

WORKING PAPERS

N° 1436

May 2023

“On a Markov construction of couplings”

Persi Diaconis and Laurent Miclo

On a Markov construction of couplings

Persi Diaconis[†] and Laurent Miclo[‡]

[†] Department of Mathematics
Department of Statistics
Stanford University

[‡] Toulouse School of Economics
Institut de Mathématiques de Toulouse
CNRS and University of Toulouse

Abstract

For $N \in \mathbb{N}$, let π_N be the law of the number of fixed points of a random permutation of $\{1, 2, \dots, N\}$. Let \mathcal{P} be a Poisson law of parameter 1. A classical result shows that π_N converges to \mathcal{P} for large N and indeed in total variation

$$\|\pi_N - \mathcal{P}\|_{\text{tv}} \leq \frac{2^N}{(N+1)!}$$

This implies that π_N and \mathcal{P} can be coupled to at least this accuracy. This paper constructs such a coupling (a long open problem) using the machinery of intertwining of two Markov chains. This method shows promise for related problems of random matrix theory.

Keywords: Uniform random permutation, number of fixed point(s), Poisson approximation, Markov approach, intertwining, coupling.

MSC2020: primary: 60J10, secondary: 05A05, 60E15, 60J22, 60J80, 37A25.

1 Introduction

For $N \in \mathbb{N} := \{1, 2, \dots\}$, let π_N be the law of the number of fixed points of a random permutation of $\{1, 2, \dots, N\}$. Let \mathcal{P} be the Poisson law of parameter 1. A classical result, see de Montmort [5], shows that π_N converges to \mathcal{P} for large N . Indeed it is well-known (and estimates of the same order are proved below) that in total variation

$$\frac{N}{N+2} \frac{2^{N+1}}{(N+1)!} \leq \|\pi_N - \mathcal{P}\|_{\text{tv}} \leq \frac{2^{N+1} - 1}{(N+1)!} \quad (1)$$

The total variation distance can be realised by a coupling of π_N and \mathcal{P} , see e.g. Proposition 4.7 of Levin, Peres and Wilmer [18], and it has been a long open problem to give an explicit realization of such a coupling. The super-exponential errors bounds in (1) occur in other problems such as the number of k -cycles in a random permutation, which has a limiting Poisson distribution of parameter $1/k$ with super-exponential error. Similar results hold for the trace of powers of random matrices for the compact classical groups O_N , U_N and SP_{2N} , see e.g. Courteaut, Johansson and Lambert [3]. The method introduced here shows promise for finding couplings for these problems. For a history of Montmort's theorem, see Takacs [24]. For extensions and a recent literature review, see Diaconis and Fulman and Guralnick [9]. At the end of this introduction we will present several attempts, successful as well as unsuccessful, to get a proof by coupling of (1).

To present our approach, for any $N \in \mathbb{N}$ and any permutation σ in the symmetric group \mathcal{S}_N , denote $\eta_1(\sigma)$ the number of fixed point of σ :

$$\eta_1(\sigma) := |\{x \in \llbracket N \rrbracket : \sigma(x) = x\}|$$

(where $\llbracket N \rrbracket := \{1, 2, \dots, N\}$ and more generally, for any $n \leq n' \in \mathbb{Z}_+ := \{0, 1, 2, \dots\}$, we write $\llbracket n, n' \rrbracket := \{n, n+1, \dots, n'\}$). The number $\eta_2(\sigma)$ of 2-cycles of σ will also play an important role:

$$\eta_2(\sigma) := |\{(x, y) \in \llbracket N \rrbracket^2 : x < y, \sigma(x) = y \text{ and } \sigma(y) = x\}|$$

Let ν_N stands for the uniform distribution on \mathcal{S}_N , so that π_N is its image by η_1 on \mathbb{Z}_+ . To simplify the notation, we will often drop the exponent N when referring to these probability measures. As mentioned in (1), the fixed-point law π is very close to the Poisson distribution \mathcal{P} . The bounds in (1) are for instance recorded in (1.11) page 15 of Arratia, Barbour and Tavaré [1] and are deduced from computations of David and Barton [4] using properties of alternating series with decreasing terms coming from the following traditional facts.

We have

$$\forall x \in \llbracket 0, N \rrbracket, \quad \pi(x) = \frac{D_{N-x}}{(N-x)!} \frac{1}{x!} \quad (2)$$

where for any $n \in \mathbb{Z}_+$, D_n stands for the number of derangements from \mathcal{S}_n , namely the permutations of \mathcal{S}_n without fixed point (with the convention that $D_0 = 1$). The formula due to de Montmort [5] gives the number of derangements:

$$\forall n \in \mathbb{N}, \quad D_n = n! \sum_{k=0}^n \frac{(-1)^k}{k!} \quad (3)$$

leading to the explicit formula:

$$\forall x \in \llbracket 0, N \rrbracket, \quad \pi(x) = \frac{1}{x!} \sum_{k=0}^{N-x} \frac{(-1)^k}{k!} \quad (4)$$

As announced above, our purpose is to deduce bounds on $\|\pi - \mathcal{P}\|_{\text{tv}}$, of the same logarithmic order as that of (1). Here is a sketch of the proof. We use a random transposition to construct a Markov chain on the symmetric group \mathcal{S}_N . Then the intertwining-lumping procedure presented in Section 2 and some fiddling around is used to construct a monotone birth-and-death chain with the fixed point distribution π as reversible distribution. A similar construction gives a monotone birth-and-death chain with a Poisson stationary distribution. Revisiting next the classical coupling of two monotone birth-and-death chains leads to our bound. In more detail the intertwining-lumping construction produces the penta-diagonal Markov kernel P on

$$V := \llbracket 0, N-2 \rrbracket \sqcup \{N\}$$

given by

$$\forall x \in V, \quad \begin{cases} P(x, x-1) &= \frac{x(N-x)}{N(N-1)} \\ P(x, x-2) &= \frac{x(x-1)}{N(N-1)} \\ P(x, x+1) &= \frac{N-x-2p(x)}{N(N-1)} \\ P(x, x+2) &= \frac{2p(x)}{N(N-1)} \\ P(x, x) &= 1 - P(x, x-1) - P(x, x-2) - P(x, x+1) - P(x, x+2) \end{cases} \quad (5)$$

where

$$x \in V, \quad p(x) := \mathbb{E}_\nu[\eta_2 | \eta_1 = x] \quad (6)$$

(the conditional expectation is with respect to the uniform measure ν on \mathcal{S}_N). As explained in Section 2 below, this chain is a projection of Markov chains on conjugacy classes derived from multiplication from random transpositions.

Note that P does not allow to get out of V : we have $P(0, -1) = P(0, -2) = P(1, -1) = P(N, N-1) = 0$ and $P(N-3, N-1) = P(N-2, N-1) = P(N, N+1) = P(N, N+2) = 0$. For the latter equalities, we need the following observations about p : obviously we have $p(N) = 0$ and the value $p(N-2)$ is 1, since knowing that $\eta_1 = N-2$, we necessarily have $\eta_2 = 1$. Similarly, the value $p(N-3)$ is 0, since knowing that $\eta_1 = N-3$, we necessarily have $\eta_2 = 0$ (and the number of 3-cycles is equal to 1).

By our construction, the probability measure π will naturally appear to be reversible for the Markov kernel P . Furthermore the reversibility of P (without even knowing the reversible probability) in conjunction with $p(N) = p(N-3) = 0$ and $p(N-2) = 1$ are sufficient to determine P and by consequence the other values of p and those of π . These features can be translated into convenient estimates on p , leading to quantitative couplings of the Markov chains whose transitions are dictated by P with other Markov chains whose invariant measure is the conditioning of \mathcal{P} on V (more conveniently, we will restrict our attention to the state space $\llbracket 0, N-4 \rrbracket \subset V$). These bounds are carried out in Section (4) and we will deduce the convergence

$$\lim_{N \rightarrow \infty} \frac{1}{N \ln(N)} \ln(\|\pi - \mathcal{P}\|_{\text{tv}}) = -1 \quad (7)$$

of the right logarithmic order.

Let us now list several attempts to prove (1) via coupling arguments, as well as some remarks.

1.1 A failed effort

This section records a natural coupling, indeed one that extends to all the classical compact groups and their Weyl groups. By the law ‘‘natural yields right’’, this should work to give good error bounds, alas it doesn’t!

Let $(X_n)_{n \in \mathbb{N}}$ be independent $\{0, 1\}$ -valued random variables with

$$\forall n \in \mathbb{N}, \quad \mathbb{P}[X_n = 1] = \frac{1}{n} = 1 - \mathbb{P}[X_n = 0]$$

Define for all $N \in \mathbb{N}$,

$$\begin{aligned} S_N &:= X_1 X_2 + X_2 X_3 + \cdots + X_{n-1} X_N + X_N \\ S_\infty &:= X_1 X_2 + X_2 X_3 + \cdots \end{aligned}$$

In the unpublished paper of Diaconis and Mallows [6], recorded in Diaconis and Forrester [8], it is shown that for any $k \in \mathbb{Z}_+$,

$$\mathbb{P}[S_N = k] = \pi_N(k) \quad \text{and} \quad \mathbb{P}[S_\infty = k] = \mathcal{P}(k)$$

Thus the joint law of (S_N, S_∞) makes a natural coupling. Alas, $S_\infty - S_N = X_N(X_{N+1} - 1) + X_{N+1}X_{N+2} + X_{N+2}X_{N+3} + \cdots$ has typical distance of order $1/N$. For more background and details, see Diaconis and Forrester [8].

1.2 A successful and strange coupling from computer science

Jim Pitman has explained a fascinating construction of a super exponential coupling due to computer scientists Duchon and Duvignau [13] and Duchon and Duvignau [14]. Pitman's development of these ideas is unpublished [22]. We thank him for permission to state his results. The construction calls for a countable collection $(U_n)_{n \in \mathbb{N}}$ of independent random variables uniformly distributed on $[0, 1]$. Define

$$\begin{aligned} S &:= \min\{n \geq 1 : U_n < U_{n+1}\}, & \text{time of first ascent} \\ T &:= \min\{n \geq 2 : U_n > \max(U_{n-1}, U_{n+1})\}, & \text{time of first peak} \\ M &:= S - \delta_{T-S \text{ is odd}} \end{aligned}$$

Theorem 1 *The distribution of the random variable M is the Poisson law of parameter 1.*

Define further for fixed $N \in \mathbb{N}$,

$$\begin{aligned} S_N &:= \min(S, N) \\ T &:= \min(T, N) \\ M_N &:= S_N - \delta_{T_N - S_N \text{ is odd}} \end{aligned}$$

Theorem 2 *The random variable M_N has the law of the number of fixed points of a random permutation of $\llbracket N \rrbracket$.*

As a consequence of the two previous theorems, we get

Corollary 3 *For any $N \in \mathbb{N}$, we have*

$$\|\mathcal{L}(M) - \mathcal{L}(M_N)\|_{\text{tv}} \leq \mathbb{P}[T > N] \leq \frac{2^N}{(N+1)!}$$

This result seems magical and the present paper records an effort to find a proof using more standard tools which might permit generalization. We also hope to study it on its own at least to generalize to the law of the number of k -cycles.

1.3 Unstability of the super-exponential bounds

The previous super-exponential bounds are delicate. Consider for example the number of fixed points in the first $N - 1$ places of a random permutation of $\llbracket N \rrbracket$. This quantity too has an approximate Poisson distribution of parameter 1 but the total variation distance between these two laws is of order $1/N$.

Similarly, for any $\theta \in (0, 1)$, the number of fixed points in places $\llbracket \theta N \rrbracket$ has a Poisson law of parameter θ as limiting law. Indeed the point process on $[0, 1]$ which has an event at k/N if and only if a random permutation σ of \mathcal{S}_N satisfies $\sigma(k) = k$ is well approximated by a unit rate Poisson process. But these approximations are only accurate up to order $1/N$.

1.4 Equality of first N moments

For $N \in \mathbb{N}$, consider two random variables X_N and X_∞ respectively distributed according to π_N and \mathcal{P} . The high order of contact between these two laws can be captured by moments. Indeed Diaconis and Shahshahani [12] show

$$\forall k \in \llbracket 0, N \rrbracket, \quad \mathbb{E}[X_N^k] = \mathbb{E}[X_\infty^k]$$

Similar results hold for the joint mixed moments of the number of k -cycles and for compact classical groups.

