Available online at www.sciencedirect.com
SCIENCE DIRECT®
Discrete Mathematics 294 (2005) 225–239
DISCRETE MATHEMATICS
www.elsevier.com/locate/disc

Gap-free compositions and gap-free samples of geometric random variables

Paweł Hitczenko[a,1], Arnold Knopfmacher[b,2]

[a] *Departments of Mathematics and Computer Science, Drexel University, 3141 Chestnut Str., Philadelphia, PA 19104, USA*

[b] *The John Knopfmacher Centre for Applicable Analysis and Number Theory, Department of Computational and Applied Mathematics, University of the Witwatersrand, P.O. Wits, 2050 Johannesburg, South Africa*

Received 5 February 2004; received in revised form 17 February 2005; accepted 23 February 2005
Available online 7 April 2005

Abstract

We study the asymptotic probability that a random composition of an integer n is *gap-free*, that is, that the sizes of parts in the composition form an interval. We show that this problem is closely related to the study of the probability that a sample of independent, identically distributed random variables with a geometric distribution is likewise gap-free.

Keywords: Integer compositions; Random compositions; Samples of geometric random variables; Analytic combinatorics

1. Introduction

A composition of a natural number n is said to be *gap-free* if the part sizes occurring in it form an interval. In addition if the interval starts at 1, the composition is said to be *complete*.

E-mail addresses: phitczenko@mcs.drexel.edu (P. Hitczenko), arnoldkn@cam.wits.ac.za (A. Knopfmacher).
URLS: http://www.mcs.drexel.edu/~phitczen, http://www.wits.ac.za/science/number_theory/arnold.htm.

[1] Supported in part by the NSA Grant MSPF-02G-043. The research was conducted during a visit to the John Knopfmacher Centre for Applicable Analysis and Number Theory at the University of Witwatersrand, Johannesburg, South Africa. He would like to thank the Centre for the invitation and for financial support.

[2] This material is based upon work supported by the National Research Foundation under Grant no. 2053740.

0012-365X/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.disc.2005.02.008

Example. Of the 32 compositions of $n=6$, there are 21 gap-free compositions arising from permuting the order of the parts of the partitions

$$6, 3+3, 3+2+1, 2+2+2, 2+2+1+1, 2+1+1+1+1, \\ 1+1+1+1+1+1$$

and 18 complete compositions arising from permuting of the order of the parts in

$$3+2+1, 2+2+1+1, 2+1+1+1+1, 1+1+1+1+1+1.$$

The sequence enumerating the number of gap-free compositions for $n=1,2,3\ldots$ is

$$1, 2, 4, 6, 11, 21, 39, 71, 141, 276, 542, 1070, 2110, 4189, 8351, 16618, 33134, \\ 66129, 131937, \ldots .$$

The corresponding sequence enumerating the number of complete compositions for $n=1,2,3\ldots$ is

$$1, 1, 3, 4, 8, 18, 33, 65, 127, 264, 515, 1037, 2052, 4103, 8217, 16408, 32811, \\ 65590, 131127, \ldots .$$

In both the cases it is apparent that the nth entry is quite close to the value 2^{n-2}. Since, as is well known, there are 2^{n-1} compositions of n, this leads to the following conjecture:

Conjecture 1. *The proportion of gap-free or of complete compositions of n tends to* $\frac{1}{2}$ *as* $n \to \infty$.

This conjecture will be established below. In fact we establish a more quantitative version of this result. Namely we prove the following.

Theorem 1. *The proportion of gap-free or of complete compositions of n is*

$$\tfrac{1}{2} + \mathrm{O}(\log^{3/2} n/\sqrt{n})$$

as $n \to \infty$.

We remark that the analogous problem of enumerating gap-free partitions and gap-free set partitions has been studied in [7] and [6], respectively.

The following elementary computation shows that it is sufficient to study the case of complete compositions.

Proposition 1. *Let* $g(n)$ *be the number of gap-free but not complete compositions of n. Then* $g(n) = \mathrm{O}(\phi^n)$ *as* $n \to \infty$, *where* $\phi \approx 1.6180$ *denotes the golden ratio*.

Proof. For $n \geqslant 1$, $g(n) \leqslant h(n)$ where $h(n)$ is the number of compositions of n with no ones. Now

$$\sum_{n=1}^{\infty} h(n)x^n = \frac{x^2/(1-x^2)}{1-x^2/(1-x)} = x\,\frac{x}{1-x-x^2} = \sum_{n=1}^{\infty} F_{n-1}x^n,$$

where F_n denotes the nth Fibonacci number. From Binet's formula for F_n, $F_n = \mathrm{O}(\phi^n)$, as $n \to \infty$. $\square$

It will be convenient to adopt a probabilistic viewpoint. That is, rather than think of proportion of complete compositions we will equip the set of all compositions of n with the uniform probability measure and will be interested in the probability that a randomly chosen composition of n is complete. In that setting, compositions of n are closely related to the special case $p = \frac{1}{2}$ of *geometric random variables* as shown in [8,9] and again in the next section. This led us to consider the same question for samples of geometric random variables. Specifically, let $\Gamma_1, \Gamma_2, \Gamma_3, \ldots,$ be independent identically distributed geometric random variables with parameter p, that is, $\mathbb{P}(\Gamma_1 = j) = pq^{j-1}$, $j = 1, 2, \ldots$, with $p + q = 1$. We will be interested in the probability that a random sample of n such variables is gap-free. Once again we can restrict our attention to the probability p_n that a sample of length n is complete, since the probability that a geometric sample of length n has no ones is exponentially small.

