ON SOME FACTORIZATIONS OF RANDOM WORDS

PHILIPPE CHASSAING

INSTITUT ELIE CARTAN

&

ELAHE ZOHOORIAN-AZAD

DAMGHAN UNIVERSITY

Maresias, AofA'08

GLOSSARY

- Alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_k < \cdots\}$
- n-letters long words $\omega = \omega_1\omega_2\ldots\omega_n,\ \omega_i \in \mathcal{A},\ |\omega| = n,\ \omega \in \mathcal{A}^n$
- Language $\mathcal{A}^* = \{\emptyset\} \cup \mathcal{A}^1 \cup \mathcal{A}^2 \cup \mathcal{A}^3 \cup \ldots$
- u is a factor of w $\exists\, r, s \in \mathcal{A}^\star$ such that $w = rus$
- u is a Prefix of w $r = \emptyset$
- u is a Suffix of w $s = \emptyset$
- Rotation
- Necklace, circular word
- Primitive word

GLOSSARY

- Alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_k < \cdots\}$
- n-letters long words $\omega = \omega_1\omega_2\ldots\omega_n,\ \omega_i \in \mathcal{A},\ |\omega| = n,\ \omega \in \mathcal{A}^n$
- Language $\mathcal{A}^* = \{\emptyset\} \cup \mathcal{A}^1 \cup \mathcal{A}^2 \cup \mathcal{A}^3 \cup \ldots$
- u is a factor of w $\exists\, r, s \in \mathcal{A}^\star$ such that $w = rus$
- u is a Prefix of w $r = \emptyset$
- u is a Suffix of w $s = \emptyset$
- Rotation $\omega = \omega_1\omega_2\ldots\omega_n \longrightarrow \tau\omega = \omega_2\omega_3\ldots\omega_n\omega_1$
- Necklace, circular word
- Primitive word

GLOSSARY

- Alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_k < \cdots\}$
- n-letters long words $\omega = \omega_1\omega_2\ldots\omega_n,\ \omega_i \in \mathcal{A},\ |\omega| = n,\ \omega \in \mathcal{A}^n$
- Language $\mathcal{A}^* = \{\emptyset\} \cup \mathcal{A}^1 \cup \mathcal{A}^2 \cup \mathcal{A}^3 \cup \ldots$
- u is a factor of w $\exists\, r, s \in \mathcal{A}^\star$ such that $w = rus$
- u is a Prefix of w $r = \emptyset$
- u is a Suffix of w $s = \emptyset$
- Rotation $\omega = \omega_1\omega_2\ldots\omega_n \rightarrow \tau\omega = \omega_2\omega_3\ldots\omega_n\omega_1$
- Necklace, circular word $\langle\omega\rangle = \left\{\tau^k\omega \mid k \in \mathbb{Z}\right\}$
- Primitive word

GLOSSARY

- Alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_k < \cdots\}$
- n-letters long words $\omega = \omega_1\omega_2\ldots\omega_n,\ \omega_i \in \mathcal{A},\ |\omega| = n,\ \omega \in \mathcal{A}^n$
- Language $\mathcal{A}^* = \{\emptyset\} \cup \mathcal{A}^1 \cup \mathcal{A}^2 \cup \mathcal{A}^3 \cup \ldots$
- u is a factor of w $\exists\, r, s \in \mathcal{A}^\star$ such that $w = rus$
- u is a Prefix of w $r = \emptyset$
- u is a Suffix of w $s = \emptyset$
- Rotation $\omega = \omega_1\omega_2\ldots\omega_n \to \tau\omega = \omega_2\omega_3\ldots\omega_n\omega_1$
- Necklace, circular word $\langle\omega\rangle = \{\tau^k\omega \mid k \in \mathbb{Z}\}$
- Primitive word $\#\langle\omega\rangle = |\omega|$

GLOSSARY

- Alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_k < \cdots\}$
- n-letters long words $\omega = \omega_1\omega_2\ldots\omega_n,\ \omega_i \in \mathcal{A},\ |\omega| = n,\ \omega \in \mathcal{A}^n$
- Language $\mathcal{A}^* = \{\emptyset\} \cup \mathcal{A}^1 \cup \mathcal{A}^2 \cup \mathcal{A}^3 \cup \ldots$
- u is a factor of w $\exists\, r, s \in \mathcal{A}^\star$ such that $w = rus$
- u is a Prefix of w $r = \emptyset$
- u is a Suffix of w $s = \emptyset$
- Rotation $\omega = \omega_1\omega_2\ldots\omega_n \rightarrow \tau\omega = \omega_2\omega_3\ldots\omega_n\omega_1$
- Necklace, circular word $\langle\omega\rangle = \{\tau^k\omega \mid k \in \mathbb{Z}\}$
- Primitive word $\#\langle\omega\rangle = |\omega|$

LYNDON WORDS

- Lexicographic Order

LYNDON WORDS

- Lexicographic Order

$\mathbf{a} \prec \mathbf{b}$

$$\text{if} \quad \begin{cases} \text{either } \exists\, p, \alpha, \beta \in \mathcal{A}^\star, a_i, a_j \in \mathcal{A} \text{ s.t. } \; i < j, \; \begin{cases} \mathbf{a} = p a_i \alpha, \\ \mathbf{b} = p a_j \beta, \end{cases} \\[2em] \text{or } \mathbf{a} \text{ is a prefix of } \mathbf{b} \end{cases}$$

