2^n} = \frac{n!}{k!(n-k)!2^n}.$$  

Here is an example with $n = 3$:

<table>
<thead>
<tr>
<th>$s$</th>
<th>$A_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HHH</td>
<td>${1, 2, 3}$</td>
</tr>
<tr>
<td>HHT</td>
<td>${1, 2}$</td>
</tr>
<tr>
<td>HTT</td>
<td>${1}$</td>
</tr>
<tr>
<td>THH</td>
<td>${2, 3}$</td>
</tr>
<tr>
<td>THT</td>
<td>${2}$</td>
</tr>
<tr>
<td>TTH</td>
<td>${3}$</td>
</tr>
<tr>
<td>TTT</td>
<td>$\emptyset$</td>
</tr>
</tbody>
</table>

For $k = 2$, the set of points $s \in S$ 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.

<table>
<thead>
<tr>
<th>$s$</th>
<th>$A_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HHH</td>
<td>${1, 2, 3}$</td>
</tr>
<tr>
<td>HHT</td>
<td>${1, 2}$</td>
</tr>
<tr>
<td>HTT</td>
<td>${1}$</td>
</tr>
<tr>
<td>THH</td>
<td>${2, 3}$</td>
</tr>
<tr>
<td>THT</td>
<td>${2}$</td>
</tr>
<tr>
<td>TTH</td>
<td>${3}$</td>
</tr>
<tr>
<td>TTT</td>
<td>$\emptyset$</td>
</tr>
</tbody>
</table>

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

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

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

(order matters)

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

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.

How many different 5-card poker hands are there?

(order does not matter)

$$\binom{52}{5} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 2,598,960.$$
How many different deals?

How many different deals of 5-card poker hands for 7 players are there? (order of hands matters, but order of cards within hands does not),

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

Each succeeding hand has 5 fewer cards to choose from, the others being used by the earlier hands.

How many 5-card poker hands are flushes?

To get a flush all five cards must be of the same suit. There are thirteen ranks in each suit, so there \(\binom{13}{5}\) distinct flushes from a given suit. There are four suits, so there

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

(This includes straight flushes.)

So what is the probability of a flush?

$$\frac{4 \binom{13}{5}}{\binom{52}{5}} = \frac{5148}{2,598,960} \approx .002$$

Sampling with and without replacement

Suppose you have an urn \(U\) with \(N\) balls, of which \(n\) are red and the remaining \(N - n\) are green.

- If \(k \leq n\) balls are drawn without replacement, what is the probability that all are red?

Here the random experiment is to choose a subset of size \(k\).

1. First argument: There are \(\binom{N}{k}\) distinct subsets of size \(k\). But in order for all of them to be red, they must in fact be drawn from the set of red balls. There are only \(\binom{n}{k}\) distinct size-\(k\) subsets of red balls. Thus the probability is

$$\frac{\binom{n}{k}}{\binom{N}{k}}.$$

2. Second argument: (This argument is a little trickier, and tacitly relies on some intuitive properties of conditional probability that I’ll gloss over until later.)

In order to draw a size-\(k\) set of red balls, the first time a ball is drawn there are \(n\) red balls out of \(N\) total, so the probability of the first being red is \(n/N\). If the first is red, the probability of the second is red is \((n-1)/(N-1)\), as there are \(n-1\) remaining red balls out \(N-1\) remaining. We shall see in a bit that we should multiply these probabilities, etc., so the probability is

$$\frac{n}{N} \frac{n-1}{N-1} \cdots \frac{n-k+1}{N-k+1}.$$ \(k\) terms

Fortunately, these two arguments give the same answer, since

$$\frac{\binom{n}{k}}{\binom{N}{k}} = \frac{n!}{k!(n-k)!} = \frac{n!}{(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{N(N-1)\cdots(N-k+1)}$$
• For $k > n$ what is the probability that all $k$ are red? Zero.

• If $k$ balls are drawn with replacement, and the draws are independent, what is the probability that all are red?

This is easier. Let $E_i$ be the event that the $i^{th}$ ball drawn is red. Each $E_i$ has probability $n/N$ and they are independent. The event that all $k$ are red is the intersection $E_1E_2\cdots E_k$, which has probability

$$\left(\frac{n}{N}\right)^k.$$ 

• With replacement, for $k > n$ what is the probability that all $k$ are red? $(n/N)^k$, same as before.

It is intuitive that with replacement the probability of all red balls should be greater than without replacement. This is certainly true for $k > n$, where we are comparing $(n/N)^k$ to 0. But even for $k \leq n$, think of it this way:

The first ball has an $n/N$ chance of being red, and if its, then the second ball has a smaller chance, $(n-1)/(N-1)$, of being red, and the next one has an even smaller chance, etc., But with replacement each ball has an $n/N$ chance of being red. Algebraically, $(n-i)/(N-i) < n/N$, so

$$\frac{n\cdot(n-1)\cdots(n-k+1)}{N\cdot(N-1)\cdots(N-k+1)} < \left(\frac{n}{N}\right)^k.$$ 

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 [7].)