1.5 The Markov approach

It is related to Stein's method, see e.g. Diaconis and Holmes [10] or Section 4 of Chatterjee, Diaconis and Meckes [2], but the underlying philosophy is quite old. Assume we would like to investigate some features of a given probability measure π . The Markov approach consists in introducing and studying a Markov process (in continuous time) or chain (in discrete time) encapsulating the "relevant characteristics" of the underlying state space and admitting π as invariant probability (sometimes no effort is required in this introduction, as π is already defined as an invariant probability). An example of this situation is the investigation of absence of phase transition, exponential decay of correlations, or analyticity of correlation of Gibbs measures, which was done via the use of stochastic Ising processes leaving these Gibbs measures invariant, see Holley and Stroock [15] or Chapter 4 of the book of Liggett [19]. Our goal here is to give a new illustration of this Markov approach by recovering the right order of (1).

The plan of the paper is as follows. In the next section we present a general procedure producing a Markov chain by projection of another Markov chain. Reversibility is preserved by such projections. In Section 3, the transposition random walk on \mathcal{S}_N is projected in this way through η_1 to get the Markov kernel P on V . In Section 4, we deduce the a priori bounds on p that are applied in Section 5 to control our couplings of Markov chains, leading to desired upper bound on the approximation of π by \mathcal{P} . In a spirit similar to that of Section 4, in Appendix A, we directly recover (4), giving an alternative proof to the classical inclusion-exclusion argument. In Appendix B, some complements are given about the conditional expectation p .

Acknowledgments: We thank Jim Pitman for telling us about Section 1.2. Diaconis is funded by NSF grant 1954042. Miclo is funded by grants ANR-17-EURE-0010 and AFOSR-22IOE016.

2 Projections of Markov chains

We present in this short section a general procedure of projection of Markov chains. We will restrict our attention to finite state spaces to simplify the exposition and since latter we will work only with such sets, but the underlying principle is much more general.

Consider a Markov chain $(X, Y) := (X_n, Y_n)_{n \in \mathbb{Z}_+}$ taking values in a product state space $V \times W$. Assume that V and W are finite and that the transition matrix Q of (X, Y) is irreducible. Denote by μ its invariant measure. Consider $r_1 : V \times W \rightarrow V$ and $r_2 : V \times W \rightarrow W$ the canonical projections and let $\mu_1 := r_1(\mu)$ be the first marginal distribution of μ . Denote by $\mu_{1,2}$ the Markov kernel from V to W corresponding to the conditional distribution of r_2 knowing r_1 . So we have the decomposition

$$\forall (x, y) \in V \times W, \quad \mu(x, y) = \mu_1(x)\mu_{1,2}(x, y)$$

Consider the Markov kernel P given on V via

$$\forall x, x' \in V \quad P(x, x') := \sum_{y, y' \in W} Q((x, y), (x', y')) \mu_{1,2}(x, y)$$

and the Markov kernel Λ from $V \times W$ to V given by

$$\forall (x, y) \in V \times W, \forall x' \in V, \quad \Lambda((x, y), x') := \mu_1(x')$$

Lemma 4 *We have the intertwining relation*

$$Q\Lambda = \Lambda P$$

Proof

On one hand, Λ can be identified with μ_1 , so that $Q\Lambda = \mu_1$.

On the other hand, we have for any $(x, y) \in V \times W$ and $x' \in V$,

$$\begin{aligned} \Lambda P((x, y), x') &= \mu_1 P(x') \\ &= \sum_{x'' \in V} \mu_1(x'') P(x'', x') \\ &= \sum_{x'' \in V} \mu_1(x'') \sum_{y', y'' \in W} Q((x'', y''), (x', y')) \mu_{1,2}(x'', y'') \\ &= \sum_{y' \in W} \sum_{(x'', y'') \in V \times W} \mu(x'', y'') Q((x'', y''), (x', y')) \\ &= \sum_{y' \in W} \mu(x', y') \\ &= \mu_1(x') \end{aligned}$$

namely

$$\Lambda P = \mu_1 = Q\Lambda$$

■

In particular, $\mu\Lambda$ is invariant for P , i.e. μ_1 is invariant for P (in fact this is just the above proof). We also have:

Lemma 5 *Assume that μ is reversible for Q , then μ_1 is reversible for P .*

Proof

Consider $f, g \in \mathbb{R}^V$. We have

$$\begin{aligned} \mu_1[fP[g]] &= \sum_{x \in V} \mu_1(x) f(x) \sum_{y \in W} Q[g \circ r_1](x, y) \mu_{1,2}(x, y) \\ &= \sum_{(x, y) \in V \times W} f \circ r_1(x, y) Q[g \circ r_1](x, y) \mu(x, y) \\ &= \mu[f \circ r_1 Q[g \circ r_1]] \\ &= \mu[g \circ r_1 Q[f \circ r_1]] \\ &= \mu_1[gP[f]] \end{aligned}$$

■

The construction above corresponds to a lumping procedure. More generally, let W be a finite (or denumerable) set and $(Q(w, w'))_{w, w' \in W}$ be a Markov kernel on W admitting $(\pi(w))_{w \in W}$ as stationary distribution. Given a partition of the state space $W = \bigsqcup_{v \in V} A_v$ into non-empty subsets, reporting which A_v contains the current state of a Markov chain associated to Q gives a ‘‘lumped process’’. As is well-known, see e.g. Theorem 6.3.2 of Kemeny and Snell [17] or Pang [21], this may not be a Markov chain. The analogous projected Markov kernel $(P(v, v'))_{v, v' \in V}$ on V can be defined as

$$\forall v, v' \in V, \quad P(v, v') := \sum_{w \in A_v, w' \in A_{v'}} \frac{\pi(w)}{\pi(A_v)} Q(w, w')$$

Arguing as above, the probability measure $(\pi(A_v))_{v \in V}$ is invariant for P (and reversible when π is reversible for Q). Defining

$$\forall w \in W, \forall v \in V, \quad \Lambda(w, v) := \pi(A_v)$$

we get the intertwining relation $Q\Lambda = \Lambda P$. If the classical Dynkin condition holds, namely for any $v, v' \in V$, $Q(w, A_{v'})$ does not depend on the choice of $w \in A_v$, then the projected chain P agrees with the usual lumped chain.

3 A penta-diagonal and two birth and death Markov chains

Here we apply the abstract projection scheme of the previous section in the setting of the symmetric group \mathcal{S}_N . It is related to Chapter 12 of Stein’s book [23], which studies the law of the numbers of the cycles of length l , for all $l \in \llbracket N \rrbracket$, under the uniform distribution on \mathcal{S}_N , using a random transposition to build a reversible Markov chain.

Consider the transposition random walk on the symmetric group \mathcal{S}_N , whose transition matrix T is given by

$$\forall \sigma, \sigma' \in \mathcal{S}_N, \quad T(\sigma, \sigma') = \begin{cases} \frac{2}{N(N-1)} & , \text{ if there exists a transposition } \tau \text{ such that } \sigma' = \tau\sigma \\ 0 & , \text{ otherwise} \end{cases}$$

(where permutations are seen as bijective mappings from $\llbracket N \rrbracket$ and the product corresponds to the composition).

The Markov kernel T is reversible with respect to the uniform probability distribution ν on \mathcal{S}_N .

Generalizing η_1 and η_2 , for any $l \in \llbracket N \rrbracket$ and $\sigma \in \mathcal{S}_N$ define $\eta_l(\sigma)$ as the number of cycles of order l in σ (singleton cycles corresponding to fixed points). In particular we have

$$\forall \sigma \in \mathcal{S}_N, \quad \eta_1(\sigma) + 2\eta_2(\sigma) + \dots + N\eta_N(\sigma) = N$$

Let $(\sigma(n))_{n \in \mathbb{Z}_+}$ be a Markov chain with transitions dictated by T and denote

$$\forall n \in \mathbb{Z}_+, \quad \eta(n) := (\eta_l(\sigma(n)))_{l \in \llbracket N \rrbracket}$$

It is well-known that $\eta := (\eta(n))_{n \in \mathbb{Z}_+}$ is also a Markov chain whose transition matrix is denoted Q and is reversible with respect to μ the image of ν in the mapping $\mathcal{S}_N \ni \sigma \mapsto (\eta_l(\sigma))_{l \in \llbracket N \rrbracket}$. Indeed this is the classical coagulation-fragmentation chain of statistical mechanics, see Diaconis, Mayer-Wolf, Zeitouni and Zerner [11].

The Markov chain η can be written under the form (X, Y) with

$$\begin{aligned} X &:= \eta_1 \\ Y &:= \eta_{[2, N]} := (\eta_2, \eta_3, \dots, \eta_N) \end{aligned}$$

We are thus in position to apply Lemmas 4 and 5.

Our next goal is to describe the corresponding Markov kernel P . Note that the corresponding state space is

$$V = \llbracket 0, N \rrbracket \setminus \{N-1\} \quad (8)$$

already met in the introduction (it is not possible for a permutation to have $N-1$ fixed points).

Consider a permutation $\sigma \in \mathcal{S}_N$. Denote f_1, f_2, \dots, f_k its fixed points (so that $\eta_1(\sigma) = k$) and let C_1, C_2, \dots, C_l be the other cycles of σ .

Consider a transposition $\tau = (i, j)$. Let us describe $\eta_1(\sigma')$ with $\sigma' := \tau\sigma$.

- If both i and j are fixed points of σ , then the fixed points of σ' are the elements of $\{f_1, f_2, \dots, f_k\} \setminus \{i, j\}$ and its non-singleton cycles are C_1, C_2, \dots, C_l and (i, j) . Thus we have $\eta_1(\sigma') = \eta_1(\sigma) - 2$.

- If i is a fixed point of σ and $j \in C_r$, with $r \in \llbracket l \rrbracket$, then the fixed points of σ' are the elements of $\{f_1, f_2, \dots, f_k\} \setminus \{i\}$ and its non-singleton cycles are C_m with $m \neq r$ in addition to a new cycle containing C_r and j . Thus we have $\eta_1(\sigma') = \eta_1(\sigma) - 1$.

- If $i \in C_r$ and $j \in C_s$ with $r \neq s$, then the cycles and fixed points of σ' are the same as those of σ , except that C_s and C_r are merged into a new cycle. In particular we have $\eta_1(\sigma') = \eta_1(\sigma)$.

- The last situation is when i and j belong to the same cycle C_r . We consider three subcases:

- When $C_r = (i, j)$, then the fixed points of σ' are $\{f_1, f_2, \dots, f_k, i, j\}$ and its non-singleton cycles are the C_s , for $s \in \llbracket l \rrbracket \setminus \{r\}$. We deduce $\eta_1(\sigma') = \eta_1(\sigma) + 2$.

- When there exists $x \in C_r$ such that $\{i, j\} = \{x, \sigma(x)\}$, assume for instance that $i = x$ and $j = \sigma(x)$ and $C_r \neq (i, j)$. Then i is a new fixed point of σ' and its non-singleton cycles are the same as those of σ , except that the point i has been removed from C_r . We deduce $\eta_1(\sigma') = \eta_1(\sigma) + 1$.

- When there does not exist $x \in C_r$ such that $\{i, j\} = \{x, \sigma(x)\}$ (in particular the cardinal of C_r is at least 4), then σ' has the same fixed points as σ and the only difference in its non-singleton cycles is that C_r has been divided into two new non-singleton cycles. We deduce $\eta_1(\sigma') = \eta_1(\sigma)$.

Integrating these observations with respect to τ uniformly distributed among all transpositions, we end up with the kernel P given in (5), with

$$p(x) = \int \eta_2 \mu_{1,2}(x, d\eta_{\llbracket 2, N \rrbracket})$$

namely the mean of η_2 knowing $\eta_1 = x$ when $\eta_{\llbracket 1, N \rrbracket}$ is distributed according to μ (the above integral is in fact a sum, but the integral notation is more convenient). Note that this formulation is equivalent to (6).

The distribution π of the number of fixed points of the uniform permutation is equal to μ_1 , with the notation of Section 2. According to Lemma 5, π is reversible for P , as announced in the introduction.

