



HAL
open science

On counting untyped lambda terms

Pierre Lescanne

► **To cite this version:**

| Pierre Lescanne. On counting untyped lambda terms. 2011. ensl-00578527v8

HAL Id: ensl-00578527

<https://ens-lyon.hal.science/ensl-00578527v8>

Preprint submitted on 16 Feb 2012 (v8), last revised 5 Oct 2012 (v9)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On counting untyped lambda terms

Pierre Lescanne
University of Lyon, ENS de Lyon,
46 allée d'Italie, 69364 Lyon, France

Abstract

We present several results on counting untyped lambda terms, i.e., on telling how many terms belong to such or such class, according to the size of the terms and/or to the number of free variables.

1 Introduction

This paper presents several results on counting untyped lambda terms, i.e., on telling how many terms belong to such or such class, according to the size of the terms and/or to the number of free variables. In addition to the inherent interest of these results from the mathematical point of view, we expect that a knowledge on the distribution of terms will improve the implementation of reduction and that results on asymptotic distributions of terms will give a better insight of the lambda calculus. For counting more easily lambda terms we adopt de Bruijn indices that are a well-known coding of bound variables by natural numbers. First we give recurrence formulas for the number of terms (and of normal forms) of size n having m variables. These recurrence formulas are not familiar in combinatorics and not amenable to a classical treatment by generating functions. In a first part of the paper, we examine the formulas for terms and normal forms when n is fixed and m varies, which are polynomials. We give the expressions of the first coefficients of those polynomials since an expression for the generic coefficients seems out of reach and no regularity appears. However this shows that these expressions are clearly connected to Catalan numbers. If we would find an explicit expression for the last coefficients, this would be an explicit expression for the closed terms. In the second part, we give formulas for the generating functions showing the difficulty of a mathematical treatment. The results presented here are a milestone in describing probabilistic properties of lambda terms with answers to questions like: How does a random lambda term looks like? How does a random normal form look like? How to generate a random lambda term (a random normal form)?

Related works

Previous works on counting lambda terms were by O. Bodini et al. [2], R. David et al. [4] and J. Wang [13]. Related works are on counting types and/or counting tautologies [14, 9, 5, 10]. Complexity of rewriting was studied by Choppy et al. [3].

2 Untyped lambda terms with de Bruijn indices

I am dedicating this book to N. G. “Dick” de Bruijn, because his influence can be felt on every page. Ever since the 1960s he has been my chief mentor, the main person who would answer my question when I was stuck on a problem that I had not been taught how to solve.

Donald Knuth in preface of [11]

The λ -calculus [1] is a logic formalism to describe functions, for instance, the function $f \mapsto (x \mapsto f(f(x)))$, which takes a function f and applies it twice. For historical reason, this function is written $\lambda f.\lambda x.f(fx)$, which contains the two variables f and x , bounded by λ .

In this paper we represent terms by de Bruijn indices [7], this means that variables are represented by numbers $\underline{1}, \underline{2}, \dots, \underline{m}, \dots$, where an index, for instance \underline{k} , is the number of λ 's, above the location of the index and below the λ that binds the variable, in a representation of λ -terms by trees. For instance, the term with variables $\lambda x.\lambda y.x y$ is represented by the term with de Bruijn indices $\lambda \lambda \underline{2} \underline{1}$. The variable x is bound by the top λ . Above the occurrence of x , there are two λ 's, therefore x is represented by $\underline{2}$ and from the occurrence of y , we count just the λ that binds y ; so y is represented by $\underline{1}$. In what follows we will call *terms*, the untyped terms¹ with de Bruijn indices. Let us call $\mathcal{T}_{n,m}$, the set of terms of size n , with m de Bruijn indices, i.e., with indices in $\mathcal{I}(m) = \{\underline{1}, \underline{2}, \dots, \underline{m}\}$. A term in $\mathcal{T}_{n,m}$, is either a de Bruijn index or an abstraction on a term with $m+1$ indices i.e., a term in $\mathcal{T}_{n,m+1}$ or an application of a term in $\mathcal{T}_{n,m}$ on a term in $\mathcal{T}_{n,m}$. We can write, using $@$ as the application symbol,

$$\mathcal{T}_{n+1,m} = \lambda \mathcal{T}_{n,m+1} \uplus \bigoplus_{k=0}^n \mathcal{T}_{n-k,m} @ \mathcal{T}_{k,m}.$$

Moreover terms of size 1 are only made of de Bruijn indices, therefore

$$\mathcal{T}_{1,m} = \mathcal{I}(m).$$

From this we get:

$$T_{n+1,m} = T_{n,m+1} + \sum_{k=0}^n T_{n-k,m} \cdot T_{k,m}$$

¹Roughly speaking, typed terms are terms consistent with properties of the domain and the codomain of the function they represent.

$$\begin{aligned} T_{1,m} &= m \\ T_{0,m} &= 0 \end{aligned}$$

$\mathcal{T}_{n,0}$ is the set of closed terms (terms with no non bound indices) of size n . Notice that

$$T_{n+1,m} = T_{n,m+1} + \sum_{k=1}^{n-1} T_{n-k,m} \cdot T_{k,m}$$

Let us illustrate this result by the array of closed terms up to size 5:

| n | terms | $T_{n,0}$ |
|-----|---|-----------|
| 1 | none | 0 |
| 2 | $\lambda \underline{1}$ | 1 |
| 3 | $\lambda \lambda \underline{1}, \lambda \lambda \underline{2}$ | 2 |
| 4 | $\lambda \lambda \lambda \underline{1}, \lambda \lambda \lambda \underline{2}, \lambda \lambda \lambda \underline{3}, \lambda(\underline{1}, \underline{1})$ | 4 |
| 5 | $\lambda \lambda \lambda \lambda \underline{1}, \lambda \lambda \lambda \lambda \underline{2}, \lambda \lambda \lambda \lambda \underline{3}, \lambda \lambda \lambda \lambda \underline{4}, \lambda \lambda(\underline{1}, \underline{1}), \lambda \lambda(\underline{1}, \underline{2}), \lambda \lambda(\underline{2}, \underline{1}), \lambda \lambda(\underline{2}, \underline{2}),$ $\lambda(\underline{1}, \lambda \underline{1}), \lambda(\underline{1}, \lambda \underline{2}), \lambda((\lambda \underline{1}), \underline{1}), \lambda((\lambda \underline{2}), \underline{1}), \lambda \underline{1}, \lambda \underline{1}$ | 13 |

The equation that defines $T_{n,m}$ allows us to compute it, since it relies on entities $T_{i,j}$ where either $i < n$ or $j < m$. Figure 1 is a table of the first values of $T_{n,m}$ up to $T_{18,7}$. We are mostly interested by the sequence of sizes of the closed terms, namely $T_{n,0}$, in other words the first column of the table. The values of $T_{n,0}$ correspond to sequence **A135501** (see <http://www.research.att.com/~nudges/sequences/A135501>) due to Christophe Raffalli, which is defined as the *number of closed lambda-terms of size n*. His recurrence formula for those numbers is more complex. He considers the values of the double sequence $f_{n,m}$. $T_{n,m}$ and $f_{n,m}$ coincide for $m = 0$, i.e., $T_{n,0} = f_{n,0}$.

$$\begin{aligned} f_{1,1} &= 1 \\ f_{0,m} &= 0 \\ f_{n,m} &= 0 \text{ if } m > 2n - 1 \\ f_{n,m} &= f_{n-1,m} + f_{n-1,m+1} + \sum_{p=1}^{n-2} \sum_{c=0}^m \sum_{l=0}^{m-c} \binom{c}{m} \binom{l}{m-c} f_{p,l+c} f_{n-p-1,m-l}. \end{aligned}$$

He adds

The last term is for the application where c is the number of common variables in both subterms. $f_{n,m}$ can be computed only using $f_{n',m'}$ with $n' < n$ and $m' \leq m + n - n'$.

