Electron. Commun. Probab. **22** (2017), no. 22, 1–7.
DOI: 10.1214/17-ECP3955
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Improved bounds for the mixing time of the random-to-random shuffle

Chuan Qin * Ben Morris †

Abstract

We prove an upper bound of $1.5321 n\log n$ for the mixing time of the random-to-random insertion shuffle, improving on the best known upper bound of $2n\log n$. Our proof is based on the analysis of a non-Markovian coupling.

Keywords: random-to-random shuffle; mixing time; non-Markovian coupling.
AMS MSC 2010: 60J10.
Submitted to ECP on November 29, 2014, final version accepted on November 26, 2015.

1 Introduction

How many shuffles does it take to mix up a deck of cards? Mathematicians have long been attracted to card shuffling problems. This is partly because of their natural beauty, and partly because they provide a testing ground for the more general problem of finding the mixing time of a Markov chain, which has applications to computer science, statistical physics and optimization.

Let X_t be a Markov chain on a finite state space V that converges to the uniform distribution. For probability measures μ and ν on V, define the *total variation distance* $||\mu-\nu|| = \frac{1}{2}\sum_{x\in V}|\mu(x)-\nu(x)|$, and define the ε-mixing time

$$T_{\mathrm{mix}}(\varepsilon) = \min\{t : ||\Pr(X_t = \cdot) - \mathcal{U}|| \le \varepsilon \text{ for all } x \in V\},$$

where $\mathcal{U}$ denotes the uniform distribution on V.

The random-to-random insertion shuffle has the following transition rule. At each step choose a card uniformly at random, remove it from the deck and then re-insert in to a random position. It has long been conjectured that the mixing time for the random-to-random insertion shuffle on n cards exhibits *cutoff* at a time on the order of $n\log n$. That is, there is a constant c such that for any $\varepsilon \in (0,1)$, the ε-mixing time is asymptotic to $cn\log n$. It has further been conjectured (see [4]) that the constant $c = \frac{3}{4}$.

Uyemura-Reyes [9] proved a lower bound of $\frac{1}{2}n\log n$. This was improved by Subag [7] to the conjectured value of $\frac{3}{4}n\log n$. However, a matching upper bound has not been found. Diaconis and Saloff-Coste [5] used comparison techniques to prove a $O(n\log n)$ upper bound. The constant was improved by Uyemura-Reyes [9] and then by Saloff-Coste and Zuniga [8], who proved upper bounds of $4n\log n$ and $2n\log n$, respectively. The main

*University of California, Davis. E-mail: lostconch@gmail.com
†University of California, Davis. Research supported by NSF grant CNS-1228828. E-mail: morris@math.ucdavis.edu

contribution of this paper is to improve the constant in the upper bound to 1.5321. We achieve this via a non-Markovian coupling that reduces the problem of bounding the mixing time to finding the second largest eigenvalue of a certain Markov chain on 10 states. We also use the technique of path coupling (see [1]).

2 Main result

For sequences a_n and b_n, we write $a_n \sim b_n$ if $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$ and $a_n \lesssim b_n$ if $\limsup_{n\to\infty} \frac{a_n}{b_n} \leq 1$. Let P be the transition matrix of the random-to-random insertion shuffle. Define

$$d(t) = \max_y ||P^t(y,\,\cdot) - \mathcal{U}|| \ .$$

When the number of cards is n, we write $d_n(t)$ for the value of $d(t)$, and $T^{(n)}_{\mathrm{mix}}(\varepsilon)$ for the ε-mixing time of the random-to-random insertion shuffle. Our main result is the following upper bound on $T^{(n)}_{\mathrm{mix}}(\varepsilon)$.

Theorem 2.1. *For any $\varepsilon \in (0,1)$ we have $T^{(n)}_{\mathrm{mix}}(\varepsilon) \lesssim 1.5321 n \log n$.*

We think of a permutation π in S_n as representing the order of a deck of n cards, with $\pi(i) =$ position of card i. Say x and x' are *adjacent*, and write $x \approx x'$, if $x' = (i,j)x$ for a transposition (i,j). We prove Theorem 2.1 using a path coupling argument (see [1]) and the following lemma.

