Around analytic inclusion-exclusion

Pierre Nicodème

CNRS, LIX - École polytechnique, INRIA - Amib

(joint work with Frédérique Bassino and Julien Clément)

12/04/2011

Inclusion-Exclusion principle: set-theoretical view

- **General set-up**

  $A \cup B = A + B - AB$

  By recurrence:

  $A_1 \cup \cdots \cup A_r = \sum_{1 \leq i \leq r} A_i - \sum_{1 \leq i_1 < i_2 \leq r} A_{i_1} A_{i_2} + \cdots + (-1)^r A_1 \ldots A_r$

Inclusion-Exclusion principle: set-theoretical view

- **General set-up**
  $A \cup B = A + B - AB$
  By recurrence:
  $A_1 \cup \cdots \cup A_r = \sum_{1 \leq i \leq r} A_i - \sum_{1 \leq i_1 < i_2 \leq r} A_{i_1} A_{i_2} + \cdots + (-1)^r A_1 \ldots A_r$
- **Derangements of $\mathfrak{S}_n$**, set $A_i = \overline{B}_i$, where
  - B_i set of permutations with **no** fixed point at position i
  - $\overline{B}_i$ set of permutations with **a** fixed point at position i.

Inclusion-Exclusion principle: set-theoretical view

- **General set-up**
  $A \cup B = A + B - AB$
  By recurrence:
  $A_1 \cup \cdots \cup A_r = \sum_{1 \leq i \leq r} A_i - \sum_{1 \leq i_1 < i_2 \leq r} A_{i_1} A_{i_2} + \cdots + (-1)^r A_1 \ldots A_r$
- **Derangements of $\mathfrak{S}_n$**, set $A_i = \overline{B}_i$, where
  - B_i set of permutations with **no** fixed point at position i
  - $\overline{B}_i$ set of permutations with **a** fixed point at position i.

  $\overline{B}_1 \cup \cdots \cup \overline{B}_r = \mathfrak{S}_n - B_1 B_2 \ldots B_r$
  $|B_1 B_2 \ldots B_r| =$
  $|\mathfrak{S}_n| - \sum_{1 \leq i \leq r} \left|\overline{B}_i\right| + \sum_{1 \leq i_1 < i_2 \leq r} \left|\overline{B}_{i_1} \overline{B}_{i_2}\right| + \cdots + (-1)^r \left|\overline{B}_1 \ldots \overline{B}_r\right|$

Inclusion-Exclusion principle: set-theoretical view

- **General set-up**
  $A \cup B = A + B - AB$
  By recurrence:
  $A_1 \cup \cdots \cup A_r = \sum_{1 \leq i \leq r} A_i - \sum_{1 \leq i_1 < i_2 \leq r} A_{i_1} A_{i_2} + \cdots + (-1)^r A_1 \ldots A_r$
- **Derangements of $\mathfrak{S}_n$**, set $A_i = \overline{B}_i$, where
  - B_i set of permutations with **no** fixed point at position i
  - $\overline{B}_i$ set of permutations with **a** fixed point at position i.

  $\overline{B}_1 \cup \cdots \cup \overline{B}_r = \mathfrak{S}_n - B_1 B_2 \ldots B_r$
  $|B_1 B_2 \ldots B_r| =$
  $|\mathfrak{S}_n| - \sum_{1 \leq i \leq r} |\overline{B}_i| + \sum_{1 \leq i_1 < i_2 \leq r} |\overline{B}_{i_1} \overline{B}_{i_2}| + \cdots + (-1)^r |\overline{B}_1 \ldots \overline{B}_r|$

  for $\overline{B}_{i_1} \overline{B}_{i_2} \ldots \overline{B}_{i_k}$ with $i_1 < i_2 < \cdots < i_k$
  - choices of indices: $\binom{n}{k}$
  - choices for other positions: $(n-k)!$

Inclusion-Exclusion principle: set-theoretical view

- **General set-up**

  $A \cup B = A + B - AB$

  By recurrence:

  $A_1 \cup \cdots \cup A_r = \sum_{1 \leq i \leq r} A_i - \sum_{1 \leq i_1 < i_2 \leq r} A_{i_1} A_{i_2} + \cdots + (-1)^r A_1 \ldots A_r$

- **Derangements of $\mathfrak{S}_n$**, set $A_i = \overline{B}_i$, where

  - B_i set of permutations with **no** fixed point at position i
  - $\overline{B}_i$ set of permutations with **a** fixed point at position i.

  $\overline{B}_1 \cup \cdots \cup \overline{B}_r = \mathfrak{S}_n - B_1 B_2 \ldots B_r$

  $|B_1 B_2 \ldots B_r| =$
  $|\mathfrak{S}_n| - \sum_{1 \leq i \leq r} \left|\overline{B}_i\right| + \sum_{1 \leq i_1 < i_2 \leq r} \left|\overline{B}_{i_1} \overline{B}_{i_2}\right| + \cdots + (-1)^r \left|\overline{B}_1 \ldots \overline{B}_r\right|$

  for $\overline{B}_{i_1} \overline{B}_{i_2} \ldots \overline{B}_{i_k}$ with $i_1 < i_2 < \cdots < i_k$

  - choices of indices: $\binom{n}{k}$
  - choices for other positions: $(n-k)!$

  $$D_n = |B_1 B_2 \ldots B_n| = n! - (n-1)! \binom{n}{1} + (n-2)! \binom{n}{2} + \cdots + (-1)^n 0! \binom{n}{n}$$

Inclusion-Exclusion principle: set-theoretical view

- **General set-up**

  $A \cup B = A + B - AB$

  By recurrence:

  $A_1 \cup \dots \cup A_r = \sum_{1 \le i \le r} A_i - \sum_{1 \le i_1 < i_2 \le r} A_{i_1} A_{i_2} + \dots + (-1)^r A_1 \dots A_r$

- **Derangements of $\mathfrak{S}_n$**, set $A_i = \overline{B}_i$, where
  - B_i set of permutations with **no** fixed point at position i
  - $\overline{B}_i$ set of permutations with **a** fixed point at position i.

  $\overline{B}_1 \cup \dots \cup \overline{B}_r = \mathfrak{S}_n - B_1 B_2 \dots B_r$

  $|B_1 B_2 \dots B_r| =$

  $|\mathfrak{S}_n| - \sum_{1 \le i \le r} |\overline{B}_i| + \sum_{1 \le i_1 < i_2 \le r} |\overline{B}_{i_1} \overline{B}_{i_2}| + \dots + (-1)^r |\overline{B}_1 \dots \overline{B}_r|$

  for $\overline{B}_{i_1} \overline{B}_{i_2} \dots \overline{B}_{i_k}$ with $i_1 < i_2 < \dots < i_k$
  - choices of indices: $\binom{n}{k}$
  - choices for other positions: $(n-k)!$

  $$D_n = |B_1 B_2 \dots B_n| = n! - (n-1)! \binom{n}{1} + (n-2)! \binom{n}{2} + \dots + (-1)^n 0! \binom{n}{n}$$

  $$\frac{D_n}{n!} = 1 - \frac{1}{2!} + \frac{1}{3!} + \dots + (-1)^n \frac{1}{n!}$$

Analytic Inclusion-Exclusion principle

Generating function point of view

- Set of *camelus genus* (camel and dromedary): each one is of size 1, the number of humps is counted by the formal variable u.

$\mathcal{P} = \{$ ,  $\}$, $\qquad P(u) = u + u^2$

Analytic Inclusion-Exclusion principle

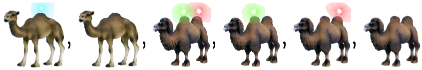
Generating function point of view

- Set of *camelus genus* (camel and dromedary): each one is of size 1, the number of humps is counted by the formal variable u.

$\mathcal{P} = \{$  $\}$, $\qquad P(u) = u + u^2$

- Distinguished set

$\mathcal{Q} =$ {"objects of $\mathcal{P}$ in which each elementary configuration (hump) is either distinguished or not"}

$= \{$  $\}$

$$Q(v) = v + 1 + v^2 + v + v + 1 = 2 + 3v + v^2$$
$$= P(1 + v)$$

Analytic Inclusion-Exclusion principle

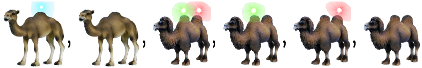
Generating function point of view

- Set of *camelus genus* (camel and dromedary): each one is of size 1, the number of humps is counted by the formal variable u.