Notice that he deals only with sequence $T_{n,0}$, whereas we consider the values for any value of m , from which we can expect to extract interesting informations. The main interesting statement we can draw from this is that considering lambda terms with explicit variables or considering lambda terms with de Bruijn indices makes no difference, at least when no β -reduction is taken into account. We feel that considering lambda terms with de Bruijn indices makes the task easier and produces more results.

3 Bounding the $T_{n,0}$'s

We can show that the $T_{n,0}$'s are bound by the Motzkin numbers. More precisely we get the following proposition.

Proposition 1 *If M_n are the Motzkin numbers, $M_n < T_{n+1,0}$.*

Proof: There is a one-to-one correspondance between unary-binary trees and lambda terms of the form λM in which all the indices are 1. Hence the results, since Motzkin numbers count unary-binary trees. \square

We conclude that the asymptotic behavior of the $T_{n,0}$'s are at least 3^n since the Motzkin numbers are asymptotically equivalent to $\sqrt{\frac{3}{4\pi n^3}} 3^n$ ([8], Example VI.3).

4 The functions $m \mapsto T_{n,m}$

Due to properties of the generating function (see Section 6) we are not able to give a simple expression for the function $n \mapsto T_{n,m}$, so we focus on the function $m \mapsto T_{n,m}$. These functions are polynomials P_n^T , defined recursively as follows:

$$P_0^T(m) = 0 \quad (1)$$

$$P_1^T(m) = m \quad (2)$$

$$P_{n+1}^T(m) = P_n^T(m+1) + \sum_{k=1}^{n-1} P_k^T(m) P_{n-k}^T(m). \quad (3)$$

See Figure 2 for the first 18 polynomials. The table below gives the coefficients of the polynomials P_n^T up to 13.

| $n \setminus m^i$ | m^8 | m^7 | m^6 | m^5 | m^4 | m^3 | m^2 | m | 1 |
|-------------------|-------|-------|--------|---------|----------|----------|----------|----------|----------|
| 1 | | | | | | | | 1 | 0 |
| 2 | | | | | | | 1 | 1 | 1 |
| 3 | | | | | | | 1 | 1 | 2 |
| 4 | | | | | | | 3 | 5 | 4 |
| 5 | | | | | | 2 | 6 | 17 | 13 |
| 6 | | | | | | 10 | 26 | 49 | 42 |
| 7 | | | | | 5 | 30 | 111 | 179 | 139 |
| 8 | | | | | 35 | 134 | 405 | 683 | 506 |
| 9 | | | | 14 | 140 | 652 | 1658 | 2629 | 1915 |
| 10 | | | | 126 | 676 | 2812 | 7122 | 10725 | 7558 |
| 11 | | | 42 | 630 | 3610 | 12760 | 30783 | 45195 | 31092 |
| 12 | | | 462 | 3334 | 17670 | 60240 | 138033 | 196355 | 132170 |
| 13 | | 132 | 2772 | 19218 | 87850 | 285982 | 635178 | 880379 | 580466 |
| 14 | | 1716 | 16108 | 104034 | 449290 | 1390246 | 2991438 | 4052459 | 2624545 |
| 15 | 429 | 12012 | 99386 | 560854 | 2308173 | 6895122 | 14436365 | 19144575 | 12190623 |
| 16 | 6435 | 76444 | 584878 | 3076878 | 12039895 | 34815210 | 71170791 | 92631835 | 58083923 |

The degrees of those polynomials increase two by two and we can describe their leading coefficients, their second leading coefficients and the third leading coefficients of the odd polynomials.

Proposition 2 $deg(P_{2p-1}^T) = deg(P_{2p}^T) = p$.

Proof: This is true for $P_1^T = m$ and $P_2^T = m+1$ which have degree 1. Assume the property true up to p . Note that all the coefficients of the P_n^T 's are positive. In

$$P_n^T(m+1) + \sum_{k=1}^{n-1} P_k^T P_{n-k}^T,$$

the degree of $P_{n+1}^T(m)$ comes from the $P_k^T P_{n-k}^T$'s. Indeed, par induction the degree of $P_n^T(m+1)$ is $n \div 2$ which is less than $(n+1) \div 2$, therefore we can consider that $P_n^T(m+1)$ does not contribute to the degree of $P_{n+1}^T(m)$. Consider the degree of $P_k^T P_{n-k}^T$ according to the parity of n and k .

$n = 2p + 1$ **and** $k = 2h - 1$. In this case, $p \geq h \geq 1$ and the degree of P_{2h-1}^T is h and the degree of $P_{2p+1-2h+1}^T$ is $p - h + 1$, hence the degree of $P_{2h-1}^T P_{2p+1-2h+1}^T$ is $p + 1$.

$n = 2p + 1$ **and** $k = 2h$. In this case, $p \geq h \geq 1$ and the degree of P_{2h}^T is h and the degree of $P_{2p+1-2h}^T$ is $p - h + 1$, hence the degree of $P_{2h}^T P_{2p+1-2h}^T$ is $p + 1$.

$n = 2p$ **and** $k = 2h - 1$. In this case, $p + 1 \geq h \geq 1$ and the degree of P_{2h-1}^T is h and the degree of $P_{2p-2h+1}^T$ is $p - h + 1$, hence the degree of $P_{2h-1}^T P_{2p-2h+1}^T$ is $p + 1$.

$n = 2p$ **and** $k = 2h$. In this case, $p + 1 \geq h \geq 1$ and the degree of P_{2h}^T is h and the degree of P_{2p-2h}^T is $p - h$, hence the degree of $P_{2h}^T P_{2p-2h}^T$ is p . These products $P_{2h}^T P_{2p-2h}^T$ do not contribute to the degree of P_{2p+1}^T .

□

In what follows, for short, we write θ_{2q+1} and θ_{2q} the leading coefficients of P_{2q+1}^T and P_{2q}^T , τ_{2q+1} and τ_{2q} the second leading coefficients of P_{2q+1}^T and P_{2q}^T , and δ_{2q+1} the third leading coefficients of P_{2q+1}^T . We also write, as usual, C_n the n^{th} Catalan number.

We define five generating functions.

$$\begin{aligned} \mathcal{O}d(z) &= \sum_{i=0}^{\infty} \theta_{2i+1} z^i & \mathcal{E}v(z) &= \sum_{i=0}^{\infty} \theta_{2i} z^i \\ \mathcal{S}od(z) &= \sum_{i=0}^{\infty} \tau_{2i+1} z^i & \mathcal{S}ev(z) &= \sum_{i=0}^{\infty} \tau_{2i} z^i \\ \mathcal{T}od(z) &= \sum_{i=0}^{\infty} \delta_{2i+1} z^i. \end{aligned}$$

Proposition 3 *The leading coefficients of P_{2q+1}^T are $\frac{1}{q+1} \binom{2q}{q}$, i.e., the Catalan numbers C_q .*

Proof: From Equation (3) and the last two steps of the proof of Proposition 2, we deduce the following relation :

$$\begin{aligned}\theta_{2q+1} &= \sum_{h=0}^{q-1} \theta_{2h+1} \theta_{2q-2h-1} \quad \text{for } q \geq 1 \\ \theta_1 &= 1.\end{aligned}$$

which says that the leading coefficient of an odd polynomial comes only from the leading coefficients in the products of odd polynomials. We get:

$$\mathcal{O}d(z) = 1 + z \mathcal{O}d(z)^2.$$

which shows that

$$\mathcal{O}d(z) = \frac{1 - \sqrt{1 - 4z}}{2z}.$$

and $\mathcal{O}d(z) = C(z)$, the generating function of the Catalan numbers. \square

Proposition 4 *The leading coefficients of P_{2q}^T are $\binom{2q-1}{q}$, for $q \geq 1$.*

