Shuffling permutations by swapping random pairs

by Pablo Rotondo

October 26, 2025

1 Introduction

A classical newbie error in programming is two shuffle an array or permutation by applying successive random transpositions as follows: [C code]

```
// initialize the array, but we will work with 0...N-1 instead in C.
for (int i = 0; i < N; i++)
    a[i] = i;
// for a number of iterations K that should depend on N
for (int t = 0; t < K; t++)
{
    int i = random(0,N-1); // pick element i uniformly from {0,...,N-1}
    int j = random(0,N-1); // independently of i
    swap(a,i,j); // swap positions i and j in a.
}
return a;
```

This procedure does not produce a uniform permutation of the array, and must be calibrated in order to approximate one by choosing K sufficiently large in order to ensure a certain approximation of a uniform permutation.

The evolution of the previous algorithm is modeled by a Markov Chain: that is, the namely the transition probabilities depend only on the current state of the permutation. We explain this. Let π_k be the permutation of $[N]:=\{1,\ldots,N\}$ at time k (a random variable), then the definition is:

- at time 0 we have $\pi_0=[1,2,3,\ldots,N]$.
- at time $t+1\geq 1$ we throw a random pair $(i,j)\in[N]\times[N]$, uniformly, and we swap the contents of positions i and j, namely $\pi_{t+1}(k):=\pi_t(k)$ for $k\neq i,j$ and $\pi_{t+1}(i):=\pi_t(j)$, $\pi_{t+1}(j):=\pi_t(i)$. In this case we write $\pi_{t+1}=(i,j)[\pi_t]$ to denote that we swap the entries.

Then $(\pi_t)_t$ is a Markov Chain in which $\Pr(\pi_{t+1}=\nu\,|\pi_t=\sigma)=0$ if ν and σ differ in more than two entries, if $\nu=\sigma$ we have $\Pr(\pi_{t+1}=\nu\,|\pi_t=\pi)=1/N$, and $\Pr(\pi_{t+1}=\nu\,|\pi_t=\sigma)=2/N^2$ if they differ in exactly two entries. Thus $p_{\sigma,\nu}=\Pr(\pi_{t+1}=\nu|\pi_t=\sigma)$ depends only on the permutations σ and ν.

Important. If the random pair (i,j) is chosen, the effect is the same as applying the transposition $(\pi(i)\quad\pi(j))$. Thus we view the problem as shuffling by transpositions: at each time we choose $(i,j)\in[N]\times[N]$ uniformly at random and apply the transposition $(i\quad j)$.

We want to study the convergence of the distribution of π_t to the uniform distribution over all permutations. Let Q_k be the distribution after k steps, while we let U be the uniform distribution. Recall that, if we define the transition matrix $P=[p_{\sigma,\nu}]_{\sigma,\nu\in S_N}$, then $Q_k=Q_0P^k$, where Q_0 is thought of as a line vector $Q_0\in\mathcal{M}_{1\times n!}(\mathbb{R})$.

Definition 1. (Total Variation Distance) *The total variation distance between to distributions P and Q over the same finite set of states S is defined by*

$$\|Q-P\|_{\mathrm{TV}}=\frac{1}{2}\sum_{s\in S}|Q(s)-P(s)|.$$

Equivalently, $\|Q-P\|_{\mathrm{TV}}=\max_{S'\subseteq S}|Q(S')-P(S')|$, the maximum difference.

We are going to prove that the cut-off happens around $K=\Theta(N\log N)$, more precisely:

Theorem 2. *As $N \to \infty$, if $K \ll (N/2)\log N$ we have* $\liminf \| Q_K - U \|_{\mathrm{TV}} \geq 1 - e^{-1}$, *while, if* $K \gg 2N\log N$ *we have* $\| Q_K - U \|_{\mathrm{TV}} \to 0$.

The first statement tells us that $\frac{N}{2}\log N$ shuffles are necessary, while the second one tells us that a bit more than $2N\log N$ are enough. In this note we give a simple proof of this fact, based on the ideas in [1]. More precise results exist. In fact, it is known that the exact cut-off happens around $\frac{1}{2}N\log N$, see [2] for more.