In the sequel it will sometimes be more convenient to work with tri-diagonal kernels than with the penta-diagonal kernel P , so let us extract two birth and death kernels from P .

The first one, denoted \tilde{P} , is given by

$$\forall x \neq y \in V, \quad \tilde{P}(x, y) := \frac{1}{N(N-1)} \begin{cases} x(N-x) & , \text{ if } x \neq N \text{ and } y = x-1 \\ N-x-2p(x) & , \text{ if } x \neq N-2 \text{ and } y = x+1 \\ 2 & , \text{ if } x = N-2 \text{ and } y = N \\ N(N-1) & , \text{ if } x = N \text{ and } y = N-2 \\ 0 & , \text{ otherwise} \end{cases} \quad (9)$$

This Markov kernel is obtained by removing all transitions of the form $(x, x+2)$ and $(x, x-2)$ from P , except for $(N-2, N)$ and $(N, N-2)$ (because $N-1$ is not a value taken by η_1), and putting their weights to the diagonal. For the corresponding Markov chains, it amounts to forbid the jumps of size two and keep the current position instead (except for the transitions between $N-2$ and N).

From the fact that π is reversible for P , we deduce that π is also reversible for \tilde{P} , since the property of being reversible is preserved by removing transitions (when the transitions in both directions along an edge are removed together). As announced, \tilde{P} corresponds to a birth-and-death Markov transition on V .

The second birth-and-death Markov transition \hat{P} will be useful in Appendix A. It is obtained by ordering V as $N - 3, N - 5, \dots, 3, 1, 0, 2, 4, \dots, N - 2, N$ when N is even. When N is odd, rather order V as $N - 3, N - 5, \dots, 2, 0, 1, 3, \dots, N - 2, N$, the following construction leads to similar results in this case, so let us only consider the situation where N is even.

Thus we define for $i \in \llbracket 0, N - 1 \rrbracket$,

$$z_i := \begin{cases} 2i & , \text{ if } i \in \llbracket 0, N/2 - 2 \rrbracket \\ 2 + 2i - N & , \text{ if } i \in \llbracket N/2 - 1, N - 1 \rrbracket \end{cases}$$

The Markov kernel \hat{P} is given on $\llbracket 0, N - 1 \rrbracket$ by

$$\forall i \neq j \in V, \quad \hat{P}(i, j) := \begin{cases} P(z_i, z_j) & , \text{ if } |i - j| = 1 \\ 0 & , \text{ otherwise} \end{cases} \quad (10)$$

This construction of \hat{P} is somewhat supplementary to that of \tilde{P} : only the transitions of size two are kept, all transitions of size 1 being removed, except those between 0 and 1, to insure irreducibility.

For the same reason as for \tilde{P} , the kernel \hat{P} admits $\hat{\pi} := (\hat{\pi}(i))_{i \in \llbracket 0, N - 1 \rrbracket}$ for reversible measure, where

$$\forall i \in \llbracket 0, N - 1 \rrbracket, \quad \hat{\pi}(i) := \pi(z_i) \quad (11)$$

4 An a priori estimate

A drawback of Definition (5) of the Markov kernel P is that the quantities $p(x)$, for $x \in V$, are a priori unknown. We will give an explicit formula for them in Appendix (A), but the control of the couplings of next section only requires an a priori bound about them, presented in Proposition 6 below.

We have seen in the introduction that $p(N - 3) = p(N) = 0$ and that $p(N - 2) = 1$. These equalities and the fact that P is reversible are sufficient knowledges to deduce the following bound:

Proposition 6 *We have for the mapping p defined in (6),*

$$\forall x \in \llbracket 0, N - 2 \rrbracket, \quad |2p(x) - 1| \leq \frac{1}{(N - x - 2)!}$$

Proof

Recall that Kolmogorov criterion for reversibility, see e.g. the book of Kelly [16], asserts that for any finite sequence (x_0, x_1, \dots, x_n) from V with $n \in \mathbb{N}$, we have

$$P(x_0, x_1)P(x_1, x_2) \cdots P(x_{n-1}, x_n)P(x_n, x_0) = P(x_0, x_n)P(x_n, x_{n-1}) \cdots P(x_2, x_1) \cdots P(x_1, x_0)$$

For given $x \in \llbracket 0, N - 4 \rrbracket$, assuming $N \geq 4$, let us apply this formula with

$$\begin{aligned} x_0 &= x \\ x_1 &= x + 1 \\ x_2 &= x + 2 \end{aligned}$$

We get

$$P(x, x + 1)P(x + 1, x + 2)P(x + 2, x) = P(x, x + 2)P(x + 2, x + 1)P(x + 1, x)$$

namely

$$\begin{aligned} & (N - x - 2p(x))(N - x - 1 - 2p(x + 1))(x + 2)(x + 1) \\ &= 2p(x)(x + 2)(N - x - 2)(x + 1)(N - x - 1) \end{aligned}$$

i.e., since $x + 1 > 0$,

$$(N - x - 2p(x))(N - x - 1 - 2p(x + 1)) = 2p(x)(N - x - 2)(N - x - 1)$$

To simplify notations, let us write $k(x) = 2p(x)$, for any $x \in \llbracket 0, N - 2 \rrbracket$. The above formula is equivalent to the downward iteration, for $x \in \llbracket 0, N - 4 \rrbracket$,

$$k(x) = \frac{(N - x)(N - x - 1 - k(x + 1))}{(N - x - 1)^2 - k(x + 1)} \quad (12)$$

Starting from $k(N - 3) = 2p(N - 3) = 0$, we deduce iteratively $k(N - 4)$, $k(N - 5)$, ... down to $k(0)$.

For $x \in \llbracket 0, N - 4 \rrbracket$, denote F_x the rational function

$$\forall r \in \mathbb{R} \setminus \{(N - x - 1)^2\}, \quad F_x(r) := \frac{(N - x)(N - x - 1 - r)}{(N - x - 1)^2 - r}$$

so that $k(x) = F_x(k(x + 1))$.

For any $x \in \llbracket 0, N - 4 \rrbracket$, 1 is a fixed point of F_x (the only one in fact), since

$$\begin{aligned} F_x(1) &= \frac{(N - x)(N - x - 1 - 1)}{(N - x - 1)^2 - 1} \\ &= 1 \end{aligned}$$

Thus (12) can be written in the convenient form

$$\begin{aligned} k(x) - 1 &= F_x(k(x + 1)) - F_x(1) \\ &= \int_1^{k(x+1)} F'_x(s) ds \end{aligned} \quad (13)$$

which suggests computing:

$$\forall s \in \mathbb{R} \setminus \{(N - x - 1)^2\}, \quad F'_x(s) = -\frac{(N - x)(N - x - 1)(N - x - 2)}{((N - x - 1)^2 - s)^2} \quad (14)$$

These observations lead to a proof of the bound of Proposition 6 by a backward iteration. Indeed, for $x = N - 2$ and $x = N - 3$, the bound is true, since it is respectively implied by

$$\begin{aligned} 2p(N - 2) - 1 &= 2 - 1 \\ &= 1 \\ &= \frac{1}{0!} \\ &= \frac{1}{(N - 2 - (N - 2))!} \end{aligned}$$

and

$$\begin{aligned} 2p(N - 3) - 1 &= 0 - 1 \\ &= -1 \\ &= -\frac{1}{1!} \\ &= -\frac{1}{(N - 2 - (N - 3))!} \end{aligned}$$

Consider $x \in \llbracket 0, N-4 \rrbracket$, we have

$$N-x \geq N-x-1 \geq N-x-2 \geq 2 \quad (15)$$

so that $F'_x < 0$. This observation and (13) imply that if $k(x+1) > 1$, then $k(x) < 1$ and conversely, if $k(x+1) < 1$, then $k(x) > 1$, namely the sequence $(k(z)-1)_{z \in \llbracket 0, N-2 \rrbracket}$ is alternating.

Let us consider separately the first case: $x = N-4$. Since $k(N-3) = 0 < 1$, we deduce from (14) with $x = N-4$, that for $s \in [k(N-3), 1] = [0, 1]$,

$$|F'_{N-4}(s)| \leq \frac{4 \times 3 \times 2}{(3^2 - 1)^2} = \frac{3}{8} \leq \frac{1}{2}$$

It follows from (13) that

$$|k(N-4) - 1| \leq \frac{1}{2} |k(N-3) - 1| \leq \frac{1}{2} = \frac{1}{(N - (N-4) - 2)!}$$

Let us now assume the bound of Proposition 6 is true for some $x+1 \in \llbracket 1, N-4 \rrbracket$ and let us prove it for x .

Note that we have

$$k(x+1) \leq 1 + \frac{1}{(N-x-2)!} \leq 1 + \frac{1}{2!} = \frac{3}{2}$$

so (15) and (14) imply that for $s \in [1, F_x(k(x+1))]$ (or $s \in [F_x(k(x+1)), 1]$ if $F_x(k(x+1)) \leq 1$),

$$|F'_x(s)| \leq \frac{(N-x)(N-x-1)(N-x-2)}{((N-x-1)^2 - 4/3)^2}$$

Let us show that the r.h.s. is bounded above by $1/(N-x-2)$. To simplify notation, write $y := N-x-1 \geq 3$, so that the desired bound amounts to

$$\frac{(y+1)y(y-1)}{(y^2 - 4/3)^2} \leq \frac{1}{y-1} \quad (16)$$

namely

$$(y^2 - 1)(y-1)y \leq (y^2 - 4/3)^2$$

i.e.

$$y^4 - y^3 - y^2 + y \leq y^4 - \frac{8}{3}y^2 + \frac{16}{9}$$

or $g(y) \geq 0$, where

$$\forall y \geq 3, \quad g(y) := y^3 - \frac{5}{3}y^2 - y + \frac{16}{9}$$

We compute

$$\forall y \geq 3, \quad g'(y) = 3y^2 - \frac{10}{3}y - 1$$

and the largest zero of the r.h.s is

$$\frac{1}{18}(10 + \sqrt{208}) < 3$$

It follows that g is increasing on $[3, +\infty)$ and we compute

$$g(3) = 27 - 15 + 3 + \frac{16}{9} > 0$$

showing the validity of (16).

We deduce from (13) that

$$\begin{aligned} |k(x) - 1| &\leq \left| \int_1^{k(x+1)} \frac{1}{N-x-2} ds \right| \\ &\leq \frac{1}{N-x-2} |k(x+1) - 1| \\ &\leq \frac{1}{(N-x-2)!} \end{aligned}$$

where we took into account the iteration assumption, namely $|k(x+1) - 1| \leq 1/((N-x-3)!)$. ■

Remark 7 The observation made after (15) implies more precisely that for $x \in \llbracket 0, N-2 \rrbracket$, $2p(N-2-x) - 1$ is positive for even x and negative for odd x . □

These computations, especially the iteration relation (12), also show that the reversible couple (π, P) is well-defined by $p(N-3) = 0 = p(N)$ and $p(N-2) = 1$: no further information are needed for its investigation, in particular not the interpretation of p as a conditional expectation on the larger space \mathcal{S}_N . Namely we can work only on V .

Let us state this construction formally:

Remark 8 Consider \bar{P} defined in (5) with the $p(x)$ replaced by some $\bar{p}(x) \geq 0$, under the constraint that \bar{P} is a Markov kernel on $\llbracket 0, N \rrbracket$. Add the constraints $\bar{p}(N-1) = \bar{p}(N) = 0$ (in our previous case, $N-1$ becomes a transient point, with $P(N-1, N) = 1/(N(N-1))$ and $P(N-1, N-1) = 1 - P(N-1, N)$). Assume furthermore that the values $\bar{p}(x)$ satisfy the iteration (12). Thus \bar{P} is a function of $\bar{p}(N-2)$ and $\bar{p}(N-3)$. Then taking $\bar{p}(N-2) = 1/2$ and $\bar{p}(N-3) = 1/2$ (implying $\bar{p}(x) = 1/2$ for all $x \in \llbracket 0, N-4 \rrbracket$, due to the fact that 1 is a fixed point of F_x), we end up with a Markov kernel \bar{P} which is reversible with respect to the restriction of the Poisson distribution on $\llbracket 0, N \rrbracket$. □

This observation is at the heart of the couplings presented in next section.