Proof: Without loss of generality, we assume that $\theta_0 = 0$. From Equation (3), we get, for $q \geq 1$,

$$\begin{aligned}\theta_{2q+2} &= \theta_{2q+1} + \sum_{k=0}^{2q+1} \theta_k \theta_{2q+1-k} \\ &= \theta_{2q+1} + \sum_{h=0}^q \theta_{2h} \theta_{2q+1-2h} + \sum_{h=0}^q \theta_{2h+1} \theta_{2q-2h} \\ &= \theta_{2q+1} + 2 \sum_{h=0}^q \theta_{2h} \theta_{2q-2h+1}.\end{aligned}$$

which says that the leading coefficient of an even polynomial comes from the leading coefficient of the preceding odd polynomial and of the products of the leading coefficients of the products of the smaller polynomials. We get:

$$\mathcal{E}v(z) = z\mathcal{O}d(z) + 2z\mathcal{E}v(z)\mathcal{O}d(z),$$

hence

$$\mathcal{E}v(z) = \frac{z\mathcal{O}d(z)}{1 - 2z\mathcal{O}d(z)} = \frac{1 - \sqrt{1 - 4z}}{2\sqrt{1 - 4z}} = \frac{\sqrt{1 - 4z}}{2(1 - 4z)} - \frac{1}{2}$$

which is the generating function of the sequence $\binom{2q+1}{q}$. \square

Proposition 5 *The second leading coefficients of P_{2q+1}^T are $(2q - 1)\binom{2(q-1)}{q-1}$.*

Proof: From the proof of Proposition 2, we see that the monomial of second highest degree of P_{2q+1} is made as the sum:

- of the monomial of highest degree of P_{2q} ,
- of the products of the monomials of highest degree from the P_i 's with even indices and
- the products of monomials of highest degree with monomials of second highest degree from the P_i 's with odd indices.

We get for $q \geq 1$:

$$\tau_{2q+1} = \theta_{2q} + \sum_{h=0}^q \theta_{2h} \theta_{2q-2h} + \sum_{h=0}^{q-1} \theta_{2h+1} \tau_{2q-2h-1} + \sum_{h=0}^{q-1} \tau_{2h+1} \theta_{2q-2h-1}.$$

We notice that $\tau_1 = 0$. Therefore we get:

$$\mathcal{S}od(z) = \mathcal{E}v(z) + \mathcal{E}v(z)^2 + 2z\mathcal{O}d(z)\mathcal{S}od(z).$$

Then

$$\mathcal{S}od(z) = \frac{\mathcal{E}v(z) + \mathcal{E}v(z)^2}{1 - 2z\mathcal{O}d(z)} = \frac{z}{\sqrt{1-4z}(1-4z)} = \frac{z\sqrt{1-4z}}{(1-4z)^2}$$

which is the generating function of $(2q-1)\binom{2(q-1)}{q-1}$. \square

Proposition 6 *The second leading coefficients of P_{2q}^T are $\tau_0 = 0$, $\tau_1 = 1$, $\tau_2 = 5$ and for $q \geq 3$,*

$$\tau_{2q} = 4^{q-1} + \frac{2(2q-5)(2q-3)(2q-1)}{3(q-2)} \binom{2(q-3)}{q-3}.$$

Proof: From Equation (3), we get

$$\begin{cases} \tau_{2q+2} &= (q+1)\theta_{2q+1} + \tau_{2q+1} \\ &+ 2\sum_{i=1}^q \theta_{2i-1}\tau_{2q-2i+2} + 2\sum_{i=1}^q \theta_{2i}\tau_{2q-2i+1} \\ \tau_0 &= 0 \end{cases}$$

The second leading coefficient of an even polynomial P_{2m+2}^T is made of four components:

- the coefficient of degree q in $\theta_{2q+1}(m+1)^{q+1}$, namely $(q+1)\theta_{2q+1}$,
- the coefficient of degree q in $\tau_{2q+1}(m+1)^q$, namely τ_{2q+1} ,
- the sum of the products of the leading coefficients of the odd polynomials and the second leading coefficients of the even polynomials (this occurs twice, once in product $P_{2i-1} P_{2q-2i+2}$ and once in product $P_{2i} P_{2q-2i+1}$),
- the sum of the products of the leading coefficients of the even polynomials and the second leading coefficients of the odd polynomials (twice).

From the above induction, $\mathcal{S}ev$ fullfils the following functional equation:

$$\mathcal{S}ev(z) = z\mathcal{O}d(z) + z^2\mathcal{O}d'(z) + z\mathcal{S}od(z) + 2z\mathcal{O}d(z)\mathcal{S}ev(z) + 2z\mathcal{E}v(z)\mathcal{S}od(z).$$

Therefore

$$\begin{aligned} \mathcal{S}ev(z) &= \frac{z\mathcal{O}d(z) + z^2\mathcal{O}d'(z) + z\mathcal{S}od(z) + 2z\mathcal{E}v(z)\mathcal{S}od(z)}{\sqrt{1-4z}} \\ &= \frac{(1 - \sqrt{1-4z})}{2\sqrt{1-4z}} \\ &\quad + \frac{z}{1-4z} - \frac{1 - (\sqrt{1-4z})}{2\sqrt{1-4z}} \\ &\quad + \frac{z^2}{(1-4z)^2} \\ &\quad + \frac{z(1 - \sqrt{1-4z})}{(1-4z)^2\sqrt{1-4z}} \\ &= \frac{z}{1-4z} + \frac{z^2}{(1-4z)^2} + \frac{z^2(1 - \sqrt{1-4z})}{(1-4z)^2\sqrt{1-4z}} \\ &= \sum_{q=1}^{\infty} 4^{q-1}z^q + \sum_{q=2}^{\infty} (q-1)4^{q-2}z^q + \sum_{q=3}^{\infty} 2a_{q-3}z^q \end{aligned}$$

where $(a_n)_{n \in \mathbb{N}}$ is sequence A029887 of the *On-Line Encyclopedia of Integer Sequences* whose value is:

$$\frac{(2n+1)(2n+3)(2n+5)}{3}C_n - (n+2)2^{2n+1}.$$

Hence

$$\begin{aligned} \mathcal{S}ev(z) &= \sum_{q=1}^{\infty} 4^{q-1}z^q + \sum_{q=3}^{\infty} \frac{2(2q-5)(2q-3)(2q-1)}{3}C_{q-3}z^q \\ &= \frac{z}{1-4z} + \frac{z^2}{(1-4z)^2\sqrt{1-4z}}. \end{aligned}$$

□

Proposition 7 *The third leading coefficients of P_{2q+1}^T are*

$$q 2^{2q-1} + \frac{q(q-1)(q-2)}{120} \binom{2q}{q} + \frac{(q+1)q(q-1)}{120} \binom{2(q+1)}{q+1}.$$

Proof: Since $\deg(P_{2n}^T) = \deg(P_{2n+1}^T) - 1$, the third coefficient is the sum of seven items:

- the second coefficient of $\theta_{2q}(m+1)^q$, namely $q\theta_{2q}$,

- the first coefficient of $(m+1)^{q-1}$, namely τ_{2q} ,
- the sum of products of leading coefficients and second leading coefficients for even polynomials (twice),
- the sum of leading coefficients and third leading coefficients for odd polynomials (twice),
- the sum of second leading coefficients with second leading coefficients.