Lemma 2.2. *If n is sufficiently large and x and x' are adjacent permutations in S_n, then there exist positive constants c and α such that*

$$||P^t(x,\,\cdot) - P^t(x',\,\cdot)|| \leq \frac{c}{n^{1+\alpha}} \qquad \text{for all } t > 1.5321 n \log n \ .$$

The proof of Lemma 2.2, which uses a non-Markovian coupling, is deferred to Section 3.

Proof of Theorem 2.1. Suppose that $t > 1.5321 n \log n$. By convexity of the l^1-norm, and since $\mathcal{U} = \frac{1}{n!}\sum_{z\in S_n} P^t(z,\,\cdot)$, it follows that for any state y we have

$$||P^t(y,\,\cdot) - \mathcal{U}|| \leq \max_z ||P^t(y,\,\cdot) - P^t(z,\,\cdot)|| \ . \tag{2.1}$$

Since any permutation in S_n can be written as a product of at most $n-1$ transpositions, by the triangle inequality the quantity on the righthand side of (2.1) is at most

$$(n-1) \max_{x\approx x'} ||P^t(x,\,\cdot) - P^t(x',\,\cdot)|| \ . \tag{2.2}$$

By (2.1), (2.2), and Lemma 2.2, if n is sufficiently large, there exist positive constants c and α such that

$$d(t) = \max_y ||P^t(y,\,\cdot) - \mathcal{U}|| \leq \frac{c(n-1)}{n^{1+\alpha}},$$

which tends to zero as $n \to \infty$. □

3 Proof of Lemma 2.2

Recall that we think of a permutation π in S_n as representing the order of a deck of n cards, with $\pi(i) =$ position of card i. Let $M_{i,j} : S_n \to S_n$ be the operation on permutations that removes the card of label i from the deck and re-inserts it

$$\begin{cases} \text{to the right of the card of label } j & \text{if } i \neq j; \\ \text{to the leftmost position} & \text{if } i = j. \end{cases}$$

We call such operations *shuffles.* If $\langle M_1, \dots, M_k \rangle$ is sequence of shuffles, we write $xM_1M_2 \cdots M_k$ for $M_k \circ M_{k-1} \cdots M_1(x)$.

The transition rule for the random-to-random insertion shuffle can now be stated as follows. If the current state is x, choose a shuffle M uniformly at random (that is, choose a and b uniformly at random and let $M = M_{a,b}$) and move to xM.

We call the numbers in $\{1, \dots, n\}$ *cards.* If a shuffle M removes card c from the deck and then re-inserts it, we call M a c-move.

If $\mathcal{P} = \langle M_1, M_2, \dots \rangle$ is a sequence of shuffles, we write $(\mathcal{P}x)_t$ for the permutation $xM_1 \cdots M_t$. Note that if $\mathcal{P}$ is a sequence of independent uniform random shuffles, then $\{(\mathcal{P}x)_t : t \geq 0\}$ is the random-to-random insertion shuffle started at x.

3.1 The Non-Markovian coupling

Fix a permutation x and $i, j \in \{1, 2, \dots, n\}$. The aim of this subsection is to define a coupling of the random-to-random insertion shuffle starting from x and $(i, j)x$, respectively. Suppose that we couple the processes so that the same labels are chosen for each shuffle. Note that if there is an i-move (respectively, j-move) followed at some point by a j-move (respectively, i-move), then the processes will couple at the time of the j-move (respectively, i-move) provided that any cards placed to the right of card j (respectively, i) at any intermediate time (and any cards placed to the right of those cards, and so on) were subsequently removed. We keep track of these "problematic" cards using a process we call the *queue*.

