The first occurrence of a pattern in a random sequence

Yixin (Kathy) Lin
Joint work with Sergi Elizalde

Dartmouth College

June 17, 2022

Penney's game

In Penney's game, Alina selects a binary word of length $n \geq 3$, then Ben selects another binary word of the same length. A fair coin is tossed repeatedly.

Ex., Alina: HHH, Ben: THH, THTTHTHTHH.

Penney's game

In Penney's game, Alina selects a binary word of length $n \geq 3$, then Ben selects another binary word of the same length. A fair coin is tossed repeatedly.

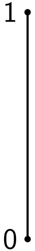
Ex., Alina: HHH, Ben: THH, THTTHTHTHH.

It is known that, for any word picked by Alina, Ben can always pick a word will be more likely to appear first.

Penney's game for permutations

Let $\mathbf{X} = X_1, X_2, \ldots$ be a sequence of i.i.d. continuous random variables. A **consecutive occurrence** of a permutation $\sigma \in \mathcal{S}_k$ is a subsequence $X_i, X_{i+1}, \ldots, X_{i+k-1}$ whose entries are in the same relative order as $\sigma_1, \sigma_2, \ldots, \sigma_k$.

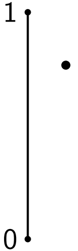
Ex., Sydney: 123, Tom: 213.



Penney's game for permutations

Let $\mathbf{X} = X_1, X_2, \ldots$ be a sequence of i.i.d. continuous random variables. A **consecutive occurrence** of a permutation $\sigma \in \mathcal{S}_k$ is a subsequence $X_i, X_{i+1}, \ldots, X_{i+k-1}$ whose entries are in the same relative order as $\sigma_1, \sigma_2, \ldots, \sigma_k$.

Ex., Sydney: 123, Tom: 213.



Penney's game for permutations

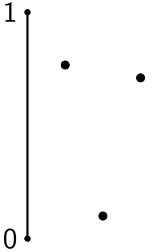
Let $\mathbf{X} = X_1, X_2, \ldots$ be a sequence of i.i.d. continuous random variables. A **consecutive occurrence** of a permutation $\sigma \in \mathcal{S}_k$ is a subsequence $X_i, X_{i+1}, \ldots, X_{i+k-1}$ whose entries are in the same relative order as $\sigma_1, \sigma_2, \ldots, \sigma_k$.

Ex., Sydney: 123, Tom: 213.

Penney's game for permutations

Let $\mathbf{X} = X_1, X_2, \ldots$ be a sequence of i.i.d. continuous random variables. A **consecutive occurrence** of a permutation $\sigma \in \mathcal{S}_k$ is a subsequence $X_i, X_{i+1}, \ldots, X_{i+k-1}$ whose entries are in the same relative order as $\sigma_1, \sigma_2, \ldots, \sigma_k$.

Ex., Sydney: 123, Tom: 213.



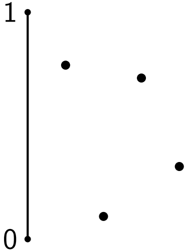
21

312

Penney's game for permutations

Let $\mathbf{X} = X_1, X_2, \ldots$ be a sequence of i.i.d. continuous random variables. A **consecutive occurrence** of a permutation $\sigma \in \mathcal{S}_k$ is a subsequence $X_i, X_{i+1}, \ldots, X_{i+k-1}$ whose entries are in the same relative order as $\sigma_1, \sigma_2, \ldots, \sigma_k$.

Ex., Sydney: 123, Tom: 213.



21

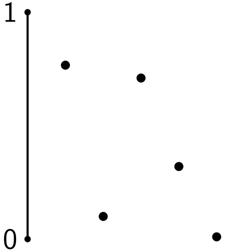
312

4132

Penney's game for permutations

Let $\mathbf{X} = X_1, X_2, \ldots$ be a sequence of i.i.d. continuous random variables. A **consecutive occurrence** of a permutation $\sigma \in \mathcal{S}_k$ is a subsequence $X_i, X_{i+1}, \ldots, X_{i+k-1}$ whose entries are in the same relative order as $\sigma_1, \sigma_2, \ldots, \sigma_k$.