The formula for δ_{2q+1} is:

$$\begin{aligned} \delta_{2q+1} &= q\theta_{2q} + \tau_{2q} + \sum_{i=0}^q \tau_{2i}\theta_{2q-2i} + \sum_{i=0}^q \theta_{2i}\tau_{2q-2i} + \\ &\quad \sum_{i=0}^{q-1} \theta_{2i+1}\delta_{2q-2i-1} + \sum_{i=0}^{q-1} \delta_{2i+1}\theta_{2q-2i-1} + \sum_{i=0}^{q-1} \tau_{2i+1}\tau_{2q-2i-1}, \end{aligned}$$

which gives the following equation on generating functions:

$$\begin{aligned} \mathcal{T}od(z) &= z\mathcal{E}v'(z) + \mathcal{S}ev(z) + 2z\mathcal{E}v(z)\mathcal{S}ev(z) + \\ &\quad 2z\mathcal{O}d(z)\mathcal{T}od(z) + z\mathcal{S}ev(z)^2. \end{aligned}$$

which yields:

$$\begin{aligned} \mathcal{T}od(z) &= \frac{z\mathcal{E}v'(z) + \mathcal{S}ev(z) + 2z\mathcal{E}v(z)\mathcal{S}ev(z) + z\mathcal{S}ev(z)^2}{1 - 2z\mathcal{O}d(z)} \\ &= \frac{1}{\sqrt{1-4z}} \left(\frac{z}{(1-4z)\sqrt{1-4z}} + \right. \\ &\quad \left. \frac{z}{1-4z} + \frac{z^2}{(1-4z)^2\sqrt{1-4z}} + \right. \\ &\quad \left. \frac{1-\sqrt{1-4z}}{\sqrt{1-4z}} \left(\frac{z}{1-4z} + \frac{z^2}{(1-4z)^2\sqrt{1-4z}} \right) + \right. \\ &\quad \left. \frac{z^3}{(1-4z)^3} \right) \\ &= \frac{2z}{(1-4z)^2} + \frac{z^2+z^3}{(1-4z)^3\sqrt{1-4z}}. \end{aligned}$$

The first part corresponds to sequence A002699 which expression is $q2^{2q-1}$. $1/(1-4z)^3\sqrt{1-4z}$ corresponds to sequence A144395. Therefore the second part yields the expression

$$\frac{q(q-1)(q-2)}{120} \binom{2q}{q} + \frac{(q+1)q(q-1)}{120} \binom{2(q+1)}{q+1}.$$

□

Hence typically if we pose

$$\begin{aligned}\tau_{2q} &= 4^{q-1} + \frac{2(2q-5)(2q-3)(2q-1)}{3}C_{q-3} \\ \delta_{2q+1} &= q 2^{2q-1} + \frac{(q+1)q(q-1)(q-2)}{120}C_q + \frac{(q+2)(q+1)q(q-1)}{120}C_{q+1}\end{aligned}$$

we have in general:

$$\begin{aligned}P_{2q}^T(m) &= (2q-1)C_{q-1}m^q + \tau_{2q}m^{q-1} + \dots + T_{2q,0} \\ P_{2q+1}^T(m) &= C_qm^{q+1} + \frac{2q(2q-1)}{2}C_{q-1}m^q + \delta_{2q+1}m^{q-1} + \dots + T_{2q+1,0}\end{aligned}$$

showing the prominent role of Catalan numbers. The relations for the other coefficients are more convoluted² and have not been computed.

It should be interesting to study the connection with the derivatives of the generating function $C(z)$ of the Catalan numbers [12].

5 Normal forms

Lest us call \mathcal{F}_m the set of normal forms with $\{\underline{1}, \dots, \underline{m}\}$ de Bruijn indices and \mathcal{G}_m the sets of normal forms with no head λ and also de Bruijn indices in $\{\underline{1}, \dots, \underline{m}\}$. The combinatorial structure equations are

$$\begin{aligned}\mathcal{G}_m &= \mathcal{I}(m) \uplus \mathcal{G}_m \circledast \mathcal{F}_m \\ \mathcal{F}_m &= \lambda \mathcal{F}_{m+1} \uplus \mathcal{G}_m\end{aligned}$$

Let $G_{n,m}$ be the number of normal forms of size n with no head λ and with de Bruijn indices in $\mathcal{I}(m)$ and let $F_{n,m}$ be the number of normal forms of size n with de Bruijn indices in $\mathcal{I}(m)$. The relations between $G_{n,m}$ and $F_{n,m}$ are

$$\begin{aligned}G_{0,m} &= 0 \\ G_{1,m} &= m \\ G_{n+1,m} &= \sum_{k=0}^n G_{n-k,m} F_{k,m} \\ F_{0,m} &= 0 \\ F_{1,m} &= m = G_{1,m} \\ F_{n+1,m} &= F_{n,m+1} + G_{n+1,m}\end{aligned}$$

whereas the relations between generating functions are

$$\begin{aligned}G_m(z) &= mz + zG_m(z)F_m(z) \\ F_m(z) &= zF_{m+1}(z) + G_m(z).\end{aligned}$$

The coefficients $F_{n,m}$ are given in Figure 3.

²Like τ_{2q} and δ_{2q+1} , they correspond to non studied sequences according to the *On-Line Encyclopedia of Integer Sequences*.

The functions $m \mapsto F_{n,m}$

Like for $m \mapsto T_{n,m}$, the functions $m \mapsto F_{n,m}$ are polynomials of degree $n+1 \div 2$, which we write P_n^{NF} and which are given in Figure 4. The coefficients of polynomials P_n^{NF} enjoy properties somewhat similar to those proved for polynomials P_n^T . In this section, we write $P_n(m)$ the polynomial $P_n^{NF}(m)$, $Q_n(m)$ the polynomial associated with $G_{n,m}$, φ_n the leading coefficient of P_n , $\overline{\varphi}_n$ the leading coefficient of Q_n , ψ_n the second leading coefficient of P_n and $\overline{\psi}_n$ the second leading coefficient of Q_n . We have the equations

$$P_{n+1}(m) = P_n(m+1) + Q_{n+1}(m) \quad (4)$$

$$Q_{n+1}(m) = \sum_{k=0}^n Q_{n-k}(m)P_k(m) \quad (5)$$

Proposition 8 $\deg(P_{2p-1}) = \deg(P_{2p}) = \deg(Q_{2p-1}) = \deg(Q_{2p}) = p$.

Proof: Here also the coefficients are positive. The degree of P_n is the degree of Q_n by (4). One notices that $\deg P_0 = \deg Q_0 = 0$ and $\deg P_1 = \deg Q_1 = 1$. The general step can be mimicked from this of Prop 2. \square

We define eight generating functions:

$$\begin{aligned} \mathcal{F}ev(z) &= \sum_{i=0}^{\infty} \varphi_{2i} z^i & \overline{\mathcal{F}ev}(z) &= \sum_{i=0}^{\infty} \overline{\varphi}_{2i} z^i \\ \mathcal{F}od(z) &= \sum_{i=0}^{\infty} \varphi_{2i+1} z^i & \overline{\mathcal{F}od}(z) &= \sum_{i=0}^{\infty} \overline{\varphi}_{2i+1} z^i \\ \mathcal{S}\mathcal{F}ev(z) &= \sum_{i=0}^{\infty} \psi_{2i} z^i & \overline{\mathcal{S}\mathcal{F}ev}(z) &= \sum_{i=0}^{\infty} \overline{\psi}_{2i} z^i \\ \mathcal{S}\mathcal{F}od(z) &= \sum_{i=0}^{\infty} \psi_{2i+1} z^i & \overline{\mathcal{S}\mathcal{F}od}(z) &= \sum_{i=0}^{\infty} \overline{\psi}_{2i+1} z^i \end{aligned}$$

Proposition 9 *The leading coefficients of P_{2q+1}^{NF} are Catalan numbers.*

Proof: We see easily that $\varphi_{2q+1} = \overline{\varphi}_{2q+1}$ by (4). By (5), we see that

$$\begin{aligned} \varphi_{2q+1} &= \sum_{h=0}^{q-1} \overline{\varphi}_{2q+1} \varphi_{2q-2h-1} \\ \varphi_1 &= \overline{\varphi}_1 = 1. \end{aligned}$$

