The Kelev

Shalosh B. EKHAD and Doron ZEILBERGER

Keeping track of the number of occurrences

So far we described how to find the exact number, let's call it $a_S(n)$ of compositions of length n, that **avoid** (i.e. do not contain) any members of the set of compositions $S = \{C_1, ..., C_r\}$. We gave an efficient algorithm, implemented in the Maple package `Compositions.txt` to explicitly find the generating function, let's call $f_S(x)$

$$f_S(x) := \sum_{n=0}^{\infty} a(n)\, x^n \quad .$$

Note, that out of laziness, so far we only implemented the case where all the members of the offending set, $S = \{C_1, ..., C_r\}$, are all of the same length.

But is very hard to stay out of trouble. Suppose that you want to find the exact number, let's call it $A_S(n; c_1, \dots, c_r)$ of compositions of n that contain

C_1, c_1 times, C_2, c_2 times, $\dots$, C_r, c_r times.

Of course, our former quantity, $a_S(n)$ is just the special case $c_1 = 0, c_2 = 0, \dots, c_r = 0$, i.e.

$$a_s(n) \;=\; A_S(n; 0, 0, \dots, 0) \quad .$$

All the information about the discrete function $A_S(n; c_1, \dots, c_r)$, with $1 + r$ *discrete* variables, is **encapsulated** in the multi-variable rational function, with $1+r$ '*continuous*' variables $x, X_1, \dots, X_r$

$$F_S(x; X_1, \dots, X_r) := \sum_{n=0}^{\infty} \sum_{c_1=0}^{\infty} \dots \sum_{c_r=0}^{\infty} A(n; c_1, \dots, c_r)\, x^n {X_1}^{c_1} \cdots {X_r}^{c_r} \quad .$$

Of course $f_S(x) = F_S(x; 0, 0, \dots, 0)$.

The beauty of the cluster method is that a very tiny tweak in the former algorithm yields a way to compute F_S. Rather then use the deep identity $0 = 1 + (-1)$ we use the only slightly deeper identity $X = 1 + (X - 1)$ applied to $X = X_1$, $X = X_2$, $\dots$, $X = X_r$. This entails redefining the *Poids* of a cluster x to the power the sum of its Skyline (as before)

times $(X_1 - 1)$ to the power the number of times C_1 shows up in the cluster,

times $(X_2 - 1)$ to the power the number of times C_2 shows up in the cluster,

$\dots$

times $(X_r - 1)$ to the power the number of times C_r shows up in the cluster.

The set of equations is modified accordingly, and Maple solves it. Of course, now it takes much longer, since we have so many more symbols, but the principle is the same.

This is implemented (still under the simplifying assumption of the members of the offending set all of the same length) in the Maple package `CompositionsPlus.txt` also available from the front of this article, where there are also sample input and output files.

Using the method of [Z], (that have been included in this package) one can do *statistical analysis*. One can, using the multi-variable generating function F_S, find expressions for the expectation, variance, (these are always linear) and higher moments (certain polynomials in n) of the random variable 'number of occurrences (as subcomposition) of C' defined on the *sample space* of the set of compositions of n. One can also find mixed moments for any set of such random variables, in particular, the *asymptotic correlation*, and confirm that for any such pair, these random variables are *joint asymptotically normal*, alas (of course) , **not** independently so. Using the asymptotic correlation one can confirm this by computing the mixed moments of the corresponding bi-variate normal distribution with correlation ρ,

$$\frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-x^2/2-y^2/2+\rho\, x\, y} \quad .$$

That our Maple package does automatically.

Just to cite one example, typing

```
InfoX2V([2,3,4],[4,3,2],x,X,Y,n,6);
```

in the Maple package `CompositionsPlus.txt` yield the following theorem

Theorem: The following statements are true

• Let $a(n)$ be the number of compositions of n that contain neither 234 nor 432, then

$$\sum_{n=0}^{\infty} a(n)\, x^n \;=\; -\frac{x^{16}+x^{15}+x^{12}+2\,x^{10}-x^7+x^5-x^4-x^2+2\,x-1}{x^{17}+x^{16}+x^{13}+2\,x^{11}-x^{10}+x^9-x^8+x^7-2\,x^5+x^4+2\,x^2-3\,x+1} \quad .$$

• $a(n)$ is asymptotic to $(0.548269839581\ldots)\cdot(1.976902834153\ldots)^n$

• Let $A(n;c,d)$ be the number of compositions of n that contain exactly c occurrences of 234 and d occurrences of 432, then

$$\sum_{n=0}^{\infty}\sum_{c=0}^{n-3}\sum_{d=0}^{n-3} A(n;c,d)\, x^n\, X^c\, Y^d \;=\; \frac{Numer(x,X,Y)}{Denom(x,X,Y)} \quad ,where$$