Ex., Sydney: 123, Tom: 213.

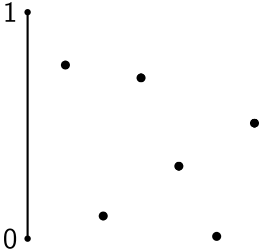


21
312
4132
52431

Penney's game for permutations

Let $\mathbf{X} = X_1, X_2, \ldots$ be a sequence of i.i.d. continuous random variables. A **consecutive occurrence** of a permutation $\sigma \in \mathcal{S}_k$ is a subsequence $X_i, X_{i+1}, \ldots, X_{i+k-1}$ whose entries are in the same relative order as $\sigma_1, \sigma_2, \ldots, \sigma_k$.

Ex., Sydney: 123, Tom: 213.

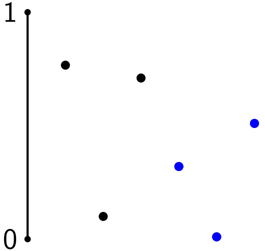


21
312
4132
52431

Penney's game for permutations

Let $\mathbf{X} = X_1, X_2, \ldots$ be a sequence of i.i.d. continuous random variables. A **consecutive occurrence** of a permutation $\sigma \in \mathcal{S}_k$ is a subsequence $X_i, X_{i+1}, \ldots, X_{i+k-1}$ whose entries are in the same relative order as $\sigma_1, \sigma_2, \ldots, \sigma_k$.

Ex., Sydney: 123, Tom: 213.



21
312
4132
52431
625314

Expected time to see the first occurrence of a pattern

For $\sigma \in \mathcal{S}_k$, define the random variable T_σ as the smallest j such that $X_1, \ldots, X_j$ contains an occurrence of σ. Let $\alpha_n(\sigma)$ be the number of permutations in $\mathcal{S}_n$ that avoid the pattern σ, and denote the corresponding exponential generating function by

$$P_\sigma(z) = \sum_{n \geq 0} \alpha_n(\sigma) \frac{z^n}{n!}.$$

The expectation $\mathbb{E}\, T_\sigma$ has a surprisingly simple expression in terms of this generating function.

Theorem

For every σ,

$$\mathbb{E}\, T_\sigma = P_\sigma(1).$$

Expected time to see the first occurrence of a pattern

Expressions for $P_\sigma(z)$ for various σ have been obtained by Elizalde and Noy (2003, 2012). For example, it follows from above theorem that

$$\mathbb{E}T_{12} = e,$$

$$\mathbb{E}T_{123} = \frac{\sqrt{3e}}{2\cos(\frac{\sqrt{3}}{2} + \frac{\pi}{6})} \approx 7.924,$$

$$\mathbb{E}T_{132} = \frac{1}{1 - \int_0^1 e^{-t^2/2} dt} \approx 6.926.$$

Probability of seeing one pattern before another

Given two permutations σ and τ, we would like to compute the probability that σ appears before τ in $\mathbf{X}$, which we denote by $\Pr(\sigma \prec \tau)$.

Observation

For the decreasing pattern $\rho_k = k(k-1)\cdots 21$,

$$Pr(\rho_k \prec 12) = \frac{1}{k!}.$$

Probability that 12 precedes ρ_k

Define $F_{\sigma \to \tau}$ as the number of further steps to see the pattern τ after the first occurrence of σ, assuming that σ occurs before τ in $\mathbf{X}$.