$\mathcal{P} = \{$  $\}, \qquad P(u) = u + u^2$

- Distinguished set

$\mathcal{Q} = \{$"objects of $\mathcal{P}$ in which each elementary configuration (hump) is either distinguished or not"$\}$

$= \{$  $\}$

$$Q(v) = v + 1 + v^2 + v + v + 1 = 2 + 3v + v^2$$
$$= P(1 + v)$$

- **Inclusion-Exclusion principle**

  $Q(v)$ easy to get, gives $P(u) = Q(u - 1)$.

Goulden-Jackson book (1983)

Back to Derangements

$\mathcal{P}$: set of all permutations.

Given a permutation $(2, 5, 3, 4, 1) \in \mathfrak{S}_5$

consider a "super" set $\mathcal{Q}$ of "super" permutations where some fixed points are marked.

$(2, 5, 3, 4, 1) \rightsquigarrow \{(2, 5, 3, 4, 1), (2, 5, 3, 4, 1), (2, 5, 3, 4, 1), (2, 5, 3, 4, 1)\}$

Back to Derangements

$\mathcal{P}$: set of all permutations.

Given a permutation $(2, 5, 3, 4, 1) \in \mathfrak{S}_5$

consider a "super" set $\mathcal{Q}$ of "super" permutations where some fixed points are marked.

$(2, 5, 3, 4, 1) \rightsquigarrow \{(2, 5, 3, 4, 1), (2, 5, 3, 4, 1), (2, 5, 3, 4, 1), (2, 5, 3, 4, 1)\}$

- The marked fixed points form a **set** $\mathcal{S}$ of positions
- **removing** the **marked fixed points** leaves a permutation of $\mathcal{P}$

$$\mathcal{Q} \cong \mathcal{S} \star \mathcal{P} \quad \Longrightarrow \quad Q(z, v) = e^{zv} \frac{1}{1-z}$$

Back to Derangements

$\mathcal{P}$: set of all permutations.

Given a permutation $(2, 5, 3, 4, 1) \in \mathfrak{S}_5$

consider a "super" set $\mathcal{Q}$ of "super" permutations where some fixed points are marked.

$(2, 5, 3, 4, 1) \rightsquigarrow \{(2, 5, 3, 4, 1), (2, 5, 3, 4, 1), (2, 5, 3, 4, 1), (2, 5, 3, 4, 1)\}$

- The marked fixed points form a **set** $\mathcal{S}$ of positions
- **removing** the **marked fixed points** leaves a permutation of $\mathcal{P}$

$$\mathcal{Q} \cong \mathcal{S} \star \mathcal{P} \quad \Longrightarrow \quad Q(z, v) = e^{zv} \frac{1}{1-z}$$

Then

$$P(z, u) = Q(z, u-1) \quad \Longrightarrow \quad D_n = [z^n] Q(z, -1) = [z^n] \frac{e^{-z}}{1-z}$$

Rises and ascending runs in permutations - Philippe's book

- **Rises** or ascending runs of length 1 (Eulerian numbers)

$$A(z,u) = \frac{u-1}{u-e^{z(u-1)}}$$

  - mean number for permutations of size n: $\frac{1}{2}(n-1)$
  - variance: $\sim \frac{1}{12}n$
- **Ascending runs**
  - mean number of ascending runs of length $\ell - 1$: $\frac{1}{\ell!}(n-l+1)$
- **Permutations without ℓ-ascending runs**

Goulden-Jackson book (1983), Elizalde, Noy, ...

Analytical approach to Word Counting

- Probabilistic methods [Prum, Rodolphe, de Turkheim 95], [Schbath 97], [Apostolico, Bock, Xuyan 98], [Reinert, Schbath, Waterman 00], ...

Analytical approach to Word Counting

- Probabilistic methods [Prum, Rodolphe, de Turkheim 95], [Schbath 97], [Apostolico, Bock, Xuyan 98], [Reinert, Schbath, Waterman 00], ...
- Combinatorial and algorithmic methods - Generating functions of probabilities [Goulden, Jackson 83], [Régnier, Szpankowski 98], [Noonan, Zeilberger 99], [Flajolet, P.N., Salvy 99], [Bassino, Clément, P.N. (to appear)]...

Analytical approach to Word Counting

- Probabilistic methods [Prum, Rodolphe, de Turkheim 95], [Schbath 97], [Apostolico, Bock, Xuyan 98], [Reinert, Schbath, Waterman 00], ...
- Combinatorial and algorithmic methods - Generating functions of probabilities [Goulden, Jackson 83], [Régnier, Szpankowski 98], [Noonan, Zeilberger 99], [Flajolet, P.N., Salvy 99], [Bassino, Clément, P.N. (to appear)]...
- Large deviations [Denise, Régnier 04]

Analytical approach to Word Counting

- Probabilistic methods [Prum, Rodolphe, de Turkheim 95], [Schbath 97], [Apostolico, Bock, Xuyan 98], [Reinert, Schbath, Waterman 00], ...
- Combinatorial and algorithmic methods - Generating functions of probabilities [Goulden, Jackson 83], [Régnier, Szpankowski 98], [Noonan, Zeilberger 99], [Flajolet, P.N., Salvy 99], [Bassino, Clément, P.N. (to appear)]...
- Large deviations [Denise, Régnier 04]

See also Lothaire vol.3 "Applied Combinatorics on Words" with a chapter by Reinert, Schbath, Waterman and another by Jacquet, Szpankowski.

Inclusion-Exclusion: one word

- A text $\mathcal{P} = abaaaabb$ and a pattern $\mathcal{U} = \{u = aaa\}$. Text with **all** occurrences marked:

  $$ab\,aa\overset{❶}{a}\overset{❶}{a}\,bb.$$

  $P(z,x) = \pi(a)^5\pi(b)^3z^8x^2$ (where x counts occurrences of u, and z the length of the text).

Inclusion-Exclusion: one word

- A text $\mathcal{P} = abaaaabb$ and a pattern $\mathcal{U} = \{u = aaa\}$. Text with **all** occurrences marked:

  $$ab\,aa\overset{❶}{a}\overset{❶}{a}\,bb.$$

  $P(z, x) = \pi(a)^5 \pi(b)^3 z^8 x^2$ (where x counts occurrences of u, and z the length of the text).

- Set of decorated texts (**some** occurrences marked)

$$\begin{aligned}
\mathrm{Q} &= \{ab\,aa\overset{❶}{a}\overset{❶}{a}\,bb,\ ab\,aa\overset{❶}{a}\,abb,\ aba\,aa\overset{❶}{a}\,bb,\ abaaaabb\} \\
Q(z, t) &= \textstyle\sum_{\mathrm{w} \in \mathrm{Q}} \pi(\mathrm{w}) z^{|\mathrm{w}|} t^{\#\text{distinguished occurrences}} \\
&= \pi(a)^5 \pi(b)^3 z^8 (t^2 + t + t + 1),
\end{aligned}$$

(where the variable t counts the distinguished occurrences).

$Q(z, t) = P(z, 1 + t)$ or $P(z, x) = Q(z, x - 1)$.

Inclusion-Exclusion: one word

- A text $\mathcal{P} = abaaaabb$ and a pattern $\mathcal{U} = \{u = aaa\}$. Text with **all** occurrences marked:

$$ab\,aa\overset{❶}{a}\overset{❶}{a}\,bb.$$

  $P(z,x) = \pi(a)^5\pi(b)^3z^8x^2$ (where x counts occurrences of u, and z the length of the text).

