Available online at www.sciencedirect.com
ScienceDirect
Discrete Mathematics 307 (2007) 2871–2890
DISCRETE MATHEMATICS
www.elsevier.com/locate/disc

Note

Gaps in samples of geometric random variables

William M.Y. Goh[a], Pawel Hitczenko[b,1]

[a]Department of Mathematics, Drexel University, Philadelphia, PA 19104, USA
[b]Departments of Mathematics and Computer Science, Drexel University, Philadelphia, PA 19104, USA

Received 23 March 2006; received in revised form 8 September 2006; accepted 22 January 2007
Available online 15 February 2007

Abstract

In this note we continue the study of gaps in samples of geometric random variables originated in Hitczenko and Knopfmacher [Gap-free compositions and gap-free samples of geometric random variables. *Discrete Math.* 294 (2005) 225–239] and continued in Louchard and Prodinger [The number of gaps in sequences of geometrically distributed random variables, Preprint available at ⟨http://www.ulb.ac.be/di/mcs/louchard/⟩ (number 81 on the list) or at ⟨http://math.sun.ac.za/∼ prodinger/pdffiles/gapsAPRIL27.pdf.⟩] In particular, since the notion of a gap differs in these two papers, we derive some of the results obtained in Louchard and Prodinger [The number of gaps in sequences of geometrically distributed random variables, Preprint available at ⟨http://www.ulb.ac.be/di/mcs/louchard/⟩ (number 81 on the list) or at ⟨http://math.sun.ac.za/∼prodinger/pdffiles/gapsAPRIL27.pdf.⟩] for gaps as defined in Hitczenko and Knopfmacher [Gap-free compositions and gap-free samples of geometric random variables. *Discrete Math.* 294 (2005) 225–239].

Keywords: Gaps; Geometric random variables; Asymptotic analysis; Mellin transform

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*. In [2] gap-free and complete compositions were studied and it was shown in particular, that the proportion of gap-free or of complete compositions of n tends to $\frac{1}{2}$ as $n \to \infty$. In probabilistic language this can be expressed as follows; if by a random composition of n we mean a composition picked according to the uniform discrete probability measure on the set of all compositions of n, then the probability that a randomly chosen composition of n is gap-free or complete is asymptotically $\frac{1}{2}$, as $n \to \infty$.

Since integer compositions are closely related to samples of i.i.d. geometric random variables with parameter $\frac{1}{2}$ (GEOM($\frac{1}{2}$)), analogous results hold for such samples. In fact, that was the approach taken in [2]; it was shown that the

[1] This author is supported in part by the NSA grant H98230-05-0016.
E-mail addresses: wgoh@math.drexel.edu (W.M.Y. Goh), phitczenko@math.drexel.edu (P. Hitczenko).
URL: http://www.math.drexel.edu/∼ phitczen (P. Hitczenko).

0012-365X/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.disc.2007.01.013

probabilities of being complete or gap-free are asymptotically the same for compositions as for samples of GEOM($\frac{1}{2}$) random variables, and it was shown that the probability that such a sample is complete is *exactly* $\frac{1}{2}$. Of course, once one is interested in samples of geometric random variables, there is no particular reason to restrict considerations to the particular value $p=\frac{1}{2}$ of the parameter p. This was done in many papers and, in fact, in [2] as well although it should be added that the case $p\neq\frac{1}{2}$ differs from $p=\frac{1}{2}$ in that that the probability of being gap-free does not have a limit as n tends to infinity, but exhibits an oscillatory behavior).

This study was continued by Louchard and Prodinger [5] who considered a more general question of the distribution of the number of gaps in a sample of i.i.d. GEOM(p) random variables. Although they used the term "the number of gaps" a closer inspection of their definitions and arguments reveals that their notion is different than the definition used in [2] and might be more appropriately called "the number of missed values". For example a composition (3, 4, 5, 3) has no gaps according to the definition used in [2], but has two gaps (namely 1 and 2) according to the notion used in [5] (as a matter of fact the difference between "the number of gaps is 0" and "the number of missed values is 0" is exactly the distinction between "gap-free" and "complete" as defined in [2]). The same notion of "gap-free" in the context of set partitions was used as early as 1992 in [1]. To conform to the definition used in [1,2] we say that the composition (or a sample of geometric random variables) has r gaps if the range of part sizes occurring in it is a union of $r+1$ disjoint intervals, and $r+1$ is the smallest number with that property. According to that distinction, a composition (2, 5, 5, 2, 8, 2, 6) of 30 has two gaps, namely the "intervals" [3, 4] and [7], while it misses four values: 1, 3, 4, and 7; thus the two quantities are clearly different.

Louchard and Prodinger carried out the detailed study of the number of values missed in a sample of geometric random variables. Among other things they found the limiting distribution and asymptotics of moments. In particular, they proved.