Hence the result $\mathcal{F}od(z) = \overline{\mathcal{F}od}(z) = C(z)$ (see proof of Proposition 3). \square

Proposition 10 *The leading coefficients of the P_n^{NF} 's for n even, are $P_0^{NF} = 0$, $P_2^{NF} = 1$ and $P_{2q+4}^{NF} = 2\binom{2q+1}{q}$, i.e., $P_{2q+4}^{NF} = 2P_{2q+2}^T$.*

Proof: From Equations (4) and (5), we get:

$$\begin{aligned}\varphi_{2(q+1)} &= \varphi_{2q+1} + \overline{\varphi}_{2(q+1)} \\ \overline{\varphi}_{2(q+1)} &= \sum_{i=0}^q \overline{\varphi}_{2q+1-2i} \varphi_{2i} + \sum_{i=0}^q \overline{\varphi}_{2q-2i} \varphi_{2i+1}\end{aligned}$$

Hence

$$\begin{aligned}\mathcal{F}ev(z) &= z\mathcal{F}od(z) + \overline{\mathcal{F}ev}(z) \\ \overline{\mathcal{F}ev}(z) &= z\overline{\mathcal{F}od}(z)\mathcal{F}ev(z) + z\overline{\mathcal{F}ev}(z)\mathcal{F}od(z)\end{aligned}$$

from which we get

$$\overline{\mathcal{F}ev}(z) = \frac{z\mathcal{F}od(z)\mathcal{F}ev(z)}{1 - z\mathcal{F}od(z)} \quad (6)$$

then

$$\mathcal{F}ev(z) = z\mathcal{F}od(z) + \frac{z\mathcal{F}od(z)\mathcal{F}ev(z)}{1 - z\mathcal{F}od(z)}$$

and

$$\mathcal{F}ev(z) - z\mathcal{F}od(z)\mathcal{F}ev(z) = z\mathcal{F}od(z) - z^2\mathcal{F}od(z)^2 + z\mathcal{F}od(z)\mathcal{F}ev(z)$$

and

$$\begin{aligned}\mathcal{F}ev(z) &= \frac{z\mathcal{F}od(z) - z^2\mathcal{F}od(z)^2}{1 - 2z\mathcal{F}od(z)} \\ &= \frac{z}{\sqrt{1-4z}} = \frac{z\sqrt{1-4z}}{1-4z} \\ &= \frac{z}{1-2zC(z)}.\end{aligned}$$

Hence $\mathcal{F}ev(z)$ is the generating function of the sequence $\varphi_0 = 0$, $\varphi_2 = 1$ and $\varphi_{2q+4} = 2\binom{2q+1}{q}$. \square

Corollary 1 $\overline{\mathcal{F}ev}(z) = \frac{1-2z-\sqrt{1-4z}}{2\sqrt{1-4z}}$

Proof:

$$\begin{aligned}\overline{\mathcal{F}ev}(z) &= \mathcal{F}ev(z) - zC(z) \\ &= \frac{z}{\sqrt{1-4z}} - \frac{1-\sqrt{1-4z}}{2} = z^2C'(z).\end{aligned}$$

\square

Proposition 11 *The second leading coefficients of the P_{2q+1}^{NF} 's are $\psi_0 = 0$, $\psi_3 = 1$ and $\psi_{2q+5} = (q+3)\binom{2q+1}{q}$.*

Proof: From the proof of Proposition 8,

$$\begin{aligned}\psi_{2q+1} &= \varphi_{2q} + \bar{\psi}_{2q+1} \\ \bar{\psi}_1 &= 0 \\ \bar{\psi}_{2q+3} &= \sum_{i=0}^{q+1} \bar{\varphi}_{2i} \varphi_{2q-2i} + \sum_{i=0}^q \bar{\psi}_{2i+1} \varphi_{2q-2i+1} + \sum_{i=0}^q \psi_{2i+1} \bar{\varphi}_{2q-2i+1},\end{aligned}$$

from which we get

$$\begin{aligned}\mathcal{SFod}(z) &= \mathcal{Fev}(z) + \overline{\mathcal{SFod}}(z) \\ \overline{\mathcal{SFod}}(z) &= \overline{\mathcal{Fev}}(z)\mathcal{Fev}(z) + z\overline{\mathcal{SFod}}(z)\mathcal{Fod}(z) + z\mathcal{SFod}(z)\overline{\mathcal{Fod}}(z).\end{aligned}$$

Then we get

$$\overline{\mathcal{SFod}}(z)(1 - z\mathcal{Fod}(z)) = \mathcal{Fev}(z)\overline{\mathcal{Fev}}(z) + z\mathcal{SFod}(z)\overline{\mathcal{Fod}}(z).$$

We know that $1 - z\mathcal{Fod}(z) = 1 - zC(z) = 1/C(z)$, then

$$\overline{\mathcal{SFod}}(z) = \frac{z}{\sqrt{1-4z}} z^2 C'(z)C(z) + z\mathcal{SFod}(z)C(z)^2$$

and

$$\mathcal{SFod}(z) = \mathcal{Fev}(z) + \frac{z^3 C(z)C'(z)}{\sqrt{1-4z}} + z\mathcal{SFod}(z)C(z)^2.$$

We know $1 - zC(z)^2 = C(z)\sqrt{1-4z}$, then

$$\begin{aligned}\mathcal{SFod}(z) &= \left(\frac{z}{\sqrt{1-4z}} + \frac{z^3 C(z)C'(z)}{\sqrt{1-4z}} \right) \frac{1}{C(z)\sqrt{1-4z}} \\ &= \frac{z}{C(z)(1-4z)} + \frac{z^3 C'(z)}{1-4z} \\ &= \frac{z^2}{(1-4z)\sqrt{1-4z}} + \frac{z}{\sqrt{1-4z}}.\end{aligned}$$

which is the generating function of the sequence 0, 1 followed by $(q+3)\binom{2q+1}{q}$. \square

Corollary 2 $\overline{\mathcal{SFod}}(z) = \frac{z^2}{(1-4z)\sqrt{1-4z}}$

Proof:

$$\overline{\mathcal{SFod}}(z) = \mathcal{SFod}(z) - \mathcal{Fev}(z) = \frac{z^2}{(1-4z)\sqrt{1-4z}}.$$

Notice that $\overline{\mathcal{SFod}}(z) = z\mathcal{Sod}(z)$. \square

Proposition 12 *The second leading coefficients of the P_{2q}^{NF} 's are $\psi_0 = 0$, $\psi_2 = 1$, $\psi_6 = 15$ and for $q \geq 4$*

$$\begin{aligned}\psi_{2q} &= \binom{2q-3}{q-2} + 2^{2q-3} + (q-2) \binom{2q-2}{q-2} + \\ & 2 \binom{2q-5}{q-3} + \frac{(q-3)(q-2)}{3} \binom{2q-5}{q-3}.\end{aligned}$$

Proof: We have

$$\begin{aligned}\psi_{2q+2} &= (q+1)\varphi_{2q+1} + \psi_{2q+1} + \bar{\psi}_{2q+2} \\ \bar{\psi}_{2q+2} &= \sum_{i=1}^q \psi_{2i-1} \bar{\varphi}_{2q-2i+2} + \sum_{i=1}^q \varphi_{2i-1} \bar{\psi}_{2q-2i+2} + \\ & \sum_{i=1}^q \bar{\varphi}_{2i-1} \psi_{2q-2i+2} + \sum_{i=1}^q \bar{\psi}_{2i-1} \varphi_{2q-2i+2}.\end{aligned}$$

This gives the equations on generic functions.

