Iterated shuffle products

Steven N. Evans

U.C. Berkeley

June, 2016

Collaborators

Hye Soo Choi
Berkeley

Anton Wakolbinger
Frankfurt

Shuffles

- The shuffle of two words from some alphabet is the random word formed by interleaving the two words uniformly at random while maintaining the relative ordering of the letters in each word.
- Example: shuffling ab and cd we get the words $abcd$, $acbd$, $cabd$, $acdb$, $cadb$, $cdab$ — each with probability $\frac{1}{6}$.

Shuffle chains

- Build a Markov chain $(W_n)_{n \in \mathbb{N}_0}$ by at each stage shuffling a fixed word with the current word to produce a new word.
- Consider from now on the simplest nontrivial case where the alphabet is $\{a, b\}$ and the fixed word repeatedly shuffled in is ab.
- The transition probabilities describing the conditional distribution of W_3 given W_2 are

$$\begin{array}{cc} & \begin{array}{ccccc} aaabbb & aababb & aabbab & abaabb & ababab \end{array} \\ \begin{array}{c} aabb \\ abab \end{array} & \begin{pmatrix} 9/15 & 4/15 & 1/15 & 1/15 & 0 \\ 0 & 4/15 & 4/15 & 4/15 & 3/15 \end{pmatrix} \end{array}$$

- The marginal distribution of W_3 is $\mathbb{P}\{W_3 = aaabbb\} = \frac{6}{15}$, $\mathbb{P}\{W_3 = aababb\} = \frac{4}{15}$, $\mathbb{P}\{W_3 = aabbab\} = \frac{2}{15}$, $\mathbb{P}\{W_3 = abaabb\} = \frac{2}{15}$, and $\mathbb{P}\{W_3 = ababab\} = \frac{1}{15}$ – not uniform.

Ballot sequences and trees

- The Markov chain gives a growing sequence of ballot sequences (strings of n letters a and and n letters b such that for any $k \leqslant 2n$ the number of letters a in the first k letters is at least the number of letters b).
- Write $\mathbb{B}_n$ for the set of ballot sequences of length $2n$ ($=$ possible values of W_n). Note that $\#\mathbb{B}_n$ is the n^{th} Catalan number $\frac{1}{n+1}\binom{2n}{n}$.
- By standard bijections, the Markov chain gives "growing" sequences of various kinds of planar trees (or any of the hundreds of other objects counted by the Catalan numbers).

Compactification

- Can we use Doob–Martin theory to compactify the space of finite ballot sequences (equivalently, various spaces of objects such as planar trees)?
- Essentially, "How can the Markov chain be conditioned to behave at large times?"

Words to matchings

- A matching of $[2n]$ is a partition of $[2n]$ into subsets of size 2.
- Given a word $w = w_1 \dots w_{2n} \in \mathbb{B}_n$, a matching $\mathcal{M}$ of $[2n]$ is an admissible associated matching if for every block $\{i, j\}$ of $\mathcal{M}$ with $i < j$ we have $w_i = a$ and $w_j = b$.
- Example: if $n = 3$ and $w = aababb$, then the admissible associated matchings of the word w are $\{\{1,3\},\{2,5\},\{4,6\}\}$, $\{\{1,3\},\{2,6\},\{4,5\}\}$, $\{\{1,5\},\{2,3\},\{4,6\}\}$, and $\{\{1,6\},\{2,3\},\{4,5\}\}$.

Matchings to words

- Any matching of $\mathcal{M}$ of $[2n]$ defines a word $w \in \mathbb{B}_n$ for which $\mathcal{M}$ is an admissible associated matching: if $\{i, j\}$ is a block of $\mathcal{M}$ with $i < j$, then place a letter a in the position i and a letter b in the position j.
- Denote this word by $\Phi(\mathcal{M})$.
- Example: $\Phi(\{\{1,3\},\{2,5\},\{4,6\}\}) = aababb$.
- Let $\Lambda(w) := \#\{\mathcal{M} : \Phi(\mathcal{M}) = w\}$ be the number of admissible associated matchings of a word w.
- Example: $\Lambda(aababb) = 4$.