Theorem 1. *If X_n is the number of values missed in the sample of n i.i.d. GEOM($\frac{1}{2}$) random variables, then, as $n\to\infty$,*

$$\mathsf{E}X_n \sim 1 \quad \textit{and} \quad \mathsf{var}(X_n)\sim 2. \tag{1}$$

Furthermore

$$\Pr(X_n=r)=\frac{1}{2^{r+1}}, \quad r=0,1,2\ldots . \tag{2}$$

In this note, aside from pointing out the difference in terminology, we describe a few results on the number of gaps (defined consistently with [2]) in a sample of geometric random variables. Before continuing we should emphasize, that Louchard and Prodinger consider the general case $0<p<1$. To a large extent we will do the same, but we would like to emphasize that the contrast is most clear for the special case $p=\frac{1}{2}$. In particular, while the results given in (1) and (2) do not exhibit oscillatory behavior (but as was shown in [5] they do if $p\neq\frac{1}{2}$), our analogs do have oscillations, even for $p=\frac{1}{2}$.

2. Expected value and the variance

In this section we prove the following counterpart of (1):

Proposition 2. *If Y_n is the number of gaps in the sample of n i.i.d. GEOM(p) random variables, then, as $n\to\infty$,*

$$\mathsf{E}Y_n=-\frac{\ln(1+q)}{\ln q}+\eta_E(\ln n)+\mathrm{o}(1), \tag{3}$$

where

$$\eta_E(x):=\frac{2}{\ln q}\sum_{k=1}^{\infty}\Re\left(\frac{1-(1+q)^{2k\pi\mathrm{i}/\ln q}}{(p(1+q))^{2k\pi\mathrm{i}/\ln q}}\Gamma\left(\frac{2k\pi\mathrm{i}}{\ln q}\right)\mathrm{e}^{(2k\pi\mathrm{i}/\ln q)x}\right)$$

is a periodic function with small amplitude. Furthermore,

$$\mathsf{var}(Y_n) = \frac{2}{\ln q} \ln\left(\frac{(1+2q)(1+q(1+q))(2+q)}{2(1+q)^4}\right) + \eta_V(\ln n) + \mathrm{o}(1), \tag{4}$$

where $\eta_V(x)$ is a function with features similar to those of η_E.

Proof. Let $\Gamma_1, \Gamma_2, \ldots$ be a sequence of i.i.d. GEOM(p) random variables. Our proof rests on an observation that there is a gap beginning at $j+1$, $j \geqslant 1$, if and only if

$$\exists k\, \Gamma_k = j \quad \text{and} \quad \forall m\ \Gamma_m \neq j+1.$$

This is actually true for all j except the one corresponding to the largest observed value in the sample. Thus, if we set

$$G_j := \left\{ \bigcup_{k=1}^{n} \{\Gamma_k = j\} \cap \bigcap_{m=1}^{n} \{\Gamma_m \neq j+1\} \right\},$$

and we follow the custom of identifying sets with their indicator functions, we see that

$$Y_n + 1 = \sum_{j \geqslant 1} G_j,$$

and consequently

$$\mathsf{E}Y_n + 1 = \sum_{j=1}^{\infty} \mathsf{Pr}(G_j). \tag{5}$$

Now,

$$G_j = \bigcap_{m=1}^{n} \{\Gamma_m \neq j+1\} \backslash \bigcap_{m=1}^{n} \{\Gamma_m \neq j+1, j\},$$

so that

$$\mathsf{Pr}(G_j) = (1-pq^j)^n - (1-pq^{j-1}-pq^j)^n. \tag{6}$$

To proceed further we will need some elementary estimates which we isolate in the following:

Lemma 3. *Let $x, y \geqslant 0$ satisfy $x + y \leqslant 1$. Then for every $n \geqslant 1$ we have*

(i) $|\mathrm{e}^{-nx} - (1-x)^n| = O(1/n)(nx)^2\mathrm{e}^{-nx}$,
(ii) $|(1-x)^n(1-y)^n - (1-x-y)^n| = O(1)nxy\mathrm{e}^{-nx}\,\mathrm{e}^{-ny}$.

Proof. For (i) write

$$\begin{aligned}
0 &\leqslant \mathrm{e}^{-nx} - (1-x)^n = (\mathrm{e}^{-x})^n - (1-x)^n \\
&= (\mathrm{e}^{-x} - (1-x)) \sum_{k=0}^{n-1} \mathrm{e}^{-kx}(1-x)^{n-1-k} \\
&\leqslant (\mathrm{e}^{-x} - (1-x)) \sum_{k=0}^{n-1} \mathrm{e}^{-kx}\mathrm{e}^{-(n-1-k)x} = O(x^2)n\mathrm{e}^{-(n-1)x} \\
&= O(n^{-1})(nx)^2\mathrm{e}^{-nx}.
\end{aligned}$$

