A permutation statistic with a very left-skewed distribution

Anant Godbole, High Point Univ & ETSU

with

Isabel Byrne, University of Delaware

Verónica Borrás-Serrano, Univ Puerto Rico, Mayagüez

Nathaniel Veiman, Microsoft Azure

Full Disclosure ...... (sigh)

In a paper with Hannah Swickheimer (DMTCS 2023) we showed that the expected total no. of consecutive patterns of all lengths $1 \le k \le n$ was close to $\frac{n^2}{2}$. Thus random permutations pack consecutive patterns almost perfectly

$$E(X) \ge \frac{n^2}{2}\left(1 - \frac{17 \ln n}{n}\right)$$

We used the Stein-Chen approach to Poisson approximation as in Barbour et al (1992)

At PP 2024, we stated that the dependencies between indicator (0-1) variables were too great – in the non-consecutive case

Here we circumvent these issues and once again establish Poisson approximation for the distribution of any individual pattern count

This is then "multiplied by k!" to give the main result

4

The focus on $P(U_j \geq 1)$ where

U_j is the no. of occurrences of

a fixed pattern allows us to use the

same trick as in G & Swickheimer

to "factor out" an $\binom{n}{k}$. Details

follow.

Moreover, the "coupling approach" to

Poisson approximation is employed as

opposed to the "local approach"

as in, e.g., Arratia, Goldstein &

Gordon (1992)

In Crane & da Salvo (DMTCS, 2018)

the local approach was used

yielding Poisson convergence for

the number of occurrences of a

pattern but with little focus on

error rates

Their focus was on $P(u_j = 0)$

(pattern avoidance in a

random permutation) Ours is on

pattern containment — with

error rates

Improving on previous estimates

(Coleman, Albert et al) Miller (JCTA 2009)

proved a definitive set of

lower & upper bounds for the

Wilf (PPI, 2003)

maximum number of patterns of

all lengths contained in a

permutation:

$$2^n - O(n^2 2^{n-\sqrt{2n}}) \le \max(X_n) \le 2^n - \Theta(n 2^{n-\sqrt{2n}})$$

In the case of words (say binary)

$$\max X \sim \gamma^n$$

$$E(X) \sim 1.5^n$$

different rates

In the case of patterns of all sizes contained in a random permutation, our main result is that

$$E(X) \geq 2^n(1 - o(1))$$

will be specified

so that

$$E(X) \text{ is close to } \max(X)$$

thus motivating the title of this talk.

Stirling's approximation is used in a gentle way without $\sqrt{n}$ factors,

$$n! \geq \left(\frac{n}{e}\right)^n, \text{ so}$$

$$\binom{n}{k} \leq \frac{n^k}{k!} \leq \left(\frac{ne}{k}\right)^k$$

$$E(X_k) = E\sum_{j=1}^{k!} I_j \begin{cases} 1 \\ 0 \end{cases}$$

$I_j = 1$ if the jth pattern occurs at least once

no of occurrences of jth pattern

$$= \sum_{j=1}^{k!} P(u_j \geq 1)$$

$$\stackrel{?}{=} k! \sum_{1}^{k!} P(u_j \geq 1)$$

error

assuming for now that if doesn't depend on j

$$\geq k!\,\Big(\underbrace{1-e^{-\lambda}}_{P(\text{Poisson}(\lambda)\geq 1)} - \varepsilon_{n,k}\Big)$$

In our case $\lambda = E(u_j) = \dfrac{\binom{n}{k}}{k!}$

$$\sum_{j=1}^{\binom{n}{k}} E(I_j) \quad = \quad u_j = \sum_{j=1}^{\binom{n}{k}} P(u_j \text{ occurs at the } j\text{th spot})$$

so $E(u_j) = \dfrac{\binom{n}{k}}{k!}$

What is ε_{nk}?