Counting admissible associated matchings

- Given a word $w = w_1 \dots w_{2n} \in \mathbb{B}_n$, the number of admissible associated matchings is

$$\Lambda(w) = \prod_{1 \leqslant k \leqslant 2n,\, w_k = a} \left(\#\{1 \leqslant i \leqslant k : w_i = a\} - \#\{1 \leqslant j \leqslant k : w_j = b\}\right).$$

- That is, if we write i_p for the index of the p^{th} letter a in w, then

$$\Lambda(w) = \prod_{p=1}^{n} h(i_p),$$

where

$$h(t) := \#\{1 \leqslant i \leqslant t : w_i = a\} - \#\{1 \leqslant j \leqslant t : w_j = b\}$$

is the height at time t of the path that makes a $+1$ step for each a and a -1 step for each b. Note that $h(t) \geqslant 0$ for $0 \leqslant t \leqslant 2n$ and $h(0) = h(2n) = 0$.

Matching chain

- Write $I_n, J_n \in [2n]$ for the positions into which a, b are shuffled to produce W_n.
- Construct as follows a Markov chain $(M_n)_{n \in \mathbb{N}_0}$ such that M_n is an admissible associated matching of W_n: in going from M_n to M_{n+1} match I_{n+1} and J_{n+1} (that is, make $\{I_{n+1}, J_{n+1}\}$ a block of the partition M_{n+1}) and define the remaining blocks by taking each block $\{k, \ell\}$ of M_n with $k < \ell$ and transforming it into the block $\{p, q\}$ of M_{n+1}, where
  - $p = k$ and $q = \ell$ if $k < \ell < I_{n+1} < J_{n+1}$,
  - $p = k$ and $q = \ell + 1$ if $k < I_{n+1} < \ell + 1 < J_{n+1}$,
  - $p = k + 1$ and $q = \ell + 1$ if $I_{n+1} < k + 1 < \ell + 1 < J_{n+1}$,
  - $p = k + 1$ and $q = \ell + 2$ if $I_{n+1} < k + 1 < J_{n+1} < \ell + 2$,
  - $p = k + 2$ and $q = \ell + 2$ if $I_{n+1} < J_{n+1} < k + 2 < \ell + 2$.

  That is, M_{n+1} with the block $\{I_{n+1}, J_{n+1}\}$ removed is a matching of $[2n+2] \backslash \{I_{n+1}, J_{n+1}\}$ obtained by pushing M_n forwards using the increasing bijection from $[2n]$ to $[2n+2] \backslash \{I_{n+1}, J_{n+1}\}$.

Uniformity

- For each $n \in \mathbb{N}_0$, the random matching M_n is uniformly distributed over the $\frac{1}{n!}\prod_{k=1}^{n}\binom{2k}{2} = (2n-1)(2n-3)\cdots 3\cdot 1 = (2n-1)!!$ matchings of $[2n]$.
- For each $n \in \mathbb{N}_0$ and $w \in \mathbb{B}_n$,

$$\mathbb{P}\{W_n = w\} = \mathbb{P}\{\Phi(M_n) = w\} = \frac{\Lambda(w)}{(2n-1)!!}.$$

“Filtering”

- For each $n \in \mathbb{N}_0$ and $w \in \mathbb{B}_n$, the conditional distribution of M_n given $W_n = w$ is uniform on the $\Lambda(w)$ admissible associated matchings of w.