In the case $p = \frac{1}{2}$ this probability turns out to be exactly $\frac{1}{2}$. The case of $p \neq \frac{1}{2}$ is more interesting. In fact, in that case the sequence (p_n) does not have a limit, but exhibits small oscillations. An asymptotic expression for p_n in the case of $p \neq \frac{1}{2}$ is derived in Section 3.

Some of the previous studies relating to combinatorics of geometric random variables are as follows. In [16] the number of left-to-right maxima was investigated in the model of *words* (strings) $a_1, \ldots, a_n$, where the letters $a_i \in \mathbb{N}$ are independently generated according to the geometric distribution. Hwang and his collaborators obtained further results about this limiting behaviour in [3]. The two parameters 'value' and 'position' of the rth left-to-right maximum for geometric random variables were considered in a subsequent paper [13]. Other combinatorial questions have been considered by Prodinger in e.g. [17,18].

The combinatorics of geometric random variables has gained importance because of their applications in computer science. We mention just two areas: *skiplists* [4,15,19] and *probabilistic counting* [5,12].

The rest of this paper is organized as follows. In the forthcoming section we will show how to reduce the case of compositions to that of samples of geometric random variables. This is the main step in the proof of Theorem 1. In Section 3 we obtain a recurrence relation (3.1) for the probability that a sample of n geometric variables with parameter p is complete. Since for $p = \frac{1}{2}$ this recurrence can be trivially solved, the proof of Theorem 1 is complete once this recurrence is derived. The rest of Section 3 is devoted to the analysis of the recurrence (3.1) for $p \neq \frac{1}{2}$. We close with a short Section 4 containing a few remarks and comments.

2. Reduction to geometric samples

The starting point is the following representation of compositions of n (see e.g. [2]). Consider sequences of n black and white dots subject to the following constraints:

(i) the last dot is always black,
(ii) each of the remaining $n - 1$ dots is black or white.

Then there is a 1-1 correspondence between all such sequences and compositions of n. Namely, part sizes in a composition correspond to "waiting times" for occurrences of black dots. For example, the sequence

$$\underbrace{\bullet}_{1}\ \underbrace{\circ\quad\circ\quad\bullet}_{3}\ \underbrace{\circ\ \ \bullet}_{2}\ \underbrace{\bullet}_{1}\ \underbrace{\bullet}_{1}\ \underbrace{\circ\ \ \bullet}_{2}\ \underbrace{\circ\ \ \bullet}_{2}$$

represents the composition of 12 into parts $(1, 3, 2, 1, 1, 2, 2)$. As discussed e.g. in [8,9] this leads to the following representation of random compositions. Let $p = \frac{1}{2}$ and define

$$\tau = \tau_n = \inf\{k \geqslant 1 : \Gamma_1 + \Gamma_2 + \cdots + \Gamma_k \geqslant n\}.$$

Then a randomly chosen composition κ of n has distribution given by

$$\kappa = \left(\Gamma_1, \Gamma_2, \ldots, \Gamma_{\tau-1}, n - \sum_{j=1}^{\tau-1} \Gamma_j\right) := (\tilde{\Gamma}_1, \tilde{\Gamma}_2, \ldots, \tilde{\Gamma}_\tau).$$

Furthermore, τ has known distribution, namely,

$$\tau \stackrel{\mathrm{d}}{=} 1 + \mathrm{Bin}\left(n-1, \tfrac{1}{2}\right),$$

where $\mathrm{Bin}(m, p)$ denotes a binomial random variable with parameters m and p and $\stackrel{\mathrm{d}}{=}$ stands for equality in distribution. Hence, τ is heavily concentrated around its mean. Specifically, since $\mathrm{var}(\tau) = \mathrm{var}(\mathrm{Bin}(n-1, 1/2)) = (n-1)/4$, for every $t > 0$ we have (see [1, Section A.1])

$$\mathbb{P}(|\tau - \mathsf{E}\tau| \geqslant t) \leqslant 2 \exp\left\{-\frac{2t^2}{n-1}\right\}.$$

In particular, for $t_n \sim \sqrt{cn \ln n}$,

$$\mathbb{P}(|\tau - \mathsf{E}\tau| \geqslant t_n) = \mathrm{O}\left(\frac{1}{n^{2c}}\right),$$

for any $c > 0$.

Let $\mathbb{P}(\kappa \in \mathscr{C})$ be the probability that a random composition is complete. We proceed by series of refinements. First, we have

$$\mathbb{P}(\kappa \in \mathscr{C}) = \mathbb{P}(\{\kappa \in \mathscr{C}\} \cap \{|\tau - \mathsf{E}\tau| < t_n\}) + \mathrm{O}\left(\frac{1}{n^{2c}}\right). \tag{2.1}$$

Indeed, the upper bound follows from

$$\begin{aligned}\mathbb{P}(\kappa \in \mathscr{C}) &= \mathbb{P}(\{\kappa \in \mathscr{C}\} \cap \{\tau - \mathsf{E}\tau| < t_n\}) + \mathbb{P}(\{\kappa \in \mathscr{C}\} \cap \{\tau - \mathsf{E}\tau| \geqslant t_n\}) \\ &\leqslant \mathbb{P}(\{\kappa \in \mathscr{C}\} \cap \{\tau - \mathsf{E}\tau| < t_n\}) + \mathrm{O}\left(\frac{1}{n^{2c}}\right),\end{aligned}$$

and since clearly

$$\mathbb{P}(\kappa \in \mathscr{C}) \geqslant \mathbb{P}(\{\kappa \in \mathscr{C}\} \cap \{|\tau - \mathsf{E}\tau| < t_n\}),$$

the lower bound holds, too. We now set m_n^- and m_n^+ to be

$$m_n^- = \left\lfloor \frac{n+1}{2} - t_n \right\rfloor, \quad m_n^+ = \left\lceil \frac{n+1}{2} + t_n \right\rceil.$$