5 A monotone coupling

Our purpose here is to prove by coupling an upper bound of the Poisson approximation of π , of the same logarithmic order as that of (1). It would be possible to push further the computations, but our main emphasis is placed on the method rather than on sharp estimates.

More precisely, we want to show (7) by only using that π is reversible with respect to the Markov kernel \tilde{P} defined in (9) and the a priori estimate given in Proposition 6.

Instead of working on V or \mathbb{Z}_+ , we can restrict our attention to $\llbracket 0, N-4 \rrbracket$ (assuming $N \geq 4$). Indeed, denote ζ the conditioning of \mathcal{P} to $\llbracket 0, N-4 \rrbracket$, since

$$\lim_{N \rightarrow \infty} \frac{1}{N \ln(N)} \ln(\mathcal{P}(\llbracket N-3, \infty \rrbracket)) = -1$$

we easily deduce that

$$\lim_{N \rightarrow \infty} \frac{1}{N \ln(N)} \ln(\|\zeta - \mathcal{P}\|_{\text{tv}}) = -1 \tag{17}$$

Furthermore it is not difficult to see that

$$\lim_{N \rightarrow \infty} \frac{1}{N \ln(N)} \ln(\pi(\{N-3, N-2, N\})) = -1$$

since by a direct investigation, we get that

$$\begin{aligned} \pi(N) &= \frac{1}{N!} \\ \pi(N-2) &= \frac{1}{2} \frac{1}{N!} \\ \pi(N-3) &= \frac{1}{3} \frac{1}{N!} \end{aligned} \tag{18}$$

It follows that

$$\lim_{N \rightarrow \infty} \frac{1}{N \ln(N)} \ln(\|\pi - \tilde{\pi}\|_{\text{tv}}) = -1 \tag{19}$$

where $\tilde{\pi}$ is the conditioning of π to $\llbracket 0, N-4 \rrbracket$.

These limiting behaviors imply that (7) amounts to

$$\limsup_{N \rightarrow \infty} \frac{1}{N \ln(N)} \ln(\|\tilde{\pi} - \zeta\|_{\text{tv}}) \leq -1 \tag{20}$$

Indeed, on one hand, from (18) we get

$$\begin{aligned} \liminf_{N \rightarrow \infty} \frac{1}{N \ln(N)} \ln(\|\pi - \mathcal{P}\|_{\text{tv}}) &\geq \liminf_{N \rightarrow \infty} \frac{1}{N \ln(N)} \ln\left(\frac{1}{2} |\pi(N) - \mathcal{P}(N)|\right) \\ &= \liminf_{N \rightarrow \infty} \frac{1}{N \ln(N)} \ln((1 - e^{-1})/(N!)) \\ &= -1 \end{aligned}$$

and on the other hand, from (17), (19) and (20),

$$\begin{aligned} &\limsup_{N \rightarrow \infty} \frac{1}{N \ln(N)} \ln(\|\pi - \mathcal{P}\|_{\text{tv}}) \\ &\leq \max\left(\limsup_{N \rightarrow \infty} \frac{1}{N \ln(N)} \ln(\|\pi - \tilde{\pi}\|_{\text{tv}}), \limsup_{N \rightarrow \infty} \frac{1}{N \ln(N)} \ln(\|\tilde{\pi} - \zeta\|_{\text{tv}}), \limsup_{N \rightarrow \infty} \frac{1}{N \ln(N)} \ln(\|\zeta - \mathcal{P}\|_{\text{tv}})\right) \\ &= -1 \end{aligned}$$

Thus it remains to prove (20).

By reversibility of \check{P} with respect to π , we have that $\tilde{\pi}$ is reversible with respect to the birth and death Markov kernel \check{P} given by

$$\forall x \neq y \in \llbracket 0, N-4 \rrbracket, \quad \check{P}(x, y) := \frac{1}{N(N-1)} \begin{cases} x(N-x) & , \text{ if } y = x-1 \\ N-x-2p(x) & , \text{ if } y = x+1 \\ 0 & , \text{ otherwise} \end{cases}$$

(as usual the diagonal entries are deduced by the fact that the rows sum to 1).

Consider the birth and death Markov kernel R given by

$$\forall x \neq y \in \llbracket 0, N-4 \rrbracket, \quad R(x, y) := \frac{1}{N(N-1)} \begin{cases} x(N-x) & , \text{ if } y = x-1 \\ N-x-1 & , \text{ if } y = x+1 \\ 0 & , \text{ otherwise} \end{cases}$$

which amounts to replacing $p(x)$ by $1/2$ in the kernel \check{P} , see Remark 8 above.

It is immediate to check that ζ is reversible for R , since we have for any $x \in \llbracket 0, N - 5 \rrbracket$,

$$\begin{aligned} \frac{\zeta(x)R(x, x+1)}{\zeta(x+1)R(x+1, x)} &= \frac{(x+1)R(x, x+1)}{R(x+1, x)} \\ &= \frac{(x+1)(N-x-1)}{(x+1)(N-x-1)} \\ &= 1 \end{aligned}$$

To simplify the notations, from now on, $\tilde{\pi}$ and \tilde{P} will be written π and P , we hope it will not bring confusion with the previous π and P .

Consider $X := (X(n))_{n \in \mathbb{Z}_+}$ a stationary Markov chain whose transitions are given by P and whose initial law is π . Similarly let $Y := (Y(n))_{n \in \mathbb{Z}_+}$ be a stationary Markov chain whose transitions are given by R and whose initial law is ζ . We couple them in a monotone way: namely at any time $n \in \mathbb{Z}_+$, the transition from $(X(n), Y(n))$ to $(X(n+1), Y(n+1))$ is given by sampling an independent uniform random variable $U(n)$ on $[0, 1]$ and by deciding that

$$X(n+1) = \begin{cases} X(n) - 1 & , \text{ if } U(n) < P(X(n), X(n) - 1) \\ X(n) & , \text{ if } P(X(n), X(n) - 1) \leq U(n) < P(X(n), X(n) - 1) + P(X(n), X(n)) \\ X(n+1) & , \text{ if } P(X(n), X(n) - 1) + P(X(n), X(n)) \leq U(n) \end{cases}$$

and

$$Y(n+1) = \begin{cases} Y(n) - 1 & , \text{ if } U(n) < R(Y(n), Y(n) - 1) \\ Y(n) & , \text{ if } R(Y(n), Y(n) - 1) \leq U(n) < R(Y(n), Y(n) - 1) + R(Y(n), Y(n)) \\ Y(n+1) & , \text{ if } R(Y(n), Y(n) - 1) + R(Y(n), Y(n)) \leq U(n) \end{cases}$$

The corresponding Markov kernel on $\llbracket 0, N - 2 \rrbracket^2$ will be denoted S , namely we have

$$\begin{aligned} \forall (x, y), (x', y') \in \llbracket 0, N - 2 \rrbracket^2, \\ S((x, y), (x', y')) &= \mathbb{P}[(X(n+1), Y(n+1)) = (x', y') | (X(n), Y(n)) = (x, y)] \end{aligned}$$

Consider, traditionally τ the coupling time

$$\tau := \inf\{n \in \mathbb{Z}_+ : X(n) = Y(n)\}$$

but also the auxiliary random chain $Z := (Z(n))_{n \in \mathbb{Z}_+}$

$$\forall n \in \mathbb{Z}_+, \quad Z(n) := \sum_{k=0}^{n-1} \mathbb{1}_{\{X(k)=Y(k), X(k+1) \neq Y(k+1)\}}$$

Their interest is that for any time $n \in \mathbb{Z}_+$, we have

$$\begin{aligned} \|\pi - \zeta\|_{\text{tv}} &\leq \mathbb{P}[X(n) \neq Y(n)] \\ &\leq \mathbb{P}[\tau > n] + \mathbb{P}[Z(n) > 0] \end{aligned} \tag{21}$$

By choosing n of order $N^4 \ln(N)$, we will get an estimate of $\|\pi - \zeta\|_{\text{tv}}$ of the order we are looking for.

This resort to coupling is different from its traditional use in the quantitative investigation of convergence to equilibrium, where different lines of the same transition kernel are coupled. The bound (21) is neither good for short or long times n , it is interesting only for certain times, enabling us to estimate the difference between the invariant probabilities of two different transition kernels.

To illustrate the difference between these approaches, let us evaluate the new term in (21):

Lemma 9 For any $n \in \mathbb{Z}_+$, we have

$$\mathbb{P}[Z(n) > 0] \leq \frac{2^N n}{N!}$$

Proof

For any given $n \in \mathbb{Z}_+$, we have

$$\begin{aligned} \mathbb{P}[Z(n) > 0] &\leq \mathbb{E}[Z(n)] \\ &= \sum_{k=0}^{n-1} \mathbb{E}[\mathbf{1}_{X(k)=Y(k), X(k+1) \neq Y(k+1)}] \\ &= \sum_{k=0}^{n-1} \mathbb{E}[\mathbf{1}_{X(k)=Y(k)} S((X(k), Y(k)), A)] \end{aligned} \quad (22)$$

where

$$A := \{(x', y') \in \llbracket 0, N-4 \rrbracket^2 : x' \neq y'\}$$

Taking into account Proposition 6, we have

$$\begin{aligned} \forall x \in \llbracket 0, N-4 \rrbracket, \quad S((x, x), A) &\leq \frac{|1 - 2p(x)|}{N(N-1)} \\ &\leq \frac{1}{(N-x-2)!} \frac{1}{N(N-1)} \\ &\leq \frac{1}{(N-x)!} \end{aligned}$$

It follows that for any $k \in \llbracket 0, n \rrbracket$,

$$\begin{aligned} \mathbb{E}[\mathbf{1}_{X(k)=Y(k)} S((X(k), Y(k)), A)] &= \sum_{x=0}^{N-4} \mathbb{P}[X(k) = x = Y(k)] S((x, x), A) \\ &\leq \sum_{x=0}^{N-4} \mathbb{P}[Y(k) = x] S((x, x), A) \\ &\leq \sum_{x=0}^{N-4} \frac{1}{Z_N x!} \frac{1}{(N-x)!} \\ &\leq \frac{1}{Z_N} \sum_{x=0}^N \frac{1}{x!} \frac{1}{(N-x)!} \\ &= \frac{1 \times 2^N}{Z_N N!} \end{aligned}$$

where we used that $(Y(k))_{k \in \mathbb{Z}_+}$ is stationary with common distribution ζ and where

$$Z_N = \sum_{x=0}^{N-4} \frac{1}{x!} \geq 1$$

The desired result follows by remembering (22). ■

Note that the bound of the above lemma will be small even if we choose a time n exponential large in N .

In view of (21) and Lemma 9, our next task is to get an estimate on $\mathbb{P}[\tau > n]$ for given $n \in \mathbb{Z}_+$. To go in this direction, we will need two other auxiliary random chains $\tilde{Z} := (\tilde{Z}(n))_{n \in \mathbb{Z}_+}$ and $\hat{Z} := (\hat{Z}(n))_{n \in \mathbb{Z}_+}$, defined respectively through

$$\forall n \in \mathbb{Z}_+, \quad \begin{cases} \tilde{Z}(n) & := \sum_{k=0}^{n-1} \mathbb{1}_{X(k) \leq Y(k), X(k+1) > Y(k+1)} \\ \hat{Z}(n) & := \sum_{k=0}^{n-1} \mathbb{1}_{X(k) \geq Y(k), X(k+1) < Y(k+1)} \end{cases}$$

as well as the hitting times of zero by X and Y :

$$\begin{aligned} \tau_0^X &:= \inf\{n \in \mathbb{Z}_+ : X(n) = 0\} \\ \tau_0^Y &:= \inf\{n \in \mathbb{Z}_+ : Y(n) = 0\} \end{aligned}$$

Indeed, it is clear that

$$\forall n \in \mathbb{Z}_+, \quad \mathbb{P}[\tau > n] \leq \mathbb{P}[\tau_0^X > n] + \mathbb{P}[\tau_0^Y > n] + \mathbb{P}[\tilde{Z}(n) > 0] + \mathbb{P}[\hat{Z}(n) > 0] \quad (23)$$

It remains to estimate each of the terms of the r.h.s.

