Chapter 6

Mixing and Sorting

6.1 On mixing and randomness

I use the rather pedantic phrase *artifacts with physical symmetry* as a reminder that the familiar objects used as examples in introductory mathematical probability – dice, roulette wheels, playing cards, etc – are in fact very special. For playing cards the “physical symmetry” is just the fact that the different card in a deck are the same size and shape. That property is shared by innumerable manufactured items, but there are two other properties of playing cards that seem essentially unique. One is obvious: the backs are designed to be indistinguishable while the fronts are designed to be distinguishable. What is the other?

Let me start with an analogy. The fact “ice floats on water” we observe in childhood, from ice cubes in drinks if not from icebergs. Not until taking a chemistry course do we learn this is really a remarkable special feature of water; almost all other liquids become denser upon freezing, so the frozen substance would sink beneath the liquid substance. Analogously, most us us learn in later childhood

playing cards are easy to mix by shuffling.

But we rarely pause to consider that this is a remarkable special feature of playing cards.

In most everyday contexts (e.g. cooking ingredients) where we might use the word *mixing*, we don’t connect mixing with *randomness*. But in the context of playing cards (and lotteries and bingo) we are deliberately using physical mixing to achieve, at least approximately, a rather precise state of “uniform randomness” – we want each card to be equally likely to be the
top card, or each ticket equally likely to be drawn as the lottery winner. So what can we say about the connection between mixing and randomness?

To a mathematician, a relevant result is the ergodic theorem. Explaining what the ergodic theorem says in a way that is both careful and makes clear the application to physical mixing is beyond my ability (and (W) not helpful), but here is a very rough statement of its consequence. Take a collection of objects that are physically identical aside from identifying labels; specify some procedure to rearrange their positions. The key requirement is that the procedure does not look at labels—so “pick a Spade and put it on top of the deck” isn’t allowed. Repeat this procedure for long enough, and the distribution of which object is in which position will get closer and closer to the uniform distribution (each possible assignment of objects to positions is equally likely).

The ergodic theorem is a triumph of abstract mathematics, but for such great generality one pays a price: it doesn’t tell you how long you need to mix in any particular context. And aside from simplified models of card shuffling, there has been surprisingly little serious research on this question. But I am willing to assert that, for most physical objects used for randomness, to make them genuinely random you need to do much more mixing than you would intuitively think.

A famous example, discussed in many textbooks on mathematical statistics, was the 1970 draft lottery. Let me quote an article by Norton Starr\(^1\) which gives both the raw data and results of several different statistical analysis.

In an attempt to expose male youth fairly to the risk of being drafted, a lottery was held to allocate birthdates at random: 366 capsules, each containing a unique day of the year, were successively drawn from a container. The first date drawn (September 14) was assigned rank 1, the second date drawn (April 24) was assigned rank 2, and so on. Those eligible for the draft who were born on September 14 were called first for physicals, then those born on April 24 were tapped, and so on.

This lottery was a source of considerable discussion before being held on December 1, 1969. Soon afterwards a pattern of unfairness in the results led to further publicity: those with birthdates later in the year seemed to have had more than their share of low lottery numbers and hence were more likely to be drafted.

\(^1\)http://www.amstat.org/publications/JSE/v5n2/datasets.starr.html
6.2. A MATHEMATICAL MODEL FOR CARD SHUFFLING

On January 4, 1970, the New York Times ran a long article, “Statisticians Charge Draft Lottery Was Not Random,” illustrated with a bar chart of the monthly averages. It described the way the lottery was carried out, and with hindsight one can see how the attempt at randomization broke down. The capsules were put in a box month by month, January through December, and subsequent mixing efforts were insufficient to overcome this sequencing.

The (W) article *Draft lottery (1969)* adds a little detail

The days of the year, represented by the numbers from 1 to 366 (including Leap Day), were written on slips of paper and the slips were placed in plastic capsules. The capsules were mixed in a shoebox and then dumped into a deep glass jar. Capsules were drawn from the jar one at a time.

The basic moral is that putting many items in a container and then trying to mix them up is harder than you think.

6.2 A mathematical model for card shuffling

Card shuffling provides almost the only instance of real-world mixing where mathematical probability provides a prediction for how much mixing is required. What makes analysis mathematically tractible is that a card deck is “one-dimensional” in the sense that only the order of cards matters; in contrast, a model for mixing items in a container would need to pay attention to positions in three-dimensional space.

The most common method of shuffling is the *riffle shuffle*. The deck is divided into two roughly equal halves, held in each hand with thumbs inward; moving thumbs upward releases in rapid succession the bottom cards of each half-deck, positioned so they slightly interleave and can finally be pushed together to reassemble the shuffled deck.