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

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i} p(A_i)$$

$$- \sum_{i<j} P(A_iA_j)$$

$$+ \sum_{i<j<k} P(A_iA_jA_k)$$

$$\vdots$$

$$+ (-1)^k \sum_{i_1<i_2<\cdots<i_k} P(A_{i_1}A_{i_2}\cdots A_{i_k})$$

$$\vdots$$

$$+ (-1)^{n+1} P(A_1A_2\cdots A_n).$$
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 depends only on $k$. Now there are $\binom{n}{k}$ size-$k$ sets of balls. Thus the $k$ term in the formula above satisfies

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

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 $n = 1, \ldots, 10$:

- $n$: Prob(match)
- 1: 1
- 2: $\frac{1}{2} = 0.5$
- 3: $\frac{2}{3} \approx 0.666667$
- 4: $\frac{3}{5} = 0.625$
- 5: $\frac{4}{9} \approx 0.633333$
- 6: $\frac{9}{144} \approx 0.631944$
- 7: $\frac{177}{280} \approx 0.632143$
- 8: $\frac{2641}{5760} \approx 0.632118$
- 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 $\sum_{k=1}^{\infty} (-1)^{k+1} \frac{1}{k!}$, which you may recognize as $1 - (1/e)$. (See the supplementary notes on series.)

### 3.5 Bernoulli Trials

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,

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

### 3.6 The Binomial Distribution

If there are $n$ stochastically independent Bernoulli trials with the same probability $p$ of success, the probability distribution of the number of successes is called the **Binomial distribution**. A
Binomial random variable is simply the count of the number of successes in \( n \) trials. To get exactly \( k \) successes, there must be \( n - k \) failures. There are
\[
\binom{n}{k} = \frac{n!}{k!(n-k)!}
\]
such outcomes (where by convention \( 0! = 1 \)), and by independence each has probability \( p^k(1-p)^{n-k} \). N.B. In this case, since success and failure are not equally likely, so the points in the sample space are not equally likely. Simple counting is not going to be adequate. 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 3.4, 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.

### 3.6.1 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:

3.6.2 Proposition (Stirling’s formula) 

\[ n! = e^{-n}n^n\sqrt{2\pi n}(1+\varepsilon_n) \]

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

For a proof, see, e.g., Robbins [8], Feller [5, p. 52] or [3, 4] or Ash [1, pp.43–45], or Diaconis and Freedman [2], or the exercises in Pitman [7, p. 136].

Thus we may write

\[ \frac{(2n)!}{n! n!} = \frac{e^{-2n}(2n)^{2n}\sqrt{4\pi n}}{e^{-n}n^n\sqrt{2\pi n}\sqrt{2\pi n}}(1 + \delta_n) = \frac{2^{2n}}{\sqrt{\pi n}}(1 + \delta_n), \]

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

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

\[ \frac{2^{2n}}{\sqrt{\pi n}}2^{-2n}(1 + \delta_n) = \frac{1}{\sqrt{\pi n}}(1 + \delta_n) \to 0 \]

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 [7, p. 86].) So we can use this as an upper bound. Thus for $k \geq 1$

\[ P(\text{between } n-k \text{ and } n+k \text{ heads}) < \frac{2k + 1}{\sqrt{\pi n}}(1 + \delta_n) \to 0 \]

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.

\[ \square \]

3.7 The Multinomial Distribution

The Multinomial distribution generalizes the binomial distribution to random experiments with more than two “types” of outcomes. If there are $m$ possible outcome types and the $i^{th}$ type has probability $p_i$, then in $n$ independent trials, if $k_1 + \cdots + k_m = n$, 

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

3.7.1 Remark If you find the above claim puzzling, this may help. Recall that in Subsection 3.4 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,
Suppose you roll 9 dice. What is the probability of getting 3 aces (ones) and 6 boxcars (sixes)?

Well there are $\binom{n}{k_1}$ sets of trials of size $k_1$. But now we have to choose 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 $\binom{n}{k_1}$ 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

$$\left(\frac{n}{k_1}\right) \times \left(\frac{n-k_1}{k_2}\right) \times \left(\frac{n-k_1-k_2}{k_3}\right) \times \cdots \times \left(\frac{n-k_1-k_2-\cdots-k_{m-1}}{k_m}\right).$$

Expanding this gives

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

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

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

points $s \in S$, each of which has probability $p_1^{k_1} \cdot p_2^{k_2} \cdots \cdot p_m^{k_m}$.

We can use random vectors to describe what is happening. For each type $i = 1, \ldots, m$, let $X_i$ denote the number of outcomes of type $i$. Then the random vector $\mathbf{X} = (X_1, \ldots, X_m)$ has a distribution given by

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

### 3.7.2 Example

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

$$\frac{9!}{3!0!0!0!0!6!} \left(\frac{1}{6}\right)^9 = 94 \frac{1}{10,077,696} \approx 0.0000083.$$

(Recall that $0! = 1$.)

### 3.8 The Negative Binomial Distribution

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

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 $t$:

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

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**.

**Warning:** The definition of the negative binomial distribution here is the same as the one in Pitman [7, p. 213] and Larsen-Marx [6, 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 `NB(t + r; r, p)`. Mathematica and R’s definition assigns positive probability to 0, ours does not.
Bibliography


http://www.jstor.org/stable/2322709

http://www.jstor.org/stable/2315671

http://www.jstor.org/stable/2314719


http://www.jstor.org/stable/2308012