LYNDON WORDS

- Lexicographic Order

$\mathbf{a} \prec \mathbf{b}$

$$\text{if} \quad \begin{cases} \text{either } \exists\, p, \alpha, \beta \in \mathcal{A}^\star, a_i, a_j \in \mathcal{A} \text{ s.t. } \; i < j, \; \begin{cases} \mathbf{a} = p a_i \alpha, \\ \mathbf{b} = p a_j \beta, \end{cases} \\ \\ \text{or } \mathbf{a} \text{ is a prefix of } \mathbf{b} \end{cases}$$

- w is a Lyndon word if w is primitive, and is the smallest word in its necklace

LYNDON WORDS

- Lexicographic Order

$\mathbf{a} \prec \mathbf{b}$

$$\text{if} \quad \begin{cases} \text{either } \exists\, p, \alpha, \beta \in \mathcal{A}^\star, a_i, a_j \in \mathcal{A} \text{ s.t. } \; i < j, \; \begin{cases} \mathbf{a} = p a_i \alpha, \\ \mathbf{b} = p a_j \beta, \end{cases} \\ \\ \text{or } \mathbf{a} \text{ is a prefix of } \mathbf{b} \end{cases}$$

- w is a Lyndon word if w is primitive, and is the smallest word in its necklace

- cbaa, baac, aacb, acba: aacb is a Lyndon word,

LYNDON WORDS

- Lexicographic Order

$\mathbf{a} \prec \mathbf{b}$

$$\text{if} \quad \begin{cases} \text{either } \exists\, p, \alpha, \beta \in \mathcal{A}^\star, a_i, a_j \in \mathcal{A} \text{ s.t. } \; i < j, \; \begin{cases} \mathbf{a} = p a_i \alpha, \\ \mathbf{b} = p a_j \beta, \end{cases} \\ \\ \text{or } \mathbf{a} \text{ is a prefix of } \mathbf{b} \end{cases}$$

- w is a Lyndon word if w is primitive, and is the smallest word in its necklace

  - cbaa, baac, aacb, acba: aacb is a Lyndon word,
  - aabaab, baac are not

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization $w=uv$ is often called the standard factorization of w.

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization $w=uv$ is often called the standard factorization of w.
  - $w=abaabbabaabb$ $u=abaabbab$ $v=aabb$

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization $w=uv$ is often called the standard factorization of w.
  - $w=abaabbabaabb$ $u=abaabbab$ $v=aabb$
  - $w=abaabbabaabb$ $u'=ab$ $v'=aabbabaabb$ $v<v'$

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization w=uv is often called the standard factorization of w.
  - w=abaabbabaabb u=abaabbab v=aabb
  - w=abaabbabaabb u'=ab v'=aabbabaabb v<v'
- Theorem (Lyndon, 1954) Any word w may be written uniquely as a non-increasing product of Lyndon words (by iteration of the standard factorization).

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization $w=uv$ is often called the standard factorization of w.
  - $w=abaabbabaabb$ $u=abaabbab$ $v=aabb$
  - $w=abaabbabaabb$ $u'=ab$ $v'=aabbabaabb$ $v<v'$
- **Theorem** (Lyndon, 1954) Any word w may be written uniquely as a non-increasing product of Lyndon words (by iteration of the standard factorization).

$$aabbaaababbaaabaa$$

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization w=uv is often called the standard factorization of w.
  - w=abaabbabaabb u=abaabbab v=aabb
  - w=abaabbabaabb u'=ab v'=aabbabaabb v<v'
- Theorem (Lyndon, 1954) Any word w may be written uniquely as a non-increasing product of Lyndon words (by iteration of the standard factorization).

$$aabbaaababbaaabaa$$

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization w=uv is often called the standard factorization of w.
  - w=abaabbabaabb u=abaabbab v=aabb
  - w=abaabbabaabb u'=ab v'=aabbabaabb v<v'
- Theorem (Lyndon, 1954) Any word w may be written uniquely as a non-increasing product of Lyndon words (by iteration of the standard factorization).

$$aabbaaababbaaabaa$$

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization w=uv is often called the standard factorization of w.
  - w=abaabbabaabb u=abaabbab v=aabb
  - w=abaabbabaabb u'=ab v'=aabbabaabb v<v'
- Theorem (Lyndon, 1954) Any word w may be written uniquely as a non-increasing product of Lyndon words (by iteration of the standard factorization).

$$aabbaaababbaaabaa$$

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization w=uv is often called the standard factorization of w.
  - w=abaabbabaabb u=abaabbab v=aabb
  - w=abaabbabaabb u'=ab v'=aabbabaabb v<v'
- Theorem (Lyndon, 1954) Any word w may be written uniquely as a non-increasing product of Lyndon words (by iteration of the standard factorization).

$$aabbaaababbaaabaa$$

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization $w=uv$ is often called the standard factorization of w.
  - $w=abaabbabaabb$ $u=abaabbab$ $v=aabb$
  - $w=abaabbabaabb$ $u'=ab$ $v'=aabbabaabb$ $v<v'$
- **Theorem** (Lyndon, 1954) Any word w may be written uniquely as a non-increasing product of Lyndon words (by iteration of the standard factorization).

$$aabbaaababbaaabaa$$

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization w=uv is often called the standard factorization of w.
  - w=abaabbabaabb u=abaabbab v=aabb
  - w=abaabbabaabb u'=ab v'=aabbabaabb v<v'
- Theorem (Lyndon, 1954) Any word w may be written uniquely as a non-increasing product of Lyndon words (by iteration of the standard factorization).

$$aabbaaababbaaabaa$$

- The standard factorization of a Lyndon word is the first step in the construction of some basis of the free Lie algebra over A

