



# The number of distinct part sizes of some multiplicity in compositions of an Integer. A probabilistic Analysis

Guy Louchard

## ► To cite this version:

Guy Louchard. The number of distinct part sizes of some multiplicity in compositions of an Integer. A probabilistic Analysis. Discrete Random Walks, DRW'03, 2003, Paris, France. pp.155-170, 10.46298/dmtcs.3347 . hal-01183943

**HAL Id: hal-01183943**

**<https://inria.hal.science/hal-01183943>**

Submitted on 12 Aug 2015

**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.

# The number of distinct part sizes of some multiplicity in compositions of an Integer. A probabilistic Analysis

Guy Louchard

Université Libre de Bruxelles, Département d'Informatique, CP 212, Boulevard du Triomphe, B-1050 Bruxelles, Belgium

[Louchard@ulb.ac.be](mailto:Louchard@ulb.ac.be)

---

Random compositions of integers are used as theoretical models for many applications. The degree of distinctness of a composition is a natural and important parameter. A possible measure of distinctness is the number  $X$  of distinct parts (or components). This parameter has been analyzed in several papers. In this article we consider a variant of the distinctness: the number  $X(m)$  of distinct parts of multiplicity  $m$  that we call the  $m$ -distinctness. A first motivation is a question asked by Wilf for random compositions: what is the asymptotic value of the probability that a randomly chosen part size in a random composition of an integer  $v$  has multiplicity  $m$ . This is related to  $\mathbb{E}(X(m))$ , which has been analyzed by Hitczenko, Rousseau and Savage. Here, we investigate, from a probabilistic point of view, the first full part, the maximum part size and the distribution of  $X(m)$ . We obtain asymptotically, as  $v \rightarrow \infty$ , the moments and an expression for a continuous distribution  $\phi$ , the (discrete) distribution of  $X(m, v)$  being computable from  $\phi$ .

**Keywords:** Mellin transforms, urns models, Poissonization, saddle point method, generating functions

---

## 1 Introduction

Let us first recall some well-known results. Let us consider the composition of an integer  $v$ , i.e.  $v = \sum_{i=1}^N x_i$ ,  $x_i : \text{integer} > 0$ . Considering all compositions as equiprobable, we know (see [HL01]) that the number of parts  $N$  is asymptotically Gaussian,  $v \rightarrow \infty$ :

$$N \sim \mathcal{N}\left(\frac{v}{2}, \frac{v}{4}\right), \quad (1)$$

and that the part sizes are asymptotically id  $\text{GEOM}(1/2)$  and *independent*. Consider now  $n$  random variables (R.V.),  $\text{GEOM}(1/2)$  and define the indicator R.V.<sup>†</sup>

$$Y_i := \llbracket \text{value } i \text{ appears among these } n \text{ R.V.} \rrbracket$$

Then, asymptotically,  $n \rightarrow \infty$ , the  $Y_i$  are independent. The first empty part value, i.e. the first  $k$  such that  $Y_k = 0$ , is of order  $O(\log n)$ . Here and in the sequel,  $\log := \log_2$ ,  $L := \ln 2$ . Similarly, the maximum part

---

<sup>†</sup> Here we use the indicator function notation proposed by Knuth et al. [GKP89].

size is also of order  $O(\log n)$ , as well as the number  $Y$  of distinct values (part sizes):  $Y = \sum_1^\infty Y_i$ . The asymptotic distributions and moments of these R.V. are also given in [HL01]. We know (see Hwang and Yeh [HY97]) that

$$\mathbb{E}(Y) \sim \log n + \gamma/L - 1/2 + \beta(\log n) + O(1/n)$$

where  $\beta$  is a small periodic function of  $\log n$ , and the distribution of  $Y$  is highly concentrated around its mean, with a  $O(1)$  range. All these distributions depend on  $\log n$ . Hence, with (1), the same R.V. related to  $v$  are asymptotically equivalent by replacing  $\log n$  by  $\log v - 1$  (see [HL01]).

In this article we consider a variant of the distinctness: the number  $X(m)$  of distinct parts of multiplicity  $m$  that we call the  $m$ -distinctness. A first motivation is a question asked by Wilf for random compositions: what is the asymptotic value of the probability  $P(m, v)$  that a randomly chosen part size in a random composition of an integer  $v$  has multiplicity  $m$ . (The corresponding problem for random partitions has been analyzed in Corteel et al. [CPSW99]). Of course, here,

$$P(m, v) = \mathbb{E}(X(m, v) / Y(v)),$$

where we explicitly show the dependence on  $v$ . But, as already mentioned,  $Y(v)$  has asymptotically the same distribution as  $Y$  (with  $\log n$  replaced by  $\log v - 1$ ). On the other side,  $Y$  is highly concentrated around its mean. Hence, asymptotically, as shown in Hitczenko, Savage [HS99] and Hitczenko et al [HRS02], for  $m = O(1)$ ,

$$P(m, v) \sim \mathbb{E}(X(m, v)) / \mathbb{E}(Y(v)).$$

Here, we investigate, from a probabilistic point of view, the first full part, the maximum part size and the distribution of  $X(m, v)$ . We obtain asymptotically, as  $v \rightarrow \infty$ , the moments and an expression for a continuous distribution  $\phi$ , the (discrete) distribution of  $X(m, v)$  being computable from  $\phi$ . We will see that, again, all asymptotic distributions for some multiplicity  $m$  depend only on  $\log n$ . Hence, the same R.V. related to  $v$  are again simply obtained by replacing  $\log n$  by  $\log v - 1$ . The paper is organized as follows: in Section 2, we consider a fixed multiplicity  $m = O(1)$ . We analyze the moments, the first full part, the maximum part size, and the distribution of  $X(m)$ . Section 3 is devoted to large multiplicity  $m$ . Section 4 concludes the paper. Due to length constraints, some proofs have been briefly presented.

In this section, we are interested in the properties of the R.V.:

$X_i(m) := \llbracket \text{value } i \text{ appears among the } n \text{ GEOM}(1/2) \text{ R.V. with multiplicity } m, \text{ for fixed } m = O(1) \rrbracket$ .

Of course,

$$\Pr[X_i(m) = 1] = \binom{n}{m} (1/2^i)^m (1 - 1/2^i)^{n-m}. \quad (2)$$

We immediately see that the dominant range is given by  $i = \log n + O(1)$ . To the left and the right of this range,  $\Pr[X_i(m) = 1] \sim 0$ . Within the range,  $\Pr[X_i(m) = 1]$  is asymptotically equivalent to a Poisson distribution:

$$\Pr[X_i(m) = 1] \sim \frac{1}{m!} (n/2^i)^m \exp(-n/2^i),$$

and, with  $X(m) := \sum_1^\infty X_i(m)$ ,

$$\mathbb{E}(X(m)) \sim G(n, m),$$

where, using the "sum splitting technique" as described in Knuth [Knu73], p.131,

$$G(n, m) := \frac{1}{m!} \sum_{i=1}^\infty (n/2^i)^m \exp(-n/2^i),$$

which, for large  $n$ , can be analyzed using Mellin transforms: see Flajolet et al. [FGD95]. It is well known that the dominant value is given by some constant. The oscillatory part has a very small amplitude, usually of order  $10^{-5}$ . Indeed, set  $f(y) := y^m e^{-y}$ . We obtain

$$G(n, m) = \frac{1}{m!} \sum_{i=1}^{\infty} f(n/2^i),$$

the Mellin transform of which is

$$G^*(s) = \frac{\Gamma(m+s)}{m!} \frac{2^s}{1-2^s},$$

defined in the fundamental strip  $< -m, 0 >$ . To the right of this strip, the poles of  $G^*(s)$  are a simple pole at  $s = 0$ , and simple poles at  $s = \chi_k := 2k\pi i/L$  ( $k \neq 0$ ). The singular expansion of  $G^*(s)$  is given by  $\ddagger$

$$G^*(s) \asymp \left[ \frac{\Gamma(m)}{Lm!s} \right] + \sum_{k \neq 0} \frac{\Gamma(m + \chi_k)}{Lm!(s - \chi_k)}.$$