$$\begin{aligned}\mathcal{S}\mathcal{F}ev(z) &= z\mathcal{F}od(z) + z^2\mathcal{F}od'(z) + z\mathcal{S}\mathcal{F}od(z) + \overline{\mathcal{S}\mathcal{F}ev}(z) \\ \overline{\mathcal{S}\mathcal{F}ev}(z) &= z\mathcal{S}\mathcal{F}od(z)\overline{\mathcal{F}ev}(z) + z\mathcal{F}od(z)\overline{\mathcal{S}\mathcal{F}ev}(z) + \\ & z\mathcal{S}\mathcal{F}ev(z)\overline{\mathcal{F}od}(z) + z\mathcal{F}ev(z)\overline{\mathcal{S}\mathcal{F}od}(z).\end{aligned}$$

Hence

$$\overline{\mathcal{S}\mathcal{F}ev}(z) = \frac{z\mathcal{S}\mathcal{F}od(z)\overline{\mathcal{F}ev}(z) + z\mathcal{S}\mathcal{F}ev(z)\overline{\mathcal{F}od}(z) + z\mathcal{F}ev(z)\overline{\mathcal{S}\mathcal{F}od}(z)}{1 - zC(z)}$$

which yields

$$\begin{aligned}\mathcal{S}\mathcal{F}ev(z) &= \mathcal{F}od(z) + z^2\mathcal{F}od'(z) + z\mathcal{S}\mathcal{F}od(z) + \\ & C(z)(z\mathcal{S}\mathcal{F}od(z)\overline{\mathcal{F}ev}(z) + z\mathcal{F}ev(z)\overline{\mathcal{S}\mathcal{F}od}(z)) \\ & zC(z)\mathcal{S}\mathcal{F}ev(z)\overline{\mathcal{F}od}(z).\end{aligned}$$

and

$$\begin{aligned}\mathcal{S}\mathcal{F}ev(z) &= \frac{\mathcal{F}od(z) + z^2\mathcal{F}od'(z) + z\mathcal{S}\mathcal{F}od(z)}{1 - zC(z)^2} + \\ & \frac{zC(z)\mathcal{S}\mathcal{F}od(z)\overline{\mathcal{F}ev}(z) + zC(z)\mathcal{F}ev(z)\overline{\mathcal{S}\mathcal{F}od}(z)}{1 - zC(z)^2} \\ &= \frac{z}{\sqrt{1-4z}} + \frac{z^2C'(z)}{C(z)\sqrt{1-4z}} + \\ & \frac{z}{C(z)\sqrt{1-4z}} \left(\frac{z^2}{(1-4z)\sqrt{1-4z}} + \frac{z}{\sqrt{1-4z}} \right) + \\ & \left(\frac{z}{\sqrt{1-4z}} - \frac{1-\sqrt{1-4z}}{2} \right) \left(\frac{z^2}{(1-4z)^2} + \frac{z^2}{1-4z} \right) + \\ & \frac{z^4}{(1-4z)^2\sqrt{1-4z}}.\end{aligned}$$

Notice that

$$\frac{z^2 C'(z)}{C(z)\sqrt{1-4z}} = \frac{z}{2(1-4z)} - \frac{z}{2\sqrt{1-4z}}.$$

and

$$\begin{aligned} & \frac{z}{C(z)\sqrt{1-4z}} \left(\frac{z^2}{(1-4z)\sqrt{1-4z}} + \frac{z}{\sqrt{1-4z}} \right) + \\ & \left(\frac{z}{\sqrt{1-4z}} - \frac{1-\sqrt{1-4z}}{2} \right) \left(\frac{z^2}{(1-4z)^2} + \frac{z^2}{1-4z} \right) = \frac{2z^3}{\sqrt{1-4z}(1-4z)} + \frac{z^2}{\sqrt{1-4z}} + \\ & \frac{z^4}{\sqrt{1-4z}(1-4z)^2} \end{aligned}$$

Hence

$$\begin{aligned} S\mathcal{F}ev(z) &= \frac{z}{2\sqrt{1-4z}} + \frac{z}{2(1-4z)} + \\ & \frac{2z^3}{\sqrt{1-4z}(1-4z)} + \frac{z^2}{\sqrt{1-4z}} + \frac{2z^4}{\sqrt{1-4z}(1-4z)^2} \end{aligned}$$

We summarize the result in the following table.

| <i>gen. fonct.</i> | <i>coefficients</i> | <i>up to</i> | <i>why?</i> |
|------------------------------------|---|--------------|----------------|
| $\frac{z}{2\sqrt{1-4z}}$ | $\binom{2q-3}{q-2}$ | $q \geq 2$ | Proposition 10 |
| $\frac{z}{2(1-4z)}$ | 2^{2q-3} | $q \geq 2$ | |
| $\frac{2z^3}{\sqrt{1-4z}(1-4z)}$ | $(q-2)\binom{2q-2}{q-2}$ | $q \geq 2$ | |
| $\frac{z^2}{\sqrt{1-4z}}$ | $2\binom{2q-5}{q-3}$ | $q \geq 3$ | |
| $\frac{2z^4}{\sqrt{1-4z}(1-4z)^2}$ | $\frac{(q-3)(q-2)}{3}\binom{2q-5}{q-3}$ | $q \geq 4$ | A002802 |

Hence we have for $q \geq 4$:

$$\begin{aligned} \psi_q &= \binom{2q-3}{q-2} + 2^{2q-3} + (q-2)\binom{2q-2}{q-2} + \\ & 2\binom{2q-5}{q-3} + \frac{(q-3)(q-2)}{3}\binom{2q-5}{q-3}. \end{aligned}$$

□

Recall what we have computed for *plain terms*:

| coefficients | generating functions | values | equivalents | |
|------------------|----------------------|---|--|---|
| $P_{2q+1,q+1}^T$ | $\mathcal{O}d(z)$ | $\frac{1-\sqrt{1-4z}}{2z}$ | C_q | $4^q \sqrt{\frac{1}{\pi q^3}}$ |
| $P_{2q+1,q}^T$ | $\mathcal{S}od(z)$ | $\frac{z}{(1-4z)\sqrt{1-4z}}$ | $(2q-1) \binom{2(q-1)}{q-1}$ | $4^q \frac{1}{2} \sqrt{\frac{q}{\pi}}$ |
| $P_{2q+1,q-1}^T$ | $\mathcal{T}od(z)$ | $\frac{\frac{2z}{(1-4z)^2} + \frac{z^2+z^3}{(1-4z)^3\sqrt{1-4z}}}{(1-4z)^3\sqrt{1-4z}}$ | $q 2^{2q-1} + \frac{q(q-1)(q-2)}{120} \binom{2q}{q} + \frac{(q+1)q(q-1)}{120} \binom{2(q+1)}{q+1}$ | $4^q \frac{1}{24} \sqrt{\frac{q^5}{\pi}}$ |
| $P_{2q,q}^T$ | $\mathcal{E}v(z)$ | $\frac{4z-1+\sqrt{1-4z}}{2(1-4z)}$ | $\binom{2q-1}{q}$ | $4^q \frac{1}{2} \sqrt{\frac{1}{\pi q}}$ |
| $P_{2q,q-1}^T$ | $\mathcal{S}ev(z)$ | $\frac{\frac{z}{1-4z} + \frac{z^2}{(1-4z)^2\sqrt{1-4z}}}{(1-4z)^2\sqrt{1-4z}}$ | $4^{q-1} + \frac{2(2q-5)(2q-3)(2q-1)}{3(q-2)} \binom{2(q-3)}{q-3}$ | $4^q \frac{1}{12} \sqrt{\frac{q^3}{\pi}}$ |