- Set of decorated texts (**some** occurrences marked)

$$\mathsf{Q} = \{ab\,aa\overset{❶}{a}\overset{❶}{a}\,bb,\ ab\,aa\overset{❶}{a}\,abb,\ aba\,aa\overset{❶}{a}\,bb,\ abaaaabb\}$$

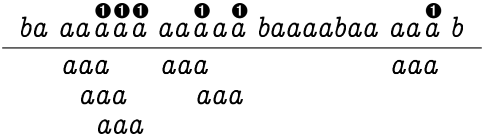
$$Q(z,t) = \sum_{\mathsf{w}\in\mathsf{Q}} \pi(\mathsf{w})z^{|\mathsf{w}|}t^{\#\text{distinguished occurrences}}$$
$$= \pi(a)^5\pi(b)^3z^8(t^2 + t + t + 1),$$

  (where the variable t counts the distinguished occurrences).
  $Q(z,t) = P(z, 1+t)$ or $P(z,x) = Q(z, x-1)$.

We need to compute the generating function of decorated texts!!!

Combinatorial description of decorated texts

Consider the text $w = baaaaaaaaaabaaaabaaaaab$, the pattern $\mathcal{U} = \{aaa\}$ and a particular decorated text



Definition (Cluster)

A cluster c with respect to a pattern $\mathcal{U}$ is a decorated text such that

- all positions are covered by at least a distinguished occurrence,
- and, either there is only one distinguished occurrence, or any distinguished occurrence has an overlap with another distinguished occurrence.

Combinatorial description of decorated texts

Consider the text $w = baaaaaaaaaabaaaabaaaab$, the pattern $\mathcal{U} = \{aaa\}$ and a particular decorated text

	c_1	c_2		c_3	
	❶❶❶	❶ ❶		❶	
ba	$aaaaa$	$aaaaa$	$baaaabaa$	aaa	b
	aaa	aaa		aaa	
	aaa	aaa			
	aaa				

Definition (Cluster)

A cluster c with respect to a pattern $\mathcal{U}$ is a decorated text such that

- all positions are covered by at least a distinguished occurrence,
- and, either there is only one distinguished occurrence, or any distinguished occurrence has an overlap with another distinguished occurrence.

Combinatorial description of decorated texts

The set of decorated texts T decomposes as sequences of either arbitrary letters of the alphabet $\mathcal{A}$ or clusters,

$$\mathsf{T} = (\mathcal{A} + \mathsf{C})^\star.$$

Combinatorial description of decorated texts

The set of decorated texts T decomposes as sequences of either arbitrary letters of the alphabet $\mathcal{A}$ or clusters,

$$\mathsf{T} = (\mathcal{A} + \mathsf{C})^{\star}.$$

Now, let us assume that we know how to compute the generating function $\xi(z, \mathbf{t})$ of the set of clusters C,

$$\xi(z, \mathbf{t}) = \sum_{\mathsf{w} \in \mathsf{C}} \pi(\mathsf{w}) z^{|\mathsf{w}|} \mathbf{t}^{\boldsymbol{\tau}(\mathsf{w})}, \text{ where } \boldsymbol{\tau}(\mathsf{w}) = (|\mathsf{w}|_1, \ldots, |\mathsf{w}|_r) \text{ ("type").}$$

Combinatorial description of decorated texts

The set of decorated texts T decomposes as sequences of either arbitrary letters of the alphabet $\mathcal{A}$ or clusters,

$$\mathsf{T} = (\mathcal{A} + \mathsf{C})^\star.$$

Now, let us assume that we know how to compute the generating function $\xi(z, \mathbf{t})$ of the set of clusters C,

$$\xi(z, \mathbf{t}) = \sum_{\mathsf{w} \in \mathsf{C}} \pi(\mathsf{w}) z^{|\mathsf{w}|} \mathbf{t}^{\boldsymbol{\tau}(\mathsf{w})}, \text{ where } \boldsymbol{\tau}(\mathsf{w}) = (|\mathsf{w}|_1, \ldots, |\mathsf{w}|_r) \text{ ("type")}.$$

From general principles the g.f. $T(z, \mathbf{t})$ of all decorated texts is

$$T(z, \mathbf{t}) = \frac{1}{1 - A(z) - \xi(z, \mathbf{t})}.$$

and the sought generating function is

$$F_{\mathcal{U}}(z, \mathbf{x}) = \frac{1}{1 - A(z) - \xi(z, \mathbf{x} - \mathbf{1})}.$$

Clusters: the simple case of one word

Take $\mathcal{U} = \{aaa\}$, the set of clusters is

$$\mathsf{C} = aa\overset{❶}{a} \cdot \left(\overset{❶}{a} + a\overset{❶}{a}\right)^{\star}.$$

The bivariate generating function $\xi(z, t)$ of C is obtained from this expression by counting the distinguished occurrences, *i.e.*, symbols ❶, with the variable t.

$$\xi(z, t) = \frac{t\pi(a)^3 z^3}{1 - t(\pi(a)z + \pi(a)^2 z^2)},$$

where t counts the number of distinguished occurrences.

Clusters: the simple case of one word

Take $\mathcal{U} = \{aaa\}$, the set of clusters is

$$\mathsf{C} = aa\overset{❶}{a} \cdot \left(\overset{❶}{a} + a\overset{❶}{a}\right)^{\star}.$$

The bivariate generating function $\xi(z, t)$ of C is obtained from this expression by counting the distinguished occurrences, *i.e.*, symbols ❶, with the variable t.

$$\xi(z, t) = \frac{t\pi(a)^3 z^3}{1 - t(\pi(a)z + \pi(a)^2 z^2)},$$

where t counts the number of distinguished occurrences.
Then, posing $\pi(a) = \pi(b) = 1$ (to get the enumerative generating function), we obtain

$$F(z, x) = \frac{1}{1 - A(z) - \xi(z, x-1)} = \frac{1}{1 - 2z - \dfrac{(x-1)z^3}{1 - (x-1)(z + z^2)}}.$$

Patterns as set of words

- **Reduced pattern**: **no word** of the pattern is **factor** of another word of the pattern

$$\mathcal{U} = \{baaab, aaaaa, aabb\}$$

- **Non-reduced patterns** (general case): no conditions

$$\mathcal{U} = \{ba\,aa\,b, aaaaa, aa, ba\}$$

Avoiding an “infinite” pattern - Zeilberger (2000)

- pattern $\mathcal{P} = \sum_{m \geq 0} ab^{m+1}c^{m+1}a = \{abca, abbcca, \ldots,\}$

Avoiding an "infinite" pattern - Zeilberger (2000)

- pattern $\mathcal{P} = \sum_{m \geq 0} ab^{m+1}c^{m+1}a = \{abca, abbcca, \ldots,\}$
- autocorrelation pattern $\mathcal{C} = \sum_{k \geq 0} b^{k+1}c^{k+1}a$

Avoiding an "infinite" pattern - Zeilberger (2000)

- pattern $\mathcal{P} = \sum_{m \geq 0} ab^{m+1}c^{m+1}a = \{abca, abbcca, \ldots, \}$
- autocorrelation pattern $\mathcal{C} = \sum_{k \geq 0} b^{k+1}c^{k+1}a$
- counting generating functions
  - $P(z) = \dfrac{z^4}{1-z^2}, \qquad C(z) = \dfrac{z^3}{1-z^2}$
  - $\xi(z,t) = \dfrac{tP(z)}{1-tC(z)}, \qquad F(z,x) = \dfrac{1}{1-3z-\frac{(x-1)P(z)}{1-(x-1)C(z)}}$

Avoiding an "infinite" pattern - Zeilberger (2000)

- pattern $\mathcal{P} = \sum_{m \geq 0} ab^{m+1}c^{m+1}a = \{abca, abbcca, \ldots,\}$
- autocorrelation pattern $\mathcal{C} = \sum_{k \geq 0} b^{k+1}c^{k+1}a$
- counting generating functions
  - $P(z) = \dfrac{z^4}{1-z^2}, \qquad C(z) = \dfrac{z^3}{1-z^2}$
  - $\xi(z,t) = \dfrac{tP(z)}{1-tC(z)}, \qquad F(z,x) = \dfrac{1}{1-3z-\frac{(x-1)P(z)}{1-(x-1)C(z)}}$

We get

$$F(z,0) = \frac{1}{1-3z+\frac{P(z)}{1+C(z)}} = \frac{1-z^2+z^3}{1-3z-z^2+4z^3-2z^4}$$

$$= 1+3z+9z^2+27z^3+80z^4+237z^5+701z^6+2074z^7+6135z^8+\ldots$$

Self-Avoiding walks (finite memory) - Noonan (1998)