This leads, by converse mapping, to

$$G(n, m) \sim \frac{1}{mL} + \beta_0(\log n) + O(1/n), \quad (3)$$

where  $\beta_0$  is a small periodic function of  $\log n$ :

$$\beta_0(\log_2 n) := \sum_{k \neq 0} \frac{\Gamma(m + \chi_k)}{Lm!} n^{-\chi_k} = \sum_{k \neq 0} \frac{\Gamma(m + \chi_k)}{Lm!} e^{-2\pi i k \log n}.$$

In the sequel,  $\beta.(\log n)$  will always denote (small) periodic functions. As  $n \sim \mathcal{N}(\frac{v}{2}, \frac{v}{4})$ , we just have to replace  $\log n$  by  $\log v - 1$ . So we recover the mean already computed in Hitczenko and Savage, [HS99] and Hitczenko, Rousseau and Savage, [HRS02]. To compute all moments, we must check that the  $X_i$  are asymptotically independent. We could proceed as was done in [HL01] for the  $Y_i$ , but we follow here another route. Let us consider  $\Pi_n = \mathbb{E}(z^X)$ . We obtain

**Theorem 1.1.**

$$\Pi_n \sim \prod_{l=1}^{\infty} \left[ \left( 1 - \frac{1}{m!} (n/2^l)^m e^{-n/2^l} \right) + z \frac{1}{m!} (n/2^l)^m e^{-n/2^l} \right], n \rightarrow \infty.$$

*Proof.* We use an urn model, as in Sevastyanov and Chistyakov, [SČ64] and Chistyakov, [Chi67], and the Poissonization method (see, for instance Jacquet and Szpankowski [JS98] for a general survey). If we Poissonize, with parameter  $\tau$ , the number of balls (i.e the number  $n$  of R.V. here), the generating function of  $X_l$  is given from (2), by

$$\left( 1 - \frac{1}{m!} (\tau/2^l)^m e^{-\tau/2^l} \right) + z \frac{1}{m!} (\tau/2^l)^m e^{-\tau/2^l},$$

---

$\ddagger$  The symbol  $\asymp$  is used to denote the fact that two functions are of the same asymptotic order.

and we have independency of cells occupation. This leads to

$$e^{-\tau} \sum_n \frac{\tau^n}{n!} \Pi_n = \prod_{l=1}^{\infty} \left[ \left( 1 - \frac{1}{m!} (\tau/2^l)^m e^{-\tau/2^l} \right) + z \frac{1}{m!} (\tau/2^l)^m e^{-\tau/2^l} \right].$$

Hence, by Cauchy, we obtain  $\Pi_n = \frac{n!}{2\pi i} \int_{\Gamma} \exp\{nf(\tau)\} d\tau/\tau$ , where  $\Gamma$  is inside the analyticity domain of the integrand and encircles the origin, and

$$f(\tau) := -\log \tau + \tau/n + \frac{1}{n} \sum_{l=1}^{\infty} \ln \left[ \left( 1 - \frac{1}{m!} (\tau/2^l)^m e^{-\tau/2^l} \right) + z \frac{1}{m!} (\tau/2^l)^m e^{-\tau/2^l} \right].$$

By standard saddle-point method (see, for instance, Flajolet and Sedgewick, [FS94]), we look for  $\tau^*$  such that  $f'(\tau^*) = 0$ , with

$$f'(\tau) = -1/\tau + 1/n - \frac{z-1}{n\tau} \sum_{l=1}^{\infty} \frac{(\tau/2^l)^{m+1} - m(\tau/2^l)^m}{m! \exp(\tau/2^l) - (\tau/2^l)^m + z(\tau/2^l)^m}.$$

But, again by Mellin, for fixed  $z > 0$ ,

$$\sum_{l=1}^{\infty} \frac{(\tau/2^l)^{m+1} - m(\tau/2^l)^m}{m! \exp(\tau/2^l) - (\tau/2^l)^m + z(\tau/2^l)^m} \sim C + \beta \cdot (\log \tau).$$

with

$$C := \int_0^{\infty} \frac{y^{m+1} - my^m}{m! \exp(y) - y^m + zy^m} dy / L.$$

Hence  $\tau^* \sim n + C$ . It is easily checked that  $C = 0$ . Finally,  $\Pi_n \sim \frac{n! e^{nf(\tau^*)}}{\sqrt{2\pi\tau^*} \sqrt{nf''(\tau^*)}}$ , and, by Stirling, we easily derive the theorem.  $\square$

Theorem 1.1 confirms the asymptotic independence assumption.

### 1.1 The moments of $X(m)$

We now have all necessary ingredients to compute the moments. The variance of  $X(m)$  is now easily derived: we obtain, by Mellin,

$$\begin{aligned} \text{VAR}(X(m)) &\sim \frac{1}{m!} \sum_1^{\infty} (n/2^i)^m \exp(-n/2^i) \left[ 1 - \frac{1}{m!} (n/2^i)^m \exp(-n/2^i) \right] \\ &\sim \int_0^{\infty} e^{-y} \frac{y^m}{m!} (1 - e^{-y} \frac{y^m}{m!}) \frac{dy}{Ly} + \beta_1(\log_2 n) \\ &= \frac{1}{mL} - \frac{(2m-1)!}{Lm!^2 2^{2m}} + \beta_1(\log_2 n). \end{aligned}$$

The other moments can be derived as follows. We obtain, setting  $z = e^s$ ,

$$\begin{aligned} \ln(\Pi_n) \sim S_2 &= \sum_{l=1}^{\infty} \ln \left[ 1 + (e^s - 1) \frac{1}{m!} (n/2^l)^m \exp(-n/2^l) \right] \\ &= \sum_{i=1}^{\infty} \frac{(-1)^{i+1} (e^s - 1)^i V_i}{i}, \text{ with} \\ V_i &:= \sum_{l=1}^{\infty} \left[ \frac{1}{m!} (n/2^l)^m \right]^i \exp(-in/2^l). \end{aligned}$$