We will argue that with overwhelming probability, κ is complete if and only if the first m_n^- of its parts are complete. Let $A_n = \{|\tau - \mathsf{E}\tau| < t_n\}$ and for any $m \geqslant 1$ let $\tilde{V}_m$ be a (random) set of values taken by the first m of $\tilde{\Gamma}$'s. That is

$$\tilde{V}_m(\omega) = \{j \geqslant 1 : \exists 1 \leqslant k \leqslant m : \tilde{\Gamma}_k(\omega) = j\}.$$

We will show that

$$\mathbb{P}(\{\kappa \in \mathscr{C}\} \cap A_n) = \mathbb{P}\left(\{\kappa \in \mathscr{C}\} \cap A_n \cap \bigcap_{k=m_n^-+1}^{m_n^+} \{\tilde{\Gamma}_k \in \tilde{V}_{m_n^-}\}\right) + \mathrm{O}\left(\frac{\ln^{3/2} n}{\sqrt{n}}\right). \quad (2.2)$$

This again follows by intersecting $\{\kappa \in \mathscr{C}\} \cap A_n$ with the set

$$\bigcap_{k=m_n^-+1}^{m_n^+} \{\tilde{\Gamma}_k \in \tilde{V}_{m_n^-}\}$$

and its complement and arguing that the latter intersection has negligible probability. To this end, note that

$$\{\tilde{\Gamma}_k \notin \tilde{V}_{m_n^-}\} \subset \{\tilde{\Gamma}_k \notin \tilde{V}_{m_n^- -1}\}$$

and that on the set

$$A_n = \{m_n^- < \tau < m_n^+\}$$

we have $\tilde{\Gamma}_j = \Gamma_j$ for $j < m_n^-$ so that, on A_n, $\tilde{V}_{m_n^- -1} = V_{m_n^- -1}$ where

$$V_m(\omega) := \{j \geqslant 1 : \exists 1 \leqslant k \leqslant m : \Gamma_k(\omega) = j\},$$

is the set of values taken by the first m geometric variables. Hence,

$$\begin{aligned} &\mathbb{P}\left(\{\kappa \in \mathscr{C}\} \cap A_n \cap \bigcup_{k=m_n^-+1}^{m_n^+} \{\tilde{\Gamma}_k \notin \tilde{V}_{m_N^-}\}\right) && (2.3)\\ &\quad \leqslant \mathbb{P}\left(\{m_n^- < \tau < m_n^+\} \cap \bigcup_{k=m_n^-+1}^{m_n^+} \{\tilde{\Gamma}_k \notin V_{m_n^-}\}\right) \\ &\quad \leqslant \mathbb{P}\left(\bigcup_{m_n^- < k \leqslant m_n^+} \{\tilde{\Gamma}_k \notin V_{m_n^-}\}\right) \\ &\quad \leqslant \sum_{k=m_n^-+1}^{m_n^+} \mathbb{P}(\tilde{\Gamma}_k \notin V_{m_n^- -1}), && (2.4) \end{aligned}$$

and it remains to estimate the probability of $\{\tilde{\Gamma}_k \notin V_{m_n^- -1}\}$. Let

$$\ell_n = \lfloor \lg(m_n^-/\ln n)\rfloor \sim \lg n.$$

We first show that with high probability the set $V_{m_n^- -1}$ contains all the integers 1 through ℓ_n. Indeed, if this is not the case then there exists a $1 \leqslant j \leqslant \ell_n$ such that for every $1 \leqslant k < m_n^-$, $\Gamma_k \neq j$. The probability of this happening is

$$\begin{aligned}
\mathbb{P}(\exists j \leqslant \ell_n, \forall i < m_n^-, \Gamma_i \neq j) &\leqslant \mathbb{P}\left(\bigcup_{j=1}^{\ell_n} \bigcap_{i=1}^{m_n^- -1} \{\Gamma_i \neq j\}\right) \leqslant \sum_{j=1}^{\ell_n} \mathbb{P}\left(\bigcap_{i=1}^{m_n^- -1} \{\Gamma_i \neq j\}\right) \\
&= \sum_{j=1}^{\ell_n} (\mathbb{P}(\Gamma_1 \neq j))^{m_n^- -1} = \sum_{j=1}^{\ell_n} \left(1 - \frac{1}{2^j}\right)^{m_n^- -1} \leqslant \sum_{j=1}^{\ell_n} \exp\left\{-\frac{m_n^- - 1}{2^j}\right\} \\
&= \sum_{j=1}^{\ell_n} \exp\left\{-\frac{m_n^- - 1}{2^{\ell_n}}\, 2^{\ell_n - j}\right\} \leqslant \sum_{k=0}^{\infty} \exp\left\{-\frac{m_n^- - 1}{2^{\ell_n}} 2^k\right\} \leqslant \sum_{k=1}^{\infty} \exp\left\{-\frac{m_n^- - 1}{2^{\ell_n}} k\right\} \\
&= \frac{\exp\{-(m_n^- - 1)/2^{\ell_n}\}}{1 - \exp\{-(m_n^- - 1)/2^{\ell_n}\}} = \mathrm{O}\left(\frac{1}{n}\right),
\end{aligned}$$

with our choice of ℓ_n. Let $B_n = \bigcap_{j=1}^{\ell_n} \{j \in V_{m_n^- -1}\}$, we have just shown that $\mathbb{P}(B_n^c) = \mathrm{O}(n^{-1})$. Intersecting the set $\{\tilde{\Gamma}_k \notin V_{m_n^- -1}\}$ with B_n and its complement we once again see that

$$\mathbb{P}(\tilde{\Gamma}_k \notin V_{m_n^- -1}) = \mathbb{P}(\{\tilde{\Gamma}_k \notin V_{m_n^- -1}\} \cap B_n) + \mathrm{O}(n^{-1}). \tag{2.5}$$

