On Coefficients of Powers of Functions

Danièle Gardy[1]

Technical Report No. CS-91-53

August, 1991

[1]L.R.I., CNRS-URA 410, Bât. 490, Université Paris XI, 91405 Orsay Cedex, France

On Coefficients of Powers of Functions

Danièle GARDY *
L.R.I., CNRS-URA 410,
Bât. 490, Université Paris XI,
91405 Orsay Cedex, France

Abstract
We study the n^{th} coefficient of a function $f(z)^d \psi(z)$ with real positive coefficients under the conditions that $d \to +\infty$ and $n = o(d)$.

1 Introduction

Generating functions of the kind $\phi(z) = f(z)^d$, where f is a function with positive coefficients and d is a parameter which tends to infinity, appear in some problems of discrete probability theory, combinatorial enumeration, coding theory, etc. These problems require to estimate the n^{th} coefficient of f^d, which we shall denote by $[z^n]\{f(z)^d\}$, for large n and d. This can frequently be done by the saddle point method, which gives an approximate value of this coefficient under general conditions on f. Previous papers have concentrated on the case where n grows at least as fast as d. In contrast, we present here results for the case where $n = o(d)$.

The plan of the paper is as follows: In order to unify the presentation, we first introduce some notations, then we recall former results and the basis of the technic we shall use (a saddle-point approximation) in Section 2. Section 3 presents our main contribution: the asymptotic approximation, for large n and d such that $n = o(d)$, of $[z^n]\{f(z)^d\}$ and of $[z^n]\{f(z)^d\ \psi(z)\}$, for a function ψ which is either independent of d or a product of "not too large" powers of functions, and which satisfies some further simple conditions. We give some examples of application in Section 4, before proving our results in Section 5.

2 Notations and related results

2.1 Notations

We consider in this paper functions of one variable with a power series expansion $f(z) = \sum_{k\geq 0} f_k z^k$, with real positive coefficients. We also assume that the radius of convergence of f is strictly positive. We present here some notations that we shall use in the sequel. We define two operators on a function f:

$$\Delta f(z) = z\frac{f'}{f}(z);$$

*This work was partially performed while the author was visiting at the Computer Science Department, Brown University, Providence R.I. 02912 (U.S.A.).

$$\delta f(z) = \frac{f''}{f}(z) - \frac{f'^2}{f^2}(z) + \frac{f'(z)}{zf(z)} = \frac{z(f(z)f''(z) - f'^2(z)) + f(z)f'(z)}{zf^2(z)}.$$

These two operators are related by: $z\delta f(z) = (\Delta f)'(z)$. As the function f has positive real coefficients, it is not difficult to show that, for all $z > 0$ for which it is defined, the value of $\delta f(z)$ is strictly positive; as a consequence *the function* Δf *is increasing.*

We give here a property of the operators Δ and δ that we shall use in the sequel:

If $f(0) \neq 0$ *and* $f'(0) \neq 0$, *then for* $z \to 0$ *we have* $\Delta f(z) = zf'(0)/f(0)(1 + O(z))$ *and* $\delta f(z) = f'(0)/(zf(0))(1 + O(z))$.

Proof: $f'(z)/(zf(z)) = f'(0)/(zf(0))(1 + o(1))$ and the two other terms of $\delta f(z)$ are $O(1)$. □

2.2 A general formula when $n = \Theta(d)$

The problem of finding the asymptotic value of $[z^n]\{f(z)^d\}$, when $n, d \to +\infty$ and n and d are roughly proportional, was studied for example by Daniels [2] and Greene and Knuth [10], mostly for probability generating functions. As noted by Good [9], this result is actually valid for a larger class of functions, such as entire functions or functions defined on an open disk; moreover it can be improved to give further terms of an asymptotic development. We give below the main result [2][9, p.868].

Theorem A:
Assume that n/d *belongs to an interval* $[a, b]$, $0 < a < b$. *Define* ρ *and* σ^2 *by* $\Delta f(\rho) = n/d$ *and* $\sigma^2 = \rho^2 \delta f(\rho)$. *Then:*

$$[z^n]\{f(z)^d\} = \frac{f(\rho)^d}{\sigma\rho^n\sqrt{2\pi d}}(1 + o(1)).$$

We can simplify Theorem A further; if $n = kd$ (i.e $\lambda = 0$ below) we get an approximation independent of n and d.

Corollary A:
If $n = kd$, *then*

$$[z^n]\{f(z)^d\} = \frac{A^d}{B\sqrt{d}}(1 + o(1))$$

and if $n = kd + \lambda$

$$[z^n]\{f(z)^d\} = \frac{A^d}{B\rho^\lambda\sqrt{d}}(1 + o(1)),$$

for suitable constants $A = f(\rho_0)\rho_0^{-k}$ *and* $B = \sigma\sqrt{2\pi}$, *and with* ρ_0 *the solution (independent of n and d) of* $\Delta f(z) = k$. *Note that* σ *too is a constant:* $\sigma^2 = \rho_0^2 \delta f(\rho_0)$.

Proof: Corollary A is obvious if $n = kd$; we just have to prove it when $n = kd + \lambda$. From Theorem A, we know that

$$[z^n]\{f^d(z)\} = \frac{f(\rho)^d}{\sigma\rho^n\sqrt{2\pi d}}(1 + o(1)),$$

with ρ defined by $\Delta f(\rho) = n/d$. Define ρ_0 by $\Delta f(\rho_0) = k$; we have $\rho_0 = \rho(1+\epsilon)$ with $\epsilon = O(1/d)$. Now the term σ_0 computed for ρ_0 is equal to $\sigma(1 + o(1))$ and

$$f(\rho_0)^d \rho_0^{-n} = f(\rho)^d \rho^{-n} e^{\epsilon(d\Delta f(\rho)-n)+O(d\epsilon^2)}.$$

But, from the definition of ρ, $d\Delta f(\rho) - n = 0$; moreover $O(d\epsilon^2) = O(1/d)$, which concludes the proof. □

The constant A is $f(\rho_0)\rho_0^{-k}$, with $k = n/d$. We can study its variation when the quotient n/d is restricted to a closed interval of $]0, +\infty[$. Let $A = A(k)$ with $k = n/d$. If we assume that $k \mapsto \rho_0 = \rho(k)$ is a differentiable function, then the derivative of A satisfies:

$$\frac{dA}{dk}(k) = A(k) \log \rho(k).$$

As $\rho(k)$ is defined by $\Delta f(\rho(k)) = k$ and as Δf is increasing, the function $k \mapsto \rho(k)$ is also increasing. We conclude that *on a suitable interval including $\Delta f(1)$, $A(k)$ is first increasing, then decreasing (i.e. is a unimodal function of k) with maximum $A(\Delta f(1)) = f(1)$.*

2.3 Function defined by an implicit equation

A recent paper by Meir and Moon [14] presents results about the coefficient of z^n in $f(z)^d$, when $d, n \to +\infty$ and with f defined by an implicit equation: $f(z) = z\phi(f(z))$ and $f(0) = 0$. Rephrased into the notations of this paper, their basic result is as follows:

Theorem B:
For $f(z)$ defined as above, and $d = \alpha n + \lambda\sqrt{n} + o(\sqrt{n})$, α a constant such that $0 \le \alpha < 1$ and λ a finite (positive or negative, possibly null) constant:

$$[z^n]\{f(z)^d\} = \frac{d}{n\sigma\sqrt{2\pi n}} e^{-\lambda^2/2\sigma^2} \rho^{d-n} \phi(\rho)^n (1 + o(1)),$$

where ρ is defined by $\Delta\phi(\rho) = 1 - \alpha$ and

$$\sigma^2 = \rho^2 \phi''(\rho)/\phi(\rho) + \alpha(1-\alpha) = \rho^2 \delta\phi(\rho).$$

Meir and Moon actually prove in passing the following result:

$$[t^n]\{\phi(t)^d\} = \frac{e^{-\lambda^2/2\sigma^2}}{\sigma\sqrt{2\pi d}} \frac{\phi(\rho)^d}{\rho^n}(1 + o(1))$$

and their range of validity is for $n = (1-\alpha)d + \lambda\sqrt{d} + o(\sqrt{d})$. For $\alpha > 0$ and $\lambda \neq 0$, this result is basically Theorem A applied to $n = kd + O(\sqrt{d})$ with a constant $k = 1 - \alpha \in]0, 1]$. Theorem B is then obtained by an application of the Lagrange inversion formula:

$$[z^n]\{f(z)^d\} = \frac{d}{n}[t^{n-d}]\{\phi(t)^n\}.$$

When $\alpha = 0$ but $\lambda > 0$, Theorem B gives an approximation valid for $d = \lambda\sqrt{n}(1 + o(1))$, i.e. $d^2 = \lambda^2 n(1 + o(1))$. If $\alpha = \lambda = 0$, the result holds for $d = o(\sqrt{n})$ i.e. $d^2 = o(n)$. Summing up, this means that Meir and Moon have results for $d \approx \alpha n$ ($\alpha \neq 0$) or for $d^2 = O(n)$ ($\alpha = 0$). This is the most general result currently available. However, it does not give bounds on $[z^n]\{f(z)^d\}$ when $n = o(d)$: Such a relation between n and d is not possible here, because $f(0) = 0$.