PROBABILISTIC MODEL

PROBABILISTIC MODEL

$$\forall a_i \in \mathcal{A}, \quad p(a_i) = p_i,$$

PROBABILISTIC MODEL

$$\forall a_i \in \mathcal{A}, \quad p(a_i) = p_i,$$

$$\forall w = w_1 w_2 \cdots w_n \in \mathcal{A}^n, \quad p(w) = \prod_{j=1}^{n} p(w_j)$$

PROBABILISTIC MODEL

$$\forall a_i \in \mathcal{A}, \quad p(a_i) = p_i,$$

$$\forall w = w_1 w_2 \cdots w_n \in \mathcal{A}^n, \quad p(w) = \prod_{j=1}^{n} p(w_j)$$

$$\mathbb{P}_n(A) = \sum_{w \in A \cap \mathcal{A}^n} p(w)$$

PROBABILISTIC MODEL

$$\forall a_i \in \mathcal{A}, \quad p(a_i) = p_i,$$

$$\forall w = w_1 w_2 \cdots w_n \in \mathcal{A}^n, \quad p(w) = \prod_{j=1}^{n} p(w_j)$$

$$\mathbb{P}_n(A) = \sum_{w \in A \cap \mathcal{A}^n} p(w)$$

- WLOG, $\{i \mid p_i > 0\}$ has no gaps and contains 1.

PROFILE OF THE DECOMPOSITION

PROFILE OF THE DECOMPOSITION

- For a word , set

$$N(w) = (N_k(w))_{k\geq 1},$$

in which $N_k(w)$ is the number of k-letters long factors in the Lyndon decomposition of w.

PROFILE OF THE DECOMPOSITION

- For a word , set

$$N(w) = (N_k(w))_{k \geq 1},$$

in which $N_k(w)$ is the number of k-letters long factors in the Lyndon decomposition of w.

$$aabbaaababbaaabaa$$

PROFILE OF THE DECOMPOSITION

- For a word , set

$$N(w) = (N_k(w))_{k\geq 1},$$

in which $N_k(w)$ is the number of k-letters long factors in the Lyndon decomposition of w.

$$aabbaaaababbaaabaa$$

- $N = (2,0,0,2,0,0,1,0,0, \ldots)$.

UNIFORM CASE

- In the uniform case ($p_i=1/q$, $1\leq i\leq q$), **Diaconis, McGrath and Pitman** (Riffle shuffles, cycles, and descents, 1995) give the exact distribution of the profile

$$N(w)=(N_k(w))_{k\geq 1}.$$

UNIFORM CASE

- In the uniform case ($p_i=1/q$, $1\leq i\leq q$), **Diaconis, McGrath and Pitman** (Riffle shuffles, cycles, and descents, 1995) give the exact distribution of the profile

$$N(w)=(N_k(w))_{k\geq 1}.$$

$$\mathbb{P}(N=\xi)=\frac{1}{q^n}\prod_{k=1}^{n}\binom{f_k(q)+\xi_k-1}{\xi_k}$$

UNIFORM CASE

- In the uniform case ($p_i=1/q$, $1\le i\le q$), **Diaconis, McGrath and Pitman** (Riffle shuffles, cycles, and descents, 1995) give the exact distribution of the profile

$$N(w)=(N_k(w))_{k\geq 1}.$$

$$\mathbb{P}(N=\xi)=\frac{1}{q^n}\prod_{k=1}^{n}\binom{f_k(q)+\xi_k-1}{\xi_k}$$

$$f_k(q)=\frac{1}{k}\sum_{d|k}\mu(d)q^{k/d}$$

UNIFORM CASE

- In the uniform case ($p_i=1/q$, $1\le i\le q$), **Diaconis, McGrath and Pitman** (Riffle shuffles, cycles, and descents, 1995) give the exact distribution of the profile

$$N(w) = (N_k(w))_{k\ge 1}.$$

$$\mathbb{P}(N=\xi) = \frac{1}{q^n}\prod_{k=1}^{n}\binom{f_k(q)+\xi_k-1}{\xi_k}$$

$$f_k(q) = \frac{1}{k}\sum_{d|k}\mu(d)q^{k/d}$$

- in which μ is the Moebius function.

ASYMPTOTICS

$$p_{q,n}(\xi) = \frac{1}{q^n} \prod_{k=1}^{n} \binom{f_k(q) + \xi_k - 1}{\xi_k}, \qquad f_k(q) = \frac{1}{k} \sum_{d|k} \mu(d) q^{k/d}$$

ASYMPTOTICS

$$p_{q,n}(\xi) = \frac{1}{q^n} \prod_{k=1}^{n} \binom{f_k(q) + \xi_k - 1}{\xi_k}, \qquad f_k(q) = \frac{1}{k} \sum_{d|k} \mu(d) q^{k/d}$$

- $p_{q,n}(\xi)$ converges, as q grows, to

ASYMPTOTICS

$$p_{q,n}(\xi) = \frac{1}{q^n} \prod_{k=1}^{n} \binom{f_k(q) + \xi_k - 1}{\xi_k}, \qquad f_k(q) = \frac{1}{k} \sum_{d|k} \mu(d) q^{k/d}$$

- ☐ $p_{q,n}(\xi)$ converges, as q grows, to

$$p_n(\xi) = \prod_{k\geq 1} \frac{1}{k^{\xi_k} \xi_k!} = \mathbb{P}(C = \xi), \qquad \sum_{k\geq 1} k\xi_k = n$$

ASYMPTOTICS

$$p_{q,n}(\xi) = \frac{1}{q^n} \prod_{k=1}^{n} \binom{f_k(q) + \xi_k - 1}{\xi_k}, \qquad f_k(q) = \frac{1}{k} \sum_{d|k} \mu(d) q^{k/d}$$

- $p_{q,n}(\xi)$ converges, as q grows, to

$$p_n(\xi) = \prod_{k \geq 1} \frac{1}{k^{\xi_k} \xi_k!} = \mathbb{P}(C = \xi), \qquad \sum_{k \geq 1} k\xi_k = n$$

- in which $c_k(w)$ is the number of k-cycles in the cycle-decomposition of the n-permutation w, and $c(w) = (c_k(w))_{k \geq 1}$.