Let us start with the last two terms. In this respect, it is useful to remark that the Markov chain Y is monotone, namely that for $x \leq y \in \llbracket 0, N-4 \rrbracket$, if $Y_x := (Y_x(n))_{n \in \mathbb{Z}_+}$ and $Y_y := (Y_y(n))_{n \in \mathbb{Z}_+}$ are Markov chain with transition kernel R starting respectively from x and y , then we can couple them in a monotone fashion (similar to the coupling of X and Y above), so that

$$\mathbb{P}[\forall n \in \mathbb{Z}_+, Y_x(n) \leq Y_y(n)] = 1$$

(see for instance the book of Lindvall [20]).

Let us prove this monotonicity of Y :

Lemma 10 *The Markov chain Y is monotone.*

Proof

Since Y is a birth and death chain, to get it is monotone, it is sufficient to check that

$$\forall x \in \llbracket 0, N-5 \rrbracket, \quad R(x, \llbracket x-1, x \rrbracket) \geq R(x+1, x)$$

(again see e.g. Lindvall [20]).

The previous bound amounts to

$$\forall x \in \llbracket 0, N-5 \rrbracket, \quad 1 - \frac{N-x-1}{N(N+1)} \geq \frac{(x+1)(N-x-1)}{N(N+1)}$$

or

$$\forall x \in \llbracket 0, N-5 \rrbracket, \quad N(N+1) \geq (x+2)(N-x-1) \quad (24)$$

The maximum of the r.h.s. as x runs in \mathbb{R} is attained at the point $x = (N-3)/2$ and replacing in the above r.h.s., the desired inequality is true if we have $N \geq (N+1)/4$, which is satisfied as soon as $N \geq 1/3$. \blacksquare

Let us come back to the quantities $\mathbb{P}[\tilde{Z}(n) > 0]$ and $\mathbb{P}[\hat{Z}(n) > 0]$, we have:

Lemma 11 *For any $n \in \mathbb{Z}_+$, we have*

$$\begin{aligned} \mathbb{P}[\hat{Z}(n) > 0] &\leq \frac{2^{N+1}n}{N!} \\ \mathbb{P}[\tilde{Z}(n) > 0] &\leq \frac{2^{N+1}n}{N!} \end{aligned}$$

Proof

We have

$$\begin{aligned} \mathbb{P}[\widehat{Z}(n) > 0] &\leq \mathbb{E}[\widehat{Z}(n)] \\ &= \sum_{k=0}^{n-1} \mathbb{E}[\mathbb{1}_{X(k) \geq Y(k), X(k+1) < Y(k+1)}] \end{aligned} \quad (25)$$

Fix some $k \in \llbracket 0, n-1 \rrbracket$. If $X(k) \geq Y(k)$ and $X(k+1) < Y(k+1)$ hold, then either $X(k) = Y(k)$ or $X(k) = Y(k) + 1$. Let us consider the latter case, we have:

$$\mathbb{E}[\mathbb{1}_{X(k)=Y(k)+1, X(k+1)<Y(k+1)}] = \sum_{x=0}^{N-4} \mathbb{P}[X(k) = x+1, Y(k) = x] S((x+1, x), A_-) \quad (26)$$

where

$$A_- := \{(x', y') \in \llbracket 0, N-4 \rrbracket : x' < y'\}$$

But for the transition from $(x+1, x)$ to B to happen, the underlying uniform random variable on $[0, 1]$ must have taken advantage of the discrepancy between $2p(x+1)$ and 1, otherwise the monotonicity of Y leads to a contradiction. We deduce

$$\begin{aligned} S((x+1, x), A_-) &\leq \frac{|1 - 2p(x+1)|}{N(N-1)} \\ &\leq \frac{1}{(N-x-1)!} \frac{1}{N(N-1)} \\ &\leq \frac{1}{(N-x+1)!} \end{aligned}$$

and it follows, as in proof of Lemma 9 that

$$\begin{aligned} \sum_{x=0}^{N-4} \mathbb{P}[X(k) = x+1, Y(k) = x] S((x+1, x), A_-) &\leq \sum_{x=0}^{N-4} \frac{1}{Z_N x!} \frac{1}{(N-x+1)!} \\ &\leq \frac{2^{N+1}}{(N+1)!} \\ &\leq \frac{2^N}{N!} \end{aligned}$$

The treatment of the cases $X(k) = Y(k)$ is similar to the proof of Lemma 9, leading to

$$\sum_{x=0}^{N-4} \mathbb{P}[X(k) = x, Y(k) = x] S((x, x), A_-) \leq \frac{2^N}{N!} \quad (27)$$

It follows that for any $k \in \llbracket 0, n-1 \rrbracket$

$$\mathbb{E}[\mathbb{1}_{X(k) \geq Y(k), X(k+1) < Y(k+1)}] \leq \frac{2^{N+1}}{N!}$$

and (25) leads to the first desired bound.

The second desired bound is obtained in a similar way, the main difference being that we have to replace, for $k \in \llbracket 0, n-1 \rrbracket$, (26) by

$$\mathbb{E}[\mathbb{1}_{X(k)=Y(k)-1, X(k+1)<Y(k+1)}] = \sum_{x=0}^{N-4} \mathbb{P}[X(k) = x+1, Y(k) = x] S((x+1, x), A_+)$$

where

$$A_+ := \{(x', y') \in \llbracket 0, N-4 \rrbracket : x' > y'\}$$

Then we rather use

$$\begin{aligned} S((x-1, x), A_+) &\leq \frac{|1-2p(x-1)|}{N(N-1)} \\ &\leq \frac{1}{(N-x-1)!} \frac{1}{N(N-1)} \\ &\leq \frac{1}{(N-x+1)!} \end{aligned}$$

leading to

$$\begin{aligned} \mathbb{E}[\mathbf{1}_{X(k)=Y(k)-1, X(k+1)<Y(k+1)}] &\leq \frac{2^{N+1}}{(N+1)!} \\ &\leq \frac{2^N}{N!} \end{aligned}$$

As in (27), we also have

$$\mathbb{E}[\mathbf{1}_{X(k)=Y(k), X(k+1)<Y(k+1)}] \leq \frac{2^N}{N!}$$

enabling us to conclude to the second desired bound. ■

We are left with the evaluation of the tails of τ_0^X and τ_0^Y in (23).

We start with the last one:

Lemma 12 *There exists a constant $c > 0$ such that for any N large enough and any $n \in \mathbb{Z}_+$, we have*

$$\mathbb{P}[\tau_0^Y > n] \leq e^{1-cn/N^3}$$

whatever the initial law of $Y(0)$.

Proof

For any time $k \in \mathbb{Z}_+$ such that $Y(k) = y \neq 0$, we compute

$$\begin{aligned} \mathbb{E} \left[\exp \left(\frac{Y_{k+1} - Y_k}{N} \right) \middle| Y_k = y \right] &= e^{-1/N} \frac{y(N-y)}{N(N-1)} + e^{1/N} \frac{N-y-1}{N(N-1)} + 1 - \frac{y(N-y) + N-y-1}{N(N-1)} \\ &= 1 + (e^{-1/N} - 1) \frac{y(N-y)}{N(N-1)} + (e^{1/N} - 1) \frac{N-y-1}{N(N-1)} \end{aligned}$$

Denoting $F_N(y)$ the r.h.s., it is a second order polynomial whose minimal value is attained at

$$\underline{y} := \frac{e^{1/N} - 1 + N}{2(1 - e^{-1/N})}$$

belonging to $\llbracket 1, N-4 \rrbracket$ for N large enough. It follows that the maximal value of $F_N(y)$ for $y \in \llbracket 1, N-4 \rrbracket$ is attained either at $y = 1$ or $y = N-4$.

We compute that

$$\begin{aligned} F_N(1) &= 1 + \frac{(e^{-1/N} - 1)(N-1) + (e^{1/N} - 1)(N-2)}{N(N-1)} \\ F_N(N-4) &= 1 + \frac{4(e^{-1/N} - 1)(N-4) + 3(e^{1/N} - 1)}{N(N-1)} \end{aligned}$$

and we deduce there exists a constant $c > 0$ such that for N large enough,

$$\max(F_N(1), F_N(N-4)) \leq 1 - \frac{c}{N^3}$$

leading to

$$\forall k \in \mathbb{Z}_+, \quad Y_k \neq 0 \Rightarrow \mathbb{E} \left[\exp \left(\frac{Y_{k+1}}{N} \right) \middle| Y_k \right] \leq \left(1 - \frac{c}{N^3} \right) \exp \left(\frac{Y(k)}{N} \right)$$

implying

$$\forall k \in \mathbb{Z}_+, \quad \mathbb{E} \left[\exp \left(\frac{Y_{k+1}}{N} \right) \middle| Y_k \right] \mathbb{1}_{Y(k) \neq 0} \leq \left(1 - \frac{c}{N^3} \right) \exp \left(\frac{Y(k)}{N} \right)$$

i.e.

$$\forall k \in \mathbb{Z}_+, \quad \mathbb{E} \left[\exp \left(\frac{Y_{k+1}}{N} \right) \mathbb{1}_{Y(k) \neq 0} \middle| \mathcal{F}^Y(k) \right] \leq \left(1 - \frac{c}{N^3} \right) \exp \left(\frac{Y(k)}{N} \right)$$

or, using the Markov property,

$$\forall k \in \mathbb{Z}_+, \quad \mathbb{E} \left[\exp \left(\frac{Y_{k+1}}{N} \right) \mathbb{1}_{Y(k) \neq 0} \middle| \mathcal{F}^Y(k) \right] \leq \left(1 - \frac{c}{N^3} \right) \exp \left(\frac{Y(k)}{N} \right)$$

where $\mathcal{F}^Y(k)$ is the sigma-field generated by $Y(0), Y(1), \dots, Y(k)$.

Iterating this relation, we get for any $k \in \mathbb{N}$,

$$\mathbb{E} \left[\mathbb{E} \left[\exp \left(\frac{Y_{k+1}}{N} \right) \mathbb{1}_{Y(k) \neq 0} \middle| \mathcal{F}^Y(k) \right] \mathbb{1}_{Y(k-1) \neq 0} \middle| \mathcal{F}^Y(k-1) \right] \leq \left(1 - \frac{c}{N^3} \right)^2 \exp \left(\frac{Y(k-1)}{N} \right)$$

Pushing further the iteration, we end up with

$$\begin{aligned} \mathbb{E} \left[\cdots \mathbb{E} \left[\mathbb{E} \left[\exp \left(\frac{Y_{k+1}}{N} \right) \mathbb{1}_{Y(k) \neq 0} \middle| \mathcal{F}^Y(k) \right] \mathbb{1}_{Y(k-1) \neq 0} \middle| \mathcal{F}^Y(k-1) \right] \cdots \mathbb{1}_{Y(0) \neq 0} \middle| \mathcal{F}^Y(0) \right] \\ \leq \left(1 - \frac{c}{N^3} \right)^k \exp \left(\frac{Y(0)}{N} \right) \end{aligned}$$

Taking into account that $Y_{k+1} \geq 0$ and that $Y(0) \leq N$, we get

$$\mathbb{E} \left[\cdots \mathbb{E} \left[\mathbb{E} \left[\mathbb{1}_{Y(k) \neq 0} \middle| \mathcal{F}^Y(k) \right] \mathbb{1}_{Y(k-1) \neq 0} \middle| \mathcal{F}^Y(k-1) \right] \cdots \mathbb{1}_{Y(0) \neq 0} \middle| \mathcal{F}^Y(0) \right] \leq e \left(1 - \frac{c}{N^3} \right)^k$$

Taking expectation and simplifying conditional expectation iteratively (starting with $\mathcal{F}^Y(0)$, next $\mathcal{F}^Y(1)$, etc.), we end up with

$$\mathbb{P}[Y(k) \neq 0, Y(k-1) \neq 0, \dots, Y(0) \neq 0] \leq e \left(1 - \frac{c}{N^3} \right)^k$$

implying

$$\begin{aligned} \mathbb{P}[\tau_0^Y > k] &\leq e \left(1 - \frac{c}{N^3} \right)^k \\ &\leq e^{1 - \frac{ck}{N^3}} \end{aligned}$$

which is desired bound, taking $k = n$. ■

The tail of τ_0^X is evaluated similarly:

Lemma 13 *There exists a constant $\tilde{c} > 0$ such that for N large enough and any $n \in \mathbb{Z}_+$, we have*

$$\mathbb{P}[\tau_0^X > n] \leq e^{1-\tilde{c}n/N^3}$$

whatever the initial law of $X(0)$.