2.4 The underlying techniques: the saddle-point approximation and the bivariate Laplace's method

The papers [2, 9, 10] use a saddle-point approximation to derive their results. Its principle relies on Cauchy's formula for analytic functions, which expresses the coefficient of such a function as an integral on a curve of the complex plane, and on approximation technics for integrals which depend on a large parameter. We refer the reader to [1, 3] for a general presentation of these asymptotic technics, in particular of the "method of steepest descent" which is the basis of the saddle-point approximation; a good presentation is also given by Hayman [11][17, Ch.5]. We sum up what we need for our purpose below.

The coefficient of z^r in an analytic function $\phi(z)$ is given by Cauchy's formula:

$$[z^n]\phi(z) = \frac{1}{2i\pi}\oint \phi(z)\frac{dz}{z^{n+1}}dz.$$

The integration contour is a closed curve around the origin of the complex plane. This is an integral of the form $(1/2i\pi)\oint e^{h(z)}dz$ for $h(z) = \log\phi(z) - (n+1)\log(z)$.

The integration contour is chosen according to the function ϕ, in such a way that a small portion of it gives the major part of the integral. When the function ϕ is entire, we may choose for integration contour a circle centered at the origin and which goes through a *saddle point.* This point ρ is defined by an equation in h, and its interest lies in the fact that, on the circle going through this point, the integrand $e^{h(z)}$ is negligible most of the time, except in the vicinity of the saddle point. We can often prove that the integral on the curve differs from the integral restricted to a neighbourhood of the saddle point by an exponentially small term, and approximate this restricted integral by a gaussian integral whose value is known. This gives:

$$\frac{1}{2i\pi}\oint e^{h(z)}dz \approx \frac{e^{h(\rho)}}{\sqrt{2\pi h''(\rho)}}.$$

It can be shown that this approximation works for extracting the coefficient of z^n in a large class of functions $\phi(z) = f^d(z)$, when n/d belongs to an interval $[a, b]$ $(0 < a < b)$ [2, 9].

The saddle-point method as exposed above is related to Laplace's method for approximating an integral, which is usually applied to integrals depending on one parameter. Fulks [6] and Pederson [15] have also studied integrals depending on two parameters, which are in the same vein as our problem, although their results cannot be directly applied.

The paper of Meir and Moon [14], on the other hand, uses a related but different technique. It also begins by computing an approximate saddle-point ρ, which is the radius of the circle chosen as integration contour (as in the classical saddle-point method), then uses an argument of dominated convergence to get an approximate value of the desired coefficient. Note that the saddle point ρ is defined by $\Delta\phi(\rho) = 1 - \alpha$, which implies that $\alpha < 1$.

3 Approximations of coefficients

3.1 Conditions on the functions

Our goal is to compute an approximate value of the integral $I_{n,d} = [z^n]\{\phi_d(z)\}$, when the function ϕ_d depends on at least one parameter d, and when both n and d grow large. We shall study functions $\phi_d = f^d$ or $\phi_d = f^d \psi$; we give below the conditions that are required on f and ψ. We shall also require that the functions f and ψ do not have singularities inside the integration contour going through the saddle point. As $n = o(d)$, and as we shall see that the saddle-point becomes o(1), this simply requires that f and ψ be analytic near the origin.

Let us write $f(z) = \sum_k f_k z^k$. We shall assume in the sequel that the function f satisfy the following property:

Assumption $\mathcal{A}_1$:
The function f has real positive coefficients, is not affine and is such that $f_0 \neq 0$ and $f_1 \neq 0$, and that the GCD of the integers k such that $f_k \neq 0$ is 1. Moreover, f has a strictly positive radius of convergence.

The condition on the GCD can be stated in an equivalent form: There exists no entire function g and no integer $m \geq 2$ with $f(z) = g(z^m)$. The restriction on f_1 is a technical one, which unfortunately we have not been able to remove in the most general case. Although it is possible to extend our results to deal with the case where $f_1 = 0$, this imposes further conditions on the relative growths of n and d, and no general form for the coefficients seems to be available, so we omit this extension.

We shall also deal with a function ψ which may depend on d or on other parameters, as long as the following property is satisfied:

Assumption $\mathcal{A}_2$:
The function ψ has positive coefficients, such that $\psi(0) \neq 0$, has a strictly positive radius of convergence, and has the general form (p is any fixed integer):

$$\psi(z) = \prod_{i=1}^{p} g_i(z)^{d_i} \qquad \text{with} \qquad d_i = o(\frac{d}{\sqrt{n}}), 1 \leq i \leq p. \tag{1}$$

Anticipating on the discussion in Section 5.7, we can justify our condition on ψ as follows: An extra factor $\psi(z)$ moves the saddle-point away from the value ρ_0 obtained for f^d; as long as the new saddle-point ρ stays within $o(1/\sqrt{n})$ of ρ_0 this does not matter. The difference $\rho - \rho_0$ is $\Theta(\rho_0(\sum_i d_i/d))$, hence the condition (1).

3.2 The case n constant

We treat this case for the sake of completeness, although it presents no real difficulty. If n is constant, the saddle-point method as exposed below does not work;[1] however a direct analysis gives some results. For example, if $f(z) = \sum_k f_k z^k$, with $f_0 \neq 0$ and $f_1 \neq 0$, then we have the following result, which simply means that the first coefficients of $f(z)^d$ behaves as those of $(f_0 + f_1 z)^d$.

Theorem 1: *If the function f satisfies the assumption $\mathcal{A}_1$, then*

$$[z^n]\{f^d(z)\} = \binom{d}{n} f_0^{d-n} f_1^n (1 + O(1/d)).$$

[1]Some of the error terms in Propositions 1 and 2 of Section 5.1 are not $o(1)$: The term $h''(\rho)$ is constant.

If moreover the function ψ satisfies the assumption $\mathcal{A}_2$, then

$$[z^n]\{f^d(z)\psi(z)\} = \binom{d}{n} f_0^{d-n} f_1^n \psi(0)(1 + O(1/\sqrt{d})).$$

Proof: We can assume w.l.o.g. that f is a polynomial of degree at most n (if not, we truncate f to get $\hat{f}(z) = \sum_{i=0}^n f_i z^i$; any term z^n in $f^d(z)$ comes from $\hat{f}(r)^d$). We can also assume that $f_0 = 1$. We expand $f^d(z)$ using the multinomial formula:

$$f^d(z) = \sum_{i_j \geq 0; i_0 + i_1 + \cdots + i_n = d} \binom{d}{i_0, i_1, \cdots, i_n} f_1^{i_1} \cdots f_n^{i_n} z^{i_1 + 2i_2 + \cdots + ni_n}.$$

Hence:

$$[z^n]\{f^d(z)\} = \sum \binom{d}{i_0, i_1, \cdots, i_n} f_1^{i_1} \cdots f_n^{i_n}.$$

The summation is on all sets of indices $i_j \geq 0$ such that $i_0 + i_1 + \cdots + i_n = d$ and $i_1 + 2i_2 + \cdots + ni_n = n$. If we take into account the fact that $i_0 = d - i_1 - \cdots - i_n$, we get the equivalent expression

$$\sum_{i_1 + \cdots + i_n \leq d; i_1 + 2i_2 + \cdots + ni_n = n} \frac{d!}{i_1! \cdots i_n! (d - i_1 - \cdots - i_n)!} f_1^{i_1} \cdots f_n^{i_n}.$$

Each i_k for $k \geq 1$ is bounded by a constant: $i_k \leq n/k$. Hence, for fixed n, the number of possible sets of indices, i.e. the number of terms in the sum, is also bounded by a constant. The factors $f_1^{i_1} \cdots f_n^{i_n}$ are also $O(1)$. Each term of the sum is a polynomial function of d, of degree $s = i_1 + \cdots + i_n$ in d. The term of highest degree is obtained when s is maximal. Now $s = i_1 + i_2 + \cdots + i_n \leq i_1 + 2i_2 + \cdots ni_n = n$, and the upper bound is obtained for $i_1 = n$ and $i_2 = \cdots = i_n = 0$. This gives the term $\binom{d}{n} f_1^n$.