$Numer(x,X,Y) \;=\; -1+x^{12}-2\,x^{11}+x^3+2\,x^5-x^6-x^4-x^7-3\,x^2+x^8+2\,x^{10}-x^{17}+x^{15}X^2-x^{13}X^2$

$$-x^{12}X-x^{12}Y-x^{13}Y^2+x^{15}Y^2-x^{17}Y^2-x^{17}X^2-2\,x^{15}X-2\,x^{15}Y+2\,x^{11}Y-2\,x^{10}X+2\,x^{13}Y+2\,x^{11}X$$

$$-2\,x^{10}Y+2\,x^{13}X+2\,x^{17}Y+2\,x^{17}X+x^{15}+3\,x+x^7XY-x^{13}+x^6XY-x^8XY+x^4XY+x^{15}X^2Y^2-x^{17}X^2Y^2$$

$$+x^{13}XY^2+x^{13}X^2Y+x^{12}XY-3\,x^{13}XY-2\,x^{11}XY+2\,x^{10}XY-2\,x^{15}X^2Y-2\,x^{15}XY^2$$

$$+4\,x^{15}XY+2\,x^{17}X^2Y-2\,x^5XY-4\,x^{17}XY+2\,x^{17}XY^2 \quad ,$$

and

$$Denom(x,X,Y) \;=\; -1+2\,x^{12}-3\,x^{11}+2\,x^3+3\,x^5-2\,x^6-x^4-x^7-5\,x^2+2\,x^8+2\,x^{10}+2\,x^{16}Y+2\,x^{16}X-2\,x^{18}Y$$

$$-2\,x^{14}X-2\,x^{14}Y-2\,x^{18}X-2\,x^{12}X-2\,x^{12}Y+x^{14}X^2+x^9X+x^{14}Y^2+x^{18}Y^2-x^{16}Y^2+x^{18}X^2+x^9Y-x^{16}X^2$$

$$+3\,x^{11}Y-2\,x^{10}X+x^{13}Y+3\,x^{11}X-2\,x^{10}Y+x^{13}X+4\,x+2\,x^{16}XY^2+x^7XY+4\,x^{18}XY-4\,x^{16}XY$$

$$-x^{13}-2\,x^{18}XY^2+2\,x^6XY-2\,x^{18}X^2Y-2\,x^8XY+2\,x^{16}X^2Y+3\,x^{14}XY+x^4XY+2\,x^{12}XY-2\,x^9+x^{14}$$

$$-x^{13}XY-3\,x^{11}XY+2\,x^{10}XY-x^{14}X^2Y-x^{14}XY^2-x^{16}X^2Y^2+x^{18}X^2Y^2-3\,x^5XY+x^{18}-x^{16} \quad .$$

• The expectation and variance of the random variables ‘number of occurrences of 234’, and ‘number of occurrences of 432’, are both (obviously they are the same)

$$\frac{n}{128} \quad , \quad \frac{147}{16384}\,n-\frac{1439}{16384} \quad .$$

• The asymptotic correlation is $\frac{71}{147} = 0.482993197\ldots$, and the joint asymptotic normality (with that correlation) is confirmed up to the sixth mixed moments (not that we had any doubts).

Encore: The asymptotic growth constants for all compositions up that of 6

This is an excerpt from the output file

`http://sites.math.rutgers.edu/~zeilberg/tokhniot/oCompositions4.txt`, ranking them according to the asymptotic growth constants of the sequences enumerating compositions that do not contain them. We only list one of them for each equivalence class.

$n=2$: 2(1)

$n=3$: 12(1), 3(1.61803398874989484820).

$n=4$: 112(1), 13(1.6180339887), 22(1.7548776662), 4(1.8392867552)

$n=5$: 1112(1), 113(1.6180339887), 212(1.7548776662), 14(1.83928675521), 23(1.86676039917), 5(1.92756197548)

$n=6$: 11112(1), 1113(1.6180339887), 2112(1.7548776662), 114(1.839286755214), 213(1.866760399), 222(1.908790738787), 15(1.92756197548), 24(1.93318498189952), 33(1.9417130342786), 6(1.965948236645)

For the ranking for the compositions of up to 11 see the above output file.

References

[Z] Doron Zeilberger, *Automated Derivation of Limiting Distributions Of Combinatorial Random Variables Whose Generating Functions are Rational*, The Personal Journal of Shalosh B. Ekhad and Doron Zeilberger, Dec. 24, 2016.
`http://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/crv.html` .