The centered moments of  $X(m)$  can be obtained by analyzing

$$S_3 := \exp[S_2 - sV_1].$$

Again, by Mellin, we obtain

$$V_i \sim B_i + \beta_i(\log n),$$

with

$$B_i = \int_0^{\infty} \left[ \frac{y^m}{m!} \right]^i e^{-iy} \frac{dy}{Ly} = \frac{(im-1)!}{m!^i L i^m},$$

and finally, the centered moments are given by

$$\begin{aligned} \tilde{\sigma}^2 := \text{VAR}(X(m)) &\sim \frac{1}{mL} - \frac{(2m)!}{2Lm!^2 2^{2m} m}, \\ \tilde{\mu}_3 := \mu_3(X(m)) &\sim \frac{1}{mL} - \frac{3(2m)!}{2Lm!^2 2^{2m} m} + \frac{2(3m)!}{3Lm!^3 3^{3m} m}, \\ \tilde{\mu}_4 := \mu_4(X(m)) &\sim \frac{1}{mL} + \frac{3}{m^2 L^2} - \frac{3(4m)!}{2Lm!^4 4^{4m} m} + \frac{4(3m)!}{Lm!^3 3^{3m} m} \\ &\quad - \frac{7(2m)!}{2Lm!^2 2^{2m} m} - \frac{3(2m)!}{3L^2 m!^2 2^{2m} m^2} + \frac{3(2m)!^2}{4L^2 m!^4 2^{4m} m^2}. \end{aligned}$$

The neglected terms are made of periodic functions  $\beta_i(\log n)$  and of  $O(\frac{1}{n})$  contributions.

Again, the centered moments (of order  $\geq 2$ ) of  $X$  related to a composition of  $v$  are given by the same expressions.

For  $n = 20000, m = 2$ , we have done a simulation (of  $T = 4000$  sets). We obtain the results of Table 1 (the probability related moments are explained later on). For an easy comparison, we give here only four significant digits.

## 1.2 The maximum part size of multiplicity $m$

The maximum part size  $\mathcal{M}_n(m)$  of multiplicity  $m$  is such that

$$\Pr(\mathcal{M}_n(m) < k) \sim \prod_{i=k}^{\infty} \left[ 1 - \frac{1}{m!} (n/2^i)^m \exp(-n/2^i) \right].$$

	Theoretical asymptotic value	Observed value	Probability related value
mean	.7213 ...	.7345...	.7214...
variance	.5861 ...	.5945...	.5863...
$\mu_3$	.3750...	.3752 ...	.3752...
$\mu_4$	1.1197...	1.1341...	1.1198...

**Tab. 1:** Moments,  $n = 20000, m = 2$ .

Set  $\eta := Lk - \ln n$ . This leads, with  $\eta = O(1)$ , to

$$\Pr(\mathcal{M}_n(m) < k) \sim \varphi_1(m, \eta),$$

with

$$\varphi_1(m, \eta) = \prod_{j=0}^{\infty} \left[ 1 - \frac{1}{m!} e^{-m(\eta+Lj)} e^{-e^{-(\eta+Lj)}} \right].$$

Figure 1 gives  $\varphi_1(m, \eta)$  for  $m = 1, \dots, 4$ , bottom to top. It appears that for  $\eta \rightarrow -\infty$ ,  $\varphi_1(m, \eta)$  seems to converge to some value, which of course corresponds to

$$P(m, 0) := \Pr(X(m) = 0),$$

but a closer view reveals the usual fluctuations, shown in Figure 2, for  $m = 2$ . Set  $\psi(n) := \log n - \lfloor \log n \rfloor$  (fractional part). With  $\eta = L(-6 - \psi(20000))$ , we obtain  $P(2, 0) = .4489079864\dots$ , which will be compared later on with a direct expression.

Similarly, we derive

$$\Pr(\mathcal{M}_n(m) = k - 1) \sim \varphi_2(m, \eta) = \varphi_1(m, \eta) e^{-m(\eta-L)} e^{-e^{-(\eta-L)}} / m!.$$

Figure 3 gives  $\varphi_2(m, \eta)$  for  $m = 1, \dots, 4$ , (more and more concentrated as  $m$  increases).

Our simulation for  $n = 20000, m = 2$  of  $T = 4000$  sets leads to Figure 4 ( $\varphi_1$ , observed = circle, asymptotic = line) and Figure 5 ( $\varphi_2$ , observed = circle, asymptotic = line). Again, for compositions, we replace  $\log n$  by  $\log v - 1$ .

### 1.3 First full part value of multiplicity $m$

Another variable of interest is the first  $k$  such that  $X_k = 1$ , i.e we are interested in the probability

$$\Pr[X_i = 0, i = 1 \dots k - 1, X_k = 1].$$

Note that this is the opposite situation of the  $Y_k$  case (see [HL01]), where we looked for the *first*  $k$  such that  $Y_k = 0$ . The probability is asymptotically given by

$$\prod_{i=1}^{k-1} \left[ 1 - \frac{1}{m!} (n/2^i)^m \exp(-n/2^i) \right] \frac{1}{m!} (n/2^k)^m \exp(-n/2^k).$$

Again, we set  $\eta := Lk - \ln n$ . This leads asymptotically, with  $\eta = O(1)$  to

$$\Pr[X_i = 0, i = 1 \dots k-1, X_k = 1] \sim \varphi_3(m, \eta),$$

with

$$\begin{aligned} \varphi_3(m, \eta) &= \varphi_4(m, \eta) \frac{1}{m!} e^{-m\eta} e^{-e^{-\eta}}, \\ \varphi_4(m, \eta) &= \prod_{j=1}^{\infty} \left[ 1 - \frac{1}{m!} e^{-m(\eta - Lj)} e^{-e^{-(\eta - Lj)}} \right]. \end{aligned}$$

Again, for compositions, we replace  $\log n$  by  $\log v - 1$ . Figure 6 gives  $\varphi_4(2, \eta)$  and Figure 7 gives  $\varphi_4(2, \eta)$  for large values of  $\eta$ . Again, this is oscillating and corresponds to  $P(2, 0)$ .

#### 1.4 Asymptotic distribution of $X(m)$

The analysis is rather similar to the one we used in [Lou87] and [HL01]. First of all we have, for any fixed  $k = O(\log n)$ ,

$$P(m, 0) \sim \varphi_4(\eta) \varphi_1(\eta).$$

Let us choose  $k = \lfloor \log n \rfloor$ . This leads to  $\eta = -L\psi(n)$  and we obtain a periodic function of  $\psi$ :

$$P(m, 0) \sim \varphi_4[-L\psi(n)] \varphi_1[-L\psi(n)],$$

shown in Figure 10 for  $m = 2$ . For  $n = 20000, m = 2$ , the numerical value of  $P(2, 0)$  is exactly the same as before. Now we turn to  $P(m, j) := \Pr(X(m) = j)$ . We take advantage of the fact that all urns are empty *before* the first occupied urn,  $k - 1$  say. Then, again with  $\eta := Lk - \ln n$ ,

$$\begin{aligned} P(m, 1) &\sim \sum_k \varphi_3(\eta - L) \varphi_1(\eta), \\ P(m, 2) &\sim \sum_k \varphi_3(\eta - L) \varphi_1(\eta) \sum_{r_1 \geq k} \left\{ \frac{1}{m!} (n/2^{r_1})^m \exp(-n/2^{r_1}) \left/ \left[ 1 - \frac{1}{m!} (n/2^{r_1})^m \exp(-n/2^{r_1}) \right] \right. \right\}, \end{aligned}$$

and more generally,

$$\begin{aligned} P(m, u+1) &\sim \sum_k \varphi_3(\eta - L) \varphi_1(\eta) \\ &\cdot \sum_{\lfloor r_1 > r_2 > \dots > r_u, r_j \geq k \rfloor} \prod_{i=1}^u \left\{ \frac{1}{m!} (n/2^{r_i})^m \exp(-n/2^{r_i}) \left/ \left[ 1 - \frac{1}{m!} (n/2^{r_i})^m \exp(-n/2^{r_i}) \right] \right. \right\} \end{aligned}$$