Theorem

For any two consecutive patterns σ and τ,

$$\Pr(\sigma \prec \tau) = \frac{\mathbb{E}F_{\tau \to \sigma} + \mathbb{E}T_\tau - \mathbb{E}T_\sigma}{\mathbb{E}F_{\tau \to \sigma} + \mathbb{E}F_{\sigma \to \tau}}.$$

Probability of seeing one pattern before another

$\Pr(\sigma \prec \tau)$ **for patterns of length 3**

$\Pr(\sigma \prec \tau)$ τ / σ	123	132	213	231	312	321
123	–	0.5	0.412	0.551	0.342	0.5
132	0.5	–	0.462	0.476	0.5	0.658
213	0.588	0.538	–	0.5	0.524	0.449
231	0.449	0.524	0.5	–	0.538	0.588
312	0.658	0.5	0.476	0.462	–	0.5
321	0.5	0.342	0.551	0.412	0.5	–

Probability of seeing one pattern before another

$\Pr(\sigma \prec \tau)$ **for patterns of length 3**

$\Pr(\sigma \prec \tau)$ (σ \ τ)	123	132	213	231	312	321
123	–	0.5	0.412	0.551	0.342	0.5
132	0.5	–	0.462	0.476	0.5	0.658
213	0.588	0.538	–	0.5	0.524	0.449
231	0.449	0.524	0.5	–	0.538	0.588
312	0.658	0.5	0.476	0.462	–	0.5
321	0.5	0.342	0.551	0.412	0.5	–

Conjecture

Assume Sydney picks $\sigma = \sigma_1 \cdots \sigma_k$, then the winning strategy for Tom would be to pick $\sigma_k \sigma_1 \cdots \sigma_{k-1}$.

Probability of seeing one pattern before another

$\Pr(\sigma \prec \tau)$ **for patterns of length 3**

$\Pr(\sigma \prec \tau)$ τ / σ	123	132	213	231	312	321
123	–	0.5	0.412	0.551	0.342	0.5
132	0.5	–	0.462	0.476	0.5	0.658
213	0.588	0.538	–	0.5	0.524	0.449
231	0.449	0.524	0.5	–	0.538	0.588
312	0.658	0.5	0.476	0.462	–	0.5
321	0.5	0.342	0.551	0.412	0.5	–

Probability of seeing one pattern before another

$\Pr(\sigma \prec \tau)$ for patterns of length 3

$\Pr(\sigma \prec \tau)$ τ / σ	123	132	213	231	312	321
123	–	0.5	0.412	0.551	0.342	0.5
132	0.5	–	0.462	0.476	0.5	0.658
213	0.588	0.538	–	0.5	0.524	0.449
231	0.449	0.524	0.5	–	0.538	0.588
312	0.658	0.5	0.476	0.462	–	0.5
321	0.5	0.342	0.551	0.412	0.5	–

Observation

$\Pr(\tau \prec \sigma) = 1 - \Pr(\sigma \prec \tau)$ assuming that neither of the patterns contains the other.

Probability of seeing one pattern before another

$\Pr(\sigma \prec \tau)$ **for patterns of length 3**

$\Pr(\sigma \prec \tau)$ τ / σ	123	132	213	231	312	321
123	–	0.5	0.412	0.551	0.342	0.5
132		–	0.462	0.476	0.5	0.658
213			–	0.5	0.524	0.449
231				–	0.538	0.588
312					–	0.5
321						–

Probability of seeing one pattern before another

$\Pr(\sigma \prec \tau)$ for patterns of length 3

$\Pr(\sigma \prec \tau)$ τ / σ	123	132	213	231	312	321
123	–	0.5	0.412	0.551	0.342	0.5
132		–	0.462	0.476	0.5	0.658
213			–	0.5	0.524	0.449
231				–	0.538	0.588
312					–	0.5
321						–

Observation

If $\bar{\sigma}$ denotes the permutation such that $\bar{\sigma}_i = k + 1 - \sigma_i$ for $1 \leq i \leq k$, then $\Pr(\sigma \prec \tau) = \Pr(\bar{\sigma} \prec \bar{\tau}) = 1 - \Pr(\bar{\tau} \prec \bar{\sigma})$.

Probability of seeing one pattern before another

$\Pr(\sigma \prec \tau)$ **for patterns of length 3**

$\Pr(\sigma \prec \tau)$ τ / σ	123	132	213	231	312	321
123	–	0.5	0.412	0.551	0.342	0.5
132		–	0.462	0.476	0.5	
213			–	0.5		
231				–		
312					–	
321						–

Probability of seeing one pattern before another

$\Pr(\sigma \prec \tau)$ **for patterns of length 3**

$\Pr(\sigma \prec \tau)$ τ / σ	123	132	213	231	312	321
123	–	0.5	0.412	0.551	0.342	0.5
132		–	0.462	0.476	0.5	
213			–	0.5		
231				–		
312					–	
321						–

Observation

$\Pr(\sigma \prec \bar{\sigma}) = 0.5.$

Probability of seeing one pattern before another

$\Pr(\sigma \prec \tau)$ **for patterns of length 3**

$\Pr(\sigma \prec \tau)$ (σ \ τ)	123	132	213	231	312	321
123	–		0.412	0.551	0.342	
132		–	0.462	0.476		
213			–			
231				–		
312					–	
321						–

Probability of seeing one pattern before another

Theorem

Let $\sigma, \tau \in \mathcal{S}_k$ be two patterns, let a_n be the number of permutations of length n that end with σ, and avoid both σ and τ elsewhere. Then the probability that σ shows up before τ is given by

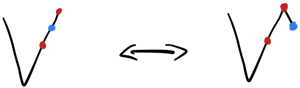
$$\begin{aligned}\Pr(\sigma \prec \tau) &= \sum_{n \geq k} \Pr(\textit{game ends at } n) \cdot \Pr(\sigma \prec \tau \,|\, \textit{game ends at } n) \\ &= \sum_{n \geq k} \frac{a_n}{n!}.\end{aligned}$$

Probability of seeing one pattern before another

$\Pr(\sigma \prec \tau)$ for patterns of length 3

$\Pr(\sigma \prec \tau)$ (σ \ τ)	123	132	213	231	312	321
123	–		0.412	0.551	0.342	
132		–	0.462	0.476		
213			–			
231				–		
312					–	
321						–

Probability that 132 precedes 231



permutations that end with 123 with at most 1 valley and no peak	$\leftrightarrow$	permutations that end with 132, and avoid 132 and 231 elsewhere

$$\Pr(132 \prec 231) = \sum_{n\geq 3} \frac{\sum_{i=0}^{n-3} \binom{n-1}{i}}{n!}$$
$$= \frac{e^2}{2} - e - \frac{1}{2} \approx 0.476.$$

Probability of seeing one pattern before another

$\Pr(\sigma \prec \tau)$ for patterns of length 3

$\Pr(\sigma \prec \tau)$ \ τ / σ	123	132	213	231	312	321
123	–		0.412	0.551	0.342	
132		–	0.462	0.476		
213			–			
231				–		
312					–	
321						–

Probability that 123 precedes 213

Lemma

Let b_n be the number of permutations of length n that end with 312, and avoid both 123 and 213 elsewhere. Then for all $n \geq 5$,

$$b_n = b_{n-1} + (n-1) \cdot b_{n-2},$$

with initial conditions

$$b_0 = b_1 = b_2 = 0, \; b_3 = 1 \text{ and } b_4 = 4.$$

Probability that 123 precedes 213

Theorem

Let c_n be the number of permutations of length n that end with 123, and avoid both 123 and 213 elsewhere. Then for all $n \geq 5$,

$$c_n = c_{n-1} + b_{n-1} + (n-1) \cdot c_{n-2},$$

with initial conditions $c_3 = 1$ and $c_4 = 2$.
Let $B(x)$ and $C(x)$ be the corresponding E.G.F. for b_n and c_n, resp. Then

$$C(x) = e^{\frac{x(x+2)}{2}} \left(\int_0^x \frac{z^2 + 2B(z)}{2} e^{-\frac{z(2+z)}{2}} dz \right).$$

$$Pr(123 \prec 213) = \sum_{n \geq 0} \frac{c_n}{n!} = C(1) \approx 0.412.$$

Probability of seeing one pattern before another

$\Pr(\sigma \prec \tau)$ **for patterns of length 3**

$\Pr(\sigma \prec \tau)$ σ \ τ	123	132	213	231	312	321
123	–		0.412	0.551	0.342	
132		–	0.462	0.476		
213			–			
231				–		
312					–	
321						–

Probability that 123 precedes 231

Theorem

Let $d(n,i)$ denote the number of permutations in S_n that end with 123, avoid both 123 and 213 elsewhere, and starting with number i. Then for all $n \geq 5$ and all $1 \leq i \leq n$,

$$d(n,i) = \sum_{j=1}^{i-1} d(n-1,j) + \sum_{j=i}^{n-2} (n-1-j) d(n-2,j),$$

with $d(3,1) = 1$, $d(3,2) = d(3,3) = 0$; and $d(4,1) = 0$, $d(4,2) = d(4,3) = d(4,4) = 1$.

$$\Pr(123 \prec 231) = \sum_{n \geq 0} \sum_{i \geq 0} \frac{d(n,i)}{n!} \approx 0.551.$$

This recurrence relation is based on 'On multi-avoidance of generalised patterns' by Kitaev and Mansour.

Probability of seeing one pattern before another

$\Pr(\sigma \prec \tau)$ for patterns of length 3

$\Pr(\sigma \prec \tau)$ τ / σ	123	132	213	231	312	321
123	–		0.412	0.551	0.342	
132		–	0.462	0.476		
213			–			
231				–		
312					–	
321						–

Probability that 132 precedes 213

Theorem

Let $s(n;i,j)$ be the number of permutations that end with 132, avoid both 132 and 213 elsewhere, and starting with numbers i and j. Then

- $s(n;i,i) = 0$ *for all* $n, i \geq 1$;
- $s(n;i,j) = \sum_{k=1}^{i-1} s(n-1;j,k)$ *if* $i > j$;
- $s(n;i,j) = \sum_{k=1}^{i-1} s(n-1;j-1,k) + \sum_{k=j}^{n-1} s(n-1;j-1,k)$ *if* $i < j$;

with initial conditions $s(3;1,3) = 1$, $s(3;i,j) = 0$ *for all other* i, j.

$$\Pr(132 \prec 213) = \sum_{n\geq 0} \sum_{i\geq 0} \sum_{j\geq 0} \frac{s(n;i,j)}{n!} \approx 0.462.$$

This recurrence relation is based on 'On multi-avoidance of generalised patterns' by Kitaev and Mansour.

Probability of seeing one pattern before another

$\Pr(\sigma \prec \tau)$ **for patterns of length 3**

$\Pr(\sigma \prec \tau)$ (σ \ τ)	123	132	213	231	312	321
123	–		0.412	0.551	0.342	
132		–	0.462	0.476		
213			–			
231				–		
312					–	
321						–

Probability that 123 precedes 312

Theorem

Let $t(n; i, j)$ be the number of permutations that end with 123, avoid both 123 and 312 elsewhere, and start with i and j. Then

- $t(n; i, i) = 0$ *for all* $n, i \geq 1$;
- $t(n; i, j) = \sum_{k=1}^{j-1} t(n-1; j-1, k)$ *if* $i < j$;
- $t(n; i, j) = \sum_{k=1}^{j-1} t(n-1; j, k) + \sum_{k=i}^{n-1} t(n-1; j, k)$ *if* $i > j$;

with $t(4; 2, 1) = 1$, *and* $t(4; i, j) = 0$ *for all other* $i, j \leq 4$.

$$\Pr(132 \prec 213) = \sum_{n \geq 0} \sum_{i \geq 0} \sum_{j \geq 0} \frac{t(n; i, j)}{n!} \approx 0.342.$$

References

- S. Collings. Coin sequence probabilities and paradoxes. *Bulletin* vol.18(11-12) (1982), p.227-232.
- S. Elizalde and M. Noy. Consecutive patterns in permutations. *Adv. in Appl. Math.* vol. 30 (2003), p.110-123.
- S. Elizalde and M. Noy. Clusters, generating functions and asymptotics for consecutive patterns in permutations. *Adv. in Appl. Math.* vol. 49 (2012), p.351-374.
- M. Gardner. On the paradoxical situations that arise from nontransitive situation. *Scientific American* (Oct 1974), p.120-125.
- S. Kitaev and T. Mansour. On multi-avoidance of generalised patterns. *Discrete Math* (2002).