The coupling approach says that (Theorem 2.B in Barbour et al (1992):

$$d_{TV}(\mathcal{L}(U_j), Po(\lambda))$$

$$\le \underbrace{\left(\frac{1-e^{-\lambda}}{\lambda}\right)}_{\text{"magic factor"}} \sum_j P(I_j = 1)\left\{P(I_j = 1) + \sum_{i \ne j} P(I_i \ne J_{ji})\right\}$$

where

$$\mathcal{L}(J_{i1}, J_{i2}, \cdots, J_{i(n)}) =$$

$$\mathcal{L}(I_1, I_2, \cdots, I_n \mid I_j = 1)$$ data: $\approx \frac{3}{20}$

It is sometimes (always? often?) convenient to use a more detailed coupling, conditioning both on $I_j = 1$ and the value of another random element Ξ

In this case we have

$$d_{TV}(\mathcal{L}(U_j), Po(\lambda))$$

$$\leq \frac{1-e^{-\lambda}}{\lambda} \sum_j \sum_\xi P(I_j = 1, \Xi = \xi) \cdot$$

$$\left\{ P(I_j = 1) + \sum_{i \neq j} P(I_i \neq J_{ji\xi}) \right\}$$

Note: $P(I_i \neq J_{ji\xi}) = 0$ if

$i \cap j = \emptyset$, because of the

coupling used:

If $I_j = 1$ and $\Xi = \xi$, i.e. the

pattern is in place in the jth spot

and $\Xi = \xi$

$\searrow$

the permutation on the $2k-r$ spots,

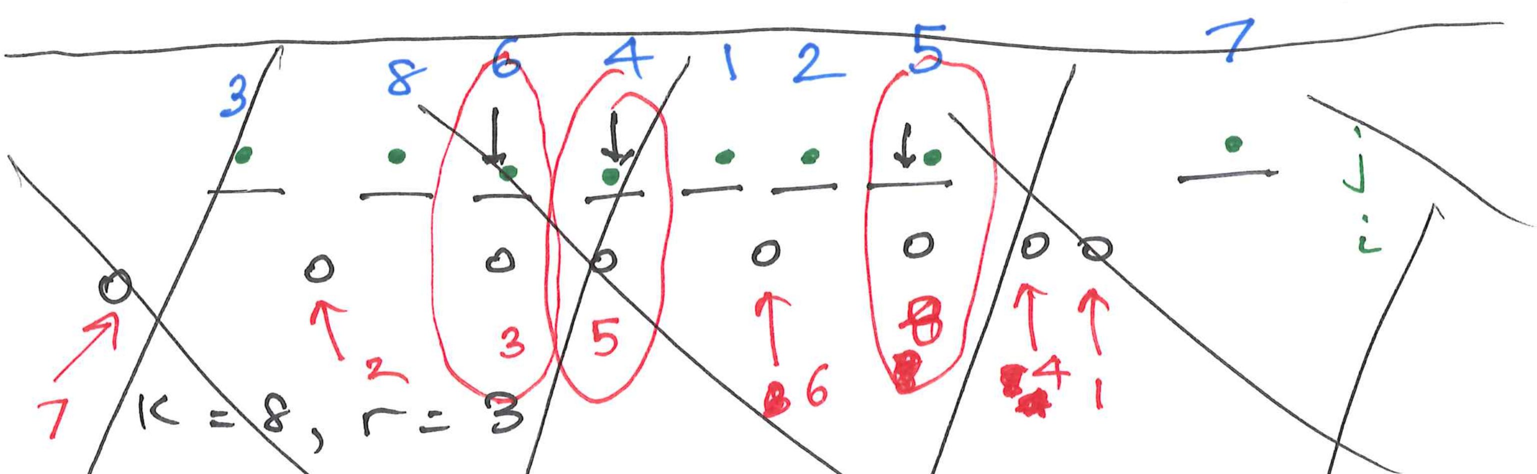
$|i \cup j| = 2k-r$, $|i \cap j| = r$ $\quad 1 \leq r \leq k-1$

....