- nearest neighbours walks on the lattice $\mathbb{Z}^d$
- **loop** of a walk: subsequence of the walk with common initial and end point
- $c_d(n)$ number of self-avoiding n-steps walks (no loops)
- $\overline{c}_d(n,k)$ number of n-steps walks with **no loops of length** $\leq k$
- By construction, $c_d(n) \leq \overline{c}_d(n,k)$

Self-Avoiding walks (finite memory) - Noonan (1998)

- nearest neighbours walks on the lattice $\mathbb{Z}^d$
- **loop** of a walk: subsequence of the walk with common initial and end point
- $c_d(n)$ number of self-avoiding n-steps walks (no loops)
- $\overline{c}_d(n,k)$ number of n-steps walks with **no loops of length** $\leq k$
- By construction, $c_d(n) \leq \overline{c}_d(n,k)$

Connectivity constant for self avoiding walks μ_d

$\overline{c}_d(m+n) \leq \overline{c}_d(m)\overline{c}_d(n) \Longrightarrow \mu_d < \lim_{n\to\infty}(\overline{c}_d(n))^{1/n}$ (Fekete lemma)

Noonan (1998)	$\mu_2 < 2.6939$
Pönitz and Tittman (2000)	$\mu_2 < 2.6792$ (**record?**)

Loop and mistakes

- ***k*-mistake**: a loop of size at most k that **contains no inner loop**
- **Steps** $= (+1, -1, +2, -2, \ldots, +d, -d)$, where $(+i, -i)$ stands for a $(+1, -1)$ increment of the ith coordinate.
- $(+1, -2, +2, -1)$ is not a mistake
- $(+1, -2, -1, +2)$ is a mistake

Loop and mistakes

- k**-mistake**: a loop of size at most k that **contains no inner loop**
- **Steps** $= (+1, -1, +2, -2, \dots, +d, -d)$, where $(+i, -i)$ stands for a $(+1, -1)$ increment of the ith coordinate.
- $(+1, -2, +2, -1)$ is not a mistake
- $(+1, -2, -1, +2)$ is a mistake

Method

- build clusters of mistakes
- use inclusion-exclusion to get the generating function of walks without k-mistakes

Remark: by construction the set of k-mistakes is a **finite reduced set**.

Equivalent mistakes

$\mathfrak{S}_d^{(s)}$: set of signed permutations of $\{\pm 1, \ldots, \pm d\}$

$\mathsf{m}_1 \equiv \mathsf{m}_2$ (m_1 and m_2 mistakes), iff

- $\exists\, \Pi^{(s)} \in \mathfrak{S}_d^{(s)}$ and $\mathsf{m}_2 = \Pi^{(s)}(\mathsf{m}_1)$
- equivalently, there is an **isometry** of $\mathbb{Z}^d$ mapping m_1 to m_2

Examples: $d = 3$

- $(1, -1) \equiv (-1, 1) \equiv (-3, 3)$
- $(1, -2, -1, 2) \equiv (2, -1, -2, 1)$

Equivalent mistakes

$\mathfrak{S}_d^{(s)}$: set of signed permutations of $\{\pm 1, \ldots, \pm d\}$

$\mathsf{m}_1 \equiv \mathsf{m}_2$ (m_1 and m_2 mistakes), iff

- $\exists\, \Pi^{(s)} \in \mathfrak{S}_d^{(s)}$ and $\mathsf{m}_2 = \Pi^{(s)}(\mathsf{m}_1)$
- equivalently, there is an **isometry** of $\mathbb{Z}^d$ mapping m_1 to m_2

Examples: $d = 3$

- $(1, -1) \equiv (-1, 1) \equiv (-3, 3)$
- $(1, -2, -1, 2) \equiv (2, -1, -2, 1)$

Property:

- $\overline{F}_{\mathsf{m}}(z)$: generating function of walks **avoiding** the **mistake** m
- $\mathsf{m}_1 \equiv \mathsf{m}_2 \implies \overline{F}_{\mathsf{m}_1}(z) = \overline{F}_{\mathsf{m}_2}(z)$ (by symmetry)

Canonical mistake of a class

- let $i \prec -i$
- sort lexicographically each class
- take the **first** mistake as **canonical** mistake

$$\begin{array}{rl} & +1,+2,-1,-2 \\ \prec & +1,-2,-1,+2 \\ \prec & +2,+1,-2,-1 \\ \prec & +2,-1,-2,+1 \\ \prec & -1,+2,+1,-2 \\ \prec & -1,-2,+1,+2 \\ \prec & -2,+1,+2,-1 \\ \prec & -2,-1,+2,-1 \end{array}$$

Canonical mistake of a class

- let $i \prec -i$
- sort lexicographically each class
- take the **first** mistake as **canonical** mistake

$$
\begin{array}{ll}
 & +1,+2,-1,-2 \\
\prec & +1,-2,-1,+2 \\
\prec & +2,+1,-2,-1 \\
\prec & +2,-1,-2,+1 \\
\prec & -1,+2,+1,-2 \\
\prec & -1,-2,+1,+2 \\
\prec & -2,+1,+2,-1 \\
\prec & -2,-1,+2,-1
\end{array}
$$

- $\mathscr{C}^{(d)}(\mathsf{m})$: class of equivalence of the mistake m in dimension d
- $|\mathscr{C}^{(2)}(+1,+2,-1,-2)| = 8$, $\quad |\mathscr{C}^{(2)}(+1,-1)| = 4$
- $|\mathscr{C}^{(d)}(+1,+2,-1,-2)| = 4d(d-1)$, $\quad |\mathscr{C}^{(d)}(+1,-1)| = 2d$

Clusters of mistakes and equations

$\Gamma_{\mathsf{m}}^{[d]}$ cluster of mistakes **finishing by** the mistake m $\qquad (|\mathsf{m}| \leq 4)$

$(+1, +2, -1, -2)$		$\Gamma^{[2]}_{(1,2,-1,-2)} = -z^4$
$\ldots\ (+1, +2, -1, -2)$ $(+2, -1, -2, +1)$ $(-1, -2, +1, +2)$ $(-2, +1, +2, -1)$ $(-2, -1, +2, +1)$	$\rightsquigarrow$	$-(z + z^2 + 2z^3)\Gamma^{[2]}_{(1,2,-1,-2)}$
$\ldots\ (+1, +2, -1, -2)$ $(-2, +2)$		$-z\Gamma^{[2]}_{(1,-1)}$

Clusters of mistakes and equations

$\Gamma_{\mathsf{m}}^{[d]}$ cluster of mistakes **finishing by** the mistake m $\quad (|\mathsf{m}| \leq 4)$

$(+1, +2, -1, -2)$

$\ldots\ (+1, +2, -1, -2)$
$(+2, -1, -2, +1)$
$(-1, -2, +1, +2)$
$(-2, +1, +2, -1)$
$(-2, -1, +2, +1)$

$\ldots\ (+1, +2, -1, -2)$
$(-2, +2)$

$\rightsquigarrow$

$\Gamma_{(1,2,-1,-2)}^{[2]} = -z^4$

$-(z+z^2+2z^3)\Gamma_{(1,2,-1,-2)}^{[2]}$

$-z\Gamma_{(1,-1)}^{[2]}$

$$\begin{cases} \Gamma_{(1,2,-1,-2)}^{[2]} = -z^4 - (z+z^2+2z^3)\Gamma_{(1,2,-1,-2)}^{[2]} - z\Gamma_{(1,-1)}^{[2]} \\ \Gamma_{(1,-1)}^{[2]} = -z^2 - z\Gamma_{(1,-1)}^{[2]} - 2z^3\Gamma_{(1,2,-1,-2)}^{[2]} \end{cases}$$

Clusters of mistakes and equations

$\Gamma^{[d]}_{\mathsf{m}}$ cluster of mistakes **finishing by** the mistake m $\qquad (|\mathsf{m}| \leq 4)$

$(+1, +2, -1, -2)$ $\qquad \Gamma^{[2]}_{(1,2,-1,-2)} = -z^4$

$\ldots (+1, +2, -1, -2)$
$(+2, -1, -2, +1)$
$(-1, -2, +1, +2)$ $\rightsquigarrow$ $-(z+z^2+2z^3)\Gamma^{[2]}_{(1,2,-1,-2)}$
$(-2, +1, +2, -1)$
$(-2, -1, +2, +1)$