ASYMPTOTICS

$$p_{q,n}(\xi) = \frac{1}{q^n} \prod_{k=1}^{n} \binom{f_k(q) + \xi_k - 1}{\xi_k}, \qquad f_k(q) = \frac{1}{k} \sum_{d|k} \mu(d) q^{k/d}$$

- $p_{q,n}(\xi)$ converges, as q grows, to

$$p_n(\xi) = \prod_{k \geq 1} \frac{1}{k^{\xi_k} \xi_k!} = \mathbb{P}(C = \xi), \qquad \sum_{k \geq 1} k\xi_k = n$$

- in which $c_k(w)$ is the number of k-cycles in the cycle-decomposition of the n-permutation w, and $c(w) = (c_k(w))_{k \geq 1}$.

- As n grows, $p_n(.)$ converges to the law of a sequence of independent Poisson random variables (with respective parameters $1/k$ for c_k).

RIFFLE SHUFFLE

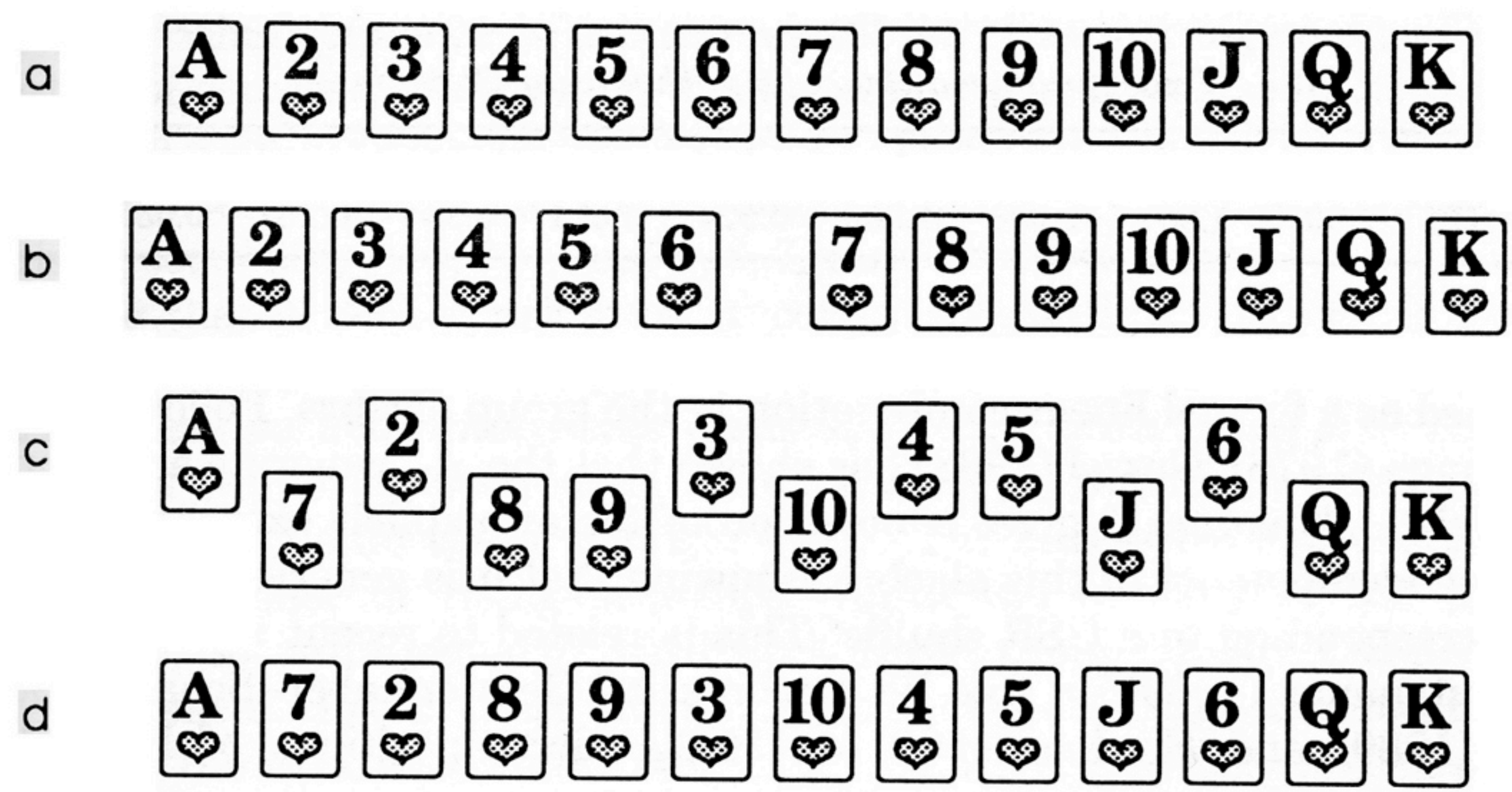
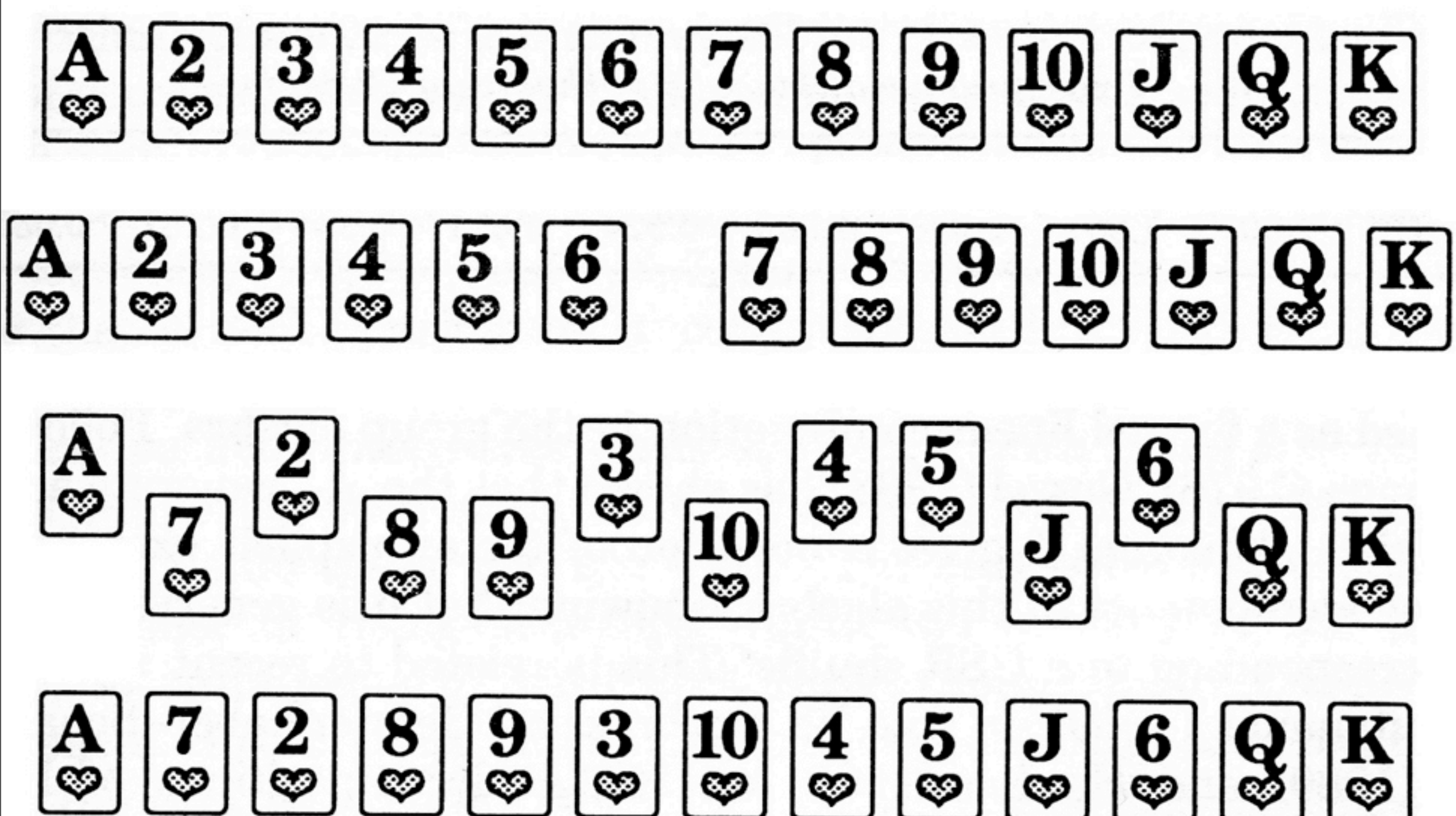
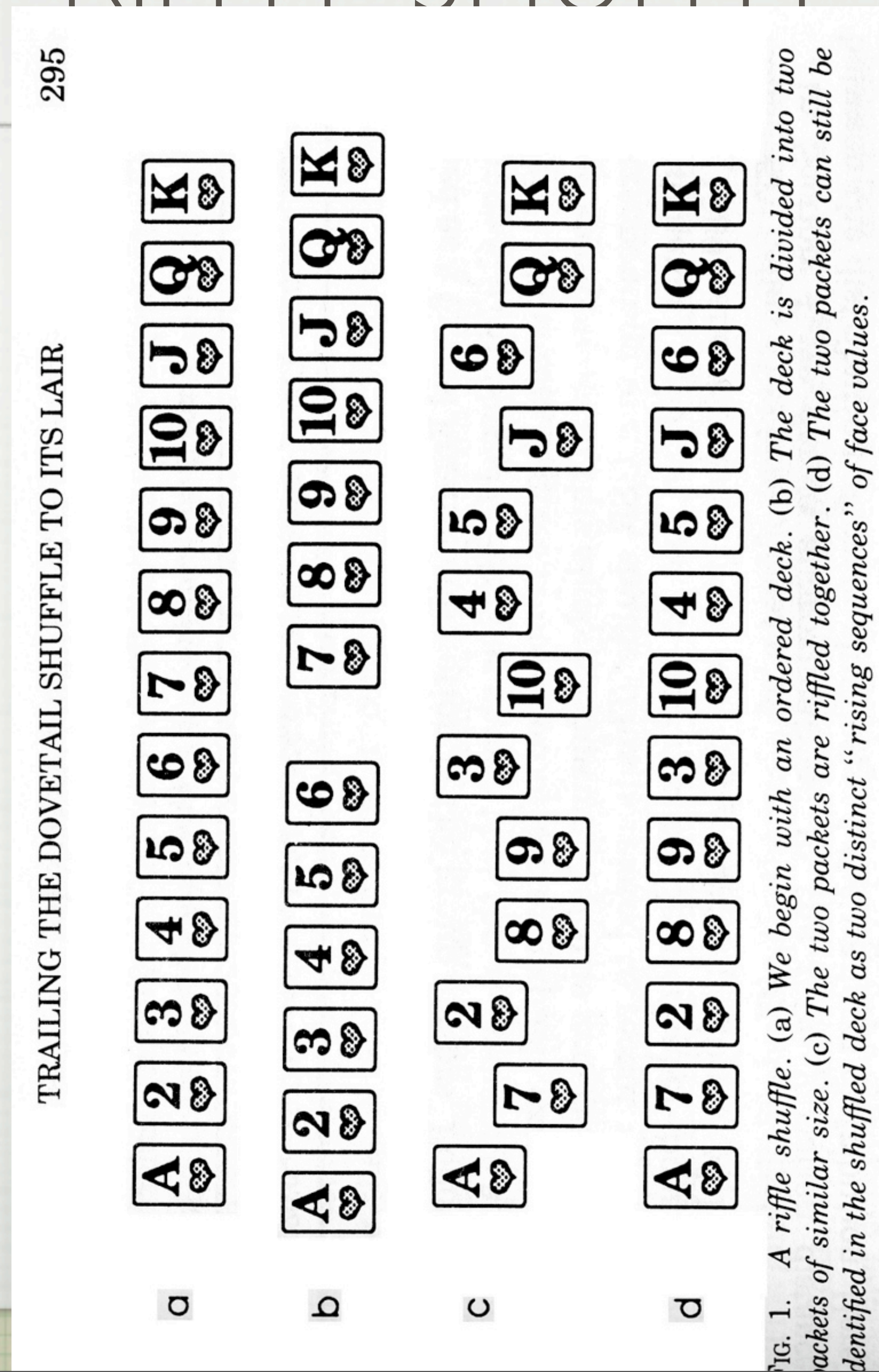


