# Maximal repetition & “Runs” conjecture

Riccardo Lo Iacono
Docente: Gabriele Fici

18 febbraio 2025

# Notazione

1. $\Sigma$ è un alfabeto ordinato e finito di simboli
2. Un elemento $S \in \Sigma^*$ è detto stringa, la cui lunghezza è denotata da $|S|$
3. Per ogni i, $1 \leq i \leq n = |S|$, S[i] indica l'i-esimo carattere in S.
4. Per ogni i, j, $1 \leq i \leq j \leq n = |S|$, S[i, j] indica la sottostringa compresa tra le posizioni i e j,

# Periodo ed esponente

**Definizione:** Data S una stringa, si definisce *periodo* un intero $p \geq 1$ tale che S[i] = S[i + p], $\forall i = 1, \ldots, |S| - p$

**Definizione:** Si definisce *esponente* exp il rapporto tra la lunghezza di S e del suo più piccolo periodo.

# Ripetizioni massimali

**Definizione:** Una coppia (i, j) è detta ripetizione massimale (o run) di una stringa S, se $\exp(S[i,j]) \geq 2$ e la periodicità non può essere estesa nè a destra ne a sinistra.

# Un esempio

Sia considerata la stringa $S = babbabbababbabbabc$; questa contiene nove run, mostrati in *Figura 1*.

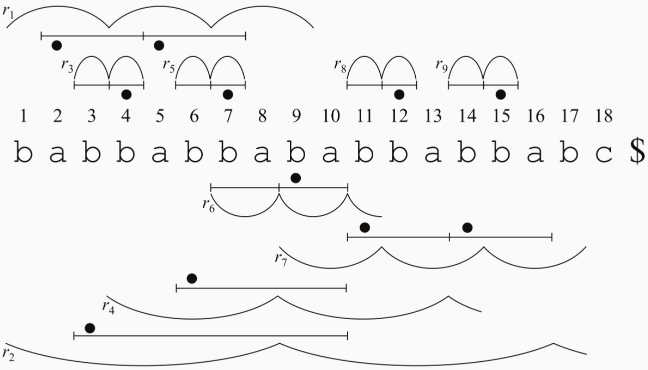


Figura 1: Illustrazione dei run nella stringa S = babbabbababbabbabc.

# Background storico

Sia $\rho(S)$ il numero dei run in una stringa S.

*Kolpakov e Kucherov*[1] dimostrano che $\rho(S) = \mathcal{O}(|S|)$.

Congettura (Runs conjecture): $\rho(S) < |S|$ per ogni stringa S.

[1] R. M. KUCHEROV AND G. KOLPAKOV, *Finding maximal repetitions in a word in linear time. FOCS, 1999*

Si puntualizza che la congettura è stata dimostrata per la prima volta da Bannai et al.[2]

La dimostrazione di Bannai et al. risulta essere complessa, per tale ragione sarà analizzata la versione proposta da Crochemore e Mercas.[3]

---

[2] HIDEO BANNAI, TOMOHIRO I, SHUNSUKE INENAGA, YUTO NAKASHIMA, MASAYUKI TAKEDA AND KAZUYA TSURATA, *The "Runs" Theorem. SODA, 2015*

[3] MAXIME CROCHEMORE AND ROBERT MERCAS, *On density of Lyndor roots in factors. Theorical computer science, 2016*

# Lyndon words

Data S una stringa e i un indice, con $1 \leq i \leq n$, la stringa S[i,n]S[1,i - 1] è detta *shift ciclico* di S (se i > 1 *shift ciclico non banale*).

**Definizione (Lyndon word):** Fissato un certo ordine, una stringa $S \in \Sigma^*, S \neq \varepsilon$, è detta Lyndon word se S è minore di ogni suo shift ciclico non banale.

Assunto $\prec_0$ ordine totale su $\Sigma$, si può definire $\prec_1$ l'ordine rovesciato. Ossia, un'ordine tale per cui presi $a, b \in \Sigma$, se

$$a \prec_0 b \implies b \prec_1 a$$

# La “Runs” conjecture

**Teorema:** $\rho(S) < |S|$, per ogni stringa S.

**Dimostrazione:** Sia S una stringa di lunghezza n, e sia (i, j), $1 \leq i < j \leq n$, un run di periodo minimo p in S.

Se j + 1 < n, ed inoltre S[j - p + 1] $\prec_0$ S[j + 1], si assegna al run un indice k per cui S[k, j] è il *greatest proper suffix* di S[i, j]. Viceversa k definisce il più grande suffisso per $\prec_1$.

Si noti che se k > i allora k > 0, e S[k, j] contiene un intero periodo del run. Inoltre, S[k, k + p - 1] è il più grande coniugato di S[i ,i + p - 1] rispetto uno dei due ordini. Pertanto è *border-free*, che è una proprietà nota delle Lyndon word.

Si intende dunque dimostrare che ogni k > 0 è posizione iniziale di al più un solo *greatest proper suffix* in un run.

Siano quindi considerati due run: (i, j) e (l, m) rispettivamente un p-run e un q-run. Inoltre, sia assunto per assurdo che i loro greatest proper suffix condividano la stessa posizione iniziale k.

Sia assunto $p \neq q$ poichè i run non possono essere distinti è avere lo stesso periodo.

1. Sia $j = l$, ma allora $S[k, j] = S[k, m]$. Assumendo per esempio $p < q$. Allora, $S[k, k + q - 1]$ ha periodo $p$, e quindi questa non è border-free, che è una contraddizione.
2. Sia asunto sensa perdita di generalità che $j < m$, e che i suffissi siano i più grandi rispetto un stesso ordine nei rispettivi run, sia questi $\prec_0$. Sia $d = S[j + 1]$, il carattere successivo il p-run.
   Per definizione si ha che $S[j - p + 1] \prec d$ e di conseguenza $S[k, k - p + 1] \prec S[k, j - p]d$. Ma poichè $S[i + p, j]$ è fattore del q-run, ciò contraddice la massimalita di $S[k, m - 1]$.

3. Sia $j \neq m$ e i suffissi risultano i più grandi secondo ordini diversi. Senza perdita di generalità sia $p < q$ e il suffisso del p-run sia il più grande rispetto $\prec_0$. Poichè $q > 1$, si ha sia che $S[k + q - 1] \prec_0 S[k]$ sia che $S[k + q - 1] = S[k - 1]$, da cui $S[k - 1] \prec_0 S[k]$.
   Si ha però che $p \not> 1$, viceversa si avrebbe $S[k] \prec_0 S[k - 1]$. Inoltre, $p \neq 1$ poichè si avrebbe che $S[k - 1] = S[k]$. Da cui un'altra contraddizione.

Ciò conclude la dimostrazione, mostrando come il numero di run non possa essere superiore agli $n - 1$ potenziali valori di k, come supposto.