When first learning this shuffle, a beginner will tend to drop a large packet of cards at once from a hand. An expert will shuffle smoothly, so that typically one one card is dropped at a time, and the drops alternate between left and right hands. Intuition says the latter should be more “efficient” at mixing the deck. But consider a “perfect shuffle” in which single drops alternate precisely between hands; in this case there’s no randomness at all! And indeed it turns out that after 52 or 8 (depending on which half-deck
starts dropping first) perfect shuffles, the deck would be back in its original order (though only a few stage magicians can actually carry this out).

\[
\begin{array}{cccc}
  a & a & c & c \\
  b & b & d & d \\
  c & c & \beta & \beta \\
  d & d & \alpha & \alpha \\
  \alpha & \gamma & \delta & \delta \\
  \beta & \beta & \gamma & \gamma \\
  \gamma & \gamma & \delta & \delta \\
  \delta & \delta & \varepsilon & \varepsilon \\
  \varepsilon & \varepsilon & \zeta & \zeta \\
  \zeta & \zeta & \delta & \delta \\
  \delta & \delta & \alpha & \alpha \\
\end{array}
\]

**Figure 1.** One riffle shuffle of a 10-card deck.

The simplest mathematical model, for a single random riffle shuffle, corresponds roughly to an intermediate skill level. When dividing the deck, suppose a Binomial(52, 1/2) number of cards go to the left hand. And suppose that at each stage, the chance that the next drop is from the left or right hand is proportional to the number of cards remaining in that hand. This is called the GSR (Gilbert-Shannon-Reeds) model.

Let’s now think what exactly is it about card shuffling that we want to understand. Probability calculations involving dealing cards (e.g. the chances of being dealt particular poker or bridge hands) assume the deck is uniformly random – all 52! orderings are equally likely. But suppose we start with a deck in known order, e.g. a brand new deck, and then do \( k \) random shuffles (from the GSR model of riffle shuffles, or some other model of some physically different shuffle). Then after the \( k \) shuffles, the deck won’t have exactly uniform distribution, but we can hope its distribution is “approximately uniform” in some sense. Instead of appealing to the general ergodic theorem, we can apply the simpler theory of convergence of finite-state Markov chains to deduce that as \( k \to \infty \) the distribution gets closer and closer to the uniform distribution (for any reasonable method of shuffling). But for a particular method, it doesn’t tell us whether 3 or 3 million shuffles are needed to get close. Answering that question requires a different calculation in each model. For the GSR model, the calculation was done in Bayer and Diaconis (1992) where the following table is given.
6.3. SOME MATH ANALYSIS OF THE GSR MODEL

Table 1. Variation distance to uniformity in the GSR model of riffle shuffling.

<table>
<thead>
<tr>
<th>Number of shuffles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>d(k)</td>
<td>1.000</td>
<td>0.924</td>
<td>0.614</td>
<td>0.334</td>
<td>0.167</td>
<td>0.085</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To explain the numbers, consider an event $A$ associated with a deck of cards. The event has some probability $P(A)$ if the deck is in uniform random order, and some probability $P_k(A)$ if the deck was obtained via $k$ random (GSR model) riffle shuffles of a new deck. So treating the deck as “completely random” is an error of magnitude $|P_k(A) - P(A)|$. The variation distance is defined as the “worst case” (over events $A$) of this error:

$$d(k) = \max_A |P_k(A) - P(A)|.$$

Note that there are $52!$ possible orderings of a deck, and since an event $A$ is a subset of the set of orderings, the number of events $A$ being considered equals $2^{52!}$. So one needs some math theory (rather than brute force computation) to get a fairly simple formula for $d(k)$.

This result has entered the popular science literature under phrases like “it takes $c$ shuffles to mix a deck of cards” but this is an oversimplification in several ways:

- The model isn’t completely realistic.
- The number $7$ arises as the smallest $k$ making $d(k) < 1/2$. Because $d(7)$ is (about) $1/3$, there are events that for a truly random deck have chance $2/3$ but for a 7-times-shuffled new deck have chance $1/3$, so this doesn’t quite correspond to our intuitive notion of “well-mixed”.
- Conversely, this error applies only to very special events. For more typical events $A$ relevant to real card games, one might reasonably expect 3 or 4 shuffles to be enough.

Despite these defects with the answer “7”, no-one has come up with any more satisfactory answer to the (vague) question “how many shuffles are needed to mix a deck of cards?”

6.3 Some math analysis of the GSR model

Rather than attempting to explain the exact formula leading to the Table 1 numbers, let me outline a simpler to explain inequality

$$d(k) \leq P(52 \text{ random } k\text{-bit numbers are not all different}).$$ (6.1)
Here a 7-bit number is like 0110011, with uniform random choices of 0 or 1, and we are asking whether amongst 52 such numbers there are any two which are exactly equal.