2 Model and definitions

As mentioned, what we have is clearly a Markov Chain. This chain is actually Ergodic and, thus the distribution Q_k converges to its unique stationary distribution, which can easily be checked to be the uniform distributon U.

2.1 Ergodic Theorem

We recall that a Markov Chain is irreducible if and only if there is a path of nonzero probability between any two states (in both senses). This ensures that all states are reachable.

Seond, a Markov Chain is aperiodic if and only if the greatest common divisor of the lengths of all cycles is one. By coprimality, this condition implies that there exists some L such that, for all $k \geq L$ there is a path of length k (with strictly positive) between every pair of states, or even from a state to itself.

Both conditions together ensure the convergence to a **unique** stationary distribution, see e.g., [4]. Observe that these conditions can be verified from the transition matrix P, and do not involve the initial distribution μ_0. Let us write $\mu_t = \mu_0 P^t$.

Theorem 3. *If a Markov Chain is both irreducible and aperiodic, then there is only one stationary distribution μ_∞, moreover, starting from any initial distribution μ_0 we have $\mu_t \to \mu_\infty$.*

In the case of our Markov Chain it is easy to verify that the uniform distribution is a stationary distribution. We remark that, since $(a,b)[s]$ is the only state such that $(a,b)[(a,b)[s]] = s$,

$$\mu_{t+1}(s) = \sum_{(a,b)\in[N]\times[N]} \mu_t((a,b)[s])\frac{1}{N^2}.$$

Trivially we have

$$\frac{1}{N!} = \frac{1}{N^2}\sum_{(a,b)\in[N]\times[N]} \frac{1}{N!},$$

the uniform distribution is stationary. Since it is clear that the chain is irreducible and aperiodic, we have the convergence to the uniform distribution.

3 The coupon collector: a lower bound

A very simple lower bound we can produce for our problem is the following: surely, if some position i has not yet been swapped we have $\pi_t(i) = i$. Such an i is said to be a fixed-point of the permutation π_t. It is important to note (and we are not going to prove it here) that the set of permutations without fixed points $\mathcal{D}_N$ satisfies $|\mathcal{D}_N| / |\mathcal{S}_N| \to e^{-1}$.

Unfortunately, if $K \leq (1-\varepsilon)\frac{N}{2}\log N$, we are going to show that $\Pr(\pi_K \in \mathcal{D}_N) \to 0$. This means that

$$\| Q_K - U \|_{\mathrm{TV}} = \max_{A \subseteq \mathcal{S}_N} |Q_K(A) - U(A)| \geq |Q_K(\mathcal{D}_N) - U(\mathcal{D}_N)| \to e^{-1}.$$

Hence we are far from converging to U.

In order to prove that there exists some i that has not yet been discovered by time K, we use the Coupon Collector Problem. The connection is simple: each $i \in \{1, \ldots, N\}$ is a coupon, and each pair (i,j) corresponds to drawing two new random coupons in the Coupon Collector Problem.

3.1 The coupon collector problem

The coupon collector problem reads as follows:

> Suppose we had to collect a collection of N distinct coupon's. At the beginning we have zero coupons. Each time we buy a coupon, we obtain a coupon among those N uniformly at random. How long does it take to complete the full collection?

The answer is actually not that difficult. Let $C = C_N$ be the necessary number of coupon's we have to buy.

Notation 4. *Let us denote by $X = \mathrm{Geom}(q)$ a generic geometric random variable that is 1 based, i.e.* $\Pr(X = j) = q\,(1-q)^{j-1}$ *for $j \in \mathbb{Z}_{>0}$. Unless otherwise stated, all of the geometrics are independent rv.*

With this notation we note that $C_N = \mathrm{Geom}(1) + \mathrm{Geom}((N-1)/N) + \ldots + \mathrm{Geom}(1/N)$. The following is an immedaite consequence of $\mathbb{E}[\mathrm{Geom}(p)] = \frac{1}{p}$ for $p > 0$.