On B_n, $\tilde{\Gamma}_k \notin V_{m_n^- -1}$ implies that $\tilde{\Gamma}_k > \ell_n$ which in turn implies that $\Gamma_k > \ell_n$ since for all $k \geqslant 1$, $\tilde{\Gamma}_k \leqslant \Gamma_k$. Thus,

$$\mathbb{P}(\{\tilde{\Gamma}_k \notin V_{m_n^- -1}\} \cap B_n) \leqslant \mathbb{P}(\{\Gamma_k > \ell_n\} \cap B_n) \leqslant \frac{1}{2^{\ell_n}} = \mathrm{O}\left(\frac{\ln n}{n}\right). \tag{2.6}$$

Substituting (2.6) into (2.5) and the result into (2.4) and summing up over k from $m_n^- + 1$ to m_n^+ shows that (2.3) is bounded above by

$$\frac{m_n^+ - m_n^-}{2^{\ell_n}} = \mathrm{O}(\sqrt{n \ln n})\, \mathrm{O}\left(\frac{1}{n} + \frac{\ln n}{n}\right) = \mathrm{O}\left(\frac{\ln^{3/2} n}{\sqrt{n}}\right).$$

This justifies our claim (2.2). Combining (2.2) with (2.1) yields

$$\mathbb{P}(\kappa \in \mathscr{C}) = \mathbb{P}\left(\{\kappa \in \mathscr{C}\} \cap \{m_n^- < \tau < m_n^+\} \cap \bigcap_{k=m_n^-+1}^{m_n^+} \{\tilde{\Gamma}_k \in \tilde{V}_{m_n^-}\}\right) + \mathrm{O}\left(\frac{\ln^{3/2} n}{\sqrt{n}}\right).$$

We now note that on the set

$$\{m_n^- < \tau < m_n^+\} \cap \bigcap_{k=m_n^-+1}^{m_n^+} \{\tilde{\Gamma}_k \in \tilde{V}_{m_n^-}\}$$

κ is complete if and only if $(\Gamma_1, \Gamma_2, \dots, \Gamma_{m_n^-})$ is complete. This is because on that set $\tau > m_n^-$ so that $\tilde{\Gamma}_j = \Gamma_j$ for $1 \leqslant j \leqslant m_n^-$ and the remaining $\tilde{\Gamma}_k$ do not contribute new values. This means that

$$\mathbb{P}\left(\{\kappa \in \mathscr{C}\} \cap \{m_n^- < \tau < m_n^+\} \cap \bigcap_{k=m_n^-+1}^{m_n^+} \{\tilde{\Gamma}_k \in \tilde{V}_{m_n^-}\}\right)$$
$$= \mathbb{P}\left(\{(\Gamma_1, \dots, \Gamma_{m_n^-}) \in \mathscr{C}\} \cap \{m_n^- < \tau < m_n^+\} \cap \bigcap_{k=m_n^-+1}^{m_n^+} \{\tilde{\Gamma}_k \in \tilde{V}_{m_n^-}\}\right).$$

Finally, using again virtually the same argument as before we obtain that the latter probability is

$$\mathbb{P}((\Gamma_1, \dots, \Gamma_{m_n^-}) \in \mathscr{C}) + \mathrm{O}\left(\frac{\ln^{3/2} n}{\sqrt{n}}\right)$$

which proves that

$$\mathbb{P}(\kappa \in \mathscr{C}) = \mathbb{P}((\Gamma_1, \dots, \Gamma_{m_n^-}) \in \mathscr{C}) + \mathrm{O}\left(\frac{\ln^{3/2} n}{\sqrt{n}}\right),$$

thereby reducing the problem to samples of geometric random variables.

3. Geometric samples

Consider $\Gamma = (\Gamma_1, \Gamma_2, \dots, \Gamma_n)$ a sample of n i.i.d. geometric random variables with parameter p. Let $p_n = \mathbb{P}(\Gamma \in \mathscr{C})$ be the probability that $(\Gamma_1, \dots, \Gamma_n)$ is complete. To obtain a recurrence relation we condition on the number of Γ_j's that are equal to 1. Since being complete implies that there is at least one 1 among the values of Γ_i's, by the law of total probability we find that

$$p_n = \sum_{j=1}^{n} \mathbb{P}\left(\{\Gamma \in \mathscr{C}\} \cap \left\{\sum_{\ell=1}^{n} I_{\Gamma_\ell = 1} = j\right\}\right)$$
$$= \sum_{j=1}^{n} \mathbb{P}\left(\Gamma \in \mathscr{C} \,\middle|\, \sum_{\ell=1}^{n} I_{\Gamma_\ell=1} = j\right) \mathbb{P}\left(\sum_{\ell=1}^{n} I_{\Gamma_\ell=1} = j\right).$$

We now observe that, given that j out of n Γ_k's are one, the sample is complete if and only if the remaining $n-j$ variables take on all the values between 2 and their maximum. This is the same as to say that the sample $(\Gamma_k - 1)$ of length $n-j$ is complete, given that all $n-j$ of them are at least 2. But, by the memoryless property of geometric random variables, the conditional distribution of $\Gamma_k - 1$ given that $\Gamma_k \geqslant 2$ is just that of Γ_k. Since those of Γ_k's that are at least 2 remain independent, this just means that

$$\mathbb{P}\left(\Gamma \in \mathscr{C} \,\middle|\, \sum_{\ell=1}^{n} I_{\Gamma_\ell=1} = j\right) = p_{n-j}.$$

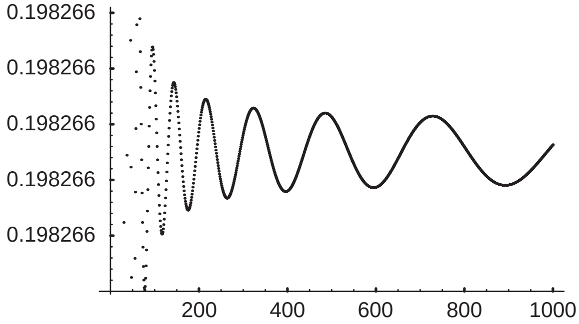

Fig. 1. Plot of p_n for $p = \frac{1}{3}$ and $1 \leqslant n \leqslant 1000$.