If $f_0 \neq 1$, then the coefficient has for asymptotic value $f_0^d \binom{d}{n} (f_1/f_0)^n = \binom{d}{n} f_0^{d-n} f_1^n$. The proof when there is an auxiliary function ψ is in the same vein. □

If we allow n to grow, both the number of terms in the sum and the factors $f_1^{i_1} \cdots f_n^{i_n}$ are unbounded, and the proof of Theorem 1 no longer holds.

3.3 The case n = o(d)

We study now the case where $n = o(d)$, but $n \to +\infty$. Theorem 2 is an extension of Theorem A to the case $n = o(d)$ and Theorem 3 deals with the extra factor ψ.

Theorem 2:
Let f satisfy the assumptions $\mathcal{A}_1$ of Section 3.1 and let $n = o(d)$, with $n, d \to +\infty$. Define ρ as the unique real positive solution of $\Delta f(z) = (n+1)/d$. Then:

$$[z^n]\{f(z)^d\} = \frac{f(\rho)^d}{\rho^n \sqrt{2\pi n}}(1 + o(1)).$$

Theorem 3:
Let f, ψ satisfy the assumptions $\mathcal{A}_1$ and $\mathcal{A}_2$ of Section 3.1 and let $n = o(d)$, with

$n, d \to +\infty$. Define ρ as the unique real positive solution of $\Delta f(\rho) + \Delta\psi(\rho)/d = (n+1)/d$. Then:

$$[z^n]\{f(z)^d \psi(z)\} = \frac{f(\rho)^d \cdot \psi(\rho)}{\rho^n \sqrt{2\pi n}}(1 + o(1)).$$

Theorem 4:
Theorems 2 and 3 still hold if we take ρ defined by $\Delta f(\rho) = n/d$.

We can simplify further Theorems 2 and 3 when we have more information on the respective growth orders of n and d. The following corollary, for example, deals with the case when $n = o(\sqrt{d})$.

Corollary 1:
If f satisfies the assumptions $\mathcal{A}_1$ of Section 3.1 and if $n = o(\sqrt{d})$, then:

$$[z^n]\{f^d(z)\} = \frac{f_0^d}{\sqrt{2\pi n}} \cdot \left(\frac{e f_1 d}{f_0 n}\right)^n (1 + o(1)).$$

Moreover, with ψ satisfying $\mathcal{A}_2$ but with the stronger conditions $d_i = o(d/n)$:

$$[z^n]\{f^d(z)\psi(z)\} = \psi(0)\frac{f_0^d}{\sqrt{2\pi n}} \cdot \left(\frac{e f_1 d}{f_0 n}\right)^n (1 + o(1)).$$

If the relationship $n = o(\sqrt{d})$ no longer holds, but some relationship $n^l = o(d^q)$ holds, it is possible to get a result similar to Corollary 1. This requires a better approximation of the saddle point ρ_0, and might become quite involved according to which coefficients of f are null, but it would be possible to work it out for a given function f. For example, we can prove the following result:

Corollary 2:
For f satisfying the assumptions $\mathcal{A}_1$ of Section 3.1 and for $n = o(d^{2/3})$, we have:

$$[z^n]\{f^d(z)\} = \frac{f_0^d}{\sqrt{2\pi n}} \left(\frac{e f_1 d}{f_0 n}\right)^n exp\left(\frac{n^2}{d}(\frac{f_2}{f_1} - \frac{1}{2})\right)(1 + o(1)).$$

If in addition ψ satisfies the assumptions $\mathcal{A}_2$ and the stronger conditions $d_i = o(d/n)$:

$$[z^n]\{f^d(z)\psi(z)\} = \psi(0)\frac{f_0^d}{\sqrt{2\pi n}} \left(\frac{e f_1 d}{f_0 n}\right)^n exp\left(\frac{n^2}{d}(\frac{f_2}{f_1} - \frac{1}{2})\right)(1 + o(1)).$$

Intuitively, Corollary 1 means that, when $n = o(\sqrt{d})$, the first two coefficients of f determine the main term in the asymptotic expression of the coefficients of f^d. This result can be compared to the relevant one for an affine function (although an affine function does not satisfy our assumptions): $[z^n]\{(f_0 + f_1 z)^d\} = \binom{d}{n} f_0^{d-n} f_1^n$; Stirling's approximation of binomial coefficients gives an approximation equivalent to Corollary 1. As n increases with respect to d, the other coefficients are progressively introduced, as in Corollary 2. However, if $n = d/\log(d)$ for example, we cannot find a relationship $n^l = o(d^q)$ and we have to take all the coefficients of f into account.

4 Applications

4.1 Binomial coefficients

We first check our results for binomial coefficients. For example, Theorem 2, applied to the function $f(z) = (1+z)^k$ for fixed $k \geq 2$, gives back the approximation of the binomial coefficient $\binom{kd}{n}$ which we obtain when we assume that kd, n and $kd - n$ all go to infinity and approximate the factorials with Stirling's approximation:

$$\binom{kd}{n} = \frac{(1+\rho)^{kd}}{\rho^n\sqrt{2\pi n}}(1+o(1)),$$

with $\rho = n/(kd-n)$, hence

$$\binom{kd}{n} = \frac{(kd)^{kd}}{n^n(kd-n)^{kd-n}\sqrt{2\pi n}}(1+o(1)).$$

We can also apply our results to $f(z) = e^z$: The saddle point is $\rho = n/d$ and we get:

$$[z^n]\{e^{dy}\} = \frac{e^n d^n}{n^n\sqrt{2\pi n}}(1+o(1)),$$

which is simply Stirling's formula again ($[z^n]\{e^{dy}\} = d^n/n!$).

4.2 Combinatorial enumeration

One of the basic constructions for obtaining combinatorial structures is to take a sequence of simpler objects. Let $f(z)$ be the generating function enumerating these basic objects according to their size; the generating function enumerating the composed objects in a sequence of d basic objects, according to their global size, is $f(z)^d$, and the coefficient $[z^n]\{f(z)^d\}$ enumerates the number of objects obtained as a sequence of d basic objects and which have size n. The same approach can also be used to analyze the abelian partitional complex, whose bivariate generating function has the form $exp(xf(y))$.

However, we have $f(0) = 0$ and $n \geq d$ (we do not count the structures of size 0). Let us define $f(y) = yg(y)$ with $g(0) \neq 0$; $[z^n]\{f^d(z)\} = [z^{n-d}]\{g^d(z)\}$. Our results can thus be applied to get the number of composed objects of size $n = d + o(d)$ formed by taking a sequence of d simpler objects.

Some examples are the Stirling numbers of first and second kind. Let $S_{n,k}$ be a Stirling number of second kind. Using either the ordinary bivariate function $\sum_{n,k} S_{n,k}x^k y^n = 1/(1 - x(e^y - 1))$ or the exponential function $\sum_{n,k} s_{n,k}x^k y^n/n! = exp(x(e^y - 1))$, we get:

$$S_{n,k} = \frac{n!}{k!}[y^{n-k}]\{f(y)^k\} \qquad \text{with} \qquad f(y) = (e^y - 1)/y.$$

So we can get an approximation for $n - k = o(k)$, i.e. $n = k + o(k)$. For $n = k + o(\sqrt{k})$ and $n - k \to +\infty$, Corollary 1 applied to $f(y) = (e^y - 1)/y = \sum_{n\geq 0} y^n/(n+1)!$ gives:

$$S_{n,k} = \frac{n!}{k!\sqrt{2\pi(n-k)}}\left(\frac{ek}{2(n-k)}\right)^{n-k}(1+o(1)),$$

i.e., using Stirling's approximation for $(n-k)!$ backwards:

$$S_{n,k} = \binom{n}{k}(k/2)^{n-k}(1+o(1)).$$

The generating function for Stirling numbers of first kind is $\hat{\Phi}(x,y) = \sum_{n,k} s_{n,k}x^k y^n/n! = exp(x\log(1/1-y))$; hence

$$s_{n,k} = \frac{n!}{k!}[y^{n-k}]\{f(y)^k\} \qquad \text{with} \qquad f(y) = \frac{1}{y}\log\frac{1}{1-y} = \sum_{n\geq 0}\frac{y^n}{n+1}.$$

We can thus get an asymptotic equivalent for $n = k + o(k)$, or equivalently $k = n - o(n)$. The saddle point ρ is approximately $2(n-k)/k$ and Corollary 1 gives, for $n = k + o(\sqrt{k})$:

$$s_{n,k} = \frac{n!}{k!\sqrt{2\pi(n-k)}}\left(\frac{ek}{2(n-k)}\right)^{n-k}(1+o(1)),$$

i.e.

$$s_{n,k} = \binom{n}{k}(k/2)^{n-k}(1+o(1)).$$

This is the same thing that we got for Stirling numbers of second kind: From Corollary 1, only f_0 and f_1 are important if $n = o(\sqrt{d})$. The average value of the $s_{n,k}$ for $0 \leq k \leq n$ is $H_n \sim \log n$, so we are here very far from the average.