Proposition 5. *The expected number of coupons we have to buy is $\mathbb{E}[C_N] = N \cdot H_N$ where $H_N = \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{N}$ are the Harmonic numbers. Moreover $\mathbb{E}[C_N] \sim N \log N$.*

3.2 Cocentration on the expected value

In this case not only is $\mathbb{E}[C_N] \sim N \log N$ but also C_N behaves like $N \log N$ with high probability. We will use the convergence (or equivalent) in probability.

Definition 6. *Let X_n be a sequence of positive random variables. We say that $X_n \to L$ in probability if and only if, for every fixed $\varepsilon > 0$ we have $\Pr(|X_n - L| \geq \varepsilon) \to 0$ as $n \to \infty$.*

Definition 7. *Let X_n be a sequence of positive random variables and let (e_n) be a sequence of real numbers. We say that $X_n \sim e_n$ if and only if X_n / e_n tends to one in probability.*

To prove concetration we use Chebyshev's inequality: if the random variable X has finite first and second moments, for any $\delta > 0$,

$$\Pr(|X - \mathbb{E}[X]| \geq \delta) \leq \frac{\sigma(X)}{\delta}, \qquad \sigma(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]}.$$

The following lemma is a direction application of Chebyshev's inequality by picking $\delta = \varepsilon\, \mathbb{E}[X_n]$:

Lemma 8. *Let X_n be a sequence of positive random variables such that $e_n := \mathbb{E}[X_n]$ tends to infinity. If $\mathbb{E}[X_n^2] \sim e_n^2$ we have that $X_n \sim e_n$ in probability.*

To better deal with the moments of our random variables C_N, we consider the probability generating functions. Observe that if X is a random variable taking values in the positive integers:

$$F_X(z) = \sum_k \Pr(X = k) z^k,$$

and then $F_X'(1) = \mathbb{E}[X]$ and $F_X''(1) = \mathbb{E}[X(X-1)]$.

In the case of C_N we simply have, due to the independence of each of the geometric random variables,

$$F(z) = F_N(z) = \prod_{i=1}^{N-1} \frac{z p_i}{1 - z(1 - p_i)}.$$

We show that we have the concentration in a more general setting for sums of geometric random variables:

Lemma 9. *For each n, define $(p_n(i))$ for $i = 1, \ldots, m(n)$ satisfying $p_n(i) \in (0,1]$, where we suppose $m(n) \to \infty$. Let $S_n = \sum_{i=1}^{m(n)} \mathrm{Geom}(p_n(i))$, then $S_n \sim \mathbb{E}[S_n]$ in probability.*

Proof. The PGF of S_n is

$$F(z) = F_n(z) = \prod_{i=1}^{m(n)} \frac{z p_n(i)}{1 - z(1 - p_n(i))}.$$

Observe that $F'(z) = m(n)\frac{F(z)}{z} + F(z)\sum_{i=1}^{m(n)} \frac{(1-p_n(i))}{1 - z(1-p_n(i))}$. Thus $F'(1) = \sum_{i=1}^{m(n)} \frac{1}{p_n(i)}$. We note that the expected value is $\mathbb{E}[S_n] = \sum_{i=1}^{m(n)} \frac{1}{p_n(i)} \geq m(n) \to \infty$.

Differentiating again,

$$F''(1) = m(n)\sum_{i=1}^{m(n)} \frac{1}{p_n(i)} - m(n) + \left(\sum_{i=1}^{m(n)} \frac{1}{p_n(i)}\right)\left(\sum_{i=1}^{m(n)} \frac{1-p_n(i)}{p_n(i)}\right) + \sum_{i=1}^{m(n)} \frac{(1-p_n(i))^2}{p_n(i)}.$$

Which we can simplify to:

$$F''(1) = -m(n) + \left(\sum_{i=1}^{m(n)} \frac{1}{p_n(i)}\right)^2 + \sum_{i=1}^{m(n)} \frac{(1-p_n(i))^2}{p_n(i)}.$$