Since

$$\mathbb{P}\left(\sum_{\ell=1}^{n} I_{\Gamma_\ell = 1} = j\right) = \binom{n}{j} p^j q^{n-j},$$

substituting these two expressions and changing the order of summation by letting $k = n - j$, we obtain the following recurrence for p_n's:

$$p_n = \begin{cases} 1 & \text{if } n = 0; \\ \sum_{k=0}^{n-1} p_k \binom{n}{k} q^k p^{n-k} & \text{if } n \geqslant 1. \end{cases} \tag{3.1}$$

Before continuing with the analysis of this recurrence, let us observe that the proof of Theorem 1 can be immediately completed by noticing that if $p = 1/2$ then the sequence $p_0 = 1$ and $p_k = 1/2$ for $k \geqslant 1$ is a solution of (3.1). Indeed, proceeding inductively we get

$$p_n = \sum_{k=0}^{n-1} p_k \binom{n}{k} q^k p^{n-k} = \frac{1}{2^n} + \sum_{k=1}^{n-1} \frac{1}{2} \binom{n}{k} \frac{1}{2^n} = \frac{1}{2} \sum_{k=0}^{n} \binom{n}{k} \frac{1}{2^n} = \frac{1}{2}.$$

The case of $p \neq 1/2$ is more interesting. In fact, in that case the sequence (p_n) does not have a limit, but exhibits small oscillations. As illustrated in Figs. 1–3, both the period and amplitude of the oscillations vary with the size of p.

To treat this case we will follow an approach that became quite common in the analysis of certain algorithms (see e.g. numerous examples in [10,20]). In our presentation we follow [11] where functional equations similar to (3.2) below were studied. We Poissonize the problem by considering the Poisson transform of the sequence (p_n), analyse its asymptotics, and then we de-Poissonize to recover the asymptotics of (p_n). To carry out this program let

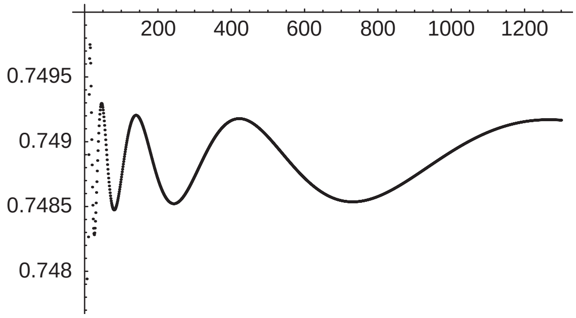

Fig. 2. Plot of p_n for $p = \frac{2}{3}$ and $1 \leqslant n \leqslant 1300$.

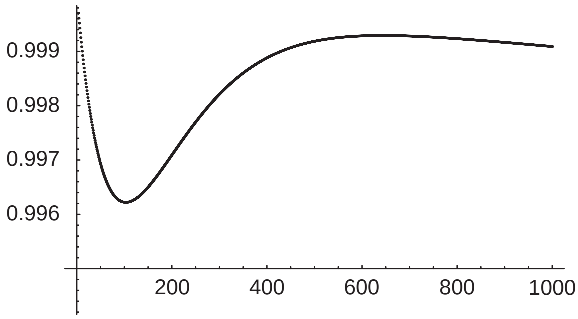

Fig. 3. Plot of p_n for $p = 0.99$ and $1 \leqslant n \leqslant 1000$.

$P(z)$, for z complex, be the Poisson transform of (p_n). That is

$$\begin{aligned}
P(z) = \sum_{n=0}^{\infty} p_n \frac{z^n}{n!} \mathrm{e}^{-z} &= p_0 \mathrm{e}^{-z} + \sum_{n=1}^{\infty} \frac{z^n \mathrm{e}^{-z}}{n!} \left\{ \sum_{\ell=0}^{n-1} p_\ell \binom{n}{\ell} q^\ell p^{n-\ell} \right\} \\
&= \mathrm{e}^{-z} + \mathrm{e}^{-z} \sum_{n=1}^{\infty} \sum_{\ell=0}^{n-1} p_\ell \frac{q^\ell z^\ell p^{n-\ell} z^{n-\ell}}{\ell!(n-\ell)!} \\
&= \mathrm{e}^{-z} + \mathrm{e}^{-z} \sum_{\ell=0}^{\infty} p_\ell \frac{q^\ell z^\ell}{\ell!} \sum_{n=\ell+1}^{\infty} \frac{p^{n-\ell} z^{n-\ell}}{(n-\ell)!} \\
&= \mathrm{e}^{-z} + \mathrm{e}^{-z} \sum_{\ell=0}^{\infty} p_\ell \frac{q^\ell z^\ell}{\ell!} \{\mathrm{e}^{pz} - 1\} \\
&= \mathrm{e}^{-z} + \mathrm{e}^{-qz} \sum_{\ell=0}^{\infty} p_\ell \frac{q^\ell z^\ell}{\ell!} \{1 - \mathrm{e}^{-pz}\} \\
&= \mathrm{e}^{-z} + (1 - \mathrm{e}^{-pz}) P(qz).
\end{aligned}$$