FIG. 1. *A riffle shuffle*. (a) *We begin with an ordered deck*. (b) *The deck is divided into two packets of similar size*. (c) *The two packets are riffled together*. (d) *The two packets can still be identified in the shuffled deck as two distinct "rising sequences" of face values*.



riffle shuffle. (a) *We begin with an ordered deck.* (b) *The deck is divided into two* imilar size. (c) *The two packets are riffled together.* (d) *The two packets can still be* the shuffled deck as two distinct "rising sequences" of face values.

RIFFLE SHUFFLE



RIFFLE SHUFFLE #2

RIFFLE SHUFFLE #2

- $RS_a * RS_b = RS_{ab}$

RIFFLE SHUFFLE #2

- $$RS_a * RS_b = RS_{ab}$$
  Doing a b-riffle-shuffle, followed by an independent a-riffle-shuffle, results in an ab-riffle-shuffle (not so obvious ...).

RIFFLE SHUFFLE #2

- $$RS_a * RS_b = RS_{ab}$$
  Doing a b-riffle-shuffle, followed by an independent a-riffle-shuffle, results in an ab-riffle-shuffle (not so obvious ...).
- Proof:

RIFFLE SHUFFLE #2

- $$RS_a * RS_b = RS_{ab}$$
  Doing a b-riffle-shuffle, followed by an independent a-riffle-shuffle, results in an ab-riffle-shuffle (not so obvious ...).
- Proof:
  - Let {x} be the fractional part of the real number x.

RIFFLE SHUFFLE #2

- $$RS_a * RS_b = RS_{ab}$$
  Doing a b-riffle-shuffle, followed by an independent a-riffle-shuffle, results in an ab-riffle-shuffle (not so obvious ...).
- Proof:
  - Let $\{x\}$ be the fractional part of the real number x.
  - Let $U = (U_k)_{1 \le k \le n}$ be n random numbers, uniform on $[0,1]$.

RIFFLE SHUFFLE #2

- $$RS_a * RS_b = RS_{ab}$$
  Doing a b-riffle-shuffle, followed by an independent a-riffle-shuffle, results in an ab-riffle-shuffle (not so obvious ...).
- Proof:
  - Let $\{x\}$ be the fractional part of the real number x.
  - Let $U = (U_k)_{1 \leq k \leq n}$ be n random numbers, uniform on $[0,1]$.
  - Map the rank of $\{aU_i\}$ in $\{aU\}$ to the rank of U_i in U: this is a realisation of an a-riffle-shuffle.

RIFFLE SHUFFLE #2

- $$RS_a * RS_b = RS_{ab}$$
  Doing a b-riffle-shuffle, followed by an independent a-riffle-shuffle, results in an ab-riffle-shuffle (not so obvious ...).
- Proof:
  - Let $\{x\}$ be the fractional part of the real number x.
  - Let $U = (U_k)_{1 \leq k \leq n}$ be n random numbers, uniform on $[0,1]$.
  - Map the rank of $\{aU_i\}$ in $\{aU\}$ to the rank of U_i in U: this is a realisation of an a-riffle-shuffle.
  - $$\{a\{bx\}\} = \{abx\}.$$

RIFFLE SHUFFLE #2

- $$RS_a * RS_b = RS_{ab}$$
  Doing a b-riffle-shuffle, followed by an independent a-riffle-shuffle, results in an ab-riffle-shuffle (not so obvious ...).
- Proof:
  - Let $\{x\}$ be the fractional part of the real number x.
  - Let $U = (U_k)_{1 \leq k \leq n}$ be n random numbers, uniform on $[0,1]$.
  - Map the rank of $\{aU_i\}$ in $\{aU\}$ to the rank of U_i in U: this is a realisation of an a-riffle-shuffle.
  - $$\{a\{bx\}\} = \{abx\}.$$
  - $\{aU_i\}$ is random uniform on $[0,1]$ and independent of $[aU_i]$.