Here we note that $\sum_{i=1}^{m(n)} \frac{(1-p_n(i))^2}{p_n(i)} \leq \sum_{i=1}^{m(n)} \frac{1}{p_n(i)} = o\left(\left(\sum_{i=1}^{m(n)} \frac{1}{p_n(i)}\right)^2\right)$ and similarly $m(n) = o\left(\left(\sum_{i=1}^{m(n)} \frac{1}{p_n(i)}\right)^2\right)$ too. Thus $\mathbb{E}[S_n(S_n-1)] = \mathbb{E}[S_n^2] - \mathbb{E}[S_n] = F''(1) \sim \left(\sum \frac{1}{p_n(i)}\right)^2 = (\mathbb{E}[S_n])^2$. □

Corollary 10. *For any fixed* $\varepsilon > 0$, $\Pr((1-\varepsilon)N\log N \leq C_N \leq (1+\varepsilon)N\log N) \to 1$.

3.3 Coupon collector and fixed points in the process

In the case of the Coupon Collector much more is known. The following is a classical inequality:

Proposition 11. $\Pr(C_N \geq N\log N + \theta N) \leq e^{-\theta}$ *for every* $\theta \in \mathbb{R}$.

Proof. Remark that $\Pr(C_N \geq M)$, for M integer, is the probability that at least one of the coupons is missing at time M. Let $A_i(M)$ be the event that coupon i is missing. Observe that $\Pr(A_i(M)) = \left(1 - \frac{1}{N}\right)^M$.

By the union bound:

$$\Pr(C_N \geq M) = \Pr\left(\bigcup_{i=1}^{N} A_i(M)\right) \leq \sum_{i=1}^{N} \Pr(A_i(M)) = N \times \left(1 - \tfrac{1}{N}\right)^M .$$

Since $1 + x \leq e^x$ for all $x \in \mathbb{R}$, we deduce $N \times \left(1 - \frac{1}{N}\right)^M \leq N e^{-M/N}$. Being C_N an integer, we deduce that $\Pr(C_N \geq N\log N + \theta N) = \Pr(C_N \geq \lceil N\log N + \theta N \rceil)$ and the result follows. □

Actually, this inequality can be made more precise by using more terms from the so-called Bonferroni inequalities (the partial sums of the inclusion-exclusion provide bounds). That is the idea behind the proof of the following result from Erdös and Rényi [3].

Theorem 12. $\Pr(C_N < N\log N + \theta N) \to \exp(-e^{-\theta})$ *as* $N \to \infty$, *for every* $\theta \in \mathbb{R}$.

The distribution function $\theta \mapsto \exp(-e^{-\theta})$ is known as a Gumbel distribution.

4 Uniform stopping rule: an upper bound

4.1 Strong uniform time and convergence

Aldous and Diaconis introduced in [1] the concept of *strong uniform time*. A uniform stopping time T for (π_t), is a stopping time, i.e., $\{T \leq t\}$ can be determined from our knowledge at time t, such that $\Pr(\pi_t = \sigma \mid T = t) = 1/N!$, i.e., that the distribution when the stopping time T tells us to stop $\mu(\sigma) = \Pr(\pi_t = \sigma \mid T = t)$ is uniform.

Of course, here we have adapted the definition to our context, but it extends easily to other contexts. The key interest of a strong uniform time is the following bound (see Lemma 1 in [1]):

Proposition 13. *Let T be a uniform stopping time and let U be the uniform distribution, then* $\| Q_k - U \|_{\mathrm{TV}} \leq \Pr(T > k)$.

That is, the total variation distance between Q_k, the distribution of π_k, and the uniform distribution U is at most $\Pr(T > k)$.

4.2 Perfectly stopping our Markov Chain

We define an increasing family of sets as follows:

- Let $S_0 = \{1\}$ (the choice of 1 is arbitrary and not important).
- Given S_t we define S_{t+1} as follows. First, $S_t \subseteq S_{t+1}$. If the next random pair (a_{t+1}, b_{t+1}) satisfies $a_{t+1} \in S_t$ then b_{t+1} is added to S_{t+1}. Else if $a_{t+1} = b_{t+1}$, then add a_{t+1} to S_{t+1}.