Hence $P(z)$ satisfies the following functional equation

$$\begin{aligned} P(z) &= P(qz) + \mathrm{e}^{-z} - \mathrm{e}^{-pz}P(qz) = P(qz) + \mathrm{e}^{-z} - \mathrm{e}^{-z}\sum_{n=0}^{\infty} p_n \frac{q^n z^n}{n!} \\ &= P(qz) - T(z), \end{aligned} \tag{3.2}$$

where $T(z) = \mathrm{e}^{-z}\sum_{n=1}^{\infty} p_n(q^n z^n/n!)$, and since it is clear from (3.1) and the binomial formula that $0 \leqslant p_n \leqslant 1$, the series converges absolutely for every z. Moreover, for $x \in \mathbf{R}$, $T(x) = \mathrm{O}(x)$ as $x \to 0_+$ and $T(x)$ has exponential decay as $x \to \infty$. Thus Mellin transform of $T(x)$ exists in the strip $\langle -1, \infty\rangle := \{s \in \mathbf{C} : -1 < \Re(s) < \infty\}$. By direct iteration we obtain for every $m \geqslant 0$

$$P(z) = P(qz) - T(z) = P(q^{m+1}z) - \sum_{j=0}^{m} T(q^j z),$$

and by passing to a limit with m,

$$P(z) = 1 - \sum_{j=0}^{\infty} T(q^j z).$$

Letting $Q(z) = P(z) - 1 = -\sum_{j=0}^{\infty} T(q^j z)$ and taking Mellin transform we obtain

$$Q^*(s) = -\sum_{j=0}^{\infty} q^{-js}T^*(s) = -\frac{T^*(s)}{1 - q^{-s}} = \frac{T^*(s)}{q^{-s} - 1},$$

provided that series converges. Since this happens for $\Re(s) < 0$, $Q^*(s)$ will exist in a strip $\langle -1, 0\rangle$. Inverting Mellin transform yields

$$Q(x) = \frac{1}{2\pi \mathrm{i}} \int_{c-\mathrm{i}\infty}^{c+\mathrm{i}\infty} Q^*(s)x^{-s}\,\mathrm{d}s,$$

for any $-1 < c < 0$. This integral can be evaluated with the aid of the residue theorem. Letting $C_{M,R}$ denote the rectangular contour with vertices $c \pm \mathrm{i}M$, and $R \pm \mathrm{i}M$, $R > 0$, $M \notin \mathbf{Z}$ and accounting for the fact that the integral is taken clockwise we get

$$\int_{C_{M,R}} Q^*(s)x^{-s}\,\mathrm{d}s = -2\pi\mathrm{i}\sum_{\chi} \mathrm{Res}_{s=\chi}(Q^*(s)x^{-s}),$$

where the sum is taken over all residues enclosed by $C_{M,R}$. Splitting the left-hand side as

$$\left(\int_{c-\mathrm{i}M}^{c+\mathrm{i}M} + \int_{\tilde{C}_{M,R}}\right) Q^*(s)x^{-s}\,\mathrm{d}s,$$

and noting that in view of the asymptotic formula

$$|\Gamma(x+\mathrm{i}y)| = \mathrm{O}(|y|^{x-1/2}\mathrm{e}^{-\pi|y|/2}) \quad \text{as } |y| \to \infty,$$

the second term becomes negligible as $M \to \infty$, $R > 0$ is fixed, and $x \to \infty$ (see e.g. [14, Section 5.2.2] for more details) we get

$$Q(x) \sim -\sum_{\chi} \mathrm{Res}_{s=\chi}(Q^*(s)x^{-s}).$$

Now, $x^{-s}Q^*(s) = x^{-s}T^*(s)/(q^{-s}-1)$ has simple poles whenever $q^{-s}=1$, i.e. at $\chi_k = 2k\pi\mathrm{i}/\ln(1/q)$, $k = 0, \pm 1, \pm 2, \ldots,$ with corresponding residues equal to

$$\lim_{s\to\chi_k}(s-\chi_k)\frac{x^{-s}T^*(s)}{q^{-s}-1} = \frac{x^{-\chi_k}T^*(\chi_k)}{q^{-\chi_k}\ln(1/q)} = \left(\frac{q}{x}\right)^{\chi_k}\frac{T^*(\chi_k)}{\ln(1/q)}.$$

The main term comes from $k=0$ and the remaining residues will contribute oscillatory terms of relatively small amplitude.

In order to complete the proof we will need to de-Poissonize this result. Once this is done we will conclude that

$$\begin{aligned} p_n &\sim P(n) = Q(n) + 1 \\ &\sim 1 - \frac{T^*(0)}{\ln(1/q)} - \frac{2}{\ln(1/q)}\,\Re\left(\sum_{k=1}^{\infty}\exp\{\chi_k\ln(q/n)\}T^*(\chi_k)\right). \end{aligned}$$

The values $T^*(\chi_k)$ are given by

$$\begin{aligned} T^*(\chi_k) &= \mathscr{M}\left(\mathrm{e}^{-z}\sum_{j=1}^{\infty}p_j\frac{q^jz^j}{j!};\chi_k\right) = \sum_{j=1}^{\infty}p_j\frac{q^j}{j!}\mathscr{M}(z^j\mathrm{e}^{-z};\chi_k) \\ &= \sum_{j=1}^{\infty}p_j\frac{q^j}{j!}\Gamma(\chi_k+j) = \sum_{j=1}^{\infty}p_jq^j\frac{\Gamma(\chi_k+j)}{\Gamma(j+1)}. \end{aligned}$$

Since the series converge geometrically fast, they can be evaluated numerically with the aid of (3.1). For example, setting $k=0$ gives the main term

$$1 - \frac{T^*(0)}{\ln(1/q)} = 1 - \frac{1}{\ln(1/q)}\sum_{j=1}^{\infty}p_j\frac{q^j}{j}.$$

Its plot as a function of p is given in Fig. 4.

