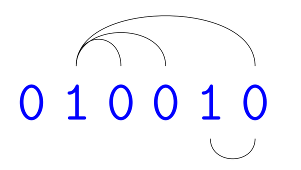
OCCURRENCES OF SUBSEQUENCES IN BINARY WORDS

Krishna Menon and Anurag Singh

$c_p(w)$

$c_p(w) =$ Number of subsequences of w that match p.

Example. For $p = 10$ and $w = 010010$, $c_p(w) = 4$.



$B_{n,p}(0)$

If p is of length l,

$$B_{n,p}(0) = \sum_{j=0}^{l-1} \binom{n}{j}.$$

Idea: How much of p can we get?

Suppose $p = 10010$

0 0 ⋯ 0

0 0 ⋯ 0 1 1 1 ⋯ 1

0 0 ⋯ 0 1 1 1 ⋯ 1 0 1 1 ⋯ 1

$$(1-p_1)^{i_1}\, p_1\, (1-p_2)^{i_2}\, p_2 \cdots (1-p_j)^{i_j}\, p_j\, (1-p_{j+1})^{i_{j+1}}$$

$B_{n,p}(k)$

$$B_{n,p}(k) = \#\{w \in \{0,1\}^n \mid c_p(w) = k\}$$

Number of binary words of length n that contain p exactly k times.

Example. If $p = 1$, then

$$B_{n,p}(k) = \binom{n}{k}.$$

Exercise. If p is of length l with r_i runs of size i, then for any $k \geq 2$,

$$B_{l+1,p}(k) = r_{k-1}.$$

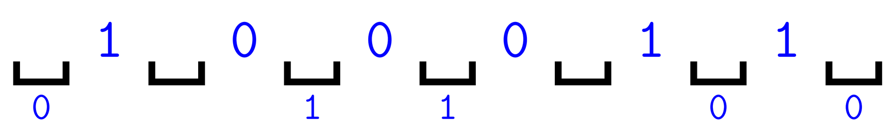
$B_{n,p}(1)$

If p is of length l and has r runs,

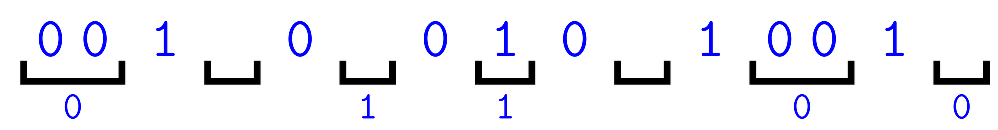
$$B_{n,p}(1) = \binom{n-r+1}{l-r+1}.$$

Idea: Insert letters into p without creating new occurrences.

Suppose $p = 100011$



One way to insert letters is



Runs

Maximal subsequence of consecutive letters that are equal.

Example. Can encode word using run lengths and first letter.

1 0 0 1 1 1 0 0

↕

$(1, 2, 3, 2)_1$

$B_{n,p}(2)$

If p has length l and r runs with r_1 of size 1,

$$B_{n,p}(2) = r_1 \binom{n-r}{l-r+1}.$$

Idea: Use runs of size 1 to make an extra occurrence. Insert letters without creating any more.

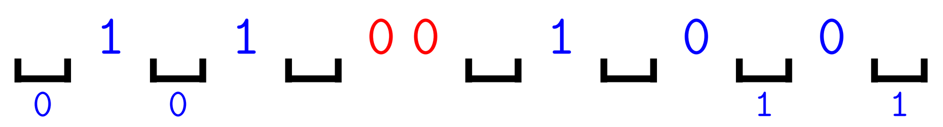
Suppose $p = 110100$

1 1 0 1 0 0

↓

1 1 0 0 1 0 0

↓



Let p be a pattern of length l with at most 3 runs, say $p = 1^i 0^j 1^k$ where $i, k \geq 0$, $j \geq 1$.

Maximum Occurrences

- $M_{n,p} = \max\{c_p(w) \mid w \in \{0,1\}^n\}$.
- $w \in \{0,1\}^n$ is p-optimal if $c_p(w) = M_{n,p}$.

1. We have $M_{n,p} = \max\left\{ \binom{a}{i}\binom{b}{j}\binom{c}{k} \mid a+b+c = n \right\}$.
2. If $1^a 0^b 1^c$ is p-optimal, then so is at least one word in $\{1^{a+1}0^b1^c, 1^a0^{b+1}1^c, 1^a0^b1^{c+1}\}$.
3. The sequence $(M_{n,p})_{n\geq 0}$ is log-concave.

Internal Zeroes

We say p has an *internal zero at* n if there exist $0 \leq k_1 < k_2 < k_3$ such that

$$B_{n,p}(k_1), B_{n,p}(k_3) \neq 0 \text{ but } B_{n,p}(k_2) = 0.$$

1. $p = 10$ does not have any internal zeroes.
2. $p = 101$ does not have an internal zero at n for $n \geq 7$.
3. If p isn't alternating, it has internal zeroes at all $n \geq l + 3$.

Future Directions

- Refine $B_{n,p}(k)$ by keeping track of statistics. Number of 1s, number of runs etc.
- We have used runs to study $B_{n,p}(k)$. Are there better ways to do this?

Conjecture

Patterns of length at most 13 ✓

For two patterns p, q, we have

$$B_{n,p}(k) = B_{n,q}(k) \text{ for all } n, k \geq 0$$

if and only if $p = q, q^r, q^c$, or q^{rc}.

Expressions for $B_{n,p}(3)$ and $B_{n,p}(4)$ in [2].
Formulas for $B_{n,p}(k)$ for $k \geq 5$?
Might be easier to study for particular patterns p.
Interesting case: Alternating words.

References

[1] M. Bóna, B. E. Sagan, and V. R. Vatter. Pattern frequency sequences and internal zeros. *Advances in Applied Mathematics*, 2002.

[2] K. Menon and A. Singh. Subsequence frequency in binary words. *arXiv:2306.07870*, 2023.

Contact Information
Krishna Menon
krishnamenon@cmi.ac.in
CHENNAI MATHEMATICAL INSTITUTE