For positive integers k we will call a sequence $\langle M_1, \dots, M_k \rangle$ of shuffles a k*-path*. For a k-path $\mathcal{P}$, define the $\mathcal{P}$*-queue* (or, simply the *queue*) as the following Markov chain $\{Q_t : t = 0, \dots, k\}$ on subsets of cards. Initially, we have $Q_0 = \emptyset$. If the queue at time t is Q_t, and the shuffle at time $t+1$ is $M_{a,b}$, the next queue Q_{t+1} is

$$\begin{cases} \{i\} & \text{if } a = j; \\ \{j\} & \text{if } a = i; \\ Q_t \cup \{a\} & \text{if } a \notin \{i, j\} \text{ and } b \in Q_t - \{a\}. \\ Q_t - \{a\} & \text{otherwise.} \end{cases}$$

We call a shuffle an i-or-j move if it is an i-move or a j-move. Note that at any time after the first i-or-j move the queue contains exactly one card from $\{i, j\}$. Let $\mathcal{P} = \langle M_1, \dots, M_k \rangle$ be a k-path. For $t < k$, we say that t is a *good time* of $\mathcal{P}$ if

1. M_t is an i-or-j move;
2. there is a time $t' \in \{t+1, \dots, k\}$ such that
   (a) $M_{t'}$ is the next i-or-j move after M_t;
   (b) the queue is a singleton at time $t'-1$ (i.e., either $\{i\}$ or $\{j\}$);
   (c) the card moved at time t' is different from the card moved at time t.

Define

$$T = \begin{cases} \max\{t < k : t \text{ is a good time of } \mathcal{P}\}, & \text{if there is a good time of } \mathcal{P}, \\ \infty, & \text{otherwise.} \end{cases}$$

and call T the *last* good time of $\mathcal{P}$. Let $\theta_{i,j}\mathcal{P}$ be the k-path obtained from $\mathcal{P}$ by reversing the roles of i and j in each shuffle before time T (that is, by replacing shuffle $M_{a,b}$ with $M_{\pi(a),\pi(b)}$, where π is a transposition of i and j). Note that $\theta_{i,j}\mathcal{P}$ has i-or-j moves at the same times as $\mathcal{P}$. Furthermore, since the queue is reset at the times of i-or-j moves, the $\theta_{i,j}\mathcal{P}$-queue will have the same values as the $\mathcal{P}$-queue at all times $t \geq T$. It follows that the last good time of $\theta_{i,j}\mathcal{P}$ is the same as the last good time of $\mathcal{P}$, and hence

$\theta_{i,j}(\theta_{i,j}(\mathcal{P})) = \mathcal{P}$. Since $\theta_{i,j}$ is its own inverse, it is a bijection and hence if $\mathcal{P}$ is a uniform random k-path, then so is $\theta_{i,j}\mathcal{P}$.

Let $x' = (i,j)x$. Let $\mathcal{P}_k$ be a uniform random k-path, and let T_k be the last good time of $\mathcal{P}_k$. Note that $T_k < k$ or $T_k = \infty$. For t with $0 \le t \le k$, define

$$x_t = (\mathcal{P}_k x)_t \qquad x'_t = ((\theta_{i,j}\mathcal{P}_k)x')_t \ .$$

It is clear that x_t and x'_t have distributions $P^t(x,\cdot)$ and $P^t(x',\cdot)$, respectively, for all $t \le k$.

Lemma 3.1. *If $x_k \neq x'_k$ then $T_k = \infty$.*

Proof. Assume that $T_k < k$. Note that at any time $t < T_k$, the permutation $(\mathcal{P}_k x)_t$ can be obtained from $((\theta_{i,j}\mathcal{P}_k)x')_t$ by interchanging the cards i and j. Suppose that the next i-or-j move after time T_k occurs at time T'_k. Without loss of generality, there is an i-move at time T_k and a j-move at time T'_k. We claim that for times t with $T_k \le t < T'_k$, the permutation x'_t can be obtained from x_t by moving only the cards in Q_t, as shown in the diagram below. (In the diagram, the mth X in the top row represents the same card as the mth X in the bottom row, and Q represents all the cards in Q_t.)

$$\begin{array}{rcccccccccc} x_t: & X & X & X & X & X & X & Q & X & X & X \\ x'_t: & X & X & X & Q & X & X & X & X & X & X \end{array}$$

To see this, note that it holds at time T_k, when the queue is the singleton $\{j\}$ (since at this time the i's are placed in the same place), and the transition rule for the queue process ensures that if it holds at time t then it also holds at time $t+1$. The claim thus follows by induction. This means that at time $T'_k - 1$ the permutations differ only in the location of card j. That is, they are of the form:

$$\begin{array}{rcccccccccc} x_{T'_k-1}: & X & X & X & X & X & X & j & X & X & X \\ x'_{T'_k-1}: & X & X & X & j & X & X & X & X & X & X \end{array}$$

Thus at time T'_k, when card j is removed and then re-inserted into the deck, the two permutations become identical, and they remain identical until time k. ☐

3.2 Tail estimate of the coupling time

Recall that T_k is the last good time of a uniform random k-path.