RIFFLE SHUFFLE: ASYMPTOTICS

RIFFLE SHUFFLE: ASYMPTOTICS

- Bonus:

$$RS_q \longrightarrow \text{uniform permutation,}$$

leading to the convergence of $M=(M_k)_{k\geq 1}$ to a Cauchy distribution, for

$$(q,n) \longrightarrow +\infty,$$

in which $M_k(w)$ is the number of cycles with length k in the permutation w.

RIFFLE SHUFFLE: ASYMPTOTICS

- Bonus:

$$RS_q \longrightarrow \text{uniform permutation},$$

leading to the convergence of $M=(M_k)_{k\geq 1}$ to a Cauchy distribution, for

$$(q,n) \longrightarrow +\infty,$$

in which $M_k(w)$ is the number of cycles with length k in the permutation w.

- Birthday paradox:

$$DV(RS_q,\text{uniform}) = O(n^2/2q).$$

RIFFLE SHUFFLE: ASYMPTOTICS

- Bonus:

$$RS_q \longrightarrow \text{uniform permutation},$$

leading to the convergence of $M=(M_k)_{k\geq 1}$ to a Cauchy distribution, for

$$(q,n) \longrightarrow +\infty,$$

in which $M_k(w)$ is the number of cycles with length k in the permutation w.

- Birthday paradox:

$$DV(RS_q,\text{uniform}) = O(n^2/2q).$$

- Bayer & Diaconis (1992):

$$DV(RS_q,\text{uniform}) = O(n^{3/2}/q).$$

GESSEL'S BIJECTION

GESSEL'S BIJECTION

- Correspondance

{random uniform words from a q-letters alphabet}

$\longleftrightarrow$

{RS_q-distributed permutations}

GESSEL'S BIJECTION

- Correspondance

{random uniform words from a q-letters alphabet}

$\longleftrightarrow$

{RS_q-distributed permutations}

- In which cycles are sent on Lyndon factors with the same length,

GESSEL'S BIJECTION

- Correspondance

{random uniform words from a q-letters alphabet}

$\longleftrightarrow$

{RS_q-distributed permutations}

- In which cycles are sent on Lyndon factors with the same length,

- And the profile of the permutation is sent on N.

NEXT ...

NEXT ...

- Diaconis et al. gives the asymptotic distribution of the lengths of the shortest factors, while the position of these factors is lost.

NEXT ...

- Diaconis et al. gives the asymptotic distribution of the lengths of the shortest factors, while the position of these factors is lost.
- What about the lengths of the longest factors ? the lengths of the last factors ?

NEXT ...

- Diaconis et al. gives the asymptotic distribution of the lengths of the shortest factors, while the position of these factors is lost.

- What about the lengths of the longest factors ? the lengths of the last factors ?

- More general distribution $p = (p_i)_{i \geq 1}$ on letters ?

MAIN RESULT

$$aabbaaababbaaabaa$$

$X_{(5)}$	$X_{(4)}$	$X_{(3)}$	$X_{(2)}$	$X_{(1)}$

MAIN RESULT

$$aabbaaababbaaabaa$$

- $x_{20} = (1,1,4,9,5,0,0,\ldots)/20$
- $x_n(k)$ is the renormalised size of the k^{th} Lyndon factor, starting from the end of the word.

- For a general alphabet $A=\{a_i\}$, and a general distribution $p=(p_i)$, x_n converges to a p_1-sticky GEM(1).

GEM(1)

[redacted]

GEM(1)

[redacted]

GEM(1)

GEM(1)

U_1

GEM(1)

GEM(1)

GEM(1)

- Terminology: Griffiths-Engen-McClosey r.v. with parameter 1, size-biased reordering of Poisson-Dirichlet(0,1) (population genetics, etc …), stickbreaking scheme …

GEM(1)

- Terminology: Griffiths-Engen-McClosey r.v. with parameter 1, size-biased reordering of Poisson-Dirichlet(0,1) (population genetics, etc ...), stickbreaking scheme ...
- The sequence of residual sizes after the k^{th} break, W_k, satisfies

  W_k/W_{k-1} are independant and uniform on [0,1].

GEM(1)

.....	$U_2(1-U_1)$	U_1

- Terminology: Griffiths-Engen-McClosey r.v. with parameter 1, size-biased reordering of Poisson-Dirichlet(0,1) (population genetics, etc ...), stickbreaking scheme ...
- The sequence of residual sizes after the k^{th} break, W_k, satisfies
  W_k/W_{k-1} are independant and uniform on [0,1].
- $W_0=1$

GEM(1)

- Terminology: Griffiths-Engen-McClosey r.v. with parameter 1, size-biased reordering of Poisson-Dirichlet(0,1) (population genetics, etc ...), stickbreaking scheme ...
- The sequence of residual sizes after the k^{th} break, W_k, satisfies W_k/W_{k-1} are independant and uniform on $[0,1]$.
- $W_0=1$
- The size X_k of the k^{th} piece of the stick is given by $X_k = W_k - W_{k-1} = U_1 U_2 \ldots U_{k-1}(1-U_k)$.

