Longest Common Patterns in Permutations

Mike Earnest, Anant Godbole, Yevgeniy Rudoy

July 4, 2013

Longest Common Subsequences (of Words)

w_1: ATTCGACGTA
w_2: CGTTATTCGA

Longest Common Subsequences (of Words)

w_1: ATTCGACGTA
w_2: CGTTATTCGA

Longest Common Subsequences (of Words)

w_1: ATTCGACGTA
w_2: CGTTATTCGA

- The sequence $w =$ ATTCGA is the Longest Common Subsequence (LCS) of w_1 and w_2.

Longest Common Subsequences (of Words)

w_1: ATTCGACGTA
w_2: CGTTATTCGA

- The sequence $w =$ ATTCGA is the Longest Common Subsequence (LCS) of w_1 and w_2.
- Can be found in $O(n^2)$ time.
- When words are randomly chosen, length of their LCS is random variable.
- With k letter alphabet, $E(\text{length of LCS}) \to \frac{2n}{\sqrt{k}}$ as $n \to \infty$. [Kiwi, Loebl, Matousek].

Longest Common Patterns

We write

$$\begin{array}{lcccccccc} \pi_1: & 1 & 4 & 2 & 5 & 6 & 3 & 8 & 7 \\ \pi_2: & 2 & 7 & 1 & 4 & 3 & 6 & 5 & 8 \end{array}$$

Longest Common Patterns

We write

$$\pi_1: \ 1 \ 4 \ 2 \ 5 \ 6 \ 3 \ 8 \ 7$$
$$\pi_2: \ 2 \ 7 \ 1 \ 4 \ 3 \ 6 \ 5 \ 8$$

- The pattern $\sigma = 1\ 2\ 4\ 3$ is a *common pattern* of π_1 and π_2.

Longest Common Patterns

We write
π_1: 1 4 2 5 6 3 8 7
π_2: 2 7 1 4 3 6 5 8

- The pattern $\sigma = 1\ 2\ 4\ 3$ is a *common pattern* of π_1 and π_2.
- Their *longest common pattern*, or LCP, is $\sigma = 1\ 3\ 2\ 4\ 5$.

Longest Common Patterns

We write

$$\pi_1: \; 1 \; 4 \; 2 \; 5 \; 6 \; 3 \; 8 \; 7$$
$$\pi_2: \; 2 \; 7 \; 1 \; 4 \; 3 \; 6 \; 5 \; 8$$

- The pattern $\sigma = 1\ 2\ 4\ 3$ is a *common pattern* of π_1 and π_2.
- Their *longest common pattern*, or LCP, is $\sigma = 1\ 3\ 2\ 4\ 5$.
- Finding legnth in general is NP-Hard, but $O(n^8)$ when one permutations is separable [Bouvel, Rossin]

LCP of Random Permutations

Definition

Let L_n be the length of the LCP of two permutations chosen randomly from S_n.

LCP of Random Permutations

Definition

Let L_n be the length of the LCP of two permutations chosen randomly from S_n.

- What is the expected value of L_n?
- How is L_n concentrated around its mean?

Upper Bound

Proposition $E(L_n) < en^{\frac{2}{3}}$.

Upper Bound

Proposition $E(L_n) < en^{\frac{2}{3}}$.

$$\pi_1: \quad 1 \quad 4 \quad 2 \quad 5 \quad 6 \quad 3 \quad 8 \quad 7$$
$$\pi_2: \quad 2 \quad 7 \quad 1 \quad 4 \quad 3 \quad 6 \quad 5 \quad 8$$

- Let S and T be length k subsequences of π_1 and π_2.

Upper Bound

Proposition $E(L_n) < en^{\frac{2}{3}}$.

$$\pi_1: \; 1 \;\; 4 \;\; \color{red}{2} \;\; 5 \;\; \color{red}{6} \;\; 3 \;\; \color{red}{8} \;\; \color{red}{7}$$
$$\pi_2: \; 2 \;\; 7 \;\; \color{blue}{1} \;\; \color{blue}{4} \;\; 3 \;\; \color{blue}{6} \;\; \color{blue}{5} \;\; 8$$

- Let S and T be length k subsequences of π_1 and π_2.
- Example: $k = 4$, $S = 2\,6\,8\,7$, and $T = 1\,4\,6\,5$.