Lemma 3.2. *Suppose that $k > 1.5321 n \log n$. Then there exist positive constants c and α such that $\mathbb{P}(T_k = \infty) \le \frac{c}{n^{1+\alpha}}$ for sufficiently large n.*

Proof. Consider a process $Y_t \in \{0,1,\dots\} \cup \infty$ that is defined as follows. The process starts in state ∞ and remains there until the first i-or-j move. From this point on, the value of Y_t is the size of the queue, until the first time that either

1. card i is moved when the queue is $\{i\}$, or
2. card j is moved when the queue is $\{j\}$.

At this point Y_t moves to state 0, which is an absorbing state. Note that $T_k = \infty$ exactly when $Y_k > 0$.

For $l = 1, 2, \dots$, define

$$q(l) = \begin{cases} \frac{1}{n} & \text{if } l = 1, \\ \frac{3n-1}{n^2} & \text{if } l = 2, \\ \frac{(l-1)(n-l+1)}{n^2} & \text{if } l \ge 3; \end{cases}$$

and define

$$p(l) = \begin{cases} \frac{n-2}{n^2} & \text{if } l = 1, \\ \frac{2n-6}{n^2} & \text{if } l = 2, \\ \frac{l(n-l-1)}{n^2} & \text{if } l \geq 3. \end{cases}$$

It is easy to check that Y_t is a Markov chain with the following transition rule. If the current state is 0, the next state is 0. If the current state is ∞ the next state is

$$\begin{cases} 1 & \text{with probability } \frac{2}{n}; \\ \infty & \text{with probability } \frac{n-2}{n}. \end{cases}$$

If the current state is $l \in \{1, 2, \dots\}$, the next state is

$$\begin{cases} l-1 & \text{with probability } q(l); \\ l+1 & \text{with probability } p(l); \\ 1 & \text{with probability } \frac{2}{n}, \text{ if } l \geq 3; \\ l & \text{with the remaining probability.} \end{cases}$$

Let $\tilde{Y}_t$ be the Markov chain on $\{0, 1, \dots, 8\} \cup \infty$ obtained from Y_t by replacing transitions to state 9 with transitions to ∞. That is, if K and $\tilde{K}$ denote the transition matrices of Y_t and $\tilde{Y}_t$, respectively, then

$$\tilde{K}(l, m) = \begin{cases} K(l, m) & \text{if } m \in \{0, 1, \dots, 8\}; \\ K(8, 9) & \text{if } l = 8 \text{ and } m = \infty. \end{cases}$$

The possible transitions of Y_t and $\tilde{Y}_t$ are indicated by the graph in Figure 1. We claim that if we start with $\tilde{Y}_0 = Y_0 = \infty$ then the distribution of $\tilde{Y}_t$ stochastically dominates the distribution of Y_t for all t. To see this, note that Y_t changes state with probability less than $\frac{1}{2}$ at each step, and when it changes state, it either makes a ± 1 move or it transitions to 1. Since for $m \in \{1, 2, \dots\} \cup \infty$, the transition probability $K(m, 1)$ is decreasing in m, it follows that Y_t is a monotone chain. (That is, $K(x, \cdot)$ is stochastically increasing in x; see [3].) The claim follows since $\tilde{Y}_t$ is obtained from Y_t by replacing moves to 9 with moves to the (larger) state of ∞.

Let $\tilde{K}_n$ be the value of the matrix $\tilde{K}$ when the number of cards is n, and $\hat{K}_n$ the matrix obtained by deleting the first row and the first column of $\tilde{K}_n$. If we write $A_n \to A$ for a sequence of matrices A_n and a fixed matrix A, it means that A_n converges to A component-wise as $n \to \infty$.