and for *normal forms*

| coefficients | generating functions | values | equivalents | |
|---------------------|-------------------------------|--|---|---|
| $P_{2q+1,q+1}^{NF}$ | $\mathcal{F}od(z)$ | $\frac{1-\sqrt{1-4z}}{2z}$ | C_q | $4^q \sqrt{\frac{1}{\pi q^3}}$ |
| $P_{2q+1,q}^{NF}$ | $\mathcal{S}\mathcal{F}od(z)$ | $\frac{z}{\sqrt{1-4z}} + \frac{z^2}{(1-4z)\sqrt{1-4z}}$ | $(q+1) \binom{2q-3}{q-2}$ | $4^q \frac{1}{8} \sqrt{\frac{q}{\pi}}$ |
| $P_{2q,q}^{NF}$ | $\mathcal{F}ev(z)$ | $\frac{z}{\sqrt{1-4z}}$ | $2 \binom{2q-3}{q-2}$ | $4^q \frac{1}{4} \sqrt{\frac{1}{\pi q}}$ |
| $P_{2q,q-1}^{NF}$ | $\mathcal{S}\mathcal{F}ev(z)$ | $\frac{\frac{z}{2\sqrt{1-4z}} + \frac{z}{2(1-4z)} + \frac{2z^3}{(1-4z)\sqrt{1-4z}} + \frac{z^2}{\sqrt{1-4z}} + \frac{2z^4}{(1-4z)^2\sqrt{1-4z}}}{(1-4z)^2\sqrt{1-4z}}$ | $\binom{2q-3}{q-2} + 2^{2q-3} + (q-2) \binom{2q-2}{q-2} + 2 \binom{2q-5}{q-3} + \frac{(q-3)(q-2)}{3} \binom{2q-5}{q-3}$ | $4^q \frac{1}{96} \sqrt{\frac{q^3}{\pi}}$ |

We notice that the coefficients of the P_n^{NF} 's have the same asymptotic behavior as the coefficients of P_n^T 's, with a slightly smaller constant, $1/8$ or $1/4$ for $1/2$ and $1/96$ for $1/12$. Notice, in particular, that the results $P_{2q,q}^{NF} \sim \frac{1}{2} P_{2q,q}^T$ and $P_{2q+1,q}^{NF} \sim \frac{1}{4} P_{2q+1,q}^T$ comes from the identities.

$$\begin{aligned} 2 \binom{2q-3}{q-2} &= \frac{q}{2q-1} \binom{2q-1}{q} \\ (q+1) \binom{2q-3}{q-2} &= \frac{q+1}{2(2q-1)} (2q-1) \binom{2(q-1)}{q-1}. \end{aligned}$$

6 Generating functions for terms

We consider the vertical generating functions which gives the $T_{n,m}$'s for each value of m .

Vertical generating functions

We see that

$$T_{n,m+1} = T_{n+1,m} - \sum_{k=0}^n T_{n-k,m} T_{k,m}.$$

Hence

$$T^{(m)}(0) = 0$$

and

$$\begin{aligned} T^{(m+1)}(z) &= \sum_{n=0}^{\infty} T_{n,m+1} z^n \\ &= \sum_{n=0}^{\infty} T_{n+1,m} z^n - \sum_{n=0}^{\infty} \sum_{k=0}^n T_{n-k,m} T_{k,m} z^n \\ &= \frac{T^{(m)}(z)}{z} - (T^{(m)}(z))^2. \end{aligned}$$

In other words

$$z(T^{(m)}(z))^2 - T^{(m)}(z) + zT^{(m+1)}(z) = 0.$$

Hence

$$T^{(m)}(z) = \frac{1 - \sqrt{1 - 4z^2 T^{(m+1)}(z)}}{2z}.$$

Moreover

$$[z]T^{(m)}(z) = \frac{dT^{(m)}}{dz}(0) = m.$$

We see that $T^{(m)}$ is defined from $T^{(m+1)}$. Like the bivariate generating function $T(z, u)$, $T^{(m)}(z)$ is also difficult to study, because we have $T^{(m)}$ defined in term of $T^{(m+1)}$.

7 Conclusion

We have given several parameters on numbers of untyped lambda terms and untyped normal forms and proved or conjectured facts about them. On another direction, it could be worth to study typed lambda terms, whereas we have only analyzed untyped lambda terms in this paper.

References

- [1] Henk P. Barendregt. *The Lambda-Calculus, its syntax and semantics*. Studies in Logic and the Foundation of Mathematics. Elsevier Science Publishers B. V. (North-Holland), Amsterdam, 1984. Second edition.

- [2] Olivier Bodini, Danièle Gardy, and Bernhard Gittenberger. Lambda-terms of bounded unary height. *2011 Proceedings of the Eighth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, 2011.
- [3] Christine Choppy, Stéphane Kaplan, and Michèle Soria Soria. Complexity analysis of term-rewriting systems. *Theoret. Comput. Sci.*, 67(2-3):261–282, October 1989.
- [4] René David, Christophe Raffalli, Guillaume Theyssier, Katarzyna Grygiel, Jakub Kozik, and Marek Zaionc. Some properties of random lambda terms. *CoRR*, abs/0903.5505, 2009.
- [5] René David and Marek Zaionc. Counting proofs in propositional logic. *Arch. Math. Log.*, 48(2):185–199, 2009.
- [6] Nicolaas Govert de Bruijn. *Asymptotic Methods in Analysis*. North-Holland, Amsterdam, 1958.
- [7] Nicolaas Govert de Bruijn. Lambda calculus with nameless dummies, a tool for automatic formula manipulation, with application to the Church-Rosser theorem. *Proc. Koninkl. Nederl. Akademie van Wetenschappen*, 75(5):381–392, 1972.
- [8] Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2008.
- [9] Hervé Fournier, Danièle Gardy, Antoine Genitrini, and Marek Zaionc. Classical and intuitionistic logic are asymptotically identical. In Jacques Duparc and Thomas A. Henzinger, editors, *CSL*, volume 4646 of *Lecture Notes in Computer Science*, pages 177–193. Springer, 2007.
- [10] Hervé Fournier, Danièle Gardy, Antoine Genitrini, and Marek Zaionc. Tautologies over implication with negative literals. *Math. Log. Q.*, 56(4):388–396, 2010.
- [11] Donald E. Knuth. *Selected Papers on Analysis of Algorithms*, volume 102 of *CSLI Lecture Notes*. Stanford, California: Center for the Study of Language and Information, 2000.
- [12] Wolfdieter Lang. On polynomials related to derivatives of the generative functions of the Catalan numbers. *The Fibonacci Quarterly*, 40(4):299–313, 2002.
- [13] Jue Wang. Generating random lambda calculus terms. Technical report, Citeseer, 2005.
- [14] Marek Zaionc. Probability distribution for simple tautologies. *Theor. Comput. Sci.*, 355(2):243–260, 2006.

| $n \setminus m$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|-----------------|------------|------------|-------------|--------------|--------------|---------------|---------------|----------------|
| 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 3 | 2 | 4 | 8 | 14 | 22 | 32 | 44 | 58 |
| 4 | 4 | 12 | 26 | 46 | 72 | 104 | 142 | 186 |
| 5 | 13 | 38 | 87 | 172 | 305 | 498 | 763 | 1112 |
| 6 | 42 | 127 | 324 | 693 | 1294 | 2187 | 3432 | 5089 |
| 7 | 139 | 464 | 1261 | 2890 | 5831 | 10684 | 18169 | 29126 |
| 8 | 506 | 1763 | 5124 | 12653 | 27254 | 52671 | 93488 | 155129 |
| 9 | 1915 | 7008 | 21709 | 57070 | 130863 | 269260 | 508513 | 896634 |
| 10 | 7558 | 29019 | 94840 | 265129 | 646458 | 1406983 | 2791564 | 5136885 |
| 11 | 31092 | 124112 | 427302 | 1264362 | 3262352 | 7502892 | 15703602 | 30429782 |
| 12 | 132170 | 548264 | 1977908 | 6168242 | 16811366 | 40776020 | 89671904 | 181746638 |
| 13 | 580466 | 2491977 | 9384672 | 30755015 | 88253310 | 225197061 | 520076012 | 1104714147 |
| 14 | 2624545 | 11629836 | 45585471 | 156409882 | 471315501 | 1263116040 | 3058077451 | 6789961206 |
| 15 | 12190623 | 55647539 | 226272369 | 810506769 | 2558249963 | 7184911623 | 18208806189 | 42244969589 |
| 16 | 58083923 | 272486289 | 1146515237 | 4275219191 | 14098296495 | 41417170373 | 109721440529 | 265618096347 |
| 17 | 283346273 | 1363838742 | 5923639803 | 22933607180 | 78832280277 | 241776779298 | 668513708207 | 1686996660888 |
| 18 | 1413449148 | 6968881025 | 31177380822 | 125027527671 | 446961983408 | 1428444131853 | 4116538065930 | 10816530842627 |