Proof

According to Proposition 6, we have $2p(x) \geq 1/2$ for all $x \in \llbracket 0, N-4 \rrbracket$, fact which suggests to consider Markov chains $\tilde{Y} := (\tilde{Y}(n))_{n \in \mathbb{Z}_+}$ associated to the transition kernel \tilde{R} given by

$$\forall x \neq y \in \llbracket 0, N-4 \rrbracket, \quad \tilde{R}(x, y) := \frac{1}{N(N-1)} \begin{cases} x(N-x) & , \text{ if } y = x-1 \\ N-x-1/2 & , \text{ if } y = x+1 \\ 0 & , \text{ otherwise} \end{cases}$$

which differs from R only the replacement of $N-x-1$ by $N-x-1/2$.

Consider the corresponding hitting time of 0:

$$\tau_0^{\tilde{Y}} := \inf\{n \in \mathbb{Z}_+ : \tilde{Y}(n) = 0\}$$

Coupling in a monotone way X and \tilde{Y} and starting with $\tilde{Y}(0) = X(0)$, we get that

$$\forall n \in \mathbb{Z}_+, \quad X(n) \leq \tilde{Y}(n)$$

at least if \tilde{Y} is monotone. This is true and is proven as for Lemma 10, where (24) has to be replaced by

$$\forall x \in \llbracket 0, N-5 \rrbracket, \quad N(N+1) \geq (x+2)(N-x-1/2)$$

We deduce that

$$\forall n \in \mathbb{Z}_+, \quad \mathbb{P}[\tau_0^X > n] \leq \mathbb{P}[\tau_0^{\tilde{Y}} > n]$$

It is thus sufficient to find a constant $\tilde{c} > 0$ such that for any $N \geq 5$ and any $n \in \mathbb{Z}_+$, we have

$$\mathbb{P}[\tau_0^{\tilde{Y}} > n] \leq e^{1-\tilde{c}n/N^3}$$

whatever the initial law of $\tilde{Y}(0)$.

This done as in the proof of Lemma 12. ■

Summarizing the previous computation, we have shown there exist two constants $c, \tilde{c} > 0$ such that for any N large enough and $n \geq 0$,

$$\begin{aligned} \|\pi - \zeta\|_{\text{tv}} &\leq \frac{2^N n}{N!} + 2 \frac{2^{N+1} n}{N!} + e^{1-cn/N^3} + e^{1-\tilde{c}n/N^3} \\ &\leq \frac{5 \times 2^N n}{N!} + 2e^{1-\tilde{c}n/N^3} \end{aligned}$$

with $\hat{c} := c \wedge \tilde{c}$.

Taking $n = N \ln(N)/\hat{c}$, we conclude (20).

A Recovering classical results on π through the Markov approach

Working in the same spirit as in Section 4, it is possible to recover the exact formula for the number of fixed points $\pi(x)$ (see (4)) from the reversibility of the Markov chain P (see (5)) with respect to π , leading to an alternative proof for Montmort's formula (3) (the traditional argument goes through the inclusion-exclusion principle, see e.g. [25] or Chapter 1 of Arratia, Barbour and Tavaré [1]).

We will use the birth and death chain \hat{P} defined in (10) above and its stationary distribution $\hat{\pi}$ defined in (11). Using that notation, the reversibility says

$$\forall i \in \llbracket 0, N-1 \rrbracket, \quad \hat{\pi}(i)\hat{P}(i, i+1) = \hat{\pi}(i+1)\hat{P}(i+1, i)$$

or

$$\forall i \in \llbracket 0, N-1 \rrbracket, \quad \pi(z_i)P(z_i, z_{i+1}) = \pi(z_{i+1})P(z_{i+1}, z_i)$$

namely

$$\pi(0)P(0, 1) = \pi(1)P(1, 0) \quad \text{and} \quad \forall x \in V \setminus \{N\}, \pi(x)P(x, x+2) = \pi(x+2)P(x+2, x) \quad (28)$$

i.e.

$$\pi(0)(N - 2p(0)) = \pi(1)(N - 1) \quad \text{and} \quad \forall x \in V \setminus \{N\}, 2\pi(x)p(x) = \pi(x+2)(x+2)(x+1)$$

The last condition implies that

$$\forall x \in V \setminus \{N\}, \quad p(x) = \frac{\pi(x+2)}{2\pi(x)}(x+2)(x+1)$$

This formula also holds for $x = N$, since both terms vanish, thus we have shown:

Lemma 14 *We have*

$$\forall x \in V, \quad p(x) = \frac{\pi(x+2)}{2\pi(x)}(x+2)(x+1)$$

Replacing this expression in the definition of the first associated birth and death kernel \tilde{P} (defined in (9)), we will deduce the following expression for the reversible probability π :

Proposition 15 *We have*

$$\forall x \in V, \quad \pi(x) = \frac{1}{x!} \sum_{k=0}^{N-x} \frac{(-1)^k}{k!}$$

Proof

From Lemma 14, we get for any $x \neq y \in V$,

$$\tilde{P}(x, y) = \frac{1}{N(N-1)} \begin{cases} x(N-x) & , \text{ if } x \neq N \text{ and } y = x-1 \\ N-x - \frac{\pi(x+2)}{\pi(x)}(x+2)(x+1) & , \text{ if } x \neq N-2 \text{ and } y = x+1 \\ 2 & , \text{ if } x = N-2 \text{ and } y = N \\ N(N-1) & , \text{ if } x = N \text{ and } y = N-2 \\ 0 & , \text{ otherwise} \end{cases}$$

Thus for $x \in \llbracket 0, N-3 \rrbracket$, the relation $\pi(x)\tilde{\mathcal{P}}(x, x+1) = \pi(x+1)\tilde{\mathcal{P}}(x+1, x)$ becomes

$$\pi(x)(N-x) - \pi(x+2)(x+2)(x+1) = \pi(x+1)(x+1)(N-x-1) \quad (29)$$

For $x = N-2$, the relation $\pi(N-2)\tilde{\mathcal{P}}(N-2, N) = \pi(N)\tilde{\mathcal{P}}(N, N-2)$ becomes

$$\pi(N-2)2 = \pi(N)N(N-1) \quad (30)$$

These relations lead us to introduce the function f on V defined by

$$\forall x \in \llbracket 0, N \rrbracket, \quad f(x) := \frac{\pi(x)}{\mathcal{P}(x)}$$

where \mathcal{P} is the Poisson distribution of parameter 1 (with the convention $f(N-1) = 0 = \pi(N-1)$). Indeed, (29) and (30) reduce to

$$\begin{aligned} \forall x \in \llbracket 0, N-3 \rrbracket, \quad f(x)(N-x) - f(x+2) &= f(x+1)(N-x-1) \\ 2f(N-2) &= f(N) \end{aligned}$$

namely

$$\forall x \in \llbracket 0, N-2 \rrbracket, \quad (f(x) - f(x+1))(N-x) = f(x+2) - f(x+1)$$

This relation leads to the introduction of the function g on V defined by

$$\forall x \in \llbracket 0, N-1 \rrbracket, \quad g(x) := f(x+1) - f(x)$$

since we get

$$\begin{aligned} \forall x \in \llbracket 0, N-2 \rrbracket, \quad g(x) &= -\frac{g(x+1)}{N-x} \\ &= \frac{g(x+2)}{(N-x)(N-x-1)} \\ &= (-1)^{N-x} \frac{g(N-1)}{(N-x)!} \\ &= (-1)^{N-x} \frac{f(N)}{(N-x)!} \end{aligned}$$

Taking into account that $g(N-2) = f(N-1) = 0$, we deduce that

$$\begin{aligned} \forall x \in \llbracket 0, N-2 \rrbracket, \quad f(x) &= -g(x) - g(x+1) - \dots - g(N-2) \\ &= f(N) \sum_{k=0}^{N-x} \frac{(-1)^k}{k!} \end{aligned}$$

a formula also valid for $x = N$, so finally

$$\forall x \in V, \quad \pi(x) = \frac{1}{A_N x!} \sum_{k=0}^{N-x} \frac{(-1)^k}{k!}$$

where $A_N = e/f(N)$. This quantity is also the normalization factor, since π is a probability, so we compute

$$\begin{aligned}
A_N &= \sum_{x \in V} \frac{1}{x!} \sum_{k=0}^{N-x} \frac{(-1)^k}{k!} \\
&= \sum_{x=0}^{N-2} \frac{1}{x!} \sum_{k=0}^{N-x} \frac{(-1)^k}{k!} + \frac{1}{N!} \\
&= \sum_{x=0}^N \frac{1}{x!} \sum_{k=0}^{N-x} \frac{(-1)^k}{k!} - \frac{1}{(N-1)!} \sum_{k=0}^1 \frac{(-1)^k}{k!} - \frac{1}{N!} \sum_{k=0}^0 \frac{(-1)^k}{k!} + \frac{1}{N!} \\
&= \sum_{x=0}^N \frac{1}{x!} \sum_{l=0}^{N-x} \frac{(-1)^{N-x-l}}{(N-x-l)!} - \frac{1}{(N-1)!} (1-1) - \frac{1}{N!} + \frac{1}{N!} \\
&= \sum_{l=0}^N \sum_{x=0}^{N-l} \frac{(-1)^{N-x-l}}{x!(N-x-l)!} \\
&= \sum_{l=0}^N \frac{1}{(N-l)!} (1-1)^{N-l} \\
&= 1
\end{aligned}$$

■

From this formula, we recover an upper bound on the total variation distance between π and \mathcal{P} almost as good as that of (1), but which is not going through a coupling. Indeed, we compute:

$$\begin{aligned}
\|\pi - \mathcal{P}\| &= \sum_{n \in \mathbb{Z}_+} (\pi(n) - \mathcal{P}(n))_+ \\
&= \sum_{n \in [0, N]} \left(\frac{D_{N-n}}{(N-n)!} - e^{-1} \right)_+ \frac{1}{n!} \\
&= \sum_{n \in [0, N]} \left(\frac{D_n}{n!} - e^{-1} \right)_+ \frac{1}{(N-n)!} \\
&= \sum_{n \in [0, N]} \left(\sum_{k=0}^n \frac{(-1)^k}{k!} - e^{-1} \right)_+ \frac{1}{(N-n)!} \\
&= \sum_{n \in [0, N]} \left(\sum_{k \geq n+1} \frac{(-1)^k}{k!} \right)_+ \frac{1}{(N-n)!} \\
&\leq \sum_{n \in [0, N], n \text{ odd}} \frac{1}{(n+1)!} \frac{1}{(N-n)!}
\end{aligned}$$

(where we used the alternance of the terms of the series $\sum_{k \geq 0} \frac{(-1)^k}{k!}$).