The point of (6.1) is that it allows us to quote the textbook analysis of the birthday paradox (W), which in abstract terms says that the chance that \( m \) independent uniform picks from an \( N \)-element set is \( \approx \exp(-m^2/(2N)) \). In our setting \( m = 52 \) and \( N = 2^k \) so the bound in (6.1) becomes approximately \( \exp(-2^{10.4-k}) \) which starts getting small\(^2\) at \( k = 12 \).

The argument is interesting (to mathematicians, anyway) because it makes a connection with an apparently quite different topic, sorting.

We can rewrite the shuffle in Figure 1 by writing a 0 on each card in the left (top) pile and a 1 on each card in the right pile. The order of numbers in the shuffled deck is 1011100110, and this sequence records the outcome of that particular shuffle.

\[
\begin{array}{cccc}
- & 0 & - & 1 \\
- & 0 & - & 0 \\
- & 0 & - & 1 \\
- & 1 & - & 1 \\
- & 1 & - & 0 \\
- & 1 & - & 0 \\
- & 1 & - & 1 \\
- & 1 & - & 0 \\
\end{array}
\]

Figure 2. Re-drawing the riffle shuffle from Figure 1.

It turns out that in the GSR model, every possible sequence of 0s and 1s is equally likely. This allows us to consider (for purposes of doing the math analysis) the notion of a reversed shuffle. Interpret Figure 1 in the right-to-left sense. That is, first assign randomly 0 or 1 to each card in the right pile, then collect the “0” cards into a top pile and the “1” cards into a bottom pile, preserving relative order, and finally out the top pile on top of the bottom pile. This constitutes one reversed shuffle.

Figure 3, read left to right, shows a sequence of four reversed shuffles. Instead of generating random bits on the fly we can imagine the four random bits on each card are generated at the start; the first reversed shuffle uses

---

\(^2\)Getting the answer 12 instead of 7 doesn’t mean we’ve done something wrong, just that we used an inequality instead of an exact formula.
6.4. PROJECTS ON SHUFFLING AND MIXING

the rightmost bit, the second uses the second-rightmost bit, and so on.

\[
\begin{array}{cccccc}
1001 & 0100 & 0100 & 1001 & 0010 \\
0100 & 0010 & 1001 & 0010 & 0100 \\
1011 & 1001 & 1101 & 1011 & 1001 \\
0010 & 1011 & 0010 & 0100 & 1011 \\
1101 & 1101 & 1011 & 1101 & 1101 \\
\end{array}
\]

Figure 3. Left-to-right represents 4 radix sort steps; right-to-left represents 4 raffle shuffles.

Now regard the four bits as the binary representation of an integer. In the left deck of Figure 3 the integers are in haphazard order (from top to bottom: 9, 4, 11, 2, 13). But in the right deck they have been sorted into increasing order (2, 4, 9, 11, 13). It is easy to see this must happen – the procedure is called \textit{radix sort} (W).

So if we view Figure 3 in right-to-left order, and recall the bits are uniform random, then we are seeing 4 random raffle shuffles of a 5-card deck following the GSR model. Consider the case of \( k \) shuffles of a 52-card deck. If the 52 \( k \)-bit numbers are all distinct then the left-to-right process produces a uniform random permutation of the cards, and therefore the right-to-left process does also. Thus (the reader may wish to think how to put this verbal argument into math symbols) the only non-uniformity arises from the non-distinctness event, leading to the stated inequality (6.1).

6.4 Projects on shuffling and mixing

There is a moderate amount of technical literature giving theoretical analysis of models of random raffle shuffling and other card shuffling schemes, though little comparison with experimental data. More notably, I know of no comparable analysis of realistic mixing for physical objects essentially different from playing cards (I mean a quantitative analysis of how much mixing is required). Let me suggest a course project addressing each point.

1. One can readily find the theoretical probabilities for the different “shapes” (4333, etc) of bridge hands. But what are the experimental frequencies, if you start (as would be typical in bridge) with a deck consisting mostly of packets of 4 same-suited cards, and then shuffle \( k \) times?

\footnote{See Assaf - Diaconis - Soundararajan \textit{A rule of thumb for raffle shuffling} for related theory}
2. Imagine you want to run a lottery with say 200 names, and you do this by writing names on pieces of paper (e.g. stiff paper like business cards), putting them in some container (e.g. a cardboard box) and then just shaking, turning over, etc the box for 5 minutes. Then reach in and draw out tickets from the top. My prediction is that this does a bad job of mixing – one can do statistical tests on the results that show the order of the 200 draws are significantly non random. It would be interesting to do enough experiments to estimate how the shaking time needed to mix grows with the number of tickets.