GEM(1)

- Terminology: Griffiths-Engen-McClosey r.v. with parameter 1, size-biased reordering of Poisson-Dirichlet(0,1) (population genetics, etc ...), stickbreaking scheme ...

- The sequence of residual sizes after the k^{th} break, W_k, satisfies W_k/W_{k-1} are independant and uniform on $[0,1]$.

- $W_0=1$

- The size X_k of the k^{th} piece of the stick is given by
$$X_k = W_k - W_{k-1} = U_1 \, U_2 \dots U_{k-1}(1-U_k).$$

- $W=(W_k)_{k\geq 0}$ is a Markov chain with transition kernel
$$p(x,dy)=1_{[0,x]}(y)\,dy/x.$$

STICKY GEM(1)

- The a-sticky GEM(1): the residual size W_k is a Markov chain starting from 1, with transition kernel

STICKY GEM(1)

- The a-sticky GEM(1): the residual size W_k is a Markov chain starting from 1, with transition kernel
  - $$p(x,dy)=1_{[0,x]}(y)dy/x, \qquad x\neq 1,$$

STICKY GEM(1)

- The a-sticky GEM(1): the residual size W_k is a Markov chain starting from 1, with transition kernel
  - $$p(x,dy)=1_{[0,x]}(y)\,dy/x, \qquad x\neq 1,$$
  - $$p(1,dy)=a\delta_1+(1-a)1_{[0,1]}(y)\,dy.$$

STICKY GEM(1)

- The a-sticky GEM(1): the residual size W_k is a Markov chain starting from 1, with transition kernel
  - $p(x,dy)=1_{[0,x]}(y)dy/x, \qquad x\neq 1,$
  - $p(1,dy)=a\delta_1+(1-a)1_{[0,1]}(y)dy.$
- W starts with a sequence of S 1's, $P(S=k)=a^{k-1}(1-a)$, $k\geq 1$, rather than with only $W_0=1$.

STICKY GEM(1)

- The a-sticky GEM(1): the residual size W_k is a Markov chain starting from 1, with transition kernel
  - $$p(x,dy)=1_{[0,x]}(y)dy/x, \qquad x\neq 1,$$
  - $$p(1,dy)=a\delta_1+(1-a)1_{[0,1]}(y)dy.$$
- W starts with a sequence of S 1's, $P(S=k)=a^{k-1}(1-a)$, $k\geq 1$, rather than with only $W_0=1$.
- X starts with a sequence of T 0's, $P(T=k)=a^k(1-a)$, $k\geq 0$, rather than with $X_0>0$.

STICKBREAKING OCCURENCES

- $$x_k = U_1\, U_2 \ldots U_{k-1}(1-U_k).$$
- Rearranging $x=(x_k)_{k\geq 0}$ in decreasing order gives the asymptotic distributions of the normalised sizes of cycles, or of logarithms of prime factors of integers, or of degrees of prime factors of polynomials on finite fields.

STICKBREAKING OCCURENCES

- $$X_k = U_1\, U_2 \ldots U_{k-1}(1-U_k).$$
- Rearranging $X = (X_k)_{k\geq 0}$ in decreasing order gives the asymptotic distributions of the normalised sizes of cycles, or of logarithms of prime factors of integers, or of degrees of prime factors of polynomials on finite fields.
- The distribution of max X_k is related to the Dickman function:

K. Dickman, On the frequency of numbers containing prime factors of a certain relative magnitude. Ark. Mat. Astronomi och Fysik 22, 1930, 1-14.

STICKBREAKING OCCURENCES

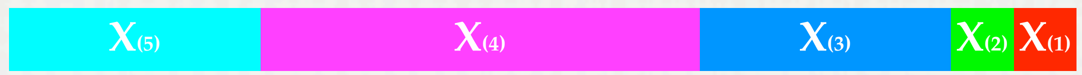
.....	$U_2(1-U_1)$	U_1

- $$X_k = U_1 U_2 \dots U_{k-1}(1-U_k).$$
- Rearranging $X = (X_k)_{k\geq 0}$ in decreasing order gives the asymptotic distributions of the normalised sizes of cycles, or of logarithms of prime factors of integers, or of degrees of prime factors of polynomials on finite fields.
- The distribution of $\max X_k$ is related to the Dickman function:

  K. Dickman, On the frequency of numbers containing prime factors of a certain relative magnitude. Ark. Mat. Astronomi och Fysik 22, 1930, 1-14.
- The normalised size of the longest factor in the Lyndon decomposition converges to the Dickman distribution, regardless of $p = (p_i)$.

RELATED RESULTS

$$aabbaaababbaaabaa$$



RELATED RESULTS

$aabbaaababbaaabaa$

- D. Bayer & P. Diaconis, Trailing the Dovetail Shuffle to Its Lair, Ann. Appl. Probability 2, 294-313, 1992.
- P. Diaconis, M.J. McGrath & J. Pitman, Riffle shuffles, cycles, and descents, Combinatorica, 15, no. 1, 11-29, 1995.
- F. Bassino, J. Clément & C. Nicaud, The standard factorization of Lyndon words: an average point of view, Discrete Mathematics, 290, 1-25, 2005.
- R. Marchand & E. Zohoorian-Azad, Limit law of the length of the standard right factor of a Lyndon word, Combinatorics, Probability and Computing, 16, 417-434, 2007.

PROOF OF THE MAIN RESULT

EXERCISES 1 & 2 ???