$\ldots (+1, +2, -1, -2)$
$(-2, +2)$ $\qquad -z\Gamma^{[2]}_{(1,-1)}$

$$\begin{cases} \Gamma^{[2]}_{(1,2,-1,-2)} = -z^4 - (z+z^2+2z^3)\Gamma^{[2]}_{(1,2,-1,-2)} - z\Gamma^{[2]}_{(1,-1)} \\ \Gamma^{[2]}_{(1,-1)} = -z^2 - z\Gamma^{[2]}_{(1,-1)} - 2z^3\Gamma^{[2]}_{(1,2,-1,-2)} \end{cases}$$

$$F^{(2)}{}_4(z) = \sum_{n\geq 0} \overline{c}_2(n,4)z^n = \frac{1}{1 - 4z - 8\Gamma^{[2]}_{(1,2,-1,-2)} - 4\Gamma^{[2]}_{(1,-1)}}$$

Clusters of mistakes and equations

$\Gamma_{\mathsf{m}}^{[d]}$ cluster of mistakes **finishing by** the mistake m $\qquad (|\mathsf{m}| \leq 4)$

$$\begin{cases} \Gamma^{[2]}_{(1,2,-1,-2)} & = -z^4 - (z+z^2+2z^3)\Gamma^{[2]}_{(1,2,-1,-2)} - z\Gamma^{[2]}_{(1,-1)} \\ \Gamma^{[2]}_{(1,-1)} & = -z^2 + z\Gamma^{[2]}_{(1,-1)} - 2z^3\Gamma^{[2]}_{(1,2,-1,-2)} \end{cases}$$

$$\begin{cases} \Gamma^{[d]}_{(1,2,-1,-2)} & = -z^4 - (z+z^2+2(d-1)z^3)\Gamma^{[d]}_{(1,2,-1,-2)} - z\Gamma^{[d]}_{(1,-1)} \\ \Gamma^{[d]}_{(1,-1)} & = -z^2 - z\Gamma^{[d]}_{(1,-1)} - 2(d-1)z^3\Gamma^{[d]}_{(1,2,-1,-2)} \end{cases}$$

$$F^{(d)}{}_4(z) = \frac{1}{1 - 2dz - 4d(d-1)\Gamma[d]_{(1,2,-1,-2)} - 2d\Gamma^{[d]}_{(1,-1)}}$$

Clusters of mistakes and equations

$\Gamma_{\mathsf{m}}^{[d]}$ cluster of mistakes **finishing by** the mistake m $\qquad (|\mathsf{m}| \leq 4)$

$$\begin{cases} \Gamma^{[2]}_{(1,2,-1,-2)} & = -z^4 - (z+z^2+2z^3)\Gamma^{[2]}_{(1,2,-1,-2)} - z\Gamma^{[2]}_{(1,-1)} \\ \Gamma^{[2]}_{(1,-1)} & = -z^2 + z\Gamma^{[2]}_{(1,-1)} - 2z^3\Gamma^{[2]}_{(1,2,-1,-2)} \end{cases}$$

$$\begin{cases} \Gamma^{[d]}_{(1,2,-1,-2)} & = -z^4 - (z+z^2+2(d-1)z^3)\Gamma^{[d]}_{(1,2,-1,-2)} - z\Gamma^{[d]}_{(1,-1)} \\ \Gamma^{[d]}_{(1,-1)} & = -z^2 - z\Gamma^{[d]}_{(1,-1)} - 2(d-1)z^3\Gamma^{[d]}_{(1,2,-1,-2)} \end{cases}$$

$$F^{(d)}{}_4(z) = \frac{1}{1 - 2dz - 4d(d-1)\Gamma[d]_{(1,2,-1,-2)} - 2d\Gamma^{[d]}_{(1,-1)}}$$

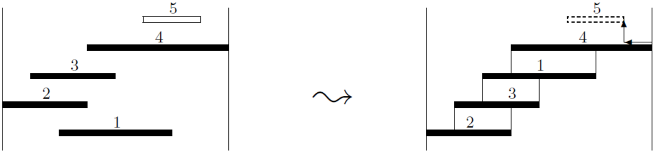
$$\overline{\mu}_k^{(d)} = \lim_{n\to\infty} \frac{\overline{c}^{(d)}(n,k)}{\overline{c}^{(d)}(n-1,k)}, \qquad \overline{c}^{(d)}(n,k) = [z^n]F^{(d)}{}_k(z)$$

General finite pattern - Reduced versus non-reduced

- Reduced case:

double staircase property

- Non-reduced case (first considered by Noonan-Zeilberger 1999)



some occurrences are factors of others

Combinatorial description of clusters

Skeleton of a cluster

Skeletization: remove factors occurrences. (The result is unique)

Combinatorial description of clusters

Skeleton of a cluster

Skeletization: remove factors occurrences. (The result is unique)

Example

Let us consider the pattern $\mathcal{U} = \{u_1 = ab, u_2 = ba, u_3 = baba\}$ and the clusters

$$\mathrm{c}_1 = abababa, \quad \mathrm{c}_2 = abababa, \quad \mathrm{c}_3 = abababa$$

c_1	c_2	c_3
ab $\;$ $baba$	ab $\;$ ba	ab $\;$ ba
$baba$	$baba$	ba
ab	$baba$	ab
	ba	$baba$

Combinatorial description of clusters

Skeleton of a cluster

Skeletization: remove factors occurrences. (The result is unique)

Example

Let us consider the pattern $\mathcal{U} = \{u_1 = ab, u_2 = ba, u_3 = baba\}$ and the clusters

$$
\mathsf{c}_1 = a\overset{1}{b}a\overset{1}{b}\overset{3}{a}b\overset{3}{a}, \quad
\mathsf{c}_2 = a\overset{1}{b}ab\overset{2,3}{a}b\overset{2,3}{a}, \quad
\mathsf{c}_3 = a\overset{1}{b}\overset{2}{a}\overset{1}{b}\overset{2}{a}b\overset{2,3}{a}
$$

$$
\begin{array}{lll}
ab \;\; baba & ab \;\; ba & ab \;\; ba \\
\;\; baba & \;\; baba & \;\; ba \\
\;\;\; ab & \;\;\;\; baba & \;\;\; ab \\
 & \;\;\;\;\; ba & \;\;\;\; baba
\end{array}
$$

We have

$$
\mathrm{Skel}(\mathsf{c}_1) = \mathrm{Skel}(\mathsf{c}_2) = a\overset{1}{b}ab\overset{3}{a}b\overset{3}{a}, \quad \mathrm{Skel}(\mathsf{c}_3) = a\overset{1}{b}\overset{2}{a}\overset{1}{b}ab\overset{3}{a}.
$$

This example illustrates that two different clusters with same support (here $abababa$) can have different skeletons.

Dual “Flip” operation

The Flip of a skeleton gives the set of all decorated clusters having the same skeleton

How?

Dual “Flip” operation

The Flip of a skeleton gives the set of all decorated clusters having the same skeleton

How?

Consider the pattern $\mathcal{U} = \{{\color{red}u_1 = ab}, {\color{blue}u_2 = ba}, {\color{green}u_3 = baba}\}$ and the skeleton:

$$\underline{c} = a\overset{❶}{b}ab\overset{❸}{a}b\overset{❸}{a},$$

the set $\mathrm{Flip}(\underline{c})$ is the set of clusters having $\underline{c}$ as skeleton and can be identified to the following *bicolored decorated word*

$$\widetilde{\underline{c}} = \mathrm{Flip}(\underline{c}) = a\overset{❶}{b}\overset{②}{a}\overset{①}{b}\overset{\overset{②}{❸}}{a}\overset{①}{b}\overset{\overset{②}{❸}}{a}$$

Dual "Flip" operation

The Flip of a skeleton gives the set of all decorated clusters having the same skeleton

How?