Define $C_n := n(\hat{K}_n - I)$, where I is the identity matrix. A straightforward calculation shows that $C_n \to C$ where

$$C = \begin{bmatrix} -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & -5 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & -7 & 3 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 3 & -9 & 4 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 4 & -11 & 5 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 5 & -13 & 6 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 6 & -15 & 7 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 7 & -17 & 8 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 \end{bmatrix}_{9\times 9}$$

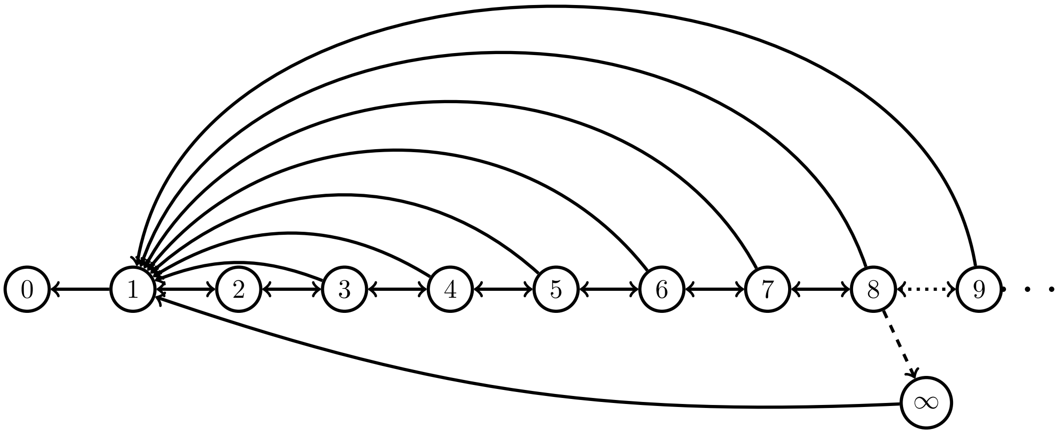

Figure 1: Graph indicating the possible transitions of Y_t and $\tilde{Y}_t$. The dotted edge indicates a possible transition of Y_t and the dashed edge indicates a transition of $\tilde{Y}_t$. (Self loops are not included.)

and that the eigenvalues of C are real and distinct (and hence C is diagonalizable), and negative. Denote the largest eigenvalue of C by $-\lambda$, where $\lambda = 0.652703\ldots$. (We can improve the eigenvalue marginally by considering a Markov chain with more than 10 states. For example with 35 states we get an eigenvalue of $-0.6527363\ldots$. However, we can't improve on this by more than 10^{-7} even if we use up to 100 states. Therefore, for simplicity we shall stick to our 10-state chain as a reasonable approximation to Y_t.)

Since $C^\top$ is diagonalizable, there exists an invertible 9×9 matrix Q such that $Q^{-1}C^\top Q = D$, where D is a diagonal matrix whose diagonal entries are the eigenvalues of C. Let $D_n = Q^{-1}C_n^\top Q$, and note that $D_n \to D$. For matrices A, let $\|A\|$ denote matrix norm induced by the l^1 norm on vectors. By continuity of the matrix exponential function and matrix norm, we have $\lim_{n\to\infty} \|e^{D_n}\| = \|e^D\| = e^{-\lambda}$. Since $\lambda > 0.6527$, it follows that $\|e^{D_n}\| \le e^{-0.6527}$ for sufficiently large n. Since $k/n > 1.5321 \log n$, submultiplicativity of operator norms implies that for sufficiently large n we have

$$\|e^{\frac{k}{n}D_n}\| \le e^{-0.6527\times 1.5321 \log n} \le \frac{1}{n^{1+\alpha}} \qquad \text{for some } \alpha > 0. \tag{3.1}$$

Since for any nonnegative integer j we have $(C_n^\top)^j = QD_n^jQ^{-1}$, it follows that

$$e^{\frac{1}{n}kC_n^\top} = Qe^{\frac{1}{n}kD_n}Q^{-1}. \tag{3.2}$$

Let X be a Poisson random variable with mean k that is independent of everything else. Then

$$e^{\frac{k}{n}C_n} = e^{k(\hat{K}_n - I)} = \sum_{j=0}^{\infty} e^{-k}\frac{k^j}{j!}\hat{K}_n^j = \sum_{j=0}^{\infty} \mathbb{P}(X=j)\hat{K}_n^j\,. \tag{3.3}$$