Upper Bound

Proposition $E(L_n) < en^{\frac{2}{3}}$.

$$\pi_1: \; 1 \;\; 4 \;\; 2 \;\; 5 \;\; 6 \;\; 3 \;\; 8 \;\; 7$$
$$\pi_2: \; 2 \;\; 7 \;\; 1 \;\; 4 \;\; 3 \;\; 6 \;\; 5 \;\; 8$$

- Let S and T be length k subsequences of π_1 and π_2.
- Example: $k = 4$, $S = 2\,6\,8\,7$, and $T = 1\,4\,6\,5$.
- Define $X_{S,T}$ be the event that the subsequences S and T are order isomorphic.

Upper Bound

Proposition $E(L_n) < en^{\frac{2}{3}}$.

$$\pi_1: \quad 1 \quad 4 \quad 2 \quad 5 \quad 6 \quad 3 \quad 8 \quad 7$$
$$\pi_2: \quad 2 \quad 7 \quad 1 \quad 4 \quad 3 \quad 6 \quad 5 \quad 8$$

- Let S and T be length k subsequences of π_1 and π_2.
- Example: $k = 4$, $S = 2\ 6\ 8\ 7$, and $T = 1\ 4\ 6\ 5$.
- Define $X_{S,T}$ be the event that the subsequences S and T are order isomorphic.

$$P(L_n \geq k)$$

Upper Bound

Proposition $E(L_n) < en^{\frac{2}{3}}$.

$$\pi_1: \quad 1 \quad 4 \quad {\color{red}2} \quad 5 \quad {\color{red}6} \quad 3 \quad {\color{red}8} \quad {\color{red}7}$$
$$\pi_2: \quad 2 \quad 7 \quad {\color{blue}1} \quad {\color{blue}4} \quad 3 \quad {\color{blue}6} \quad {\color{blue}5} \quad 8$$

- Let S and T be length k subsequences of π_1 and π_2.
- Example: $k = 4$, $S = 2\,6\,8\,7$, and $T = 1\,4\,6\,5$.
- Define $X_{S,T}$ be the event that the subsequences S and T are order isomorphic.

$$P(L_n \geq k) = P\left(\bigcup X_{S,T}\right)$$

Upper Bound

Proposition $E(L_n) < en^{\frac{2}{3}}$.

$$\pi_1: \quad 1 \quad 4 \quad 2 \quad 5 \quad 6 \quad 3 \quad 8 \quad 7$$
$$\pi_2: \quad 2 \quad 7 \quad 1 \quad 4 \quad 3 \quad 6 \quad 5 \quad 8$$

- Let S and T be length k subsequences of π_1 and π_2.
- Example: $k = 4$, $S = 2\,6\,8\,7$, and $T = 1\,4\,6\,5$.
- Define $X_{S,T}$ be the event that the subsequences S and T are order isomorphic.

$$P(L_n \geq k) = P\left(\bigcup X_{S,T}\right) \leq \sum P(X_{S,T})$$

Upper Bound

Proposition $E(L_n) < en^{\frac{2}{3}}$.

$$\begin{array}{lcccccccc} \pi_1: & 1 & 4 & 2 & 5 & 6 & 3 & 8 & 7 \\ \pi_2: & 2 & 7 & 1 & 4 & 3 & 6 & 5 & 8 \end{array}$$

- Let S and T be length k subsequences of π_1 and π_2.
- Example: $k = 4$, $S = 2\ 6\ 8\ 7$, and $T = 1\ 4\ 6\ 5$.
- Define $X_{S,T}$ be the event that the subsequences S and T are order isomorphic.

$$P(L_n \geq k) = P\left(\bigcup X_{S,T}\right) \leq \sum P(X_{S,T}) = \frac{1}{k!}$$

Upper Bound

Proposition $E(L_n) < en^{\frac{2}{3}}$.

$$\pi_1: \; 1 \;\; 4 \;\; 2 \;\; 5 \;\; 6 \;\; 3 \;\; 8 \;\; 7$$
$$\pi_2: \; 2 \;\; 7 \;\; 1 \;\; 4 \;\; 3 \;\; 6 \;\; 5 \;\; 8$$

- Let S and T be length k subsequences of π_1 and π_2.
- Example: $k = 4$, $S = 2\,6\,8\,7$, and $T = 1\,4\,6\,5$.
- Define $X_{S,T}$ be the event that the subsequences S and T are order isomorphic.

$$P(L_n \geq k) = P\left(\bigcup X_{S,T}\right) \leq \sum P(X_{S,T}) = \frac{1}{k!}\binom{n}{k}^2$$

Upper Bound

Proposition $E(L_n) < en^{\frac{2}{3}}$.