Consider the pattern $\mathcal{U} = \{u_1 = ab, u_2 = ba, u_3 = baba\}$ and the skeleton:

$$\underline{c} = a\overset{❶}{b}ab\overset{❸}{a}b\overset{❸}{a},$$

the set $\mathrm{Flip}(\underline{c})$ is the set of clusters having $\underline{c}$ as skeleton and can be identified to the following *bicolored decorated word*

$$\widetilde{\underline{c}} = \mathrm{Flip}(\underline{c}) = a\overset{❶}{b}\overset{②}{a}\overset{①}{b}\overset{②}{\overset{❸}{a}}\overset{①}{b}\overset{②}{\overset{❸}{a}}$$

equivalent to $2^5 = 32$ decorated texts: each factor occurrence or ① (resp. ②) can be distinguished or not, becoming ❶ (resp. ❷) or nothing, without modifying the skeleton.

Integrity rule

Two **distinct** skeletons **cannot give rise** to the **same decorated text** (integrity rule).

$\mathcal{U} = \{aaa, aaaaaaa\}$

$$\underline{c}_1 = \begin{array}{l} aa\overset{❶}{a}aaaaaa\overset{❷}{a} \\ \hline aaa \\ \quad aaaaaaa \end{array} \Bigg| \implies \mathrm{Flip}(\underline{c}_1) = aa\overset{❶}{a} \cdot a\overset{①}{a}\overset{①}{a}\overset{①}{a}\overset{①}{a}\overset{❷}{\overset{①}{a}}. \qquad (1)$$

Integrity rule

Two **distinct** skeletons **cannot give rise** to the **same decorated text** (integrity rule).

$\mathcal{U} = \{aaa, aaaaaaa\}$

$$\underline{c}_1 = \begin{array}{l} aa\overset{❶}{a}aaaaa\overset{❷}{a} \\ \hline aaa \\ \quad aaaaaaa \end{array} \Bigg| \implies \mathrm{Flip}(\underline{c}_1) = aa\overset{❶}{a} \cdot a\overset{①}{a}\overset{①}{a}\overset{①}{a}\overset{①}{a}\overset{❷}{\overset{①}{a}}. \qquad (1)$$

the fourth position has no label ① signaling a factor occurrence aaa; considering a **factor** occurrence aaa at this position would break the integrity rule and correspond to a skeleton $\underline{c}_2$

$$\underline{c}_2 = \begin{array}{l} aa\overset{❶}{a}\overset{❶}{a}aaaa\overset{❷}{a} \\ \hline aaa \\ \; aaa \\ \quad aaaaaaa \end{array} \Bigg| \implies \mathrm{Flip}(\underline{c}_2) = aa\overset{❶}{a} \cdot \overset{❶}{a} \cdot \overset{①}{a}\overset{①}{a}\overset{①}{a}\overset{①}{a}\overset{❷}{\overset{①}{a}}. \qquad (2)$$

General strategy for clusters

Two steps

- Describe clusters with respect to their skeletons
- Reinject all possible factor occurrences (with the “Flip” operation)

We must ensure that all (decorated) clusters are generated exactly once !

First step: how do we extend a skeleton?

(Auto)-Correlation Set

- Auto-correlation
  $\mathcal{C}_{h,h} = \{\ w,\ h \cdot w = r \cdot h \text{ and } |w| < |h|\}, \quad ababa \rightsquigarrow \begin{array}{r|l} ababa & \\ \hline ababa & \varepsilon \\ aba & ba \\ a & baba \end{array}$
- Correlation set of two words

$$\mathcal{C}_{u,v} = \{\ w, \quad u \cdot w = r \cdot v \quad \text{and} \quad |w| < |v|\ \}$$

$$u = baba,\ v = abaaba \ \rightsquigarrow\ \mathcal{C}_{baba,abaaba} = \{aba, baaba\}$$

Problem: not rigorously defined in the non-reduced case!

First step: how do we extend a skeleton?

(Auto)-Correlation Set

- Auto-correlation
  $\mathcal{C}_{h,h} = \{\ w,\ h \cdot w = r \cdot h \text{ and } |w| < |h|\}$, $\quad ababa \rightsquigarrow \begin{array}{r|l} ababa & \\ \hline ababa & \varepsilon \\ aba & ba \\ a & baba \end{array}$

- Correlation set of two words

$$\mathcal{C}_{u,v} = \{\ w, \quad u \cdot w = r \cdot v \quad \text{and} \quad |w| < |v|\ \}$$

$$u = baba,\ v = abaaba \ \rightsquigarrow\ \mathcal{C}_{baba,abaaba} = \{aba, baaba\}$$

Problem: not rigorously defined in the non-reduced case!
The notion of **right extension set** of two words u and v is a generalization of the **correlation set** of two words but **differs** in that:

(i) overlapping not allowed to start at the beginning of u;
Ex: $a^3 \rightsquigarrow a^7$; $\quad$ a^4 is not in the right extension set of a^3 to a^7

(ii) extension has to add some letters to the right of u; (forbid ε)

First step: how do we extend a skeleton?

(Auto)-Correlation Set

- Auto-correlation
  $\mathcal{C}_{h,h} = \{\ w,\ h \cdot w = r \cdot h \text{ and } |w| < |h|\},\quad ababa \rightsquigarrow \dfrac{ababa|}{\begin{array}{r|l} ababa & \varepsilon \\ aba & ba \\ a & baba \end{array}}$

- Correlation set of two words

$$\mathcal{C}_{u,v} = \{\ w,\quad u \cdot w = r \cdot v \quad \text{and} \quad |w| < |v|\ \}$$

$$u = baba,\ v = abaaba \rightsquigarrow \mathcal{C}_{baba,abaaba} = \{aba, baaba\}$$

Problem: not rigorously defined in the non-reduced case!
The notion of **right extension set** of two words u and v is a generalization of the **correlation set** of two words but **differs** in that:

(i) overlapping not allowed to start at the beginning of u;
Ex: $a^3 \rightsquigarrow a^7$; $\quad a^4$ is not in the right extension set of a^3 to a^7

(ii) extension has to add some letters to the right of u; (forbid ε)

These two conditions prevent from considering factor occurrences.

First step: how do we extend a skeleton?

(Auto)-Correlation Set

- Auto-correlation
  $\mathcal{C}_{h,h} = \{\ w,\ h \cdot w = r \cdot h \text{ and } |w| < |h|\}$, $\quad ababa \rightsquigarrow$ $\begin{array}{r|l} ababa & \\ \hline ababa & \varepsilon \\ aba & ba \\ a & baba \end{array}$

- Correlation set of two words

$$\mathcal{C}_{u,v} = \{\ w, \quad u \cdot w = r \cdot v \quad \text{and} \quad |w| < |v|\ \}$$

$$u = baba,\ v = abaaba \rightsquigarrow \mathcal{C}_{baba,abaaba} = \{aba, baaba\}$$

Problem: not rigorously defined in the non-reduced case!
The notion of **right extension set** of two words u and v is a generalization of the **correlation set** of two words but **differs** in that:

(i) overlapping not allowed to start at the beginning of u;
Ex: $a^3 \rightsquigarrow a^7$; $\quad$ a^4 is not in the right extension set of a^3 to a^7

(ii) extension has to add some letters to the right of u; (forbid ε)

These two conditions prevent from considering factor occurrences.

To **extend** a **skeleton**: **start** from a word of $\mathcal{U}$ and **iteratively concatenate** a word of the **right extension set**.

Second step: factor occurrences

Factor occurrences must not change the skeleton and must be considered within the last occurrence constituting the skeleton.
This is simply done by considering bicolored versions of right extensions.

For $\mathcal{U} = \{ab, aba\}$, we have $\mathcal{E} = \begin{pmatrix} \emptyset & \emptyset \\ b & ba \end{pmatrix}$.

$$\mathrm{u}_1 = a\overset{\text{❶}}{b}, \quad \mathrm{u}_2 = ab\overset{\text{❷}}{a}, \quad \mathrm{Flip}(\mathrm{u}_1) = \{a\overset{\text{❶}}{b}\} \text{ and } \mathrm{Flip}(\mathrm{u}_2) = \{a\overset{\text{①}}{b}\overset{\text{❷}}{a}\}$$

The decorated right extension matrix verifies

$$\mathsf{E} = \begin{pmatrix} \emptyset & \emptyset \\ \{\overset{\text{❶}}{b}\} & \{\overset{\text{①}}{b}\overset{\text{❷}}{a}\} \end{pmatrix}.$$

Second step: factor occurrences

Factor occurrences must not change the skeleton and must be considered within the last occurrence constituting the skeleton.
This is simply done by considering bicolored versions of right extensions.