If $I_j = 1$ & $\Xi = \xi$, we "do nothing" & let $J_{ji\xi} = 1 = I_i \ \forall i$

If $I_j = 0$ & $\underset{K}{\Xi} = \xi$, we change the pattern in the "j spots" to the desired one (i.e. to conform with $I_j = 1$) and let

$J_{ji\xi} = 1$ if the pattern now is created in the ith set of positions



Pattern: 7 2 3 5 6 8 4 1

overlaps are both 2 3 1

Since the numbers in the non-overlap spots of i will not change as a result of putting the desired pattern in place in the k positions of j, they must be in the right locations. However the nos. in j can be put willy-nilly as long as they are permutations of

K

$\{1, 2, 3, 4, 5, 6, 7, 8\}$.

Lemma 1 Given any 2 sets of k-positions with an overlap of r, the nos. in $\{1, 2, \dots, 2k-r\}$ can be allotted in $\leq 2^{2k-2r}$ ways

Proof: The nos. in the r overlap spots must be $u_1+l_1-1, u_2+l_2-2, \dots, u_r+l_r-r$. The nos. in between the overlap spots can be allotted in

$$\binom{u_{s+1}+l_{s+1}-(s-1)-(u_s+l_s-s)-1}{u_{s+1}-u_s-1}$$

ways (why?).

The product of the above nos. from 1 to r can be bounded by 2^{2k-2r} $\square$

So given the patterns & the overlap spots, there are $\leq 2^{2k-r}$ perms ~~that~~ on $[2k-r]$ that respect this. NOTE: The overlaps MUST be order isomorphic

Back to I & J

The numbers in $i \cup j$ must be one of the $\leq 2^{2k-2r}$ possible.

of these, j ~~must~~ can have the the corresponding nos. in any order (since they will be put in the right order). The nos. in ~~j~~ $i \cup j$ must be correct

Example

j

4 9 3 2 6 5 1 7 8

× × × × × × × × ×

0 0 0 0 0 0 0 0 0

4 9 3 2 6 5 1 7 8

i

7 14 6 15 5 3 4 10 9 2 11 8 1 12 13

$3+4-1=6$ $\quad$ $5+6-2=9$ $\quad$ $8+7-3=12$

One of $2^{2k-2r} = 2^6$ possibilities for final outcome

One possible initial outcome out of 8! possible for $T_j = 0$

2 3 9 6 15 5 7 4 14 12 11 8 1 10 13

Lemma 2: The coupling is correct, i.e. satisfies

$$\mathcal{L}\left(J_{j\xi 1}, \dots, J_{j\xi\binom{n}{k}}\right) = \mathcal{L}\left(I_1, I_2, \dots, I_{\binom{n}{k}} \,\middle|\, I_j = 1\right)$$

Proof Omitted, but consists of showing that any sample outcome is equally likely under both of the above models.

Lemma 3 No set of numbers other than among the 2^{2k-2r} perms are correct, i.e. "things go wrong"

Pf: omitted, but here (verbally) is an example

Back to the formula:

$$d_{TV} \le \left(\frac{1-e^{-\lambda}}{\lambda}\right) \sum_{j} \sum_{\xi} P(I_j = 1, \Xi = \xi) \Big\{ P(I_j = 1) + \sum_{i \ne j} P(I_i \ne J_{ji\xi}) \Big\}$$

bound by 1 since

$\lambda = \dfrac{\binom{n}{k}}{k!}$ is small

$$\lambda \le \left(\frac{ne^2}{k^2}\right)^k \to 0$$

if e.g. $k \ge (e+1)\sqrt{n}$

**

$$\sum_{i \ne j} \Big\{ P(I_i = 1, J_i = 0) + P(I_i = 0, J_i = 1) \Big\} \Big\}$$