If the difference $n-k$ becomes bigger, but stays small relative to k, we can still get an approximate value. For example, if $n = k + o(k^{2/3})$, we have:

$$\begin{aligned} S_{n,k} &= \frac{1}{\sqrt{2\pi(n-k)}}\left(\frac{ek}{2(n-k)}\right)^{n-k} e^{-(n-k)^2/6k}(1+o(1)); \\ s_{n,k} &= \frac{1}{\sqrt{2\pi(n-k)}}\left(\frac{ek}{2(n-k)}\right)^{n-k} e^{(n-k)^2/6k}(1+o(1)). \end{aligned}$$

4.3 Urn models and database profiles

A classical occupancy problem of discrete probability theory is as follows: *We throw n balls into k urns randomly and independently; what is the number of urns with at least one ball?* Let $N_{n,d}$ be the number of ways of assigning the n balls to exactly d urns. If the balls are undistinguishable and if the urns have unbounded capacity, the associate generating function is [12]:

$$\Phi(x,y) = \sum_{n,d} N_{n,d}\, x^d\, \frac{y^n}{n!} = (1+x(e^y-1))^k.$$

Let $f(y) = (e^y-1)/y$; we have

$$N_{n,d} = n!\,\binom{k}{d}\,[y^{n-d}]\{f(y)^d\}.$$

This can be expressed using Stirling numbers of second kind: $N_{n,d} = d!\binom{k}{d}S_{n,d}$.

Variations of this problem are related to the estimation of some parameters in relational databases: What is the size of the projection of a relation of known size? In many instances, we can map the k possible values of the projection attribute to the k urns and the tuples of the relation to the balls in such a way that the number of urns with at least one ball is the size of the projection; different constraints on relations correspond to different assumptions on the balls or the urns.

The probability that a relation of size r has a projection of size d is expressed in function of $[x^d y^r]\{(1-x+x\lambda(y))^k\}$, with k a parameter of the problem and $\lambda(y)$ a function describing the constraints on the relations and generally well-behaved [7]. We always have $d \leq r$ and $\lambda(0)=1$. Hence

$$[x^d y^r]\{(1-x+x\lambda(y))^d\} = \binom{k}{d}[y^{r-d}]\{f(y)^d\} \qquad \text{with} \qquad f(y)=(\lambda(y)-1)/y.$$

The average value of the projection size is linear in r; its maximal value is r. If $r-d$ is of the same order as d, we can apply Theorem A. This gives the probability that the size of the projection is roughly a constant fraction of the size of the initial relation. Our results give the probability that the projection has a size d close to the size r of the initial relation (but we still require that $r-d=o(\sqrt{r})$ and $r-d \to +\infty$). We can extend them to $r-d=o(r^{k/l})$, by taking more terms in the expansion of the saddle point.

5 Proofs of theorems

5.1 Intermediate results

We state a succession of intermediate results, from which we shall derive our theorems. We assume in all the Lemmas and Propositions below that $n, d \to +\infty$ and that $n=o(d)$.

Lemma 1: *Let f be a function satisfying the assumptions $\mathcal{A}_1$ of Section 3.1. Then the equation $\Delta f(z)=(n+1)/d$ has a real positive solution, which is equal to $f_0(n+1)/(f_1 d)(1+O(n/d))$.*

Proof of Lemma 1: It suffices to check that $\Delta f(0)=0$ and that the function $\Delta f(z)$ is continuous and increasing for $z \in [0,+\infty[$; this holds as the function f has positive coefficients. As we assume that $n=o(d)$, Lemma 1 implies that ρ becomes smaller than any fixed number. □

Lemma 2: *Let $h(z)$ be a function defined and analytic in an open set including the closed disk $\{|z| \leq \rho\}$, such that $h'(\rho)=0$ and $h''(\rho)$ is strictly positive, and such that $h(\rho e^{i\theta})$ has a Taylor expansion in terms of θ around 0:*

$$h(\rho e^{i\theta}) = h(\rho) + \frac{\rho^2}{2}h''(\rho)(e^{i\theta}-1)^2 + O(\rho^3(e^{i\theta}-1)^3 h'''(\rho\eta)).$$

Then there exists a constant α_0 independent of h and ρ, such that, for all α in $]0,\alpha_0[$, we have:

$$\frac{1}{2i\pi}\int_{|\theta|\leq\alpha} e^{h(\rho e^{i\theta})} d(\rho e^{i\theta}) = \frac{e^{h(\rho)}}{\sqrt{2\pi h''(\rho)}}\Big(1 + O(\sqrt{h''(\rho)}\alpha^2 \rho e^{-\rho^2\alpha^2 h''(\rho)/4}) + O(||h'''||\rho^3\alpha^3) + O(e^{-\rho^2\alpha^2 h''(\rho)/4})\Big).$$

We defer the proof of Lemma 2 until Section 5.3.

Proposition 1 *Under the assumptions $\mathcal{A}_1$ of Section 3.1, define the function h by $h(z) = d\log f(z) - (n+1)\log z$, and define ρ as the real positive solution of $\Delta f(\rho) = (n+1)/d$. Then, for a suitable alpha > 0 and for $C_\rho = f_0 f_1 \rho/(4f(\rho)(f_0 + f_1\rho))$:*

$$[z^n]\{f(z)^d\} = \frac{f(\rho)^d}{\rho^{n+1}\sqrt{2\pi d \cdot \delta f(\rho)}}(1 + O(||h'''||\rho^3\alpha^3) + O(\sqrt{h''(\rho)}\alpha^2\rho e^{-\rho^2\alpha^2 h''(\rho)/4})$$
$$+O(e^{-\rho^2\alpha^2 h''(\rho)/4}) + O(\rho\sqrt{h''(\rho)}e^{-C_\rho d\alpha^2})).$$

Proposition 1 is proved in Section 5.2.

Proposition 2 *We require the assumptions $\mathcal{A}_1$ and $\mathcal{A}_2$ of Section 3.1. Define the function h by $h(z) = d\log f(z) - (n+1)\log z + \log\psi(z)$ For a suitable alpha > 0, for ρ defined by $\Delta f(\rho) + \Delta\psi(\rho)/d = (n+1)/d$, and for $C_\rho = f_0 f_1\rho/4f(\rho)(f_0 + f_1\rho)$:*

$$[z^n]\{f(z)^d\psi(z)\} = \frac{f(\rho)^d\psi(\rho)}{\rho^{n+1}\sqrt{2\pi h''(\rho)}}(1 + O(||h'''||\rho^3\alpha^3) + O(e^{-\rho^2\alpha^2 h''(\rho)/4})$$
$$+O(\sqrt{h''(\rho)}\alpha^2\rho e^{-\rho^2\alpha^2 h''(\rho)/4}) + O(\rho\sqrt{h''(\rho)}e^{-C_\rho d\alpha^2})).$$

Proposition 2 is proved in Section 5.4.

Proposition 3 *Propositions 1 and 2 are still valid if we choose for saddle point the approximate value defined by $\Delta f(z) = n/d$. In fact, they are still valid if we substitute $\rho = \rho_0(1 + o(n^{-1/2}))$ for the exact saddle point ρ_0.*

5.2 Proof of Proposition 1

The proof of Proposition 1 is very similar to that of Theorem A. However, Theorem A is valid for n and d proportional; we want to extend this when n and d both grow large, but are not necessarly proportional. Although the main term is the same, this requires more precision in the error terms, which we give here in extenso.