To de-Poissonize we use the fact that $P(z)$ satisfies (3.2) which is a special case of [20, Theorem 10.5] (see also [10]) with $\gamma_1(z)\equiv 0$, $\gamma_2(z)\equiv 1$, and $t(z) = -T(z)$. Thus we need to verify conditions (10.28)–(10.32) of [20]. But this is straightforward: (10.28) holds for any $\beta > 0$, (10.29) holds as well, since

$$|t(z)| = |T(z)| \leqslant |\mathrm{e}^{-z}|\sum_{n=1}^{\infty}p_n\frac{q^n|z|^n}{n!} \leqslant \mathrm{e}^{-\Re(z)}\sum_{n=1}^{\infty}\frac{q^n|z|^n}{n!} \leqslant \mathrm{e}^{-\Re(z)}\mathrm{e}^{q|z|},$$

which is bounded by 1 provided $\Re(z)/|z| > q$, which holds in a cone

$$\mathscr{S}_\theta := \{z = r\mathrm{e}^{\mathrm{i}\theta}:\ \cos\theta > q\}.$$

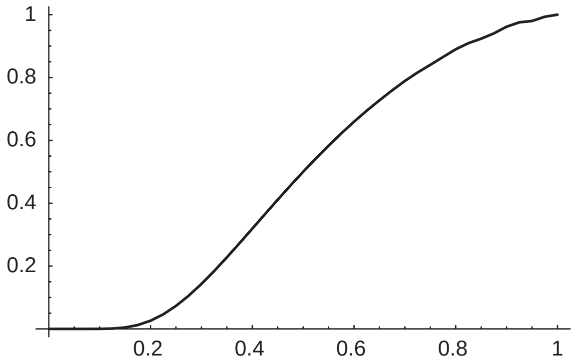

Fig. 4. Plot of the non-oscillating limit term for p_n for $0 \leqslant p \leqslant 1$.

(10.30) is trivial and (10.31) holds outside the cone since for $z \notin \mathscr{S}_\theta$, $\Re(z) < \alpha|z|$ for some $\alpha < 1$. Finally, (10.32) is true since

$$|T(z)|e^{\Re(z)} \leqslant e^{\Re(z)}|e^{-z}| \sum_{n=1}^{\infty} p_n \frac{q^n|z|^n}{n!} \leqslant e^{q|z|} \leqslant \frac{1}{3} e^{\alpha|z|},$$

as long as $q < \alpha < 1$ and $|z|$ is large enough.

4. Further remarks

(i) Since the number of compositions of n grows exponentially, a direct enumeration of the number of complete compositions for $n > 20$ is not practical. However, by using the fact that a complete composition of n arises as a permutation of the parts in a complete partition of n, it is possible to extend the exact enumeration of complete compositions quite a bit further.

The following graph of the difference between the probability that a composition of n is complete and $\frac{1}{2}$ for $n = 21, \ldots, 70$, shows what appears to be an interesting oscillating behaviour of the error that awaits further study. Moreover, the $O(\log^{3/2} n/\sqrt{n})$ error term we obtained cannot be significantly improved by our method. The picture suggests, however, that the error is of much smaller order (Fig. 5).

(ii) As shown above the special case $p = \frac{1}{2}$ does not have oscillations. This also follows from the asymptotic formula for p_n in view of the fact that for $p = \frac{1}{2}$ and $\chi_k = 2k\pi i/\ln 2$,

$$T^*(\chi_k) = \sum_{j=1}^{\infty} \frac{1}{2^{j+1}} \frac{\Gamma(\chi_k + j)}{\Gamma(j+1)} = \frac{1}{2}(-1 + 2^{\frac{2ik\pi}{\ln 2}})\Gamma\left(\frac{2ik\pi}{\ln 2}\right) = 0,$$

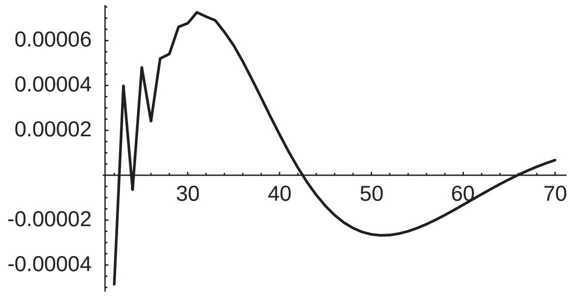

Fig. 5. Plot of the completeness probability minus $\frac{1}{2}$ for $21 \leqslant n \leqslant 70$.

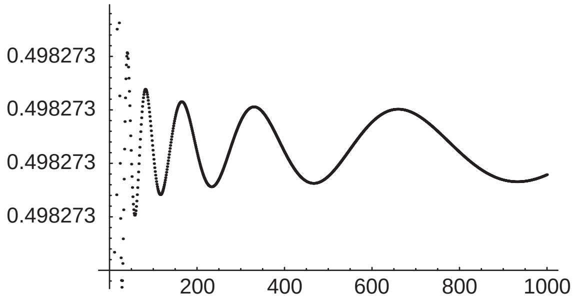

Fig. 6. Plot of p_n for $p = 0.499$ and $1 \leqslant n \leqslant 1000$.