Going backwards in time

- For $v \in \mathbb{B}_n$ and $w \in \mathbb{B}_{n+1}$,

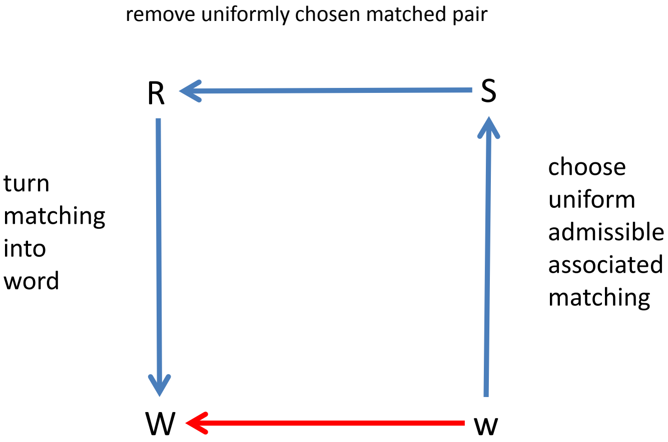
$$\mathbb{P}\{W_n = v \,|\, W_{n+1} = w\} = \frac{1}{n+1} N(v,w) \frac{\Lambda(v)}{\Lambda(w)},$$

  where $N(v,w)$ is the number of pairs (i,j) with $1 \leqslant i < j \leqslant 2(n+1)$ such that $w = v_1 \ldots v_{i-1} a v_i \ldots v_{j-2} b v_{j-1} \ldots v_{2n}$.
- The conditional distribution of M_n given $M_{n+1} = \mathcal{M}$ is the distribution of the random partition of $[2n]$ that is produced by first removing a block $\{i,j\}$ uniformly at random from the $n+1$ blocks of $\mathcal{M}$ to produce a matching of the set $[2n+2] \backslash \{i,j\}$ and then applying the unique increasing bijection from $[2n+2] \backslash \{i,j\}$ to $[2n]$ to turn this matching into a matching of $[2n]$.

Combining filtering and going backwards in time

- Consider $w \in \mathbb{B}_{n+1}$ and construct a random matching R of $[2n]$ as follows. Let S be a uniform random admissible associated matching for w and let R be such that the conditional distribution of R given $S = \mathcal{S}$ coincides with the conditional distribution of M_n given $M_{n+1} = \mathcal{S}$.
- Then, the distribution of R is the same as the conditional distribution of M_n given $\{W_{n+1} = w\}$.
- Thus, the distribution of the random word $\Phi(R)$ coincides with the conditional distribution of W_n given $\{W_{n+1} = w\}$.
- Moreover, given $\Phi(R)$ the conditional distribution of the random matching R is uniform on the set of admissible associated matchings of $\Phi(R)$.

Combining filtering and going backwards in time



$$P\{W=*\} = P\{W_n = * \mid W_{n+1} = w\}$$

Labeled matchings

- A labeled matching of $[2n]$ is a matching in which the n blocks are labeled with distinct elements of $[n]$.
- Using the same randomness that was used to construct $W_1, W_2, \ldots$ and $M_1, M_2, \ldots$, build a Markov chain $L_1, L_2, \ldots$ such that L_n is a labeled matching of $[2n]$ for $n \in \mathbb{N}$: the blocks of L_n are the same as the blocks of M_n and in going from L_n to L_{n+1} the newly created block $\{I_{n+1}, J_{n+1}\}$ is labeled with $n+1$ whilst the blocks that arise by transforming blocks already present in M_n keep their labels.
- Given a labeled matching $\mathcal{L}$ of $[2n]$, let $\Psi(\mathcal{L}) \in \mathbb{B}_n$ be the corresponding word (that is, forget about the labels and for each block $\{i, j\}$ with $i < j$ we place a letter a in position i and a letter b in position j).
- By construction, $\Psi(L_n) = W_n$.

Uniformity with labels

- For each $n \in \mathbb{N}$, the random matching L_n is uniformly distributed over the $n!(2n-1)!!$ labeled matchings of $[2n]$.

“Filtering” with labels

- For each $n \in \mathbb{N}$, the conditional distribution of L_n given M_n is uniform over the $n!$ labelings of M_n.
- The conditional distribution of L_n given W_n is uniform over the $n!\Lambda(W_n)$ labeled admissible associated matchings of W_n.