Now we set  $r_i = k + w_i, l = k - \lfloor \log n \rfloor$  and we finally derive the following theorem

**Theorem 1.2.** Set  $\psi(n) := \log n - \lfloor \log n \rfloor$ , then

$$P(m, u+1) \sim \sum_{l=-\infty}^{\infty} \varphi_5[L(l - \psi(n))],$$



with

$$\varphi_5(\eta) = \varphi_3(\eta - L)\varphi_1(\eta).$$

$$\cdot \sum \llbracket w_1 > w_2 > \dots > w_u, w_j \geq 0 \rrbracket \prod_{i=1}^u \left\{ \frac{1}{m!} e^{-m(\eta+Lw_i)} e^{-e^{-(\eta+Lw_i)}} \right\} \left/ \left[ 1 - \frac{1}{m!} e^{-m(\eta+Lw_i)} e^{-e^{-(\eta+Lw_i)}} \right] \right\}$$

Note that, for compositions, we obtain asymptotically  $\psi(n) = \psi(v)$ . We get again periodic function of  $\psi(n)$ . We give in Figure 11 and Figure 12 the sums  $\sum_{i=0}^3 P(2, i)$ ,  $\sum_{i=0}^4 P(2, i)$ . The effect of computing  $P(2, i)$  with bounded indices (we limit the values of  $w_u$  to 16) becomes apparent at the  $10^{-7}$  precision.

Figure 13 gives  $P(m, i)$ ,  $m = 1, \dots, 4$ , (from top to bottom to the right of  $i = 2$ ). The distributions become more concentrated as  $m$  increases.

Finally, we compare the observed distribution of  $X(2)$  with the asymptotic one in Figure 14 (observed = circle, asymptotic = line). Apart from  $i = 0$  the fit is quite good. The "Probability related values" moments given in Table 1 are computed with the distribution  $P(2, i)$ .

## 2 Large multiplicity $m$

### 2.1 Fixed number of parts $n$

It is now clear that large  $m$  are related to small integer values  $i$ . More precisely, the number  $M_i$  of integers equal to  $i$  is asymptotically given by a Gaussian:

$$\Pr(M_i = m) \sim \exp\left\{-\frac{(m - n/2^i)^2}{[2n/2^i(1 - 1/2^i)]}\right\} / \sqrt{2\pi n/2^i(1 - 1/2^i)}. \quad (4)$$

The means  $n/2^i, i = 1, 2, \dots$  are given by  $n/2, n/4, \dots$ , separated by  $n/4, n/8, \dots$  which shows that the Gaussians (4) are asymptotically exponentially distinct in the sense that some common intervals, for instance  $m \in [3n/2^{i+2} - n/2^{i+3}, 3n/2^{i+2} + n/2^{i+3}]$  have asymptotically small probability measures. So for any large value  $m$ , only one value

$$i = \text{round}[\log(n/m)] \quad (5)$$

is related to  $m$  and  $X(m)$  has only two possible values:  $\{0, 1\}$ . The following events are equivalent:  $\llbracket X_i(m) = 1 \rrbracket \equiv \llbracket M_i = m \rrbracket$ . The probability (4) is small, of order at most  $O(1/\sqrt{m})$ . Figure 15 gives  $\Pr(X_i(m) = 1)$  for  $n = 2000$  (first three ranges,  $i = 1, 2, 3$ ) and Figure 16 gives the corresponding distribution functions, together with the observed values provided by a simulation of  $T = 2000$  sets (observed = circle, asymptotic = line).

An interesting check would be to recover the dominant term of the mean of  $Y$ :  $\mathbb{E}(Y) \sim \log n$ . Choose  $\tilde{j} := \alpha \log n, 0 < \alpha < 1$  which corresponds, by (5), to  $\tilde{m} = n^{1-\alpha}$ . For each  $i \leq \tilde{j}$ , by Euler-McLaurin,

$$\sum_{m=\lfloor 3n/2^{i+2} \rfloor}^{\lfloor 3n/2^{i+1} \rfloor} \exp\left\{-\frac{(m - n/2^i)^2}{[2n/2^i(1 - 1/2^i)]}\right\} / \sqrt{2\pi n/2^i(1 - 1/2^i)} \sim 1,$$

and this contributes to  $\mathbb{E}(Y)$  by  $S_1 = \tilde{j}$ . On the other side, each  $m < \tilde{m}$  contributes, by (3), with  $\frac{1}{mL}$ , with a total contribution

$$S_2 = \frac{1}{L} \sum_{i=1}^{\tilde{m}} 1/m \sim \frac{1}{L} \ln \tilde{m}.$$

The quantity  $S_1 + S_2 \sim \log n$  as expected.

## 2.2 Composition of $\mathbf{v}$ .

Now the number of parts  $N$  is such that (see(1))

$$N \sim \mathcal{N}\left(\frac{\mathbf{v}}{2}, \frac{\mathbf{v}}{4}\right).$$

We obtain

$$\mathbb{E}(M_k) = \frac{\mathbf{v}}{2} \frac{1}{2^k}. \quad (6)$$

The asymptotic distribution of  $M_k$  is obtained as follows. We derive, setting  $\tilde{M}_k := (M_k - n/2^k)/\sqrt{\mathbf{v}}$ ,

$$\begin{aligned} \mathbb{E}[\exp[iM_k\theta/\sqrt{\mathbf{v}}]] &= \mathbb{E}[\exp[in\theta/(\sqrt{\mathbf{v}}2^k) + i\tilde{M}_k\theta]] \\ &\sim \mathbb{E}[\exp[in\theta/(\sqrt{\mathbf{v}}2^k) - \theta^2 n/(2\mathbf{v}2^k)(1 - 1/2^k)]] \\ &\sim \exp\left[i\mathbf{v}\theta/(2\sqrt{\mathbf{v}}2^k) - \mathbf{v}\theta^2/(2\mathbf{v}2^k)(1 - 1/2^k) + \mathbf{v}/8[i\theta/(\sqrt{\mathbf{v}}2^k) - \theta^2/(2\mathbf{v}2^k)(1 - 1/2^k)]^2\right] \\ &\sim \exp\left[i\theta\sqrt{\mathbf{v}}/(2^k) - \theta^2/2[1/(4^k) + 1/(2^k)(1 - 1/2^k)]\right], \mathbf{v} \rightarrow \infty. \end{aligned}$$

The first term confirms (6). The second term shows that

$$M_k \sim \mathcal{N}\left(\frac{\mathbf{v}}{2} \frac{1}{2^k}, \mathbf{v}\sigma_m^2\right),$$

with

$$\sigma_m^2 = 1/(4^k) + 1/(2^k)(1 - 1/2^k).$$

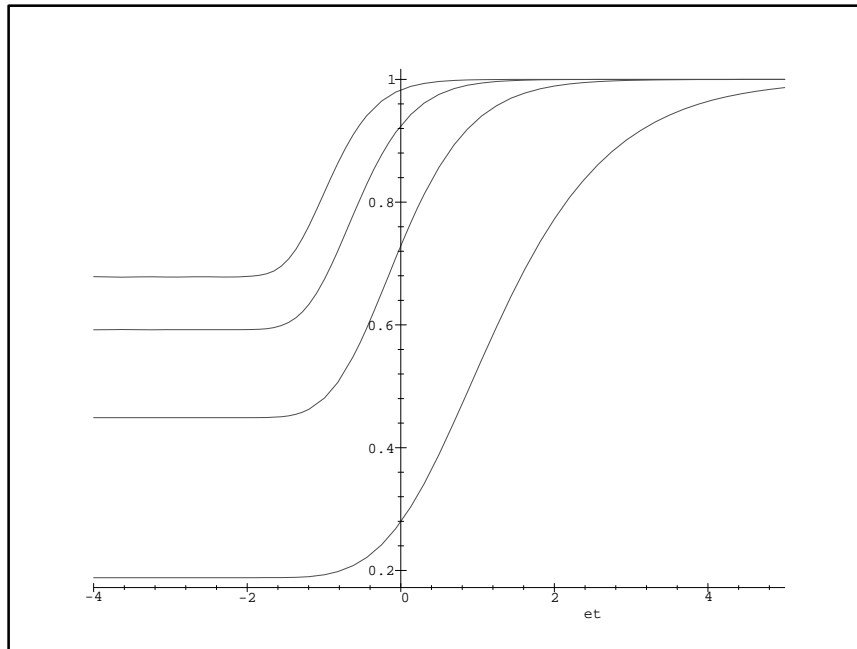
The conclusions of Sec. 2.2 are still valid.

## 3 Conclusion

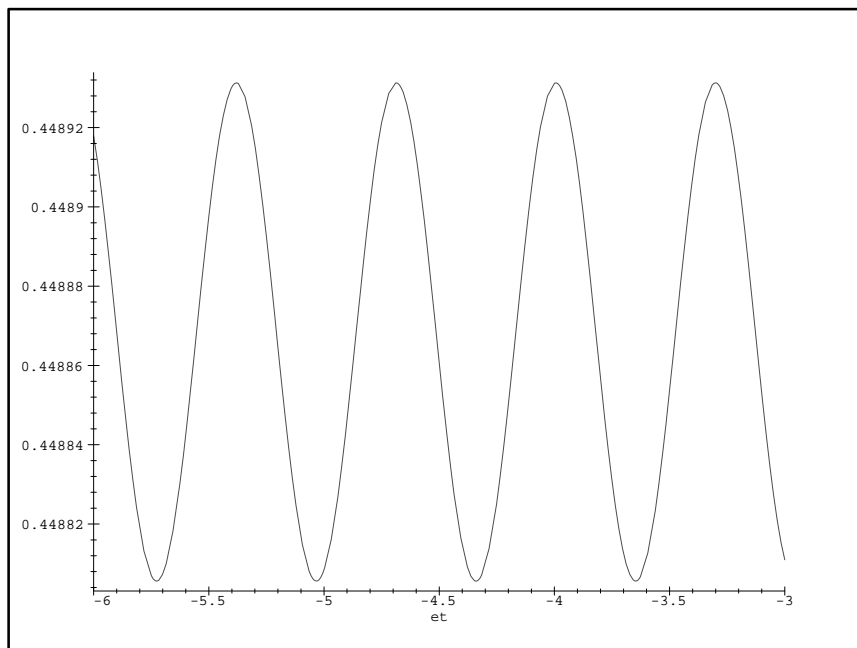
Using various techniques from analysis and probability theory, we have analyzed the stochastic properties of the  $m$ -distinctness of random compositions. An interesting open problem would be to extend our results to the Carlitz compositions, where two successive parts are different (see [LP02]).

## Acknowledgements

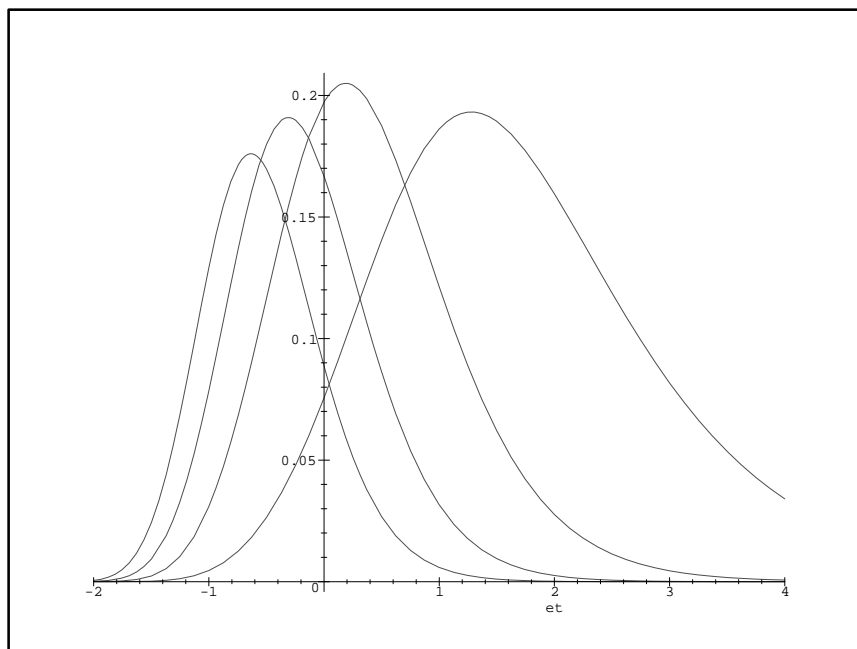
The pertinent comments of the referees led to substantial improvements in the presentation.



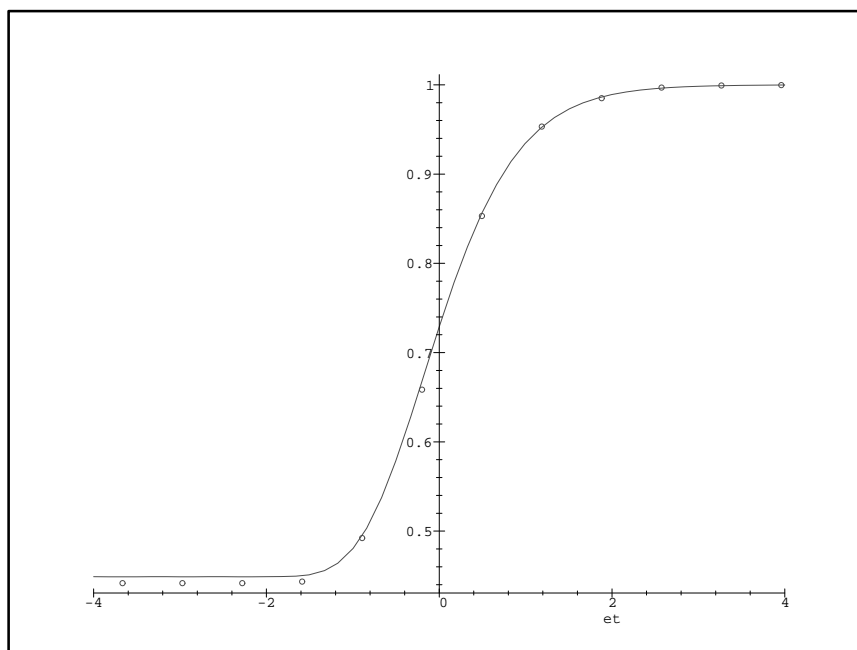
**Fig. 1:**  $\varphi_1(m, \eta)$  for  $m = 1, \dots, 4$ , bottom to top



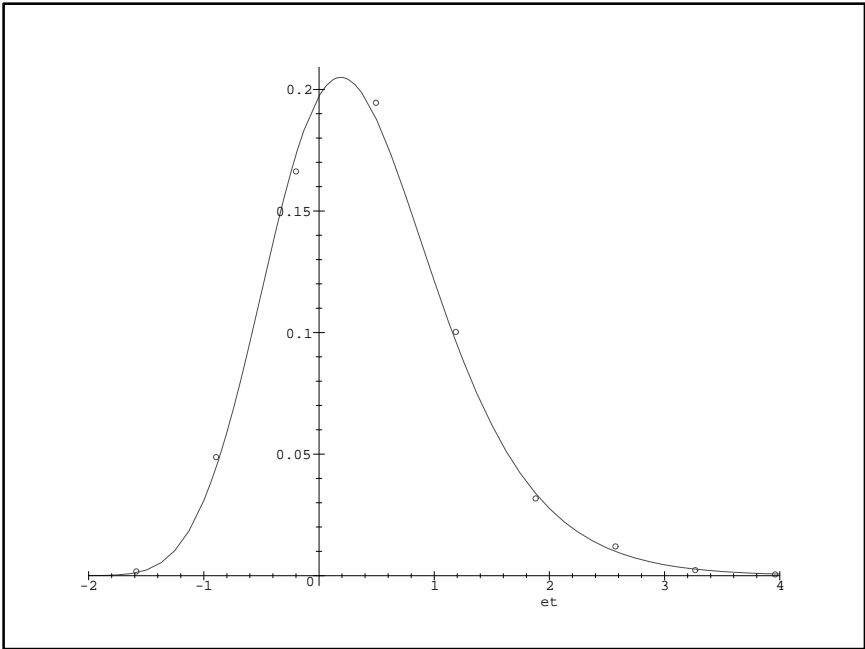
**Fig. 2:**  $\varphi_1(2, \eta)$  for large negative values of  $\eta$



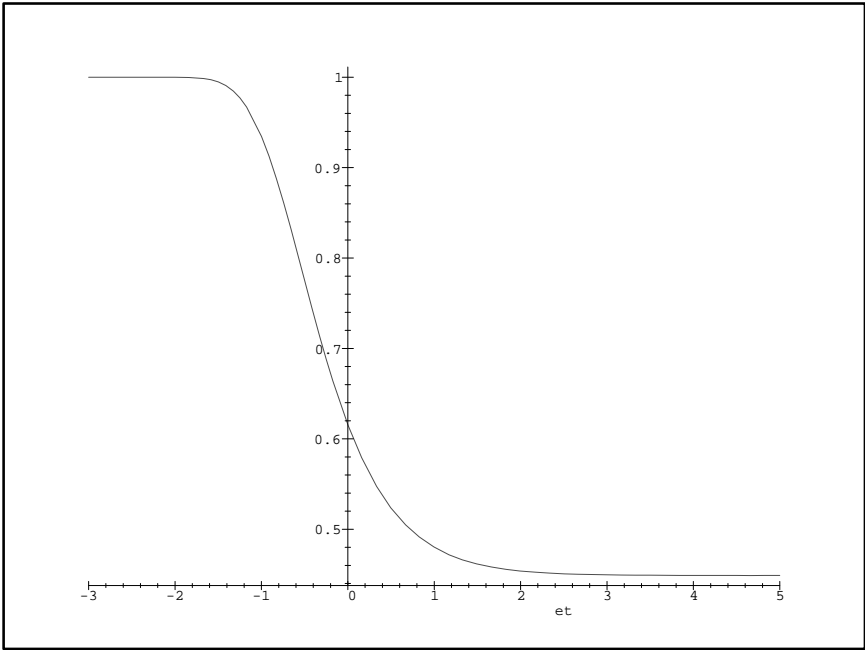
**Fig. 3:**  $\varphi_2(m, \eta)$  for  $m = 1, \dots, 4$



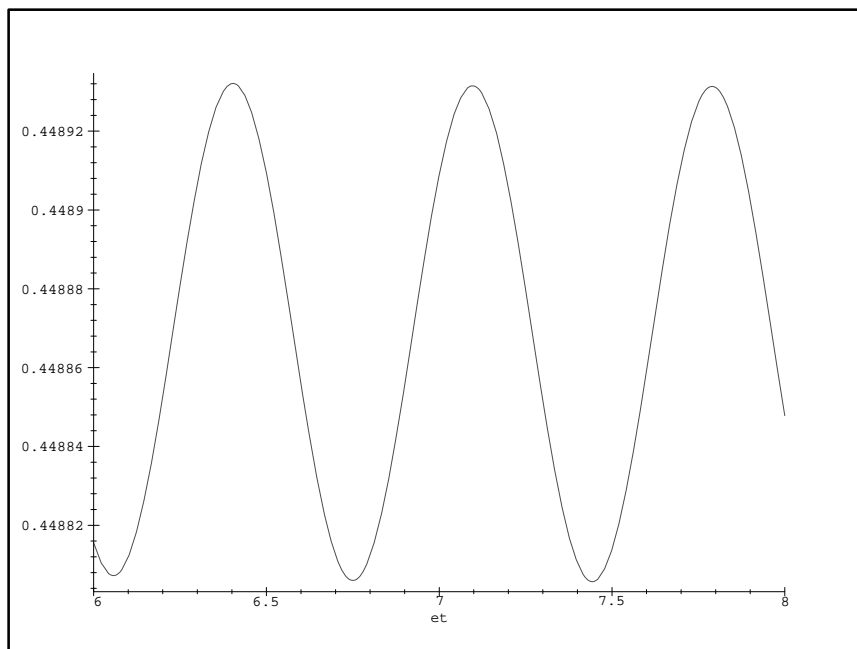
**Fig. 4:** Maximum part size distribution function ( $m = 2$ , observed = circle, asymptotic = line)