Using Sterling's Approximation,

$$P(L_n \geq k) \leq \binom{n}{k}^2 \cdot \frac{1}{k!}$$

Upper Bound

Proposition $E(L_n) < en^{\frac{2}{3}}$.

Using Sterling's Approximation,

$$P(L_n \geq k) \leq \binom{n}{k}^2 \cdot \frac{1}{k!} \leq \frac{1}{k^{\frac{3}{2}}} \cdot \left(\frac{e^3 n^2}{k^3}\right)^n$$

Upper Bound

Proposition $E(L_n) < en^{\frac{2}{3}}$.

Using Sterling's Approximation,

$$P(L_n \geq k) \leq \binom{n}{k}^2 \cdot \frac{1}{k!} \leq \frac{1}{k^{\frac{3}{2}}} \cdot \left(\frac{e^3 n^2}{k^3}\right)^n$$

When $k > en^{\frac{2}{3}}$,

Upper Bound

Proposition $E(L_n) < en^{\frac{2}{3}}$.

Using Sterling's Approximation,

$$P(L_n \geq k) \leq \binom{n}{k}^2 \cdot \frac{1}{k!} \leq \frac{1}{k^{\frac{3}{2}}} \cdot \left(\frac{e^3 n^2}{k^3}\right)^n$$

When $k > en^{\frac{2}{3}}$,

$$P(L_n \geq k) < \frac{1}{k^{\frac{3}{2}}}$$

Upper bound on L_n taking high values $\rightarrow$ upper bound on mean.

Lower Bound

Theorem

$$\liminf_{n\to\infty} \frac{E(L_n)}{n^{2/3}} \geq C$$

where $C = 2 - \log(2e - 1) \approx 0.51$.

Proof Outline of Lower Bound

Divide the permutations into blocks of size $\sim \sqrt[3]{n}$.

$$\begin{array}{l|l|l|l|l}
\pi_1: & 1\ \ 4 & 2\ \ 5 & 6\ \ 3 & 8\ \ 7 \\
\pi_2: & 2\ \ 7 & 1\ \ 4 & 3\ \ 6 & 5\ \ 8
\end{array}$$

Proof Outline of Lower Bound

Divide the permutations into blocks of size $\sim \sqrt[3]{n}$.

π_1:	1	4	2	5	6	3	8	7
π_2:	2	7	1	4	3	6	5	8

Divide the numbers $1, 2, \ldots, n$ into groups of size $\sqrt[3]{n}$:

$$1, 2, 3, 4, 5, 6, 7, 8$$

Proof Outline of Lower Bound

Divide the permutations into blocks of size $\sim \sqrt[3]{n}$.

π_1:	1	4	2	5	6	3	8	7
π_2:	2	7	1	4	3	6	5	8

Divide the numbers $1, 2, \ldots, n$ into groups of size $\sqrt[3]{n}$:

$$1, 2, 3, 4, 5, 6, 7, 8$$

Proof Outline of Lower Bound

π_1:	1	4	2	5	6	3	8	7
π_2:	2	7	1	4	3	6	5	8

- A *match* occurs when entry in π_1 is in same block and group as one in π_2.

Proof Outline of Lower Bound

$$
\begin{array}{l|cc|cc|cc|cc}
\pi_1: & \boxed{1} & 4 & 2 & 5 & 6 & 3 & 8 & 7 \\
\pi_2: & \boxed{2} & 7 & 1 & 4 & 3 & 6 & 5 & 8
\end{array}
$$

- A *match* occurs when entry in π_1 is in same block and group as one in π_2.

Proof Outline of Lower Bound

$$\pi_1: \quad \boxed{1}\ \ 4 \ \Big| \ 2 \ \ 5 \ \Big| \ 6 \ \ 3 \ \Big| \ 8 \ \ 7$$

$$\pi_2: \quad \boxed{2}\ \ 7 \ \Big| \ 1 \ \ 4 \ \Big| \ 3 \ \ 6 \ \Big| \ 5 \ \ 8$$

- A *match* occurs when entry in π_1 is in same block and group as one in π_2.
- Two matches are *compatible* if they are in different blocks and groups.