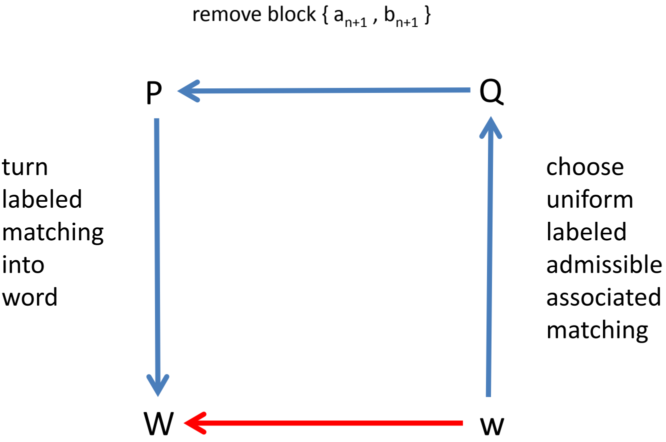
Going backwards in time with labels

- The labeled matching L_n is obtained from the labeled matching L_{n+1} by removing the block labeled $n+1$ and if this block contains the indices $\{i,j\}$ applying the unique increasing bijection from $[2n+2]\backslash\{i,j\}$ to $[2n]$ to turn this labeled matching of $[2n+2]\backslash\{i,j\}$ into a labeled matching of $[2n]$.
- Note that the backward transition dynamics are deterministic.

Combining filtering and going backwards in time with labels

- Consider $w \in \mathbb{B}_{n+1}$ and construct a random labeled matching P of $[2n]$ as follows. Let Q be a uniform random labeled admissible associated matching for w and let P be such that the conditional distribution of P given $\{Q = \mathcal{Q}\}$ coincides with the conditional distribution of L_n given $\{L_{n+1} = \mathcal{Q}\}$. Then, the distribution of P is the same as the conditional distribution of L_n given $\{W_{n+1} = w\}$.
- Thus, the distribution of the random word $\Psi(P)$ coincides with the conditional distribution of W_n given $\{W_{n+1} = w\}$.
- Moreover, given $\Psi(P)$, the conditional distribution of the random labeled matching P is uniform on the set of labeled admissible associated matchings of $\Psi(P)$.

Combining filtering and going backwards in time with labels



$$P\{W=*\} = P\{W_n = * \mid W_{n+1} = w\}$$

Bridges

- For $w \in \mathbb{B}_n$, let $(W_0^w, W_1^w, \ldots, W_n^w)$ be the bridge obtained by conditioning $(W_0, W_1, \ldots, W_n)$ on the event $\{W_n = w\}$.
- All bridges have the same backwards transition probabilities as $(W_n)_{n \in \mathbb{N}_0}$:

$$\mathbb{P}\{W_k^w = u \,|\, W_{k+1}^w = v\} = \mathbb{P}\{W_k = u \,|\, W_{k+1} = v\}.$$

- An infinite bridge is a Markov process $(W_n^\infty)_{n \in \mathbb{N}_0}$ with the same backwards transition probabilities as $(W_n)_{n \in \mathbb{N}_0}$.
- Finding the Doob–Martin compactification is essentially the same as characterizing the infinite bridges with almost surely trivial tail σ-fields (any infinite bridge is a mixture of such infinite bridges).

Lifting an infinite bridge to a process of labeled matchings

- Suppose that $(W_n^\infty)_{n\in\mathbb{N}_0}$ is an infinite bridge. Then there exists a Markov process $(L_n^\infty)_{n\in\mathbb{N}_0}$ with distribution uniquely specified by the requirements that:
  - L_n^∞ is a random labeled matching of $[2n]$ for all $n \in \mathbb{N}$,
  - the process $(\Psi(L_n^\infty))_{n\in\mathbb{N}_0}$ has the same distribution as $(W_n^\infty)_{n\in\mathbb{N}_0}$,
  - the conditional distribution of L_n^∞ given $\Psi(L_n^\infty)$ is uniform on the set of labeled associated admissible matchings of $\Psi(L_n^\infty)$ for all $n \in \mathbb{N}_0$.
- That is, we can lift up an infinite word-valued bridge to produce a nice process taking values in the space of labeled matchings that has deterministic backwards transitions.