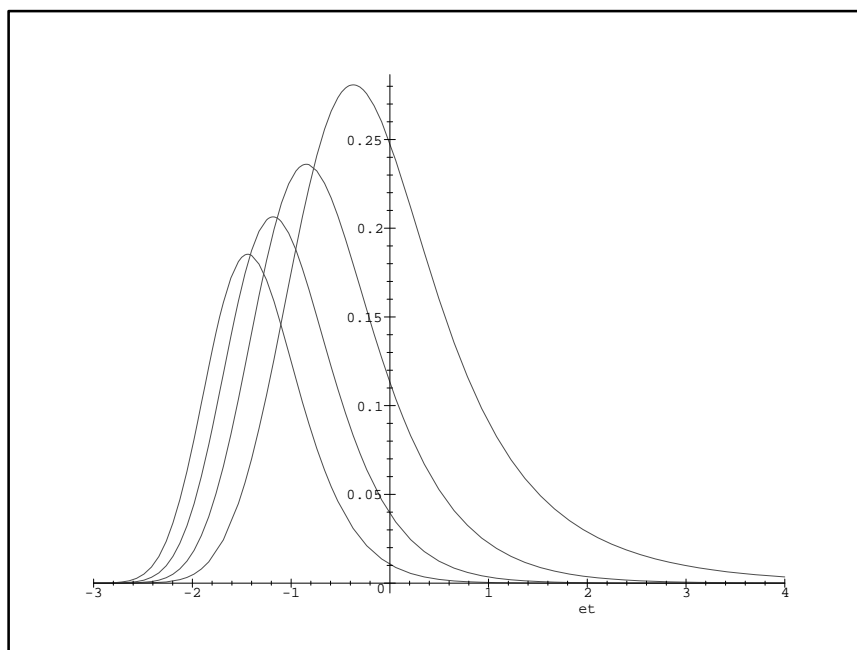
**Fig. 5:** Maximum part size distribution ( $m = 2$ , observed = circle, asymptotic = line)



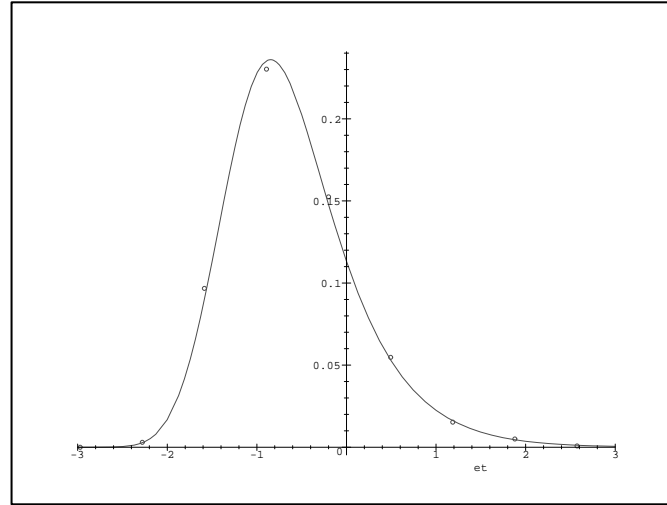
**Fig. 6:**  $\varphi_4(2, \eta)$



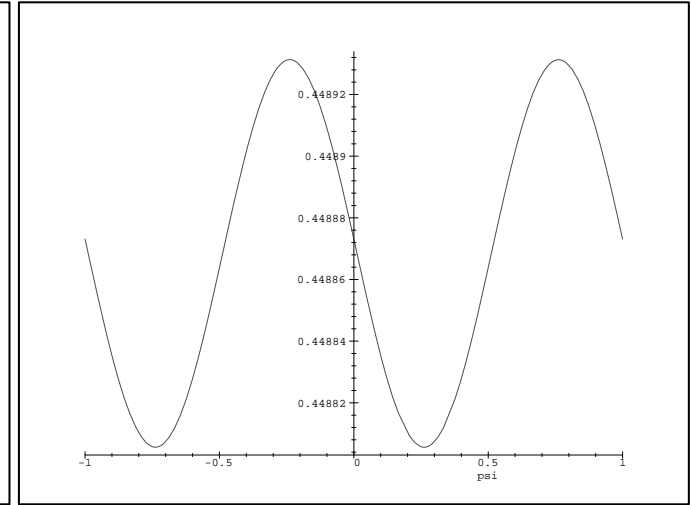
**Fig. 7:**  $\varphi_4(2, \eta)$  for large values of  $\eta$



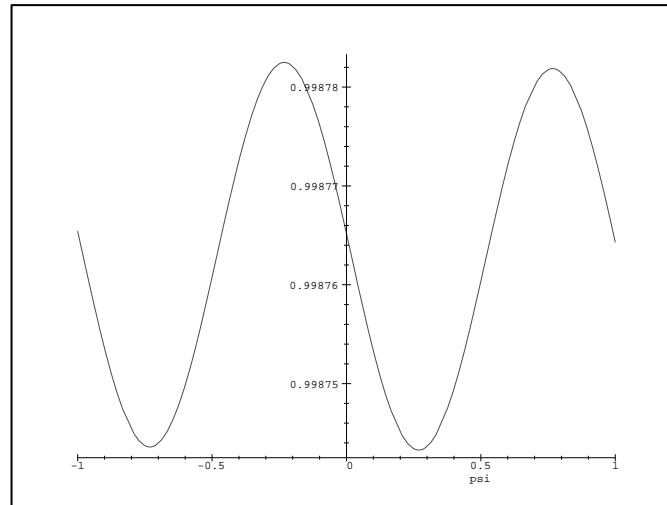
**Fig. 8:**  $\varphi_3(2, \eta)$  for  $m = 1, \dots, 4$ .



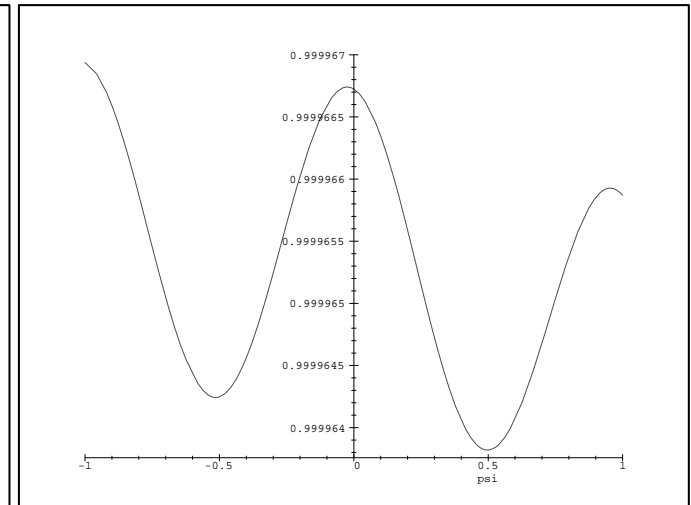
**Fig. 9:** First full part distribution ( $m = 2$ , observed = circle, asymptotic = line)



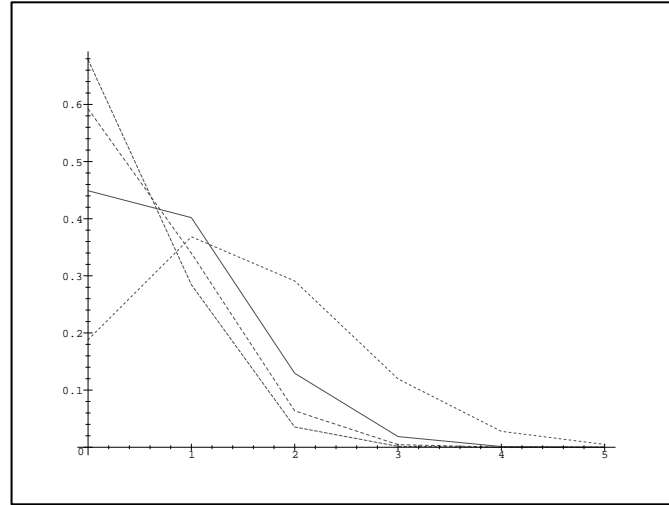
**Fig. 10:**  $P(2,0)$  as a function of  $\psi$