From Cauchy's formula, we have $I_{n,d} = [z^n]\{f^d(z)\} = (1/2i\pi)\oint f^d(z)z^{-n-1}dz$, where the integration contour is a simple curve around the origin. We shall integrate on a circle, whose radius is the saddle point defined below.

5.2.1 The saddle point

Define the function

$$h(z) = d\log f(z) - (n+1)\log z.$$

We solve the equation $h'(z) = 0$, which we first rewrite as:

$$\Delta f(z) = \frac{n+1}{d}. \tag{2}$$

From Lemma 1, this equation has a unique real positive solution ρ, with approximate value:

$$\rho = \frac{f_0(n+1)}{f_1 d}(1 + O(\frac{n}{d})).$$

We place here some computation that we shall need later on. As

$$z^2 h''(z) = z^2 d \frac{f''(z)}{f(z)} - d\Delta f(z)^2 + n + 1,$$

and as, by definition of ρ, $\Delta f(\rho) = (n+1)/d$, we have

$$\rho^2 h''(\rho) = \rho^2 d \frac{f''(\rho)}{f(\rho)} + (n+1)(1 - \frac{n+1}{d}).$$

Now $\rho^2 d f''(\rho)/f(\rho) = O(\rho^2 d) = O(n^2/d)$, and

$$\rho^2 h''(\rho) = \rho^2 d \delta f(\rho) = (n+1)(1 + O(n/d)). \tag{3}$$

5.2.2 Main part of the integral

The condition on the domain where f is analytic ensures that we can integrate f on the circle $z = \rho e^{i\theta}$, as soon as n and d are large enough for ρ to be close enough to 0.

We choose for integration contour a circle of radius ρ and center the origin:

$$I_{n,d} = \frac{1}{2i\pi} \int_{\theta \in [-\pi, +\pi]} e^{h(\rho e^{i\theta})} d(\rho e^{i\theta}).$$

Let $\alpha \in]0, \pi[$. We divide the integral in two parts: $I_{n,d} = I_1 + I_2$, with:

$$\begin{aligned} I_1 &= \frac{1}{2i\pi} \int_{|\theta| \le \alpha} e^{h(\rho e^{i\theta})} d(\rho e^{i\theta}); \\ I_2 &= \frac{1}{2i\pi} \int_{\alpha \le |\theta| \le \pi.} e^{h(\rho e^{i\theta})} d(\rho e^{i\theta}). \end{aligned}$$

We shall compute an approximation of I_1 from Lemma 2, and show that I_2 can be neglected.

By definition of ρ, $h'(\rho) = 0$. We know that the value of h'' at the point ρ defined by the equation (2) is strictly positive: $h''(\rho) = n(1 + o(1))$. Hence we can apply Lemma 2 to $h(z) = d \log f(z) - (n+1) \log z$; we get:

$$I_1 = \frac{e^{h(\rho)}}{\sqrt{2\pi h''(\rho)}} (1 + O(||h'''|| \rho^3 \alpha^3) + O(\sqrt{h''(\rho)} \alpha^2 \rho e^{-\rho^2 \alpha^2 h''(\rho)/4}) + O(e^{-\rho^2 \alpha^2 h''(\rho)/4})), \tag{4}$$

where the main term can also be written as $f(\rho)^d / (\rho^{n+1} \sqrt{2\pi h''(\rho)})$.

5.2.3 Rest of the integral

We are now interested in the remaining term $I_2 = (1/2i\pi) \int_{\alpha \le |\theta| \le \pi} e^{h(z)} dz$, for $z = \rho e^{i\theta}$. We isolate the main factor corresponding to the approximate value of I_1 we have just computed:

$$I_2 = \frac{e^{h(\rho)}}{2i\pi} \int_{\alpha \le |\theta| \le \pi} e^{h(\rho e^{i\theta}) - h(\rho)} d(\rho e^{i\theta}).$$

Recall that $h(z) = d\log f(z) - (n+1)\log z$; this gives:

$$h(\rho e^{i\theta}) - h(\rho) = d(\log f(\rho e^{i\theta}) - \log f(\rho)) - (n+1)i\theta.$$

Hence:

$$I_2 = \frac{e^{h(\rho)}}{2\pi}\rho e^{-in\theta}\int_{\alpha\leq|\theta|\leq\pi}\left(\frac{f(\rho e^{i\theta})}{f(\rho)}\right)^d d\theta$$

and

$$|I_2| \leq \frac{e^{h(\rho)}}{2\pi}\rho\int_{\alpha\leq|\theta|\leq\pi}|\frac{f(\rho e^{i\theta})}{f(\rho)}|^d d\theta. \tag{5}$$

Our problem is now to get an upper bound of $f(\rho e^{i\theta})$ valid on the set $\{\alpha \leq |\theta| \leq \pi\}$. We use the fact that the function f has positive coefficients and that the GCD of its coefficients is equal to 1, to deduce that f attains its modulus on the real axis and nowhere else. More precisely, we can show that there exists a strictly positive factor C, independent of α, such that we have for all θ satisfying $\alpha \leq |\theta| \leq \pi$: $|f(\rho e^{i\theta})| \leq f(\rho)(1 - C\alpha^2)$. However, C depends on the saddle point ρ, so we shall denote it by C_ρ.

To get an expression of C_ρ, we use the assumption $f_1 \neq 0$. Let $f(z) = \sum_{k\geq 0} f_k z^k$, with real positive coefficients; we have:

$$|f(\rho e^{i\theta})| = |\sum_k f_k\rho^k e^{ik\theta}| \leq |f_0 + f_1\rho e^{i\theta}| + |\sum_{k\geq 2} f_k\rho^k e^{ik\theta}|.$$

The term $|\sum_{k\geq 2} f_k\rho^k e^{ik\theta}|$ is bounded by $\sum_{k\geq 2} f_k\rho^k$, which is $f(\rho) - (f_0 + f_1\rho)$. This shows that:

$$|f(\rho e^{i\theta})| \leq |f_0 + f_1\rho e^{i\theta}| + f(\rho) - (f_0 + f_1\rho).$$

We next get an upper bound on $|f_0 + f_1\rho e^{i\theta}|$:

$$|f_0 + f_1\rho e^{i\theta}|^2 = (f_0 + f_1\rho)^2(1 - c_\rho(1 - \cos\theta)),$$

with $c_\rho = 2f_0f_1\rho/(f_0 + f_1\rho)^2$. The term c_ρ has an upper bound valid for all real ρ, but may become close to 0 when ρ itself is close to 0. Moreover, $1 - c_\rho(1-\cos\theta) \leq 1 - c_\rho(1 - \cos\alpha)$ for $\alpha \leq |\theta| \leq \pi$. We deduce that:

$$|f_0 + f_1\rho e^{i\theta}| = (f_0 + f_1\rho)\cdot\sqrt{1 - c_\rho(1-\cos\alpha)}.$$

Using the inequalities $1 - \cos\alpha \geq \alpha^2/4$ for α close to 0 and $\sqrt{1-y} \leq 1 - y/2$ for all y, we get:

$$\sqrt{1 - c_\rho(1-\cos\alpha)} \leq \sqrt{1 - c_\rho\frac{\alpha^2}{4}} \leq 1 - c_\rho\frac{\alpha^2}{8}.$$

Returning to f, we have:

$$|f(\rho e^{i\theta})| \leq f(\rho) - (f_0 + f_1\rho)c_\rho\frac{\alpha^2}{8},$$

i.e.

$$|f(\rho e^{i\theta})| \leq f(\rho) - \frac{f_0f_1\rho}{4(f_0 + f_1\rho)}\alpha^2.$$

This yields finally:

$$|\frac{f(\rho e^{i\theta})}{f(\rho)}| \leq 1 - \frac{f_0 f_1 \rho}{4f(\rho)(f_0 + f_1\rho)}\alpha^2.$$

Plugging this inequality into equation (5), and using the fact that $1 - x \leq e^{-x}$, we get

$$|I_2| \leq \frac{e^{h(\rho)}}{\sqrt{2\pi h''(\rho)}} \cdot O(\rho\sqrt{h''(\rho)}e^{-C_\rho d\alpha^2}) \qquad \text{with} \qquad C_\rho = \frac{f_0 f_1 \rho}{4f(\rho)(f_0 + f_1\rho)}. \tag{6}$$