Labeled random matchings and total orders

- Turn L_n^∞ into a word of length $2n$ in the alphabet $\bigcup_{k=1}^n \{a_k, b_k\}$ in which each letter appears exactly once as follows: place the letter a_p in position i if the block of L_n^∞ labeled p is of the form $\{i, j\}$ with $i < j$ and place the letter b_q in position ℓ if the block of L_n^∞ labeled q is of the form $\{k, \ell\}$ with $k < \ell$.
- This word defines a total order on $\bigcup_{k=1}^n \{a_k, b_k\}$ in the obvious way: x precedes y in the total order if the letter x comes before the letter y in the word.
- This total order is paired, by which we mean that a_r always precedes b_r.
- These paired total orders are consistent as n varies and hence define a paired total order on $\mathbb{I}_0 := \bigcup_{k \in \mathbb{N}} \{a_k, b_k\}$.

- A random paired total order $\prec$ on $\mathbb{I}_0$ is exchangeable if for every $n \in \mathbb{N}$ the induced random total order $\prec^n$ on $\bigcup_{k=1}^n \{a_k, b_k\}$ has the same distribution as the random total order $\prec_\sigma^n$ for any permutation σ of $[n] := \{1, 2, \ldots, n\}$, where $\prec_\sigma^n$ is defined as follows:
  - $a_{\sigma(i)} \prec_\sigma^n b_{\sigma(j)}$ iff $a_i \prec^n b_j$,
  - $b_{\sigma(i)} \prec_\sigma^n a_{\sigma(j)}$ iff $b_i \prec^n a_j$,
  - $a_{\sigma(i)} \prec_\sigma^n a_{\sigma(j)}$ iff $a_i \prec^n a_j$,
  - $b_{\sigma(i)} \prec_\sigma^n b_{\sigma(j)}$ iff $b_i \prec^n b_j$.
- The random paired total order associated with $(L_n^\infty)_{n \in \mathbb{N}_0}$ is exchangeable.
- Conversely, any exchangeable random paired total order is the paired total order associated with a unique infinite bridge $(W_n^\infty)_{n \in \mathbb{N}_0}$ via the corresponding $(L_n^\infty)_{n \in \mathbb{N}_0}$.

Ergodicity and tail triviality

- An measurable set A of total orders is almost invariant for a exchangeable random paired total order $\prec$ on $\mathbb{I}_0$ if $\{\prec \in A\} = \{\prec_\sigma \in A\}$ almost surely for all permutations σ that leave all but finitely many integers fixed.
- A exchangeable random paired total order $\prec$ on $\mathbb{I}_0$ is ergodic if all almost invariant sets have probability 0 or 1.
- The tail σ-field of an infinite bridge $(W_n)_{n \in \mathbb{N}_0}$ is almost surely trivial if and only if the corresponding exchangeable random paired total order is ergodic.

A generic example

- Consider a probability measure η on $\mathbb{R}^2$ that assigns all of its mass to the set $\{(s,t) \in \mathbb{R}^2 : s < t\}$ and has diffuse marginals.
- Let $((S_n, T_n))_{n \in \mathbb{N}}$ be i.i.d. with common distribution η.
- The total order $\lhd$ on $\mathbb{I}_0$ constructed by declaring that
  - $a_i \lhd a_j$ if $S_i < S_j$,
  - $b_i \lhd b_j$ if $T_i < T_j$,
  - $a_i \lhd b_j$ if $S_i < T_j$,
  - $b_i \lhd a_j$ if $T_i < S_j$,

  is paired, exchangeable and ergodic.