Proof Outline of Lower Bound

π_1:	$\boxed{1}$	4	$\boxed{2}$	5	6	3	8	7
π_2:	$\boxed{2}$	7	$\boxed{1}$	4	3	6	5	8

- A *match* occurs when entry in π_1 is in same block and group as one in π_2.
- Two matches are *compatible* if they are in different blocks and groups.

Proof Outline of Lower Bound

$$\pi_1: \quad \boxed{1}\ 4 \;\Big|\; \boxed{2}\ 5 \;\Big|\; 6\ \ 3 \;\Big|\; 8\ \ 7$$
$$\pi_2: \quad \boxed{2}\ 7 \;\Big|\; \boxed{1}\ 4 \;\Big|\; 3\ \ 6 \;\Big|\; 5\ \ 8$$

- A *match* occurs when entry in π_1 is in same block and group as one in π_2.
- Two matches are *compatible* if they are in different blocks and groups.
- To find a common pattern, search for set of pairwise compatible matches.

Proof Outline for Lower Bound

Step 1: Search for matches in first group (**black**).

π_1:	1 4	2 5	6 3	8 7
π_2:	2 7	1 4	3 6	5 8

Proof Outline for Lower Bound

Step 1: Search for matches in first group (**black**).

π_1:	1 4	2 5	6 3	8 7
	↓			
π_2:	2 7	1 4	3 6	5 8

Proof Outline for Lower Bound

Step 2: Search for matches in second group (red).

π_1:	1 4	2 5	6 3	8 7
π_2:	2 7	1 4	3 6	5 8

Proof Outline for Lower Bound

Step 3: Search for matches in third group (green).

π_1:	1 4	2 5	6 3	8 7
	↓		↙	
π_2:	2 7	1 4	3 6	5 8

Proof Outline for Lower Bound

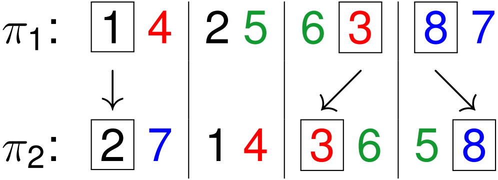
Step 4: Search for matches in fourth group (blue).

π_1: [1] 4 | 2 5 | 6 [3] | [8] 7

π_2: [2] 7 | 1 4 | [3] 6 | 5 [8]

Proof Outline for Lower Bound

Step 4: Search for matches in fourth group (blue).



Proof consists of proving that this algorithm finds, on average, at least $0.51n^{\frac{2}{3}}$ matches.

More Permutations

Let $L_{n,m}$ be the length of the longest common pattern among m permutations randomly chosen from S_n.

Proposition. $E(L_{n,m}) < en^{\frac{m}{2m-1}}$.

More Permutations

Let $L_{n,m}$ be the length of the longest common pattern among m permutations randomly chosen from S_n.

Proposition. $E(L_{n,m}) < en^{\frac{m}{2m-1}}$.

$$P(L_{n,m} \geq k) \leq \binom{n}{k}^m \frac{1}{k!^{m-1}}$$

More Permutations

Let $L_{n,m}$ be the length of the longest common pattern among m permutations randomly chosen from S_n.

Proposition. $E(L_{n,m}) < en^{\frac{m}{2m-1}}$.

$$P(L_{n,m} \geq k) \leq \binom{n}{k}^m \frac{1}{k!^{m-1}} \approx \frac{1}{k^{\frac{2m-1}{3}}} \left(\frac{e^{2m-1} n^m}{k^{2m-1}} \right)^{km}$$

More Permutations

Let $L_{n,m}$ be the length of the longest common pattern among m permutations randomly chosen from S_n.

Proposition. $E(L_{n,m}) < en^{\frac{m}{2m-1}}$.

$$P(L_{n,m} \geq k) \leq \binom{n}{k}^m \frac{1}{k!^{m-1}} \approx \frac{1}{k^{\frac{2m-1}{3}}} \left(\frac{e^{2m-1} n^m}{k^{2m-1}} \right)^{km}$$

Conjecture. $E(L_{n,m}) \in \Theta\left(n^{\frac{m}{2m-1}}\right)$.

Future Work

- We showed that, asymptotically,

$$0.51 < \frac{E(L_n)}{n^{\frac{2}{3}}} < e$$

- Conjecture: There exists a constant κ such that

$$\lim_{n\to\infty} \frac{E(L_n)}{n^{\frac{2}{3}}} = \kappa$$

- Find κ, improve bounds.
- Variations on problem (consecutive patterns, permutations with specific properties).

Thanks for listening!

Questions?