Equations (4) and (6) can now be used to get a global evaluation of $I_{n,d}$:

$$\begin{aligned} I_{n,d} &= \frac{e^{h(\rho)}}{\sqrt{2\pi h''(\rho)}}(1 + O(||h'''||\rho^3\alpha^3) + O(\sqrt{h''(\rho)}\alpha^2\rho e^{-\rho^2\alpha^2 h''(\rho)/4}) \\ &\qquad + O(e^{-\rho^2\alpha^2 h''(\rho)/4}) + O(\rho\sqrt{h''(\rho)}e^{-C_\rho d\alpha^2})). \end{aligned}$$

5.3 Proof of Lemma 2

The function $h(\rho e^{i\theta})$, considered as a function of θ, has a Taylor expansion around 0. By definition of ρ, $h'(\rho)$ is null and we have for $z = \rho e^{i\theta}$:

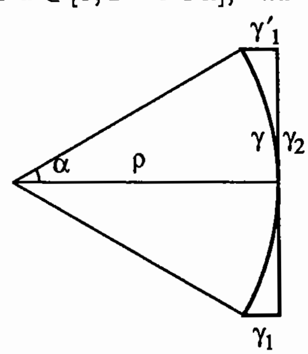
$$h(z) = h(\rho) + \frac{1}{2}(z-\rho)^2 h''(\rho) + O(||h'''||(z-\rho)^3).$$

The error term depends on the value of h''' at some point between z and ρ.

The integral $I_1 = (1/2i\pi)\int_{|\theta|\leq\alpha} e^{h(\rho e^{i\theta})}d(\rho e^{i\theta})$ can be expressed as an integral on $\gamma = \{z = \rho e^{i\theta}, |\theta| \leq \alpha\}$:

$$\begin{aligned} I_1 &= \frac{1}{2i\pi}\int_\gamma e^{h(z)}dz \\ &= \frac{e^{h(\rho)}}{2i\pi}\int_\gamma e^{\frac{1}{2}(z-\rho)^2 h''(\rho) + O(||h'''||(z-\rho)^3)}dz \\ &= \frac{e^{h(\rho)}}{2i\pi}(1 + O(||h'''||\rho^3\alpha^3))\int_\gamma e^{\frac{1}{2}(z-\rho)^2 h''(\rho)}dz. \end{aligned}$$

This last integral can then be approximated by a gaussian integral whose value is known. One way to do it consists in substituting the path $\gamma_1 \cup \gamma_2 \cup \gamma_1'$ to the arc γ: $\gamma_1 = \{z = \rho(1-u) - i\rho\sin\alpha\}$ and $\gamma_1' = \{z = \rho(1-u) + i\rho\sin\alpha\}$ for $u \in [0, 1-\cos\alpha]$, and $\gamma_2 = \{z = \rho + i\rho t\}$ for $|t| \leq \sin\alpha$.



We then show that the integrals on γ_1 and γ_1' give only error terms, and that the main part of I_1 comes from the integral on γ_2. On γ_1, $z = \rho(1-u) - i\rho\sin\alpha$ and $dz = -\rho du$, which gives:

$$\int_{\gamma_1} e^{(z-\rho)^2 h''(\rho)/2} dz = \rho \int_0^{1-\cos\alpha} e^{\rho^2(u+i\sin\alpha)^2 h''(\rho)/2} du. \tag{7}$$

The modulus of the right part of the equation (7) can be bounded by:

$$\rho e^{-\rho^2 h''(\rho)\sin^2\alpha/2} \int_0^{1-\cos\alpha} e^{\rho^2 h''(\rho) u^2/2} du.$$

Now:

$$\int_0^{1-\cos\alpha} e^{\rho^2 h''(\rho) u^2/2} du \le (1-\cos\alpha) e^{\rho^2 h''(\rho)(1-\cos\alpha)^2/2} \le O(\alpha^2 e^{\rho^2 h''(\rho)(1-\cos\alpha)^2/2}).$$

This gives an upper bound on the integral on γ_1:

$$\int_{\gamma_1} e^{(z-\rho)^2 h''(\rho)/2} dz = O(\rho\alpha^2 e^{\rho^2((1-\cos\alpha)^2-\sin^2\alpha)h''(\rho)/2}).$$

Now $(1-\cos\alpha)^2 - \sin^2\alpha = -\alpha^2 + O(\alpha^4)$, and we have an upper bound $(1-\cos\alpha)^2 - \sin^2\alpha \le -\alpha^2/2$ for α less than some constant $\alpha_0 > 0$. The error term becomes $O(\rho\alpha^2 e^{-\rho^2 h''(\rho)\alpha^2/4})$ as soon as $\alpha < \alpha_0$. As the Taylor expansion of h is valid for α close to 0, we can assume in the sequel that this condition is satisfied. The integral on γ_1' is computed in a similar way and gives the same bound.

We now compute the integral on γ_2:

$$\int_{\gamma_2} e^{(z-\rho)^2 h''(\rho)/2} dz = i\rho \int_{|t|\le\sin\alpha} e^{-\rho^2 h''(\rho)\, t^2/2} dt.$$

A change of variable $v = \rho\sqrt{h''(\rho)}\, t$ gives:

$$\int_{\gamma_2} e^{(z-\rho)^2 h''(\rho)/2} dz = (i/\sqrt{h''(\rho)} \int_{|v|\le\rho\sqrt{h''(\rho)}\sin\alpha} e^{-v^2/2} dv,$$

which can be extended to a gaussian integral as soon as $\rho\sqrt{h''(\rho)}\sin\alpha \to +\infty$. We also use the fact that $\sin^2\alpha > \alpha^2/2$ for α close to 0, to simplify the error term $O(e^{-\rho^2 h''(\rho)\sin^2\alpha/2})$, which comes from extending the integral to the real axis, into $O(e^{-\rho^2 h''(\rho)\alpha^2/4})$:

$$\int_{|v|\le\rho\sqrt{h''(\rho)}\sin\alpha} e^{-v^2/2} dv = \sqrt{2\pi} + O(e^{-\rho^2 h''(\rho)\alpha^2/4}).$$

Summing up, we get:

$$\int_\gamma e^{(z-\rho)^2 h''(\rho)/2} dz = i\sqrt{\frac{2\pi}{h''(\rho)}}(1 + O(e^{-\rho^2 h''(\rho)\alpha^2/4}) + O(\rho\alpha^2\sqrt{h''(\rho)}e^{-\rho^2 h''(\rho)\alpha^2/4})),$$

which completes the proof of Lemma 2.

5.4 Proof of Theorem 2

Proposition 1 now allows us to Prove Theorem 2. Recall that $f_1 \neq 0$ and that $n = o(d)$. The approximate value of ρ is $\rho = f_0(n+1)/(f_1 d)(1 + O(n/d))$. By equation (3), we also know that

$$\rho^2 h''(\rho) = n(1 + o(1)).$$

We use this information to simplify the error terms of Proposition 1:

- The term $O(e^{-\rho^2\alpha^2 h''(\rho)/4})$ is $O(e^{-an\alpha^2})$ for some constant $a > 0$.
- The term $O(\sqrt{h''(\rho)}\alpha^2\rho e^{-\rho^2\alpha^2 h''(\rho)/4})$ simplifies into $O(\alpha^2\sqrt{n}e^{-an\alpha^2})$.
- The expression $C_\rho = f_1\rho/(4f_0)(1 + o(1))$ is $n/(4d)(1 + o(1))$; the term $O(\rho\sqrt{h''(\rho)}e^{-C_\rho d\alpha^2})$ is $O(\sqrt{n}e^{-an\alpha^2})$ (we can choose the constant a so that it is the same in all three error terms).
- As $e^{-an\alpha^2} = O(\sqrt{n}e^{-an\alpha^2})$ and $\alpha^2\sqrt{n}e^{-an\alpha^2} = O(\sqrt{n}e^{-an\alpha^2})$, the last three error terms of Proposition 1 are $O(\sqrt{n}e^{-an\alpha^2})$.