- All ergodic exchangeable paired random total orders arise by the construction in the previous slide.
- To see why, first define an order-preserving injection $f : \mathbb{I}_0 \to [0,1]$ by

$$f(y) = \limsup_{n\to\infty} \frac{1}{2n} \#\{1 \leqslant k \leqslant n : a_k \prec y\} + \limsup_{n\to\infty} \frac{1}{2n} \#\{1 \leqslant \ell \leqslant n : b_\ell \prec y\}.$$

- Define a sequence $((X_n, Y_n))_{n\in\mathbb{N}}$ of $[0,1]^2$-valued random variables by setting $X_n := f(a_n)$ and $Y_n := f(b_n)$. Then $((X_n, Y_n))_{n\in\mathbb{N}}$ is i.i.d. with common distribution a probability measure μ that assigns all of its mass to the set $\{(x,y) \in [0,1]^2 : x < y\}$ and has diffuse marginals.
- The probability measure μ is uniquely determined by the moment formulae

$$\int_{[0,1]^2} x^m y^n \, \mu(dx, dy) = \left(\frac{1}{2}\right)^{m+n} \sum_{c \in \prod_{k=1}^{m+n} \{a_k, b_k\}} \mathbb{P}\{c_1 \prec a_{m+n+1}, \ldots, c_m \prec a_{m+n+m}, c_{m+1} \prec b_{m+n+1}, \ldots, c_{m+n} \prec b_{m+n+n}\}.$$

- An infinite bridge $(W_n^\infty)_{n \in \mathbb{N}_0}$ has an almost surely trivial tail σ-field if and only if it is the limit in distribution as $k \to \infty$ of finite bridges $(W_0^{w_k}, \ldots, W_{n_k}^{w_k})$ for some sequence $(w_k)_{k \in \mathbb{N}}$ with $w_k \in \mathbb{B}_{n_k}$.
- In Doob-Martin language, the full boundary and minimal boundary coincide. Equivalently, all points in the boundary correspond to extremal harmonic functions.

Ballot sequences to Dyck paths reminder

- Suppose that $w \in \mathbb{B}_n$.
- Write i_k (respectively, j_k) for the position of the k^{th} letter a (respectively, letter b) in the word w; that is, $\#\{1 \leqslant r \leqslant i_k : w_r = a\} = k$ and $\#\{1 \leqslant s \leqslant j_k : w_s = b\} = k$.
- For $1 \leqslant t \leqslant 2n$ set

$$\begin{aligned} h(t) &:= \#\{1 \leqslant i \leqslant t : w_i = a\} - \#\{1 \leqslant j \leqslant t : w_j = b\} \\ &= \#\{1 \leqslant p \leqslant n : i_p \leqslant t\} - \#\{1 \leqslant q \leqslant n : i_q \leqslant t\}. \end{aligned}$$

- Thus $0 = h(0), h(1), \ldots, h(2n)$ is a path that starts and ends at 0, makes steps of ± 1, and stays nonnegative (that is, a Dyck path).

An explicit example

- Suppose that η is the probability distribution of (U, V) conditional on $U < V$, where U and V are independent exponential random variables with respective rate parameters α and β.
- Then

$$\mathbb{P}\{W_n^\infty = w\}$$
$$= n! \prod_{p=1}^{n} h(i_p)$$
$$\times \frac{(\alpha+\beta)^n 2^n}{\prod_{k=1}^{2n} \left((2n-k+1-h(k-1))\alpha + (2n-k+1+h(k-1))\beta\right)}.$$

- Note: When $\alpha = \beta$, $(W_n^\infty)_{n \in \mathbb{N}_0}$ has the same distribution as $(W_n)_{n \in \mathbb{N}_0}$.
- Has anyone seen this family of probability distributions on ballot sequences (equivalently, Dyck paths)?