6.5 Sorting and randomized algorithms

*Sorting* here refers to putting a list of numbers into increasing order, which is conceptually equivalent to putting names into alphabetical order. Discussing sorting in a book about chance may seem strange, because we think of sorting as a deliberate planned process. But there are several connections. Analogous to the everyday notions of tidying and making untidy as opposites, one can think intuitively of sorting and mixing as opposites. In our section 6.3 analysis of the GSR shuffle we did see a precise relation (riffle shuffle as the reversal of radix sort), though such precise relations are rare.

Second, *sorting algorithms* (W) form one of the most “classical” topics in the theory of algorithms. Suppose we know the average-case time (number of steps required) of a given sorting algorithm. What can we deduce about worst-case time? For most algorithmic problems, one can deduce very little. But sorting is special. For arbitrary data, we could first mix them into uniform random order, and then sort. The second part (by definition) takes, on average, the average-case time. But as intuition suggests, mixing can be done much more quickly than sorting, so the first randomization step takes relatively small time (for a long list). So the bottom line is that, by adding at negligible cost an “artificial” randomization, one can eliminate any possible influence of a “bad” given data configuration, at least as regards the expectation of the time required.

A *randomized algorithm* (W) in one in which randomness is “artificially” introduced within the algorithm, which typically is designed to solve a problem that does not involve randomness at all. This may sound weird but outside of algorithms it has a long history (e.g. random choice of jurors, random sampling for opinion polls). The “mix then sort” scheme above is a general randomized algorithm scheme. Let me describe what is perhaps the best known concrete instance, the *Quicksort* (W) algorithm.
6.6 QuickSort and its analysis

[I do it in class; being a standard topic treated in many texts, I will not repeat it here].

6.7 Wrap-up and further reading

Technical work on the GSR shuffling model can be found via a Google Scholar search; and some non-probabilistic aspects of shuffling are described in the monograph *Magic Tricks, Card Shuffling and Dynamic Computer Memories* by Morris. A 2011 paper\(^4\) gives an analysis of a mechanical card shuffling machine of the kind used by casinos.

The relation between random riffle shuffling and random radix sort can be placed within a more abstract setting of ergodic theory, as the time-reversal relation between two doubly-infinite stationary processes. There is another well-known artificial example of such a time-reversal relation. Here the forwards process \( (X_i) \) with values in \((0,1)\) is defined by the deterministic rule \( X_{i+1} = 2X_i \text{ modulo } 1 \), while the backwards process \( (Y_i) \) is defined by the random rule \( Y_{i+1} = \frac{1}{2}(Y_i + B_i) \) where \( B_i \) are uniform on \( \{0,1\} \). This example is sometime cited in philosophical discussions of determinism vs randomness (e.g. in *What is Random? Chance and Order in Mathematics and Life* by Beltrami) though I am rarely impressed by purely mathematical examples.

Loosely related is the topic of *(pseudo)random number generators*, which are deterministic rules for producing sequences of numbers intended to be indistinguishable from true independent random numbers. Intuitively we look for a rule that “mixes up” the input number to produce an output number, analogous to the way a riffle shuffle “mixes up” a deck of cards. See (W) for verbal discussion and e.g. *Random Number Generation and Monte Carlo Methods* by Gentle for an introductory technical account.

A link between pseudo-random number generators and physical mixing is provided by *slot machines*, which were originally purely mechanical devices operated by pulling a lever but nowadays are purely electronic, with outcome determined by the random number; the apparent reels are merely an image. The (W) article gives a detailed account.

There is a connection between random shuffling and entropy (discussed in Lecture 2). In any simple model it is easy to calculate the entropy \( H_{n1} \)

\(^4\) *Analysis of Casino Shelf Shuffling Machines* by Diaconis - Fulman - Holmes; http://arxiv.org/abs/1107.2961
of a single shuffle. An elementary fact within the “algebra of entropy”
implies that the entropy $\mathcal{E}nt_k$ of the random configuration obtained after $k$
 successive random shuffles satisfies

$$\mathcal{E}nt_k \leq k\mathcal{E}nt_1.$$  

Because the entropy of a uniform random $n$-card deck is $\log n$, the number
of shuffles required to make the deck (approximately) uniform must be$^5$ at
least (approximately) $\log n/\mathcal{E}nt_1$.

The latter style of argument is often used by theoreticians who, instead
of a 52-card deck, like to consider an $n$-card deck as $n \to \infty$. In this context,
a perhaps memorable result is that in the GSR model of riffle shuffling the
number of shuffles required (measured by variation distance) is asymptotic
to $\frac{3}{2} \log_2 n$.

$^5$Here we are quantifying “approximately” in terms of entropy rather than the variation
distance used in Table 1. Relations between different such quantifications form part of
the technical side of the topic.