We now study h''' around ρ: $h'''(z) = d(f'/f)''(z) - 2(n+1)/z^3 = -2(n+1)/z^3 + O(d)$. This shows that $\rho^3||h'''|| = \Theta(n)$; hence $O(||h'''||\rho^3\alpha^3) = O(n\alpha^3)$. All these simplifications give finally:

$$I_{n,d} = \frac{e^{h(\rho)}}{\sqrt{2\pi h''(\rho)}}(1 + O(n\alpha^3) + O(\sqrt{n}e^{-an\alpha^2})).$$

Now we choose α such that $n\alpha^3 \to 0$ and $n\alpha^2/\log n \to +\infty$. If we take for example $\alpha = \log n/\sqrt{n}$, the error terms in the above formula are $o(1)$. We then use the relations $e^{h(\rho)} = f^d(\rho)/\rho^{n+1}$ and $\rho^2 h''(\rho) = n(1 + o(1))$ to get the final formula:

$$[z^n]\{f^d(z)\} = \frac{f^d(\rho)}{\rho^n\sqrt{2\pi n}}(1 + o(1)).$$

5.5 Proof of Proposition 2

In this section we introduce a function ψ which satisfies the assumptions $\mathcal{A}_2$ of Section 3.1. To compute $I_{n,d} = [z^n]\{f^d\psi\}$, we define

$$h(z) = d\log f(z) + \log\psi(z) - (n+1)\log z.$$

As in Section 5.2, we have $I_{n,d} = (1/2i\pi)\oint e^{h(z)}dz$. The saddle point ρ is defined by $h'(z) = 0$, i.e. by the equation

$$\Delta f(z) + \Delta\psi(z)/d = (n+1)/d. \tag{8}$$

From the assumption (1) we have $\Delta\psi(z) = \sum_{i=1}^{p} d_i\Delta g_i(z)$ and $d_i/d = o(1)$. Hence $\Delta\psi(z)/d = o(z)$; as $\Delta f(z) = f_1 z/f_0(1 + O(z))$, this is also $\Delta\psi(z)/d = o(\Delta f(z))$. The saddle point ρ is then close to the solution of $\Delta f(z) = (n+1)/d$ and is $O(n/d)$. For large n and d such that $n = o(d)$, we can assume that all singularities of ψ are further than ρ. We then choose for integration contour the circle $z = \rho e^{i\theta}$, and proceed as in Section 5.2.

For $\alpha \in]0, \pi[$, we divide the integral in two parts: $I_{n,d} = I_1 + I_2$; as in Section 5.2 I_1 is the integral on $|\theta| \leq \alpha$ and I_2 is the integral for $\alpha \leq |\theta| \leq \pi$. Denote by γ the part of the integration circle for $|\theta| \leq \alpha$. The integral on $\gamma = \{z = \rho e^{i\theta}, |\theta| \leq \alpha\}$ is:

$$\begin{aligned} I_1 &= \frac{e^{h(\rho)}}{2i\pi} \int_\gamma e^{\frac{1}{2}(z-\rho)^2 h''(\rho) + O(||h'''||(z-\rho)^3)} dz \\ &= \frac{e^{h(\rho)}}{2i\pi} (1 + O(||h'''|| \rho^3 \alpha^3)) \int_\gamma e^{\frac{1}{2}(z-\rho)^2 h''(\rho)} dz. \end{aligned}$$

This last integral can be approximated using Lemma 2. The value of h'' at the point ρ defined by the equation (8) is:

$$h''(\rho) = d\,\delta f(\rho) + \delta\psi(\rho).$$

The values of $\delta f(\rho)$ and of $\delta\psi(\rho)$ are strictly positive when the initial functions f and ψ have real positive coefficients; this shows that $h''(\rho)$ also is strictly positive. More precisely, we have $h''(\rho) = d\ \delta f(\rho)(1 + o(1)) = df_1/(\rho f_0)(1 + o(1))$ and $\rho = (n + 1)f_0/(df_1)(1 + o(1))$, hence $\rho^2 h''(\rho) = (n + 1)(1 + o(1))$.

Lemma 2 applies again and we obtain the approximation

$$I_1 = \frac{e^{h(\rho)}}{\sqrt{2\pi h''(\rho)}} (1 + O(||h'''|| \rho^3 \alpha^3) + O(\sqrt{h''(\rho)} \alpha^2 \rho e^{-\frac{\rho^2}{4}\alpha^2 h''(\rho)}) + O(e^{-\frac{\rho^2}{4}\alpha^2 h''(\rho)})).$$

As for I_2,

$$|I_2| \leq \frac{e^{h(\rho)}}{2\pi} \rho \int_{\alpha \leq |\theta| \leq \pi} |\frac{f(\rho e^{i\theta})}{f(\rho)}|^d \cdot |\frac{\psi(\rho e^{i\theta})}{\psi(\rho)}| d\theta.$$

The function ψ has positive coefficients and satisfies $|\psi(\rho e^{i\theta})| \leq |\psi(\rho)|$. The inequality (6) was proved without using the fact that the radius of the integration circle is a saddle point, and still holds. The approximation of I_1 and the bound on I_2 now give an evaluation of $I_{n,d}$.

5.6 Proof of Theorem 3

As in Section 5.4, the approximate value of ρ is $\rho = f_0(n + 1)/(f_1 d)(1 + O(n/d))$ and we have $\rho^2 h''(\rho) = n(1 + o(1))$. A similar line of reasoning shows that the last three error terms of Proposition 2 are $O(\sqrt{n} e^{-n\alpha^2})$. Again, $||h'''||$ is of order $\Theta(d^3/n^2)$ around the saddle point, and $\rho^3 ||h'''|| = \Theta(n)$; hence $O(||h'''|| \rho^3 \alpha^3) = O(n\alpha^3)$. As before, we choose $\alpha = \log n/\sqrt{n}$; then $n\alpha^3 \to 0$ and $n\alpha^2/\log n \to +\infty$; the error terms in the approximate expression of $I_{n,d}$ are $o(1)$ and

$$[z^n]\{f(z)^d \psi(z)\} = \frac{f(\rho)^d \psi(\rho)}{\rho^n \sqrt{2\pi n}} (1 + o(1)).$$

5.7 Proof of Proposition 3

Define ρ_0 as the (exact) solution of $h'(z) = 0$, and let $\rho = \rho_0(1 + \eta)$, with $\eta = o(1)$. The question is: How large can η be, for the approximations of Propositions 1 and 2 still to be valid? In other words, what is the error when we substitute ρ for ρ_0 in the main term $f(\rho_0)^d/(\rho_0^{n+1}\sqrt{2\pi h''(\rho_0)})$?

We have $\rho^{n+1} = \rho_0^{n+1} e^{(n+1)\eta + O(n\eta^2)}$ and $f(\rho)^d = f(\rho_0)^d e^{d\eta\Delta f(\rho_0) + O(dC\eta^2\rho^2)}$ with $C = f''/f - f'^2/f^2$ applied at point ρ. Recall that $\Delta f(\rho_0) = (n+1)/d$; we also have $h''(\rho) = h''(\rho_0)(1+o(1))$ and $C = O(1)$. Then:

$$\frac{f(\rho)^d}{\rho^{n+1}\sqrt{2\pi d h''(\rho)}} = \frac{f(\rho_0)^d}{\rho_0^{n+1}\sqrt{2\pi d h''(\rho_0)}}(1 + O(d\rho_0^2\eta^2) + O(n\eta^2) + o(1)). \tag{9}$$

If the substitution of ρ_0 for ρ is valid, all the error terms must be $o(1)$: $d\eta^2\rho_0^2 = o(1)$ and $n\eta^2 = o(1)$. We assumed that $n = o(d)$, so $\rho_0 = \Theta(n/d)$ and $d\eta^2\rho_0^2 = o(n\eta^2)$. We finally get the condition $n\eta^2 = o(1)$, which is simply $\eta = o(1/\sqrt{n})$.

So we can substitute $\rho = \rho_0(1 + o(1/\sqrt{n}))$ for the exact saddle point ρ_0 in Proposition 1 and get the same approximation of $[z^n]\{f(z)^d\}$. In particular, the solution ρ of the equation $\Delta f(z) = n/d$ is related to the solution ρ_0 of the equation $\Delta f(z) = (n+1)/d$ by: $\rho = \rho_0(1 + O(1/n))$ and we can use indifferently ρ or ρ_0 in Proposition 1.

If there is a function ψ, the saddle point ρ_0 in Proposition 2 is defined by the equation $\Delta f(z) + \Delta\psi(z)/d = (n+1)/d$. Let again $\rho = \rho_0(1+\eta)$; what conditions must be satisfied by ψ and η if we want to substitute ρ for ρ_0 in Proposition 2 and still get a valid approximation?