$$= T_1 + T_2 + T_3$$

T_1 $$\sum_{j} \sum_{\xi} P(I_j = 1, \Xi = \xi)\, P(I_j = 1)$$

$$\le \sum_{j} \frac{1}{k!} P(I_j = 1)$$

$$\le \binom{n}{k} \frac{1}{k!^2}$$

$$T_1 = \frac{\binom{n}{k}}{k!^2} \longrightarrow 0 \quad \text{if} \quad \left(\frac{ne}{k}\right)^k \left(\frac{e^2}{k^2}\right)^k \longrightarrow 0$$

$$\text{OR} \quad \left(\frac{ne^3}{k^3}\right)^k \longrightarrow 0 \quad \text{OR} \quad \text{if} \quad k > (e+1)\sqrt[3]{n}$$

$$T_2 = \sum_j \sum_\xi P(I_j = 1, \Xi = \xi) \underbrace{\sum_{i \neq j} P(I_i = 1, J_{ij\xi} = 0)}$$

$$\leq \sum_i P(I_i = 1)$$

$$= \frac{\binom{n}{k}}{k!}$$

$$= \sum_j \frac{\binom{n}{k}}{k!} \sum_\xi P(I_i = 1, \xi)$$

$$= \sum_j \frac{\binom{n}{k}}{k!} P(I_i = 1)$$

$$= \sum \frac{\binom{n}{k}}{k!} \frac{1}{k!}$$

$$= \frac{\binom{n}{k}^2}{k!^2} \leq \left(\frac{ne}{k}\right)^{2k} \left(\frac{e}{k}\right)^{2k} = \frac{n^2 e^4}{k^4} \longrightarrow 0 \quad \text{if} \quad k > (e+1)\sqrt{n}$$

$$T_3 = \sum_j \sum_\xi P(I_j = 1, \Xi = \xi) \sum_{i \neq j} P(I_i = 0, J_i = 1)$$

$$= \sum_j P(\Xi = \xi)\, P(I_j = 1 | \cancel{\xi}) \sum_{\substack{r=1 \\ i \cap j = r}}^{k-1} P(I_i = 0, J_{ij\xi} = 1)$$

$$= \sum_j P(I_j = 1) \sum_{r=1}^{k-1} \frac{2^{2k-2r} \binom{k}{r} \binom{n-k}{k-r} k!\,1}{(2k-r)!}$$

$$= \sum_j \frac{1}{\cancel{k!}} \sum_{r=1}^{k-1} \frac{2^{2k-2r} \binom{k}{r} \binom{n-k}{k-r} \cancel{k!}}{r!\,(2k-r)!}$$

want to sneak this in via Lemma 4

$$= \sum_j \sum_r \varphi(n, k, r)$$

$\varphi(n, k, r)$ is a <u>beast</u> to maximize since it depends on both n & k

Pause for a while $\boxed{11}$

we need the overlap spots to be isomorphic

j

i has to align itself so that

$j_0 \approx i_0$

$\uparrow$ overlap

using $r = 3$ as an example

$$P(j_0 \approx i_0) = \Big(|123 \text{ in } j|^2 + |132 \text{ in } j|^2 + \cdots |321 \text{ in } j|^2 \Big) \Big/ \binom{k}{r}^2$$

As we vary the pattern and the 2^{2k-2r} possibilities for each we are looking at

$$\frac{E|123|^2 + E|132|^2 + \cdots + E|321|^2}{\binom{k}{r}^2}$$

$$E|i_1 i_2 \dots i_r| = \frac{\binom{K}{r}}{r!}$$

patterns of size r

aha

$E|\quad|^2$

$$V(|i_1 i_2 \dots i_r|) \sim \cancel{K^{2r-1}} - \frac{\binom{K}{r}^2}{r!^2}$$

$E|\quad|^2$

$$= C\cdot K^{2r-1} \text{ by work of Bona et al}$$

$$\therefore E|\quad|^2 = \frac{\binom{K}{r}^2}{r!^2} + O(K^{2r-1})$$

needs more detail

$$\sim \frac{\binom{K}{r}^2}{r!^2}$$

$$\therefore \sum_{i_1, i_2, \dots i_r} \frac{E|i_1 \dots i_r|^2}{\binom{K}{r}^2} = \frac{r!\binom{K}{r}^2}{r!^2}\frac{1}{\binom{K}{r}^2} = \frac{1}{r!}$$