Let $x_0 = (0,0,\ldots,0,1) \in \mathbb{R}^9$. It follows from definition of $\tilde{Y}_t$ and (3.3) that

$$\mathbb{P}(\tilde{Y}_X > 0) = \sum_{j=0}^{\infty} \mathbb{P}(X=j)\left\|x_0\hat{K}_n^j\right\|_1 = \left\|\sum_{j=0}^{\infty} \mathbb{P}(X=j)x_0\hat{K}_n^j\right\|_1 = \left\|x_0 e^{\frac{k}{n}C_n}\right\|_1 .$$

By (3.2) and (3.1), there exists some $c > 0$ independent of n such that

$$\|x_0 e^{\frac{k}{n}C_n}\|_1 \le \left\| e^{\frac{k}{n}C_n^\top} \right\| = \|Q e^{\frac{k}{n}D_n} Q^{-1}\| \le \frac{C}{2}\|e^{\frac{k}{n}D_n}\| \le \frac{c}{2n^{1+\alpha}} .$$

Since Y_t is stochastically dominated by $\tilde{Y}_t$, we have

$$\mathbb{P}(Y_X > 0) \le \mathbb{P}(\tilde{Y}_X > 0) \le \frac{c}{2n^{1+\alpha}} .$$

Also, we have

$$\begin{aligned} \mathbb{P}(Y_X > 0) &= \sum_{j=0}^{\infty} \mathbb{P}(X = j)\mathbb{P}(Y_j > 0) \\ &\ge \mathbb{P}(Y_k > 0) \sum_{j=0}^{k} \mathbb{P}(X = j) \\ &\ge \frac{1}{2}\mathbb{P}(Y_k > 0), \end{aligned}$$

where the last line follows from the fact that the median of X (defined as the least integer m such that $\mathbb{P}(X \le m) \ge \frac{1}{2}$) equals $\mathbf{E}[X] = k$ (see [2]). Therefore, we have

$$\mathbb{P}(T_k = \infty) = \mathbb{P}(Y_k > 0) \le 2\mathbb{P}(Y_X > 0) \le \frac{c}{n^{1+\alpha}} \qquad \text{for sufficiently large } n. \qquad \square$$

Proof of Lemma 2.2. Recall that for any two probability measures μ and ν on a probability space Ω, we have

$$\|\mu - \nu\| = \min\{\mathbb{P}(X \ne Y) : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\} .$$

The main lemma then follows immediately from Lemma 3.1 and Lemma 3.2. □

References

[1] R. Bubley and M. Dyer, Path Coupling: A technique for proving rapid mixing in Markov Chains, Proceedings of the 38th Annual Symposium on Foundation of Computer Science, 223–231, 1997.

[2] K. P. Choi, On the Medians of Gamma Distributions and an Equation of Ramanujan, Proceedings of the American Mathematical Society 121 (1), 245–251, 1994. MR-1195477

[3] D. J. Daley, Stochastically monotone Markov chains, Z. Wahrscheinlichkeitstheorie verw. Geb. 10, 305–317, 1968. MR-0242270

[4] P. Diaconis, Mathematical developments from the analysis of riffle shuffling, In Groups, Combinatorics, Geometry (Durham 2001), 73–97, 2001.

[5] P. Diaconis and L. Saloff-Coste, Comparison techniques for random walks on finite groups, Ann. Probab. 21, 2131–2156, 1993.

[6] D. Levin, Y. Peres and E. Wilmer, Markov Chains and mixing time, American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

[7] E. Subag, A Lower Bound for the Mixing Time of the Random-to-Random Insertions Shuffle, Electron. J. Probab. 18, 1–20, 2013. MR-3035748

[8] L. Saloff-Coste and J. Zuniga, Refined estimates for some basic random walks on the symmetric and alternating groups, Latin American Journal of Probability and Mathematical Statistics 4, 359–392, 2008. MR-2461789

[9] J. Uyemura-Reyes, Random Walk, semi-direct products, and card shuffling, Ph.D. Thesis, Stanford University, 2002.

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS[1])
- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm