Some Results Concerning the Expected Number of Distinct Patterns in a Random Permutation

Verónica Borrás-Serrano, UPR, Mayaguez

Isabel Byrne, Virginia Tech

Anant Godbole, ETSU

Hannah Swickheimer, ETSU

Nathaniel Veimau, Swarthmore College

Supported by NSF DMS 1852171

In 2003, at PPI in Dunedin, Herb Wilf asked about the maximum number of distinct patterns (of all sizes) contained in a permutation on $[n]$. Of course patterns need not be in consecutive positions, but many of our results concern such a case.

Let this quantity be denoted by $\phi(\pi_n)$

- $\min_{\pi_n \in S_n} \phi(\pi_n) = n+1$
- Trivial pigeonhole
  $\max_{\pi_n \in S_n} \phi(\pi_n) \le \sum_{k=1}^{n} \min\{k!, \binom{n}{k}\} \sim 2^n$

- Coleman (2004)

$$\max \phi(\pi_n) \geq 2^{n-2\sqrt{n}+1} \quad (n=2^k)$$

Albert, Coleman, Leader, Flynn (2007)

$$\max \phi(\pi_n) \geq 2^n\left(1-6\sqrt{n}\,2^{-\sqrt{n}/2}\right)$$

Miller (2009)

$$2^n - O\left(n^2 2^{n-\sqrt{2n}}\right) \leq \max \phi(\pi_n) \leq$$

$$2^n - \Theta\left(n 2^{n-\sqrt{2n}}\right)$$

Recently, Allen, Cruz, Dobbs, Downs, Fokuoh, Godbole, Papanikolaou, Soto, and Yoshikawa (2022+) considered the <u>expected</u> no. of distinct <u>consecutive</u> patterns in a random permutation

Let $\psi(\Pi_n)$ be the no. of distinct consecutive patterns in a permutation Π_n.

- $$\max \psi(\Pi_n) \le \sum_{k=1}^{n} \min(n-k+1, k!)$$
$$\le \frac{n^2}{2}(1 + o(1))$$

- $$\sum \min(n-k+1, k!) \ge \frac{n^2}{2}(1 - o(1))$$

- The bound of $\sum \min(n-k+1, k!)$ is attained $\forall\, n \le 8$.

- Moreover the expected value of $\psi(\Pi_n)$, $E(\psi(\Pi_n)) = E(X)$ is very close to the max

Ex: For $n = 5$

$$\sum_{k=1}^{n} \min\{n-k+1, k!\} = 9 \text{ is}$$

attained by 14325 and

$E(x)$, over all $120 = 5!$ permutations

is 8.7

Theorem (Allen et al, 2022+)

$$E(x) = \frac{n^2}{2}(1 - o(1))$$

Proof Outline

$$E(x) = \sum_{k=1}^{n} E(x_k) \geq \sum_{k=b_n}^{n} E(x_k)$$

$n-k+1 = k!$ if

$k \sim \dfrac{\log n}{\log \log n}$; b_n will

be taken to be $100 \log n$ for

technical reasons

$$E(X) \geq \sum_{k=b_n}^{n} (n-k+1) - E(Y_k)$$

[Y_k = no of 'repeat' consecutive patterns (counted lexicographically)]

$$\geq \sum (n-k+1) - E(Z_k)$$

[Z_k = no. of pairs of consecutive positions that give the same pattern of length k]

The proof is technical & is skipped. One lemma is as follows:

<u>Lemma</u>: In the consecutive case, when $r \leq \lfloor \frac{k}{2} \rfloor$

$$\rho_1 \to \times\ \times\ \times\ \times\ \times\ \times\ \times\ \times$$

$$\rho_2 \to \underbrace{\circ\ \circ\ \circ}_{r}\ \underbrace{\circ\ \circ\ \circ\ \circ\ \circ}_{k-r}$$

$$P(\rho_1 \simeq \rho_2) \leq \frac{3^k}{k!}$$

Precise form of theorem in the consecutive case

$$E(X) \geq \frac{(n - \lceil 100 \ln n \rceil)^2}{2} - 1$$

The proof is technical & is skipped. One lemma is as follows:

Lemma: In the consecutive case, when $r \le \lfloor \frac{k}{2} \rfloor$

$$\rho_1 \to \times\ \times\ \times\ \times\ \times\ \times\ \times\ \times$$

$$\rho_2 \to \underbrace{\circ\ \circ\ \circ}_{r}\ \underbrace{\circ\ \circ\ \circ\ \circ\ \circ}_{k-r}$$

$$P(\rho_1 \simeq \rho_2) \le \frac{3^k}{k!}$$

Precise form of theorem in the consecutive case

$$E(X) \ge \frac{(n - \lceil 100 \ln n \rceil)^2}{2} - 1$$

In 2021, Borras, Byrne, Veimau began work with G on the non-consecutive case (the focus is still $E(x)$)

---

2 critical lemmas captured by the example below:

$5+6-2$ $\quad$ $8+7-\cancel{8}13$

$3+4-1$

4	9	3				2	6	5	1	7			8	
X	X	X				X	X	X	X	X			X	
		6						9					12	
		O	O	O	O			O			O	O	O	O
		4	9	3	2			6			5	1	7	8

$\rho = 493265178$

The numbers $1 \longrightarrow 15$ can be allotted in $\binom{5}{2} * \binom{2}{1} * \binom{3}{2} * \binom{3}{1}$ ways