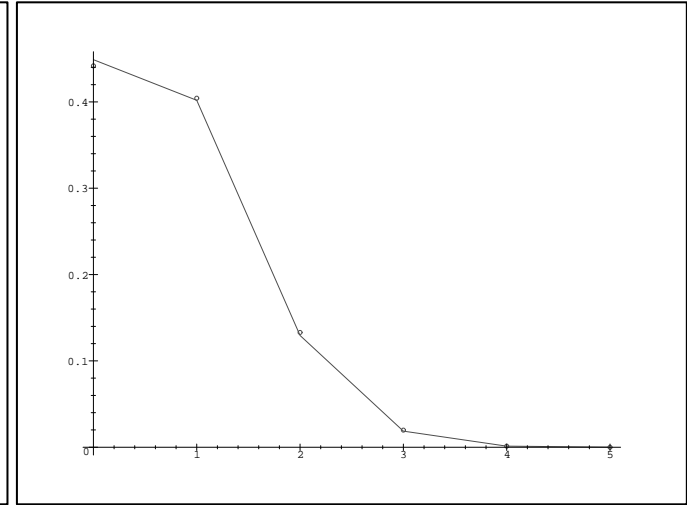
**Fig. 11:**  $\sum_{i=0}^3 P(2,i)$



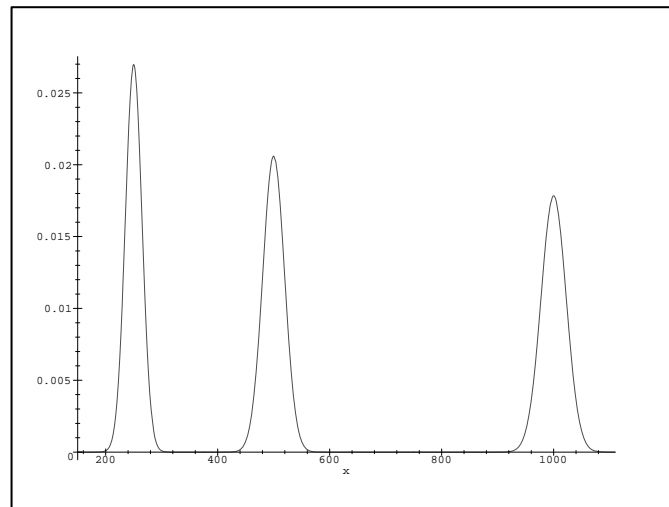
**Fig. 12:**  $\sum_{i=0}^4 P(2,i)$



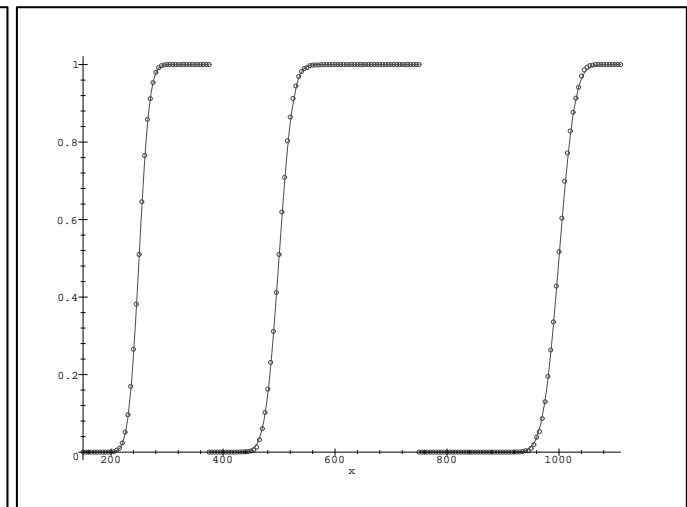
**Fig. 13:**  $P(m, i), m = 1, \dots, 4$



**Fig. 14:** Distribution of  $X(2)$  (observed = circle, asymptotic = line)



**Fig. 15:**  $\Pr(X_i(m) = 1), i = 1, \dots, 3$



**Fig. 16:** Distribution function of  $M_i, i = 1, \dots, 3$  (observed = circle, asymptotic = line)



## References

- [Chi67] V. P. Chistyakov. Discrete limit distributions in the problem of balls falling in cells with arbitrary probabilities. *Math. Notes*, 1:6–11, 1967.
- [CPSW99] Sylvie Corteel, Boris Pittel, Carla D. Savage, and Herbert S. Wilf. On the multiplicity of parts in a random partition. *Random Structures Algorithms*, 14(2):185–197, 1999.
- [FGD95] Philippe Flajolet, Xavier Gourdon, and Philippe Dumas. Mellin transforms and asymptotics: harmonic sums. *Theoret. Comput. Sci.*, 144(1-2):3–58, 1995. Special volume on mathematical analysis of algorithms.
- [FS94] P. Flajolet and R. Sedgewick. Analytic combinatorics – symbolic combinatorics: Saddle point asymptotics. Book in preparation. See also Technical Report 2376, INRIA, 1994.
- [GKP89] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete mathematics*. Addison-Wesley Publishing Company Advanced Book Program, Reading, MA, 1989. A foundation for computer science.
- [HL01] Paweł Hitczenko and Guy Louchard. Distinctness of compositions of an integer: a probabilistic analysis. *Random Structures Algorithms*, 19(3-4):407–437, 2001. Analysis of algorithms (Krynica Morska, 2000).
- [HRS02] Paweł Hitczenko, Cecil Rousseau, and Carla D. Savage. A generating functionology approach to a problem of Wilf. *J. Comput. Appl. Math.*, 142(1):107–114, 2002. Probabilistic methods in combinatorics and combinatorial optimization.
- [HS99] P. Hitczenko and C.D. Savage. On the multiplicity of parts in a random composition of a large integer. Technical report. Available at <http://www.csc.ncsu.edu/faculty/savage/>, 1999.
- [HY97] H.-K. Hwang and Y.-N. Yeh. Measures of distinctness for random partitions and compositions of an integer. *Adv. in Appl. Math.*, 19(3):378–414, 1997.
- [JS98] Philippe Jacquet and Wojciech Szpankowski. Analytical de-Poissonization and its applications. *Theoret. Comput. Sci.*, 201(1-2):1–62, 1998.
- [Knu73] Donald E. Knuth. *The art of computer programming. Volume 3*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1973. Sorting and searching, Addison-Wesley Series in Computer Science and Information Processing.
- [Lou87] G. Louchard. Exact and asymptotic distributions in digital and binary search trees. *RAIRO Inform. Théor. Appl.*, 21(4):479–495, 1987.
- [LP02] Guy Louchard and Helmut Prodinger. Probabilistic analysis of Carlitz compositions. *Discrete Math. Theor. Comput. Sci.*, 5(1):71–95 (electronic), 2002.
- [SČ64] B. A. Sevast'janov and V. P. Čistjakov. Asymptotic normality in the classical problem of balls. *Theory of Probability and Applications*, 9:198–211, 1964.