For $\mathcal{U} = \{ab, aba\}$, we have $\mathcal{E} = \begin{pmatrix} \emptyset & \emptyset \\ b & ba \end{pmatrix}$.

$$\mathrm{u}_1 = a\overset{\text{❶}}{b}, \quad \mathrm{u}_2 = ab\overset{\text{❷}}{a}, \quad \mathrm{Flip}(\mathrm{u}_1) = \{a\overset{\text{❶}}{b}\} \text{ and } \mathrm{Flip}(\mathrm{u}_2) = \{a\,b\overset{\text{①❷}}{a}\}$$

The decorated right extension matrix verifies

$$\mathsf{E} = \begin{pmatrix} \emptyset & \emptyset \\ \{\overset{\text{❶}}{b}\} & \{b\overset{\text{①❷}}{a}\} \end{pmatrix}.$$

Description for the set of clusters

$$\mathsf{C} = (\mathrm{Flip}(\mathrm{u}_1), \ldots, \mathrm{Flip}(\mathrm{u}_r)) \cdot \mathsf{E}^\star \cdot \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon \end{pmatrix}.$$

From decorated text to generating function

Essence of the symbolic method:

- symbols α of the alphabet $\mapsto \pi(\alpha) z$ (commutative weight)
- ❶, ❷, ❸, $\ldots \mapsto t_1, t_2, t_3, \ldots$
- ①, ②, ③, $\ldots \mapsto (1 + t_1), (1 + t_2), (1 + t_3), \ldots$

The translation gives

$$\xi(z, \mathbf{t}) = (U_1(z, \mathbf{t}), \ldots, U_r(z, \mathbf{t})) \cdot \Big(\mathbb{I} - \mathbb{E}(z, \mathbf{t})\Big)^{-1} \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

From decorated text to generating function

Essence of the symbolic method:

- symbols α of the alphabet $\mapsto \pi(\alpha) z$ (commutative weight)
- ❶, ❷, ❸, ... $\mapsto t_1, t_2, t_3, \ldots$
- ①, ②, ③, ... $\mapsto (1+t_1), (1+t_2), (1+t_3), \ldots$

The translation gives

$$\xi(z,\mathbf{t}) = (U_1(z,\mathbf{t}), \ldots, U_r(z,\mathbf{t})) \cdot \Big(\mathbb{I} - \mathbb{E}(z,\mathbf{t})\Big)^{-1} \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

For instance, taking $(u_1, u_2) = (ab, aba)$

$$\mathrm{Flip}(\mathrm{u}_1) = \{a\overset{❶}{b}\} \mapsto \quad U_1(z, t_1, t_2) = z^2 t_1$$

$$\mathrm{Flip}(\mathrm{u}_2) = \{a\overset{①}{b}\overset{❷}{a}\} \mapsto \quad U_2(z, t_1, t_2) = z^3 t_2 (1 + t_1)$$

$$\mathsf{E} = \begin{pmatrix} \emptyset & \emptyset \\ \left\{\overset{❷}{a}\right\} & \left\{\overset{①}{b}\overset{❷}{a}\right\} \end{pmatrix} \mapsto \quad \mathbb{E}(z, t_1, t_2) = \begin{pmatrix} 0 & 0 \\ z t_2 & z^2 t_2 (1 + t_1) \end{pmatrix}.$$

Applications (typical formulas)

Proposition. Let $\mathcal{U} = \{u_1, \ldots, u_k\}$ be a pattern. The expected value and the variance of the variable X_n counting the number of occurrences of $\mathcal{U}$ in a random text of size n satisfy

$$\mathbf{E}[X_n] = \sum_{u \in \mathcal{U}} \pi(u)(n - |u| + 1),$$

$$\frac{1}{n}\mathbf{Var}[X_n] = \pi(\mathcal{U}) - \sum_{u,v \in \mathcal{U}} \pi(u)\pi(v)(|u| + |v| - 1)$$

$$+ 2 \sum_{u,v \in \mathcal{U}} \pi(u)\pi(\mathcal{E}_{u,v}) + 2 \sum_{\substack{u,v \in \mathcal{U} \\ u \neq v}} \pi(u)|u|_v + o(1).$$

Proposition. Let $\mathcal{U} = \{u_1, \ldots, u_k\}$ and $\mathcal{V} = \{v_1, \ldots, v_j\}$ be two patterns. The covariance of the variables X_n and Y_n counting respectively the number of occurrences of $\mathcal{U}$ and $\mathcal{V}$ in a random text of size n verifies

$$\frac{1}{n}\mathbf{Cov}(X_n, Y_n) = \pi(\mathcal{U} \cap \mathcal{V}) - \sum_{u \in \mathcal{U}, v \in \mathcal{V}} \pi(u)\pi(v)\big(|u| + |v| - 1\big)$$

$$+ \sum_{u \in \mathcal{U}, v \in \mathcal{V}} \Big(\pi(u)\pi(\mathcal{E}_{u,v}) + \pi(v)\pi(\mathcal{E}_{v,u})\Big) + \sum_{\substack{u \in \mathcal{U}, v \in \mathcal{V} \\ u \neq v}} \Big(|u|_v \pi(u) + |v|_u \pi(v)\Big) + o(1)$$

Example - Covariance Matrix for a^3 and a^7

$p = \Pr(a)$, X_n and Y_n respectively count the number of occurrences of a^3 and a^7 in a random text of size n.

$$\mathbb{B}_{11} = \lim_{n\to\infty} \frac{1}{n} \mathbf{Var}(X_n), \quad \mathbb{B}_{22} = \lim_{n\to\infty} \frac{1}{n} \mathbf{Var}(Y_n),$$

$$\mathbb{B}_{12} = \mathbb{B}_{21} = \lim_{n\to\infty} \frac{1}{n} \mathbf{Cov}(X_n, Y_n)$$

$$\mathbb{B}^{(a^3,a^7)} = \begin{pmatrix} p^3+2p^3(p+p^2)-5p^6 & p^7(5+2p+2p^2-9p^3) \\ p^7(5+2p+2p^2-9p^3) & p^7+2p^7(p+p^2+p^3+p^4+p^5+p^6)-13p^{14} \end{pmatrix}$$

$$\begin{aligned}\Delta(p) &= \left|\mathbb{B}^{(a^3,a^7)}\right| \\ &= p^{10}+4p^{11}+8p^{12}+5p^{13}-25p^{14}-20p^{15}-24p^{16}+67p^{17}-16p^{20}\end{aligned}$$

$\Delta(1) = 0$ **Ouf! Degeneracy of the system**

Conclusion & Perspectives

- The inclusion-exclusion method gives the multivariate generating function of occurrences for a (arbitrary) finite set of words
  - main parameter is the number of words
  - use explicit relations between words (right extension sets) which can be built efficiently with the Aho-Corasick algorithm.
- An alternative exists using the Aho-Corasick automaton to compute the generating functions
  - main parameter is the number of states.
  - relations between words hidden in the automaton

Conclusion & Perspectives

- The inclusion-exclusion method gives the multivariate generating function of occurrences for a (arbitrary) finite set of words
  - main parameter is the number of words
  - use explicit relations between words (right extension sets) which can be built efficiently with the Aho-Corasick algorithm.
- An alternative exists using the Aho-Corasick automaton to compute the generating functions
  - main parameter is the number of states.
  - relations between words hidden in the automaton
- We would like to prove in this context that “most of the times” a multivariate normal distribution holds
- extends to a Markovian model or dynamical sources

Complexity

For a set $\mathcal{U} = \{u_1, \ldots, u_r\}$, $r = \mathrm{Card}(\mathcal{U})$.

- For the inclusion-exclusion approach, we need to compute the quasi-inverse of a $r \times r$ matrix with entries which are polynomials of degrees at most $\mathtt{max}_i(|u_i|)$ in any variables.

Complexity

For a set $\mathcal{U} = \{u_1, \ldots, u_r\}$, $r = \mathrm{Card}(\mathcal{U})$.