The invariant is the following: the restriction of π_t to S_t is a uniform permutation. This is easily proven by induction. As S_t increases in size, at some point we obtain $S_t = [N]$.

We define our stopping time as follows:

$$T(\omega) = \inf \{t: S_t(\omega) = [N]\}.$$

Proposition 14. *[T is a strong uniform time]* $\Pr(\pi_k = \sigma \,|T = k)$ *is uniformly distributed in* σ.

Proof. Suppose that $\pi_k|_{S_k}$ is a uniform permutation for some k, we will show that this also holds for $k+1$. This is obvious if $S_k = S_{k+1}$, so suppose S_{k+1} has some new element j.

This means that the random pair (a_{k+1}, b_{k+1}) was either (a, j) with a in S_k or it was (j, j). Remark that all of these transpositions have equal probability of being produced. Most importantly, as j has the same probability of being swapped with any element in $S_{k+1} = S_k \cup \{j\}$ we conclude that the resulting permutation (given that $\pi_k|_{S_k}$ is a uniform permutation for some k) is also uniform $\pi_{k+1}|_{S_{k+1}}$. □

Remark 15. It is possible to produce a perfect random permutation by keeping S_t and thus perfectly stopping our permutation. Of course, no one does this in practice, and there are much better ways to produce permutations.

4.3 Concentration of the stopping time

In this section we prove that the stopping time is concentrated around its expected value $\mathbb{E}[T] \sim 2N \log N$. Let us start by calculating the expected value. The probability of discovering a new number at time $t+1$, given that $|S_t| = i$ is given by

$$p_i = \frac{i\,(N-i) + (N-i)}{N^2} = \frac{(i+1)(N-i)}{N^2},$$

independently of the past.

Thus we have

$$T = \sum_{i=1}^{N-1} \mathrm{Geom}(p_i)$$

for certain independent geometric random variables.

Proposition 16. *The expected value satisfies* $\mathbb{E}[T] \sim 2N \log N$ *as* $N \to \infty$.

Proof. It suffices to check that $\mathbb{E}[T] = \sum_{i=1}^{N-1} \frac{1}{p_i} = \sum_{i=1}^{N-1} \frac{N^2}{(i+1)(N-i)} = N \sum_{i=1}^{N-1} \left(\frac{1}{i+1} + \frac{1}{N-i}\right)$. Here we remark that $H_{N-1} = \sum_{i=1}^{N-1} \frac{1}{N-i}$, while $\sum_{i=1}^{N-1} \frac{1}{i+1} = H_N - 1$. □

Now, applying Lemma 9 we deduce the following corollary.

Corollary 17. *For any fixed* $\varepsilon > 0$, $\Pr(T \geq (1+\varepsilon)2N\log N) \to 0$.

Finally, Proposition 13 proves that $\| Q_K - U \|_{\mathrm{TV}} \to 0$ for $K \geq (2+\varepsilon)N\log N$, for any $\varepsilon > 0$, completing the proof of our theorem.

5 Conclusions

The method proposed at the beginning is not very good; it requires $\Theta(N\log N)$ random numbers from $[N]$. A simple algorithm that is efficient, and a perfect simulation, is the following [known as Knuth's shuffle or the Fisher-Yates shuffle]

```
// initialize the array, but we will work with 0...N-1 instead.
for (int i = 0; i < N; i++)
    a[i] = i;
// for each position choose one of the not-chosen elements
for (int i = 0; i < N; i++)
{
    int pos = random(i,N-1); // pick uniformly at random from {i,...,N-1}
    swap(a,i,pos); // swap positions i and pos in a.
}
return a;
```

Bibliography

[1] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *The American Mathematical Monthly*, 93(5):333–348, 1986.

[2] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57:159–179, 1981.

[3] Pál Erdős and Alfréd Rényi. On a classical problem of probability theory. *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, 6(1-2):215–220, 1961.

[4] Olle Häggström. Finite markov chains and algorithmic applications: simulated annealing. 2002.