So $P(j_0 \sim i_0) = \dfrac{r!\binom{k}{r}^2}{r!^2\binom{k}{r}^2} = \dfrac{1}{r!}$

Lemma 4, proved

$$T_3 = \sum_j \sum_{r=1}^{k-1} \underbrace{\frac{2^{2k-2r}}{r!\,(2k-r)!} \binom{k}{r}\binom{n-k}{k-r}}_{q(n,k,r)}$$

we'll instead maximize $\dfrac{2^{2k-r}}{k!\,(2k-r)!}$

$$= \eta(k,r)$$

T_3

$$\le \sum_j \sum_r \binom{k}{r}\binom{n-k}{k-r} \max_r \underbrace{\frac{2^{2k-2r}}{r!\,(2k-r)!}}_{\eta(k,r)}$$

$$\frac{\eta(r+1)}{\eta(r)} = \frac{1}{4}\,\frac{(2k-r)}{(r+1)}$$

So easy!

It was a mess to optimize the whole expression

≥ 1 if

$$2k-r \ge 4r+4$$

$$5r+4 \le 2k$$

recall k is big $> n^{1/2}$ so can ignore the "4"

$$r \le \frac{2k}{5}$$

no floors or ceilings used

$$T_3 \le \sum_j \sum_r \binom{k}{r}\binom{n-k}{k-r} \max_r \eta(k,r)^{25}$$

$$= \sum_j \sum_r \binom{k}{r}\binom{n-k}{k-r} \frac{2^{6k/5}}{\left(\frac{2k}{5}\right)!\,\left(\frac{8k}{5}\right)!}$$

2^{2k-2r} with $r = \frac{2k}{5}$

$(2k-r)!$

$r!$

$$\le \sum_j \frac{2^{6k/5}}{\left(\frac{2k}{5}\right)!\left(\frac{8k}{5}\right)!} \underbrace{\sum_r \binom{k}{r}\binom{n-k}{k-r}}_{\binom{n}{k} \text{ by Chu-Vandermonde}}$$

$$\le \binom{n}{k}^2 \frac{2^{6k/5}}{\left(\frac{2k}{5}\right)!\left(\frac{8k}{5}\right)!}$$

$$= \frac{\binom{n}{k}^2 2^{6k/5}}{\left(\frac{2k}{5e}\right)^{2k/5}\left(\frac{8k}{5e}\right)^{8k/5}} \sim \frac{\binom{n}{k}^2}{k^{2k}} \sim \left(\frac{n^2e^2}{k^4}\right)^k \to 0 \text{ if } k > \sqrt[2]{n}$$

So $T_1 + T_2 + T_3$

$$\le \frac{\binom{n}{k}}{k!^2} + \frac{\binom{n}{k}^2}{k!^2} + \frac{\binom{n}{k}^2}{k^{2k}}$$

just a bit larger than

$$T_2 = \frac{\binom{n}{k}^2}{k!^2}$$

and thus

$$E(X_k) \ge k!\left(1 - e^{-\lambda} - \frac{\binom{n}{k}}{k!^2} - \frac{\binom{n}{k}^2}{k!^2} - \frac{\binom{n}{k}^2}{k^{2k}}\right)$$

$$= k!\left(1 - e^{-\lambda}\right) - \frac{\binom{n}{k}}{k!} - \frac{\binom{n}{k}^2}{k!} - \frac{k!\binom{n}{k}^2}{k^{2k}}$$

$$\ge k!\left(\frac{\lambda}{1+\lambda}\right)$$

$$= \frac{\cancel{k!}\,\frac{\binom{n}{k}}{\cancel{k!}}}{1+\lambda}$$

$$\ge \binom{n}{k}(1-\lambda)$$

$$= \binom{n}{k}\left(1 - \frac{\binom{n}{k}}{k!}\right)$$

$$= \binom{n}{k} - \binom{n}{k}^2 / k!$$

So...

