25 2025-03-24 | Week 10 | Lecture 24

25.1 Shuffling Cards

I mostly talked about the paper "Trailing the Dovetail Shuffle to its Lair", by Bayer and Diaconis (1992). This is a classical paper that discusses the question, "how many times must one shuffle a deck for it to be sufficiently randomized?" The short answer is "about 7", but the mathematics to get to that answer are quite sophisticated. Some of the key ideas are as follows:

The *symmetric group of order n* is the set

$$S_n = \text{the set of permtuations of } \{1, 2, \ldots, n\} \ ,$$

where we can combine permutations by composing them together.

Formally, a function f is a *permutation* of $\{1, \ldots, n\}$ if

$$f : \{1, \ldots, n\} \to \{1, \ldots, n\}$$

and f is both injective and surjective.

Informally, a *permutation* of $\{1, \ldots, n\}$ is simply any reordering of n distinc objects.

We wrote out all the permutations of $\{1, 2, 3\}$ and showed how to compose them. Composing permutations is like doing one permutation and then doing the other. The result is always a new permutation.

Key idea: shuffling a deck of 52 cards induces a probability distribution on the set S_{52}. The shuffling procedure "fully randomizes" the deck if the resulting probability distribution is close to the uniform distribution on S_{52}.

The first 6 or so riffle shuffles don't do much to randomize, but after 7 shuffles there is an inflection point and, the distribution quickly converges to uniform.

25.2 The Coupon Collector Problem

Problem statement: Suppose a dart board is divided up into n equal-sized sections, labeled 1 through n. Each dart thrown is equally likely to hit any one of the sections. How many times must one throw the dart so that the probability of all sections being hit is at least 99%?

Solution: Let m be any positive integer greater than n. Let

$$V = \text{number of times until each section is hit at least once}$$

and let

$$A_b = \text{section } b \text{ is not hit in the first } m \text{ throws.}$$

To solve the problem, we will need to determine how big m must be so that

$$\mathbb{P}\left[V > m\right] < 0.01.$$

Then, if we throw at least that many darts, the probability of all sections being hit will be at least .99.

$$\begin{aligned}
\mathbb{P}\left[V > m\right] &= \mathbb{P}\left[A_1 \cup A_2 \cup \ldots \cup A_n\right] \\
&\leq \mathbb{P}\left[A_1\right] + \mathbb{P}\left[A_2\right] + \ldots + \mathbb{P}\left[A_n\right] && \text{(the union bound)} \\
&\leq n\left(1 - \frac{1}{n}\right)^m \\
&= n\left[\left(1 - \frac{1}{n}\right)^n\right]^{\frac{m}{n}} \\
&\approx n\left[e^{-1}\right]^{\frac{m}{n}} \\
&= ne^{-m/n}
\end{aligned}$$

Setting

$$ne^{-m/n} < 0.01$$

we can solve for m:

$$e^{-m/n} < \frac{0.01}{n}$$

$$-\frac{m}{n} < \log\left(\frac{0.01}{n}\right)$$

or

$$m > -n\log\left(\frac{0.01}{n}\right)$$

or equivalently,

$$m > n\log(100n)$$

In other words, if we throw at least $n\log(100n)$ darts, then with probability at least 99%, we will hit every section on the dartboard.