- For the inclusion-exclusion approach, we need to compute the quasi-inverse of a $r \times r$ matrix with entries which are polynomials of degrees at most $\mathtt{max}_i(|u_i|)$ in any variables.
- The Aho-Corasick automaton approach considers the quasi inverse of a matrix of size N^2 (where $N = O(\sum_i |u_i|)$ is the number of states of the automaton), but it is sparse and entries are monomials of degree at most one in any variables.

Aho-Corasick automaton

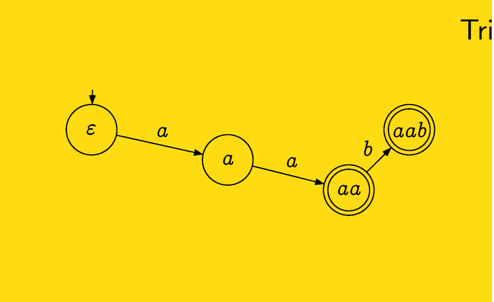
- **Input:** non-reduced set of words $\mathcal{U}$.
- **Output:** automaton $\mathcal{A}_{\mathcal{U}}$ recognizing $\mathcal{A}^*\mathcal{U}$.

Algorithm:

1. build $\mathcal{T}_{\mathcal{U}}$, the ordinary trie representing the set $\mathcal{U}$
2. build $\mathcal{A}_{\mathcal{U}} = (\mathcal{A}, Q, \delta, \varepsilon, T)$:
   - $Q = Pref(\mathcal{U})$
   - $T = \mathcal{A}^*\mathcal{U} \cap Pref(\mathcal{U})$
   - $\delta(q, x) = \lambda(qx)$
     where $\lambda(v)$ = the longest suffix of v which belongs to $Pref(\mathcal{U})$.

Aho-Corasick automaton (example)

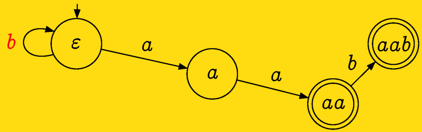
$\mathcal{U} = \{aa, aab\}$



Trie $\mathcal{T}_{\mathcal{U}}$ of $\mathcal{U}$

Aho-Corasick automaton (example)

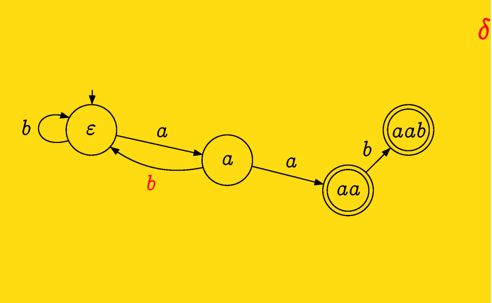
$\mathcal{U} = \{aa, aab\}$



$\delta(\varepsilon, b) = \lambda(b) = \varepsilon$

Aho-Corasick automaton (example)

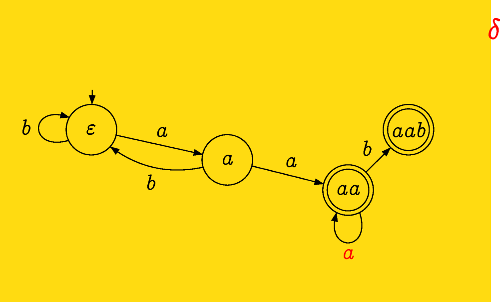
$\mathcal{U} = \{aa, aab\}$



$\delta(a, b) = \lambda(ab) = \varepsilon$

Aho-Corasick automaton (example)

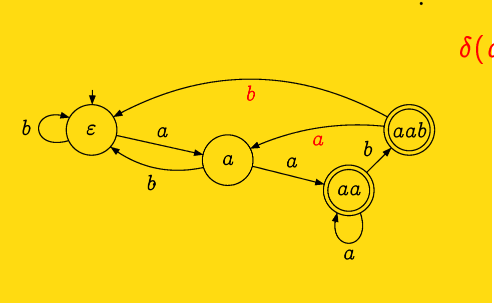
$\mathcal{U} = \{aa, aab\}$



$\delta(aa, a) = \lambda(aaa) = aa$

Aho-Corasick automaton (example)

$\mathcal{U} = \{aa, aab\}$

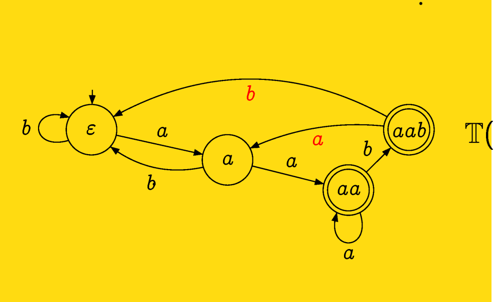


$\delta(aab, a) = \lambda(aaba) = a$

$\delta(aab, b) = \lambda(aabb) = \varepsilon$

Aho-Corasick automaton (example)

$\mathcal{U} = \{aa, aab\}$



$$\mathbb{T}(x_1, x_2) = \begin{pmatrix} b & a & 0 & 0 \\ b & 0 & a\,x_1 & 0 \\ 0 & 0 & a\,x_1 & b\,x_2 \\ b & a & 0 & 0 \end{pmatrix}$$

$$F(a, b, x_2, x_2) = (1, 0, 0, 0)(\mathbb{I} - \mathbb{T}(a, b, x_1, x_2))^{-1} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$= \frac{1 - a(x_1 - 1)}{1 - a x_1 - b + ab(x_1 - 1) - a^2 b x_1 (x_2 - 1)^2}.$$

An easy application - Bender and Kochman

Consider in a random text $\mathcal{T}_n$ of size n over $\mathcal{A} = \{a, b\}$ which **avoid** a word w
Question: **expectation** of **number** of letters a in $\mathcal{T}_n$

An easy application - Bender and Kochman

Consider in a random text $\mathcal{T}_n$ of size n over $\mathcal{A} = \{a, b\}$ which **avoid** a word w

Question: **expectation** of **number** of letters a in $\mathcal{T}_n$

Clusters:

- clusters $w.\mathcal{C}_w^\star$ of w:
- and clusters of a (limited to the word a; no overlapping)

An easy application - Bender and Kochman

Consider in a random text $\mathcal{T}_n$ of size n over $\mathcal{A} = \{a, b\}$ which **avoid** a word w

Question: **expectation** of **number** of letters a in $\mathcal{T}_n$

Clusters:

- clusters $w.\mathcal{C}_w^\star$ of w:
- and clusters of a (limited to the word a; no overlapping)

Generating function: (t_1, x_1 for w, and t_2, x_2 for a)

$$\xi(z, t_1, t_2) = \pi(w) z^{|w|} t_1 (1 + t_2)^{|w|_a} \times \frac{1}{1 - \mathcal{C}_w(z, 1 + t_2)} + \pi(a) t_2 z$$

$$F(z, x_1, x_2) = \frac{1}{1 - \xi(z, t_1 - 1, t_2 - 1)}$$

$$\Longrightarrow \mathbf{E}_n(X_{a,\overline{w}}) = [z^n] \left. \frac{\partial F(z, 0, x_2)}{\partial x_2} \right|_{x_2=1} \Big/ [z^n] F(z, 0, 1)$$

An easy application - Continued

$$\mathbf{E}_n(X_{a,\overline{w}}) = [z^n]\ \frac{\partial F(z,0,x_2)}{\partial x_2}\bigg|_{x_2=1} \Big/ [z^n]F(z,0,1)$$

$$K_{n,a,\overline{w}} = \frac{1}{n}\mathbf{E}_n(X_{a,\overline{w}})$$

Tuned distribution of letters a

$$\sum_{n\geq 1} K_{n,a,\overline{aaa}} z^n = .5z + .5z^2 + .4285714286z^3 + .4230769231z^4 + .4166666667z^5 + .4090909091z^6 + .4056437390z^7 + .4026845638z^8 \dots$$

$$\sum_{n\geq 1} K_{n,a,\overline{aab}} z^n = .5z + .5z^2 + .4761904762z^3 + .4583333333z^4 + .4400000000z^5 + .4242424242z^6 + .4100529101z^7 + .3977272727z^8 \dots$$