$$E(X_k) \geq \binom{n}{k}\left\{1 - \frac{\binom{n}{k}}{k!} - \frac{1}{k!} - \frac{\binom{n}{k}}{k!} - \frac{\binom{n}{k}}{k^k}\right\}$$

$$\geq \binom{n}{k}\left\{1 - 4\frac{\binom{n}{k}}{k!}\right\}$$

$$\frac{\binom{n}{k}}{k!} \sim \left(\frac{ne}{k^2}\right)^k \to 0 \quad \text{if } k > c\sqrt{n}$$

For specificity,

Let $k = n^{3/4}$, so $\frac{ne}{k^2} \sim \frac{e}{\sqrt{n}}$

$$E(X_k) \geq \binom{n}{k}\left\{1 - \frac{4e}{\sqrt{n}}\right\} \text{ for } k \geq n^{3/4}$$

Lemma 5

$$P\left(\mathrm{Bin}(n,\tfrac{1}{2}) \le n^{3/4}\right)$$

$$\le 2P\left(\mathrm{Bin}(n,\tfrac{1}{2}) = n^{3/4}\right)$$

$$= \frac{1}{2^{n-1}} \binom{n}{n^{3/4}}$$

$$\le \frac{1}{2^{n-1}} \left(\frac{ne}{n^{3/4}}\right)^{n^{3/4}}$$

$$= \frac{1}{2^{n-1}} \left(n^{1/4} e\right)^{n^{3/4}}$$

$$\sim \frac{1}{2^{n-1}} 2^{\frac{1}{4}\lg n \cdot n^{3/4}}$$

$$E(X) = \sum_{k=1}^{n-1} E(X_k)$$

$$\geq \sum_{k=n^{3/4}}^{n-1} E(X_k)$$

$$\geq \sum_{n^{3/4}}^{n} \binom{n}{k}\left(1 - \frac{4e}{\sqrt{n}}\right)$$

$$\geq \sum_{1}^{n} \binom{n}{k}\left(1 - \frac{4e}{\sqrt{n}}\right) - \sum_{1}^{n^{3/4}} \binom{n}{k}$$

$$\geq \left(1 - \frac{4e}{\sqrt{n}}\right) \sum_{1}^{n} \binom{n}{k} - 2^n \cdot \frac{1}{2^{n-1}} \frac{1}{4} \lg n \, n^{3/4}$$

$$\geq \left(1 - \frac{4e}{\sqrt{n}}\right) 2^n - \frac{2^n}{\sqrt{n}}$$

$$= \left(1 - \frac{5e}{\sqrt{n}}\right) 2^n$$

$$= 2^n(1 - o(1)),$$ as

claimed.

Clearly

$$2^n\left(1 - \frac{5e}{\sqrt{n}}\right) \le 2^n\left(1 - \frac{Kn^2}{2^{\sqrt{2n}}}\right)$$

Miller's Bound

since $\dfrac{Kn^2}{2^{\sqrt{2n}}} \le \dfrac{5e}{\sqrt{n}}$

Open Problems

- No point improving the $\frac{5e}{\sqrt{n}}$

- Can multivariate Poisson approximation work too?

  ($\exists$ challenges)

- Can subaddivity methods WORK?

  (Jackson & LeBlanc)

- BIG ONE

  Distribution of X. Lognormal?

  Variance of X

  Right normalization