We also have $\psi(\rho) = \prod_i g_i(\rho)^{d_i}$ and $g_i(\rho)^{d_i} = g_i(\rho_0)^{d_i} exp(d_i\eta\Delta g_i(\rho_0) + O(d_i\eta^2\rho_0^2))$. Hence

$$\psi(\rho) = \psi(\rho_0)e^{\eta\Delta\psi(\rho_0) + O(\sum_i d_i\eta^2\rho_0^2)}.$$

The approximation (9) still holds. By definition of ρ_0, $d\Delta f(\rho_0) + \Delta\psi(\rho_0) - (n+1) = 0$ and we have

$$\frac{f(\rho)^d\psi(\rho)}{\rho^{n+1}\sqrt{2\pi d h''(\rho)}} = \frac{f(\rho_0)^d\psi(\rho_0)}{\rho_0^{n+1}\sqrt{2\pi d h''(\rho_0)}}(1 + O(d\rho^2\eta^2) + O(n\eta^2) + O(\sum_i d_i\eta^2\rho^2) + o(1)).$$

As long as each d_i is $O(d)$, the last error term $O(\sum_i d_i\eta^2\rho^2)$ is $o(n\eta^2)$ and Proposition 2 is still valid if $n\eta^2 = o(1)$, i.e. if $\eta = o(1/\sqrt{n})$.

Now let ρ be again the solution of $\Delta f(z) = n/d$. Let $\epsilon = \sum_{i=1}^{p} d_i/d$; we can show that $\rho = \rho_0(1+\eta)$ with $\eta = O(1/n) + O(\epsilon)$. Hence requiring $n\eta^2 = o(1)$ implies $n\epsilon^2 = o(1)$, i.e. $d_i = o(d/\sqrt{n})$ for each d_i; this is the condition (1) of the assumption $\mathcal{A}_2$.

Theorem 4 is an obvious consequence of Proposition 3.

5.8 Proofs of Corollaries 1 and 2

We shall further simplify Theorems 2 and 3 when $n = o(\sqrt{d})$. Recall that, for $n = o(d)$, $\Delta f(z) = (f_1/f_0)z + O(z^2)$, and that the saddle point is $\rho = f_0(n+1)/(f_1 d)(1 + O(n/d))$. Hence $\rho^n = e(f_0 n/f_1 d)^d(1 + O(n^2/d))$, $f(\rho) = f_0(1 + (n+1)/d + O(n^2/d^2))$ and $f(\rho)^d = f_0^d e^{n+1}(1 + O(n^2/d))$. When $n = o(\sqrt{d})$, the error terms are $o(1)$ and we get:

$$[z^n]\{f^d(z)\} = \frac{f_0^{d-n}}{\sqrt{2\pi n}} \cdot \left(\frac{e f_1 d}{n}\right)^n (1 + o(1)).$$

Extension when $n \neq o(\sqrt{d})$ but $n = o(d^{2/3})$:
We first improve on the approximation of the saddle point: $\Delta f(z) = f_1 z/f_0 \cdot (1 + (2f_2/f1 - f_1/f_0)z + O(z^2))$. The solution ρ of $\Delta f(z) = (n+1)/d$ is then

$$\rho = \frac{f_0(n+1)}{f_1 d}(1 + (1 - 2\frac{f_0 f_2}{f_1^2})z + O(z^2)),$$

and we have

$$\rho^n = (\frac{f_0(n+1)}{f_1 d})^n exp\left((1 - 2\frac{f_0 f_2}{f_1^2})\frac{n(n+1)}{d}\right)(1 + O(\frac{n^3}{d^2})).$$

The approximate value of ρ also gives

$$f(\rho) = f_0(1 + \frac{n+1}{d} + (1 + \frac{f_2}{f_1} - 2\frac{f_0 f_2}{f_1^2})(\frac{n+1}{d})^2 + O(\frac{n^3}{d^2})),$$

hence

$$f(\rho)^d = f_0^d \, e^{n+1} exp\left((\frac{1}{2} + \frac{f_2}{f_1} - 2\frac{f_0 f_2}{f_1^2})\frac{(n+1)^2}{d}\right)(1 + O(\frac{n^3}{d^2})).$$

Putting all this together, and assuming that $n = o(d^{2/3})$, we get:

$$[z^n]\{f^d(z)\} = \frac{f_0^{d-n}}{\sqrt{2\pi n}}\left(\frac{e f_1 d}{n}\right)^n exp\left(\frac{n^2}{d}(\frac{f_2}{f_1} - \frac{1}{2})\right)(1 + o(1)).$$

If there is a factor ψ, we merely have to see that $\psi(\rho) = \psi(0)(1 + o(1))$. This comes from

$$\psi(\rho) = \psi(0)(1 + O(\rho\sum_{i=1}^{p} d_i)) = (1 + O(n\sum_{i=1}^{p} d_i/d)).$$

As long as each d_i is $o(d/n)$, the error term is $o(1)$ and

$$\psi(\rho) = \psi(0)(1 + O(\rho||\frac{\psi'}{\psi}||)) = \psi(0)(1 + o(1)).$$

6 Concluding remarks

Some possible extensions of this work are

- Remove the restriction $f_1 \neq 0$ (this seems possible, but introduces further restrictions on n and d).
- We have studied the case $n = o(d)$; Daniels and Good give results for $n = \Theta(d)$ (Section 2.2) and the result of Meir and Moon (Section 2.3) can sometimes be applied when $d = O(n^2)$. The general case $d = o(n)$ is still open.

An application area that we did not mention there is Coding Theory: See for example [8] for some problems where powers of generating functions appear; the results of that paper could be readily extended using the theorems we have presented here. Related problems also include local limit theorems, evaluating trie parameters [5], the number of lattice points in a ball [13, 16] or the analysis of a random walk on an hypercube [4].

References

[1] N. BLEISTEIN and R.A. HANDELSMAN. *Asymptotic Expansions of Integrals.* Dover, New York, 1986.

[2] H.E. DANIELS. Saddlepoint approximations in statistics. *Annals of Mathematical Statistics*, 25:631–650, 1954.

[3] N. G. deBRUIJN. *Asymptotic Methods in Analysis.* Dover, New York, 1981 (reprint).

[4] P. DIACONIS, R.L.GRAHAM, and J.A. MORRISON. Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures and Algorithms*, 1(1):51–72, 1990.

[5] P. FLAJOLET. On the performance evaluation of extendible hashing and trie searching. *Acta Informatica*, 20:345–369, 1983.

[6] W. FULKS. A generalization of Laplace's method. *American Mathematical Society Proceedings*, 2(4):613–622, August 1951.

[7] D. GARDY. Normal limiting distributions for projection and semijoin sizes. (Revised version). Technical Report 574, Laboratoire de Recherche en Informatique, Orsay, France, May 1990. To appear in SIAM Journal of Discrete Mathematics.

[8] D. GARDY and P. SOLE. Saddle-point techniques in asymptotic coding theory. In *Rencontre France-URSS de Codage*, 22-24 Juillet 1991.

[9] I.J. GOOD. Saddle-point methods for the multinomial distribution. *Annals of Mathematical Statistics*, 28:861–881, 1957.

[10] D.H. GREENE and D.E. KNUTH. *Mathematics for the Analysis of Algorithms.* Birkhäuser, 1982.

[11] W.K. HAYMAN. A generalisation of Stirling's formula. *Journal für die reine und angewandte Mathematik*, 196:67–95, 1956.

[12] N.L. JOHNSON and S. KOTZ. *Urn models and their application.* Wiley & Sons, 1977.

[13] J.E. MAZO and A.M. ODLYZKO. Lattice points in high-dimensional spheres. Technical report, AT&T Bell Laboratories, Murray Hill, USA, 1990.

[14] A. MEIR and J.W. MOON. The asymptotic behaviour of coefficients of powers of certain generating functions. *Journal Européen de Combinatoire*, 11:581–587, 1990.

[15] R.N. PEDERSON. Laplace's method for two parameters. *Pacific Journal of Mathematics*, 15(2):585–596, 1965.

[16] J.A. RUSH and N.J.A. SLOANE. Bound for packing superballs. *Mathematika*, 34:8–18, 1987.

[17] H.S. WILF. *Generatingfunctionology.* Academic Press, 1990.