The last term is also equal to

$$\begin{aligned}
\sum_{n \in [0, N], n \text{ even}} \frac{1}{n!} \frac{1}{(N+1-n)!} &\leq \sum_{n \in [0, N]} \frac{1}{n!} \frac{1}{(N+1-n)!} \\
&= \frac{1}{(N+1)!} \sum_{n \in [0, N]} \binom{N+1}{n} \\
&= \frac{2^{N+1}}{(N+1)!}
\end{aligned}$$

For completeness, let us also recall a simple proof of the well-known formula (2):

Lemma 16 *For any $x \in V$, we have*

$$\pi(x) = \frac{D_{N-x}}{(N-x)!} \frac{1}{x!}$$

Proof

Fix $x \in V$ and denote I_x the set of subsets of $\llbracket N \rrbracket$ whose cardinal is x . By symmetry we have, denoting by σ a generic permutation,

$$\begin{aligned} \pi[\eta_1 = x] &= \sum_{I \in I_x} \pi[\forall i \in I, \sigma(i) = i, \forall j \in \llbracket N \rrbracket \setminus I, \sigma(j) \neq j] \\ &= \binom{N}{x} \pi[\forall j \in \llbracket N-x \rrbracket, \sigma(j) \neq j, \forall i \in \llbracket N-x+1, N \rrbracket, \sigma(i) = i] \\ &= \binom{N}{x} \frac{D_{N-x}}{N!} \\ &= \frac{D_{N-x}}{(N-x)!} \frac{1}{x!} \end{aligned}$$

■

Montmort's formula (3) is now a consequence of the above lemma and of Proposition 15:

Corollary 17 *We have for any $n \in \mathbb{N}$,*

$$D_n = n! \sum_{k=0}^n \frac{(-1)^k}{k!}$$

Remark 18

a) It seems from (28) that we have an extra relation for $p(0)$: $\pi(0)(N - 2p(0)) = \pi(1)(N - 1)$, which amounts to

$$p(0) = \frac{N}{2} \left(1 - \frac{D_{N-1}}{D_N} (N - 1) \right)$$

Comparing with (31), which gives for $x = 0$,

$$p(0) = \frac{1}{2} \frac{D_{N-2}}{(N-2)!} \frac{N!}{D_N}$$

we deduce

$$D_N = (N - 1)(D_{N-1} + D_{N-2})$$

This is the well-known iteration formula for the derangement numbers, see e.g. [25].

b) Note that π is not close to \mathcal{P} is the separation discrepancy

$$\mathfrak{s}(\pi, \mathcal{P}) = \sup \left\{ 1 - \frac{\pi(x)}{\mathcal{P}(x)} : x \in \mathbb{Z}_+ \right\}$$

since the r.h.s. is trivially 1. But with the notations of Section 5, we even have

$$\begin{aligned} \liminf_{N \rightarrow \infty} \mathfrak{s}(\tilde{\pi}, \zeta) &\geq \liminf_{N \rightarrow \infty} 1 - \frac{\tilde{\pi}(N-4)}{\zeta(N-4)} \\ &= \lim_{N \rightarrow \infty} 1 - \frac{\pi(N-4)}{\mathcal{P}(N-4)} \\ &= 1 - e^{-\frac{D_4}{4!}} \\ &> 0 \end{aligned}$$

This fact a priori excludes a proof via strong stationary times (see Diaconis and Fill [7]) in Section 5.

□

B Complements on the conditional expectation p

Some observations about p are gathered here.

Note that Lemma 16 also leads to an expression of the quantities $p(x)$ in terms of the number of derangements, from Lemma 14:

$$\begin{aligned} \forall x \in V, \quad p(x) &= \frac{1}{2} \frac{D_{N-x-2}}{(N-x-2)!} \frac{(N-x)!}{D_{N-x}} \\ &= \frac{1}{2} \frac{\sum_{k=0}^{N-x-2} \frac{(-1)^k}{k!}}{\sum_{l=0}^{N-x} \frac{(-1)^l}{l!}} \end{aligned} \quad (31)$$

This formula leads to an estimate of our quantities of interest, the $|2p(x) - 1|$, for $x \in \llbracket 0, N-2 \rrbracket$, of the same order as that of Proposition 6:

Lemma 19 *We have*

$$\begin{aligned} \forall x \in \llbracket 0, N-2 \rrbracket, \quad |2p(x) - 1| &\leq 3 \frac{N-x-1}{(N-x)!} \\ &\leq \frac{3}{(N-x-1)!} \end{aligned}$$

and in particular we get, for $N \geq 4$,

$$\forall x \in \llbracket 0, N-4 \rrbracket, \quad \frac{1}{4} \leq p(x) \leq \frac{3}{4}$$

Proof

From (31) we deduce:

$$\begin{aligned} \forall x \in V, \quad 2p(x) &= \frac{\sum_{k=0}^{N-x-2} \frac{(-1)^k}{k!}}{\sum_{l=0}^{N-x} \frac{(-1)^l}{l!}} \\ &= 1 - \frac{\sum_{N-x-1}^{N-x} \frac{(-1)^k}{k!}}{\sum_{l=0}^{N-x} \frac{(-1)^l}{l!}} \end{aligned}$$

implying

$$\begin{aligned} \forall x \in V, \quad |2p(x) - 1| &= \frac{\left| \frac{1}{(N-x-1)!} - \frac{1}{(N-x)!} \right|}{\sum_{l=0}^{N-x} \frac{(-1)^l}{l!}} \\ &= \frac{1}{(N-x-1)!} \frac{1 - \frac{1}{N-x}}{\sum_{l=0}^{N-x} \frac{(-1)^l}{l!}} \\ &= \frac{N-x-1}{(N-x)!} \frac{1}{\sum_{l=0}^{N-x} \frac{(-1)^l}{l!}} \end{aligned}$$

Note that the series $\sum_{l=0}^n \frac{(-1)^l}{l!}$ provide alternating approximations of e^{-1} , it follows that

$$\forall x \in \llbracket 0, N-2 \rrbracket, \quad \sum_{l=0}^3 \frac{(-1)^l}{l!} \leq \sum_{l=0}^{N-x} \frac{(-1)^l}{l!} \leq \sum_{l=0}^2 \frac{(-1)^l}{l!}$$

namely

$$\forall x \in \llbracket 0, N-2 \rrbracket, \quad \frac{1}{2!} - \frac{1}{3!} \leq \sum_{l=0}^{N-x} \frac{(-1)^l}{l!} \leq \frac{1}{2!}$$

i.e.

$$\forall x \in \llbracket 0, N-2 \rrbracket, \quad \frac{1}{3} \leq \sum_{l=0}^{N-x} \frac{(-1)^l}{l!} \leq \frac{1}{2}$$

whose lower bound leads to the first desired estimate.

For the second estimate, note that

$$\begin{aligned} \forall x \in \llbracket 0, N-4 \rrbracket, \quad \frac{3}{(N-x-1)!} &\leq \frac{3}{(N-(N-4)-1)!} \\ &\leq \frac{3}{3!} = \frac{1}{2} \end{aligned}$$

■

Lemma 19 can be used similarly to Proposition 6 in Section 5, leading to the same conclusion.

Coming back to the formulation (6) of p as a conditional expectation of η_2 , it is natural to wonder if it could not be deduced from symmetry arguments. Remark it is true for the whole expectation: $\mathbb{E}_\nu[\eta_2] = 1/2$ (see the proof of Lemma 22 below with $k = 0$), in the same way one immediately gets $\mathbb{E}_\nu[\eta_1] = 1$. So to finish this appendix, let us show that symmetry arguments lead to a natural linear equation satisfied by p , even if we did not find how to use it to deduce the a priori bounds similar to those of Proposition 6 or Lemma 19.

For $k \in \llbracket 0, N \rrbracket$, denote

$$\mathcal{A}_k = \{(i_1, i_2, \dots, i_k) \in \llbracket N \rrbracket^k : m \neq n \in \llbracket k \rrbracket \Rightarrow i_m \neq i_n\}$$

In particular, we have

$$|\mathcal{A}_k| = N(N-1) \cdots (N-k+1) \tag{32}$$

(by convention, $\mathcal{A}_0 = \{\emptyset\}$ and $|\mathcal{A}_0| = 1$).

For $k \in \llbracket 0, N \rrbracket$, we define the mapping F_k on the symmetric group \mathcal{S}_N via

$$\forall \sigma \in \mathcal{S}_N, \quad F_k(\sigma) := \sum_{(i_1, \dots, i_k) \in \mathcal{A}_k} \prod_{j \in \llbracket k \rrbracket} \mathbb{1}_{\{\sigma(i_j) = i_j\}}$$

Let us check these mappings are functions of η_1 (the number of fixed points):

Lemma 20 *For any $k \in \llbracket 0, N \rrbracket$, we have*

$$F_k = \eta_1(\eta_1 - 1) \cdots (\eta_1 - k + 1) \tag{33}$$

Proof

Indeed, for any given $\sigma \in \mathcal{S}_N$, denote $\mathcal{F}(\sigma)$ the set of fixed points of σ . We have

$$\begin{aligned} F_k(\sigma) &= |\mathcal{A}_k \cap \mathcal{F}(\sigma)^k| \\ &= \eta_1(\sigma)(\eta_1(\sigma) - 1) \cdots (\eta_1(\sigma) - k + 1) \end{aligned}$$

■

Remark 21 Since F_k is a polynomial of order k in η_1 , any function of η_1 can be expressed as a linear combination of the F_k for $k \in \llbracket 0, N \rrbracket$, and even only for $k \in \llbracket 0, N - 1 \rrbracket$ or alternatively $k \in V$, since η_1 is taking N values, those of $V := \llbracket 0, N \rrbracket \setminus \{N - 1\}$. \square

It follows that if we want to prove that

$$\mathbb{E}_\nu[\eta_2|\eta_1] = f(\eta_1)$$

for a given function $f : V \rightarrow \mathbb{R}_+$, it is sufficient to check that

$$\forall k \in \llbracket 0, N \rrbracket, \quad \mathbb{E}_\nu[\eta_2 F_k] = \mathbb{E}_\nu[f(\eta_1) F_k]$$

We are thus led to compute the l.h.s.

Lemma 22 *For any $k \in \llbracket 0, N \rrbracket$, we have*

$$\mathbb{E}_\nu[\eta_2 F_k] = \begin{cases} 1/2 & , \text{ if } k \in \llbracket 0, N - 2 \rrbracket \\ 0 & , \text{ if } k \in \{N - 1, N\} \end{cases}$$

Proof

Note that

$$\begin{aligned} \forall \sigma \in \mathcal{S}_N, \quad \eta_2(\sigma) &= \frac{1}{2} \sum_{m \in \llbracket N \rrbracket} \mathbf{1}_{\{\sigma(m) \neq m, \sigma^2(m) = m\}} \\ &= \frac{1}{2} \sum_{m \neq n \in \llbracket N \rrbracket} \mathbf{1}_{\{\sigma(m) = n, \sigma(n) = m\}} \end{aligned}$$

so that

$$2\mathbb{E}_\nu[\eta_2 F_k] = \sum_{m \neq n \in \llbracket N \rrbracket} \sum_{(i_1, \dots, i_k) \in \mathcal{A}_k} \mathbb{E}_\nu \left[\mathbf{1}_{\{\sigma(m) = n, \sigma(n) = m\}} \prod_{j \in \llbracket k \rrbracket} \mathbf{1}_{\{\sigma(i_j) = i_j\}} \right]$$

Note that the above expectation vanishes if $m \in \{i_1, \dots, i_k\}$ or $n \in \{i_1, \dots, i_k\}$, so writing $i_{k+1} = m$ and $i_{k+2} = n$, we end up with

$$\begin{aligned} 2\mathbb{E}_\nu[\eta_2 F_k] &= \sum_{(i_1, \dots, i_k, i_{k+1}, i_{k+2}) \in \mathcal{A}_{k+2}} \mathbb{E}_\nu \left[\mathbf{1}_{\{\sigma(i_{k+1}) = i_{k+2}, \sigma(i_{k+2}) = i_{k+1}\}} \prod_{j \in \llbracket k \rrbracket} \mathbf{1}_{\{\sigma(i_j) = i_j\}} \right] \\ &= \sum_{(i_1, \dots, i_k, i_{k+1}, i_{k+2}) \in \mathcal{A}_{k+2}} \pi[\sigma(i_1) = i_1, \sigma(i_2) = i_2, \dots, \sigma(i_k) = i_k, \sigma(i_{k+1}) = i_{k+2}, \sigma(i_{k+2}) = i_{k+1}] \end{aligned}$$

For any $(i_1, \dots, i_k, i_{k+1}, i_{k+2}) \in \mathcal{A}_{k+2}$, the above probability can be computed by first choosing $\sigma(i_1) = i_1$, whose probability is $1/N$, next choosing $\sigma(i_2) = i_2$, whose subsequent probability is $1/(N - 1)$, etc, up to choosing $\sigma(i_{k+2}) = i_{k+1}$, whose probability is $1/(N - k - 1)$. We deduce

$$\pi[\sigma(i_1) = i_1, \sigma(i_2) = i_2, \dots, \sigma(i_k) = i_k, \sigma(i_{k+1}) = i_{k+2}, \sigma(i_{k+2}) = i_{k+1}] = \frac{1}{N(N - 1) \cdots (N - k - 1)}$$

and by consequence

$$\begin{aligned} 2\mathbb{E}_\nu[\eta_2 F_k] &= \frac{|\mathcal{A}_{k+2}|}{N(N - 1) \cdots (N - k - 1)} \\ &= 1 \end{aligned}$$

due to (32), at least when $k + 2 \leq N$. Obviously, when $k \in \llbracket 0, N - 1 \rrbracket$ satisfies $k + 2 > N$, namely when $k \in \{N - 1, N\}$, we end up with $2\mathbb{E}_\nu[\eta_2 F_k] = 0$. \blacksquare

Lemma 23 For any $k \in \llbracket 0, N \rrbracket$, we have

$$\mathbb{E}_\nu[F_k] = 1$$

Proof

Indeed, as in the above proof,

$$\begin{aligned} \mathbb{E}_\nu[F_k] &= \sum_{(i_1, \dots, i_k) \in \mathcal{A}_k} \pi[\sigma(i_1) = i_1, \sigma(i_2) = i_2, \dots, \sigma(i_k) = i_k] \\ &= \sum_{(i_1, \dots, i_k) \in \mathcal{A}_k} \frac{1}{N(N-1) \cdots (N-k+1)} \\ &= \frac{|\mathcal{A}_k|}{N(N-1) \cdots (N-k+1)} \\ &= 1 \end{aligned}$$

■

It follows that if $f : V \rightarrow \mathbb{R}$ is a function satisfying

$$\forall k \in V, \quad \mathbb{E}_\nu[f(\eta_1)F_k] = 1 \quad (34)$$

then we can conclude that $f = \mathbb{1}$, the function only taking the value 1.