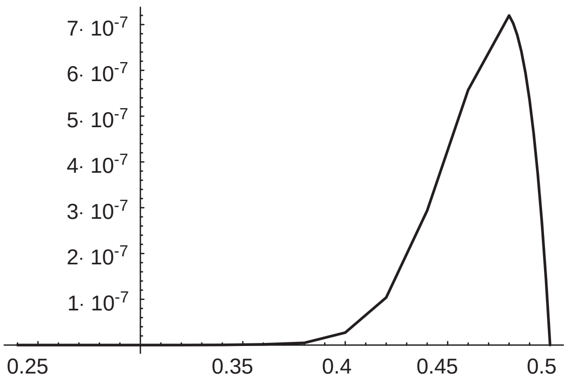

Fig. 7. Plot of the amplitude of the fluctuations for $p \leqslant 0.5$.

for $k = \pm 1, \pm 2, \pm 3, \ldots$. To see this, notice that

$$\frac{\Gamma(\chi_k + j)}{\Gamma(j+1)} = \frac{(\chi_k + j - 1)!}{j!} = (\chi_k - 1)! \binom{\chi_k + j - 1}{j} = (\chi_k - 1)! \binom{-\chi_k}{j} (-1)^j.$$

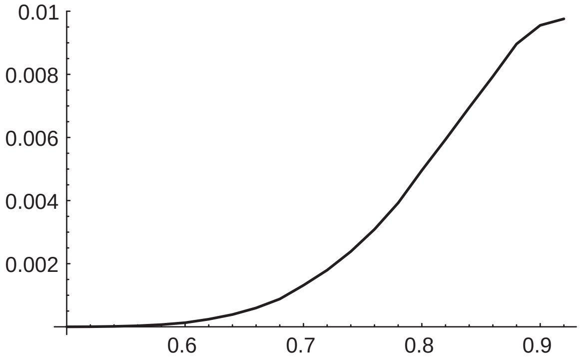

Fig. 8. Plot of the amplitude of the fluctuations for $p \geqslant 0.5$.

Thus

$$T^*(\chi_k) = \sum_{j=1}^{\infty} \frac{1}{2^{j+1}} (\chi_k - 1)! \binom{-\chi_k}{j} (-1)^j = \frac{(\chi_k - 1)!}{2} [-1 + (1 - 1/2)^{-\chi_k}]$$
$$= \frac{\Gamma(\chi_k)}{2} [-1 + 2^{\chi_k}] = 0.$$

(iii) It is interesting to investigate the amplitude of the oscillations on either side of the critical value $p = \frac{1}{2}$. As shown in Fig. 6 these become very small near to the critical value. For $0 \leqslant p \leqslant \frac{1}{2}$ the amplitude of the fluctuations increase steadily up until around $p = 0.48$ and then decrease rapidly to zero. For $p \geqslant \frac{1}{2}$ the amplitude of the fluctuations increase steadily and in general are orders of magnitude larger that for $p < \frac{1}{2}$ (Figs. 7 and 8).

References

[1] N. Alon, J.H. Spencer, The Probabilistic Method, Wiley, New York, 2000.
[2] G.E. Andrews, The Theory of Partitions, Addison-Wesley, Reading, MA, 1976.
[3] Z.-D. Bai, H.-K. Hwang, W.-Q. Liang, Normal approximations of the number of records in geometrically distributed random variables, Random Struct. Alg. 13 (1998) 319–334.
[4] L. Devroye, A limit theory for random skip lists, Ann. Appl. Probab. 2 (1992) 597–609.
[5] P. Flajolet, G.N. Martin, Probabilistic counting algorithms for data base applications, J. Comput. System Sci. 31 (1985) 182–209.
[6] W.M.Y. Goh, E. Schmutz, Gap-free set partitions, Random Struct. Algorithms 3 (1992) 9–18.
[7] P.J. Grabner, A. Knopfmacher, Analysis of some new partition statistics, Ramanujan J., to appear.
[8] P. Hitczenko, G. Louchard, Distinctness of compositions of an integer: a probabilistic analysis, Random Struct. Alg. 19 (2001) 407–437.
[9] P. Hitczenko, C.D. Savage, On the multiplicity of parts in a random composition of a large integer, SIAM J. Discrete Math. 18 (2004) 418–435.
[10] P. Jacquet, W. Szpankowski, Analytical de–Poissonization and its applications, Theoret. Comput. Sci. 201 (1998) 1–62.
[11] S. Janson, W. Szpankowski, Analysis of the asymmetric leader election algorithm, Electr. J. Combin. 4(1): research paper 17 (1997) 16pp.

[12] P. Kirschenhofer, H. Prodinger, in: E. Hlawka, R.F. Tichy (Eds.), On the Analysis of Probabilistic Counting, Lecture Notes in Mathematics, vol. 1452, 1990, pp. 117–120.

[13] A. Knopfmacher, H. Prodinger, Combinatorics of geometrically distributed random variables: value and position of the rth left-to-right maximum, Discrete Math. 226 (2001) 255–267.

[14] D. Knuth, The Art of Computer Programming, vol. 3: Sorting and Searching, Addison-Wesley, Reading, MA, 1973.

[15] T. Papadakis, I. Munro, P. Poblete, Average search and update costs in skip lists, BIT 32 (1992) 316–332.

[16] H. Prodinger, Combinatorics of geometrically distributed random variables: left-to-right maxima, Discrete Math. 153 (1996) 253–270.

[17] H. Prodinger, Combinatorics of geometrically distributed random variables: lengths of ascending runs, LATIN2000, Lecture Notes in Computer Science, vol. 1776, 2000, pp. 473–482.

[18] H. Prodinger, Combinatorics of geometrically distributed random variables: inversions and a parameter of Knuth, Ann. Combin. 5 (2001) 141–150.

[19] W. Pugh, Skip lists: a probabilistic alternative to balanced trees, Comm. ACM 33 (1990) 668–676.

[20] W. Szpankowski, Average Case Analysis of Algorithms on Sequences, Wiley, New York, 2001.