Figure 1: Values of $T_{n,m}$ up to (18, 7)

| n | $P_n^T(m)$ |
|-----|--|
| 1 | m |
| 2 | $m + 1$ |
| 3 | $m^2 + m + 2$ |
| 4 | $3m^2 + 5m + 4$ |
| 5 | $2m^3 + 6m^2 + 17m + 13$ |
| 6 | $10m^3 + 26m^2 + 49m + 42$ |
| 7 | $5m^4 + 30m^3 + 111m^2 + 179m + 139$ |
| 8 | $35m^4 + 134m^3 + 405m^2 + 683m + 506$ |
| 9 | $14m^5 + 140m^4 + 652m^3 + 1658m^2 + 2629m + 1915$ |
| 10 | $126m^5 + 676m^4 + 2812m^3 + 7122m^2 + 10725m + 7558$ |
| 11 | $42m^6 + 630m^5 + 3610m^4 + 12760m^3 + 30783m^2 + 45195m + 31092$ |
| 12 | $462m^6 + 3334m^5 + 17670m^4 + 60240m^3 + 138033m^2 + 196355m + 132170$ |
| 13 | $132m^7 + 2772m^6 + 19218m^5 + 87850m^4 + 285982m^3 + 635178m^2 + 880379m + 580466$ |
| 14 | $1716m^7 + 16108m^6 + 104034m^5 + 449290m^4 + 1390246m^3 + 2991438m^2 + 4052459m + 2624545$ |
| 15 | $429m^8 + 12012m^7 + 99386m^6 + 560854m^5 + 2308173m^4 + 6895122m^3 + 14436365m^2 + 19144575m + 12190623$ |
| 16 | $6435m^8 + 76444m^7 + 584878m^6 + 3076878m^5 + 12039895m^4 + 34815210m^3 + 71170791m^2 + 92631835m + 58083923$ |
| 17 | $1430m^9 + 51480m^8 + 502384m^7 + 3389148m^6 + 16925916m^5 + 63753310m^4 + 179178860m^3 + 358339416m^2 + 458350525m + 283346273$ |
| 18 | $24310m^9 + 357256m^8 + 3176112m^7 + 19799164m^6 + 93981244m^5 + 342274990m^4 + 938333964m^3 + 1840448776m^2 + 2317036061m + 1413449148$ |

Figure 2: The polynomials P_n^T for the function $m \mapsto T_{n,m}$

| $n \setminus m$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-----------------|-----------|-----------|------------|-------------|-------------|--------------|--------------|---------------|---------------|
| 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 2 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 3 | 2 | 4 | 8 | 14 | 22 | 32 | 44 | 58 | 74 |
| 4 | 4 | 10 | 20 | 34 | 52 | 74 | 100 | 130 | 164 |
| 5 | 10 | 25 | 58 | 121 | 226 | 385 | 610 | 913 | 1306 |
| 6 | 25 | 72 | 185 | 400 | 753 | 1280 | 2017 | 3000 | 4265 |
| 7 | 72 | 223 | 614 | 1497 | 3244 | 6347 | 11418 | 19189 | 30512 |
| 8 | 223 | 728 | 2195 | 5716 | 12863 | 25688 | 46723 | 78980 | 125951 |
| 9 | 728 | 2549 | 8108 | 22745 | 56360 | 125093 | 253004 | 473753 | 832280 |
| 10 | 2549 | 9254 | 31253 | 93734 | 244997 | 564854 | 1173029 | 2237558 | 3983189 |
| 11 | 9254 | 35168 | 124778 | 395720 | 1109222 | 2770904 | 6261818 | 12999728 | 25130630 |
| 12 | 35168 | 138606 | 512898 | 1720040 | 5097660 | 13347978 | 31308206 | 66902388 | 132274680 |
| 13 | 138606 | 563907 | 2174894 | 7645095 | 23948550 | 66818531 | 167837142 | 384821079 | 816168830 |
| 14 | 563907 | 2369982 | 9459993 | 34771380 | 114618495 | 335857722 | 880524117 | 2092596528 | 4571548155 |
| 15 | 2369982 | 10231830 | 42221886 | 161568762 | 558056526 | 1723895502 | 4785906510 | 12073186866 | 28016723742 |
| 16 | 10231830 | 45381558 | 192944940 | 765787548 | 2764390146 | 8947158690 | 25962816408 | 68135021640 | 163627733358 |
| 17 | 45381558 | 206266797 | 901441688 | 3701763855 | 13912595562 | 47127027713 | 143678500332 | 397091138883 | 1005324501470 |
| 18 | 206266797 | 959283300 | 4302919895 | 18223902654 | 71123969121 | 251343711032 | 799893538635 | 2302171013970 | 6046781201429 |

Figure 3: Values of $F_{n,m}$ up to $(18, 8)$

| n | $P_n^{NF}(m)$ |
|-----|---|
| 1 | m |
| 2 | $m + 1$ |
| 3 | $m^2 + m + 2$ |
| 4 | $2m^2 + 4m + 4$ |
| 5 | $2m^3 + 3m^2 + 10m + 10$ |
| 6 | $6m^3 + 15m^2 + 26m + 25$ |
| 7 | $5m^4 + 12m^3 + 49m^2 + 85m + 72$ |
| 8 | $20m^4 + 62m^3 + 155m^2 + 268m + 223$ |
| 9 | $14m^5 + 50m^4 + 240m^3 + 589m^2 + 928m + 728$ |
| 10 | $70m^5 + 263m^4 + 870m^3 + 2146m^2 + 3356m + 2549$ |
| 11 | $42m^6 + 210m^5 + 1153m^4 + 3658m^3 + 8351m^2 + 12500m + 9254$ |
| 12 | $252m^6 + 1128m^5 + 4658m^4 + 14838m^3 + 33575m^2 + 48987m + 35168$ |
| 13 | $132m^7 + 882m^6 + 5446m^5 + 21198m^4 + 63138m^3 + 137695m^2 + 196810m + 138606$ |
| 14 | $924m^7 + 4862m^6 + 24086m^5 + 93748m^4 + 275898m^3 + 587814m^2 + 818743m + 563907$ |
| 15 | $429m^8 + 3696m^7 + 25372m^6 + 117120m^5 + 429435m^4 + 1223102m^3 + 2558090m^2 + 3504604m + 2369982$ |
| 16 | $3432m^8 + 20996m^7 + 121286m^6 + 556920m^5 + 2011411m^4 + 5601948m^3 + 11448828m^2 + 15384907m + 10231830$ |
| 17 | $1430m^9 + 15444m^8 + 116892m^7 + 624768m^6 + 2717670m^5 + 9524196m^4 + 26064412m^3 + 52459126m^2 + 69361301m + 45381558$ |
| 18 | $12870m^9 + 90683m^8 + 598120m^7 + 3162562m^6 + 13513606m^5 + 46329205m^4 + 124109404m^3 + 245453736m^2 + 319746317m + 206266797$ |

Figure 4: The polynomials P_n^{NF} for the function $m \mapsto F_{n,m}$