Consider $f : V \rightarrow \mathbb{R}$ given by the conditional expectation

$$f(\eta_1) = \mathbb{E}[2\eta_2 | \eta_1]$$

According to Lemma 22, f almost satisfies (34), the only discrepancy being the case $k = N$. Of course it can not satisfy (34), otherwise we would get from Section 5 that the law of η_1 is the conditioning of the Poisson distribution of parameter 1 to V and this is not true (e.g. due to (4)).

Nevertheless, Lemma 22 leads to a linear equation for f . Denote $a := (a_k)_{k \in V}$ the vector of the coefficients in the writing

$$f(\eta_1) =: \sum_{k \in V} a_k F_k$$

we have

$$Ga = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{pmatrix}$$

i.e.

$$a = G^{-1} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{pmatrix}$$

where $G := (G_{k,l})_{k,l \in V}$ is the Gram matrix given by

$$\forall k, l \in V, \quad G_{k,l} := \mathbb{E}_\nu[F_k F_l] \quad (35)$$

In accordance with Remark 21, the family $(F_k)_{k \in V}$ is linearly independent in $\mathbb{L}^2(\pi_1)$, due to the fact that $\pi_1(x) > 0$ for all $x \in V$, which implies that $\dim(\mathbb{L}^2(\pi_1)) = N$. As a consequence, G is invertible.

As seen in Section 5, more interesting for us is the function $g := f - \mathbb{1}$ defined on V . Since Lemma 23 shows that

$$\mathbb{1} = \sum_{k \in V} b_k F_k$$

with

$$\begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{N-2} \\ b_{N-1} \end{pmatrix} = G^{-1} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}$$

we deduce that $g(\eta_1) = \sum_{k \in V} c_k F_k$, with

$$\begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_{N-2} \\ c_{N-1} \end{pmatrix} = G^{-1} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

More precisely, the computations of Section 5 show the only a priori informations we need to control our coupling constructions are estimates on expressions such as

$$\sum_{x \in \llbracket 0, N-2 \rrbracket} |g(x)| \frac{1}{ex!} \tag{36}$$

Below we compute the entries of G directly via symmetry arguments, without a priori knowledge of the law π of η_1 , nevertheless, it does not seem very helpful to estimate expressions such as (36).

Proposition 24 *The matrix G is symmetric and extending Definition (35) to any $k, l \in \llbracket 0, N \rrbracket$, we have*

$$\forall k \leq l \in \llbracket 0, N \rrbracket, \quad G_{k,l} = k! \sum_{r=0}^{k \wedge (N-l)} \frac{1}{r!} \binom{l}{k-r}$$

Proof

For any $i = (i_1, \dots, i_k) \in \mathcal{A}_k$, denote $\{i\}$ the set $\{i_1, \dots, i_k\} \subset \llbracket N \rrbracket$, as well as

$$\begin{aligned} \tilde{\mathcal{A}}_k &:= \{\{i\} : i \in \mathcal{A}_k\} \\ &= \{S \subset \llbracket N \rrbracket : |S| = k\} \end{aligned}$$

We compute, for any $k \leq l \in \llbracket N \rrbracket$,

$$\begin{aligned} \mathbb{E}_\nu[F_k F_l] &= k!! \sum_{S \in \tilde{\mathcal{A}}_k, T \in \tilde{\mathcal{A}}_l} \mathbb{P}[\forall s \in S, \sigma(s) = s, \forall t \in T, \sigma(t) = t] \\ &= k!! \sum_{S \in \tilde{\mathcal{A}}_k, T \in \tilde{\mathcal{A}}_l} \frac{1}{N(N-1) \cdots (N - |S \cup T| + 1)} \\ &= k!! \sum_{u \in \llbracket l, (l+k) \wedge N \rrbracket} A_{k,l}(u) \frac{(N-u)!}{N!} \end{aligned}$$

where

$$\forall u \in \llbracket l, (l+k) \wedge N \rrbracket, \quad A_{k,l}(u) := |\{(S, T) \in \tilde{\mathcal{A}}_k \times \tilde{\mathcal{A}}_l : |S \cup T| = u\}|$$

Note that for any fixed $T \in \tilde{\mathcal{A}}_l$ and $r \in \llbracket 0, k \wedge (N-l) \rrbracket$, we have

$$|\{S \in \tilde{\mathcal{A}}_k : |S \setminus T| = r\}| = \binom{N-l}{r} \binom{l}{k-r}$$

the r.h.s. corresponding to the number of choices of r elements in $\llbracket N \rrbracket \setminus T$ and $k-r$ elements in T .

It follows that, with the change of variable $u = l+r$,

$$A_{k,l}(l+r) = \binom{N}{l} \binom{N-l}{r} \binom{l}{k-r}$$

and by consequence

$$\begin{aligned} G_{k,l} &= k!l! \sum_{r \in \llbracket 0, k \wedge (N-l) \rrbracket} \binom{N}{l} \binom{N-l}{r} \binom{l}{k-r} \frac{(N-l-r)!}{N!} \\ &= \frac{k!l!}{N!} \binom{N}{l} \sum_{r \in \llbracket 0, k \wedge (N-l) \rrbracket} \binom{N-l}{r} \binom{l}{k-r} (N-l-r)! \\ &= \frac{k!l!}{(N-l)!} \sum_{r \in \llbracket 0, k \wedge (N-l) \rrbracket} \frac{(N-l)!}{r!(N-l-r)!} \frac{l!}{(l-k+r)!(k-r)!} (N-l-r)! \\ &= k! \sum_{r \in \llbracket 0, k \wedge (N-l) \rrbracket} \frac{1}{r!} \frac{l!}{(l-k+r)!(k-r)!} \\ &= k! \sum_{r=0}^{k \wedge (N-l)} \frac{1}{r!} \binom{l}{k-r} \end{aligned}$$

■

References

- [1] Richard Arratia, A. D. Barbour, and Simon Tavaré. *Logarithmic combinatorial structures: A probabilistic approach*. EMS Monogr. Math. Zürich: European Mathematical Society (EMS), 2003.
- [2] Sourav Chatterjee, Persi Diaconis, and Elizabeth Meckes. Exchangeable pairs and Poisson approximation. *Probab. Surv.*, 2:64–106, 2005.
- [3] Klara Courteaut, Kurt Johansson, and Gaultier Lambert. From Berry-Esseen to super-exponential. arXiv preprint, 2022.
- [4] F. N. David and D. E. Barton. *Combinatorial chance*. Hafner Publishing Co., New York, 1962.
- [5] Pierre Rémond de Montmort. Essay d’analyse sur les jeux de hazard. 3rd ed. fotogr. repr. of the 2nd ed. of 1713/14, Paris. New York: Chelsea Publishing Company. xlii, 414 p. \$ 39.50 (1980)., 1980.
- [6] P. Diaconis and C. Mallows. New representations for the characteristic polynomial of a random matrix in the classical groups. draft manuscript, 2000.
- [7] Persi Diaconis and James Allen Fill. Strong stationary times via a new form of duality. *Ann. Probab.*, 18(4):1483–1522, 1990.

- [8] Persi Diaconis and Peter J. Forrester. Hurwitz and the origins of random matrix theory in mathematics. *Random Matrices Theory Appl.*, 6(1):26, 2017. Id/No 1730001.
- [9] Persi Diaconis, Jason Fulman, and Robert Guralnick. On fixed points of permutations. *J. Algebr. Comb.*, 28(1):189–218, 2008.
- [10] Persi Diaconis and Susan Holmes, editors. *Stein’s method: expository lectures and applications. Papers from the workshop on Stein’s method, Stanford, CA, USA, 1998.*, volume 46 of *IMS Lect. Notes, Monogr. Ser.* Beachwood, OH: IMS, Institute of Mathematical Statistics, 2004.
- [11] Persi Diaconis, Eddy Mayer-Wolf, Ofer Zeitouni, and Martin P. W. Zerner. The Poisson-Dirichlet law is the unique invariant distribution for uniform split-merge transformations. *Ann. Probab.*, 32(1B):915–938, 2004.
- [12] Persi Diaconis and Mehrdad Shahshahani. On the eigenvalues of random matrices. volume 31A, pages 49–62. 1994. *Studies in applied probability.*
- [13] Philippe Duchon and Romaric Duvignau. A new generation tree for permutations, preserving the number of fixed points. In *Proceedings of the 26th international conference on formal power series and algebraic combinatorics, FPSAC 2014, Chicago, IL, USA, June 29 – July 3, 2014*, pages 679–690. Nancy: The Association. Discrete Mathematics & Theoretical Computer Science (DMTCS), 2014.
- [14] Philippe Duchon and Romaric Duvignau. Preserving the number of cycles of length k in a growing uniform permutation. *Electron. J. Comb.*, 23(4):research paper p4.22, 39, 2016.
- [15] Richard A. Holley and Daniel W. Stroock. Applications of the stochastic Ising model to the Gibbs states. *Comm. Math. Phys.*, 48(3):249–265, 1976.
- [16] F. P. Kelly. *Reversibility and stochastic networks.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2011. Revised edition of the 1979 original with a new preface.
- [17] John G. Kemeny and J. Laurie Snell. *Finite Markov chains. With a new appendix “Generalization of a fundamental matrix”.* 3rd printing. Undergraduate Texts Math. Springer, Cham, 1983.
- [18] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.
- [19] Thomas M. Liggett. *Interacting particle systems.* Classics in Mathematics. Springer-Verlag, Berlin, 2005. Reprint of the 1985 original.
- [20] Torgny Lindvall. *Lectures on the coupling method.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1992. A Wiley-Interscience Publication.
- [21] C. Y. Amy Pang. Lumpings of algebraic Markov chains arise from subquotients. *J. Theor. Probab.*, 32(4):1804–1844, 2019.
- [22] Jim Pitman. A modified Pascal triangle for uniform renewals, ascents and peaks of permutations, and Poisson matching distributions. Unpublished draft, personal communication, April 2020.
- [23] Charles Stein. *Approximate computation of expectations*, volume 7 of *IMS Lect. Notes, Monogr. Ser.* Hayward, CA: Institute of Mathematical Statistics, 1986.
- [24] Lajos Takacs. The problem of coincidences. *Arch. Hist. Exact Sci.*, 21:229–244, 1979.
- [25] Wikipedia contributors. Derangement — Wikipedia, the free encyclopedia. <https://en.wikipedia.org/w/index.php?title=Derangement&oldid=1084547992>, 2022. [Online; accessed 20-June-2022].

diaconis@math.stanford.edu
Department of Mathematics
Department of Statistics
Stanford University, USA

miclo@math.cnrs.fr
Institut de Mathématiques de Toulouse
Université Paul Sabatier, 118, route de Narbonne
31062 Toulouse cedex 9, France
Toulouse School of Economics,
1, Esplanade de l'université
31080 Toulouse cedex 06, France