In 2021, Borras, Byrne, Veimau began work with G on the non-consecutive case (the focus is still $E(X)$)

---

2 critical lemmas captured by the example below:

$3+4-1$ $\quad$ $5+6-2$ $\quad$ $8+7-3$

4 9 3 2 6 5 1 7 8

X X X X X X X X X

6 $\quad$ 9 $\quad$ 12

O O O O O O O O O

4 9 3 2 6 5 1 7 8

$\sigma = 493265178$

The numbers $1 \longrightarrow 15$ can be allotted in $\binom{5}{2} * \binom{2}{1} * \binom{3}{2} * \binom{3}{1}$ ways

Lemma 1: $2k-r$ numbers can be assigned to two k patterns that overlap in r spots iff the patterns in the overlaps are isomorphic

Ex: In the example

3	5	8
X	X	X
4	6	7
O	O	O

were both 123 patterns

Lemma 2: For a fixed pattern, the numbers in the non overlapping positions can be placed in 2^{2k-2r} ways

Proof?

- Lemmas 1 & 2 yield the fundamental formula:

P(2k−r spots yield isomorphic k patterns) is

$$\le \frac{\underbrace{\binom{n}{k}}_{\rho_1 \text{ spots}} \underbrace{\binom{k}{r}\binom{n-k}{k-r}}_{\rho_2 \text{ spots}} \underbrace{\binom{k}{r}^2 r!}_{\text{Lemma 1}} \overbrace{(k-r)!}^{\text{Rest of } \rho_1=\rho_2 \text{ pattern}} \underbrace{2^{2k-2r}}_{\text{Lemma 2}}}{(2k-r)!}$$

- The above formula, derived so as to mimic the X, Y, Z development as in the consecutive case, gives a Z that exceeds $\binom{n}{k}$

- X, Y, Z approach will not work ☹

However it did yield

<u>Theorem</u> (B, B, G, V)

The expected no. of <u>pairs</u> of non-isomorphic k-perms is at least

$$\binom{n}{k}^2 - \sum_r \binom{n}{k}\binom{k}{r}\binom{n-k}{k-r}\frac{(k-r)!\,2^{2k-2r}\binom{k}{r}^2 r!}{(2k-r)!}$$

$$\geq \binom{n}{k}^2\left\{1 - \frac{1}{2^{n(1-o(1))}}\right\}$$

analysis with suitable <u>k's</u>, $k = \frac{n}{2} \pm \sqrt{n}\log n$

So what to do next?

We tried Poisson approximation as in Barbour et al (1992)

We worked with X directly

X_K = no of patterns of size k that occur ≥ 1 times

$$\underbrace{123, 123, 123}, \underbrace{132}_{1}, \underbrace{231, 231}_{2}, \underbrace{312}_{1}$$

occurrences: 3

$$X_K = X_3 = 4$$

Let ω_K be any pattern of length K.

$$E(\omega_K) = \frac{\binom{n}{k}}{K!} = '\lambda'$$

So <u>if</u> ω_K is approximately Poisson,

$$P(\omega_K = 0) = e^{-\frac{\binom{n}{k}}{K!}} + \text{error}$$

So

$$P(\omega_K \geq 1) \geq 1 - e^{-\binom{n}{k}/k!} - \text{error}$$

The BHJ book is full of error estimates

P(the pattern occurs in any k spots) = $\frac{1}{k!}$ is small (rare event)

if k is large

The Literature on Normal Approx'n of patterns is for small k.

So...

$$P(w_k \geq 1) \geq 1 - e^{-\binom{n}{k}/k!} - \text{error}$$

$$\geq \frac{\binom{n}{k}/k!}{1 + \frac{\binom{n}{k}}{k!}} - \text{error}$$

$$\therefore E(X_k) \geq k! \left\{ \frac{\binom{n}{k}/k!}{1 + \frac{\binom{n}{k}}{k!}} - \text{error} \right\}$$

$$= \frac{\binom{n}{k}}{1 + \frac{\binom{n}{k}}{k!}} - k! \text{ error}$$

$$\geq \binom{n}{k} \left\{ 1 - \underbrace{\frac{\binom{n}{k}}{k!}}_{\text{small?}} - \underbrace{\frac{k! \text{ error}}{\binom{n}{k}}}_{\text{small?}} \right\}$$

Unfortunately one of the terms in the error is (for the monotone pattern)

$$\frac{k!^3 (n-k)!\, 2^{2k-2r}}{(r-1)!^2 (k-r)!^3 (2k-r)! (n-2k+r)!}$$

From Lemma 2

which does not tend to zero.

Work in progress: to find a large class of patterns for which the error does go to zero

---

But — we went back to the future !!

Armed with Lemmas 1 & 2 Swickheimer & G went back to the consecutive case

Theorem: In the consecutive case

$$d_{TV}\left(\mathcal{L}(W_K), \text{Poisson}\left(\lambda = \frac{n-K+1}{K!}\right)\right)$$

$$\leq \frac{n}{K!^2} + \frac{2nK}{K!^2} + \frac{4n}{(K+1)!}$$

for any W_K, which leads to

Theorem

$$\sum_{K \geq b_n} E(X_K) = \frac{n^2}{2} - n\ln n - n + \ln n$$

via Poisson approximation and

no auxiliary variables Y_K, Z_K

For the general case, can CLT, Suen, Janson, compound Poisson, multivariate Poisson, Lovász lemma OR particularly, $\binom{n}{K}^2\left\{1 - \frac{1}{2^n}\right\}$ help?