For part (ii) we clearly have

$$(1-x)^n(1-y)^n = (1-x-y+xy)^n \geqslant (1-x-y)^n.$$

On the other hand

$$\begin{aligned}(1-x-y+xy)^n - (1-x-y)^n &= xy\sum_{k=0}^{n-1}(1-x-y+xy)^k(1-x-y)^{n-1-k}\\ &\leqslant xyn(1-x-y+xy)^{n-1} = nxy(1-x)^{n-1}(1-y)^{n-1}\\ &\leqslant nxy\mathrm{e}^{-(n-1)x}\mathrm{e}^{-(n-1)y} = O(1)nxy\mathrm{e}^{-nx}\,\mathrm{e}^{-ny},\end{aligned}$$

which proves (ii). □

Using Lemma 3 (i) and (6) we see that

$$\sum_{j\geqslant 1}\Pr(G_j) = \sum_{j\geqslant 1}\mathrm{e}^{-npq^j}(1-\mathrm{e}^{-npq^{j-1}}) + O\left(\frac{1}{n}\right)\sum_{j\geqslant 1}(npq^{j-1})^2\mathrm{e}^{-npq^{j-1}}.$$

Since the sum

$$\sum_{j\geqslant 0}(npq^j)^2\mathrm{e}^{-npq^j} = \int_0^\infty (npq^x)^2\mathrm{e}^{-npq^x}\,\mathrm{d}x + O(1) = O(1),$$

it remains to evaluate the sum

$$\sum_{j\geqslant 1}\mathrm{e}^{-npq^j}(1-\mathrm{e}^{-npq^{j-1}}).$$

To that end consider

$$f(x) := \sum_{j=1}^{\infty}\mathrm{e}^{-xpq^j}\left(1-\mathrm{e}^{-xpq^{j-1}}\right)$$

and its Mellin transform

$$\begin{aligned}\mathscr{M}(f)(s) &:= \int_0^\infty f(x)x^{s-1}\,\mathrm{d}x = \sum_{j=1}^{\infty}\int_0^\infty \mathrm{e}^{-xpq^j}\left(1-\mathrm{e}^{-xpq^{j-1}}\right)x^{s-1}\,\mathrm{d}x\\ &= \sum_{j=1}^{\infty}\frac{1}{(pq^j)^s}\int_0^\infty \mathrm{e}^{-x}(1-\mathrm{e}^{-x/q})x^{s-1}\,\mathrm{d}x\\ &= \frac{1}{p^s}\frac{1}{q^s-1}\Gamma(s)\left(1-\frac{1}{(1+1/q)^s}\right) = \frac{\Gamma(s)}{p^s(1+q)^s}\frac{(q+1)^s-q^s}{q^s-1}.\end{aligned}$$

By inversion

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

where the vertical line passing through c is in the fundamental strip on which the Mellin transform exists. In our case, c may be any number in $(-1,0)$. The latter integral may be evaluated by residue theorem; the integrand

$$\frac{\Gamma(s)}{p^s(1+q)^s}\frac{(q+1)^s-q^s}{q^s-1}x^{-s}$$

has simple poles at $\chi_k = 2k\pi\mathrm{i}/\ln q$, $k = 0, \pm 1, \pm 2, \ldots$. The residue at χ_0 is

$$\frac{\ln(1+q)}{\ln q} - 1 \tag{7}$$

and the residue at χ_k, $k \neq 0$ is

$$\frac{\Gamma(\chi_k)((1+q)^{\chi_k} - 1)}{(1+q)^{\chi_k}(px)^{\chi_k}\ \ln q}.$$

Thus, by considering a large rectangular box with vertices $c \pm \mathrm{i}M$ and $R \pm \mathrm{i}M$, $R, M > 0$, $M \notin \mathbf{N}$ (and omitting the standard verification that the integrals along the horizontal line segments and the vertical segment on the right vanish as $M \to \infty$ and then $R \to \infty$) we obtain that

$$f(n) \sim 1 - \frac{\ln(1+q)}{\ln q} + \frac{2}{\ln q}\sum_{k=1}^{\infty}\Re\left(\frac{1-(1+q)^{\chi_k}}{p^{\chi_k}(1+q)^{\chi_k}}\Gamma(\chi_k)\mathrm{e}^{\chi_k\ \ln n}\right).$$

Combining this with (5) we obtain that

$$\mathsf{E}Y_n = -\frac{\ln(1+q)}{\ln q} + \frac{2}{\ln q}\sum_{k=1}^{\infty}\Re\left(\frac{1-(1+q)^{2k\pi\mathrm{i}/\ln q}}{(p(1+q))^{2k\pi\mathrm{i}/\ln q}}\Gamma\left(\frac{2k\pi\mathrm{i}}{\ln q}\right)\mathrm{e}^{(2k\pi\mathrm{i}/\ln q)\ln n}\right),$$

which proves (3).

The argument for the variance is similar:

$$\begin{aligned}\mathsf{var}(Y_n) = \mathsf{var}(Y_n+1) &= \mathsf{E}\left(\sum_{k\geqslant 1} G_k\right)^2 - \left(\mathsf{E}\sum_{k\geqslant 1} G_k\right)^2 \\ &= \mathsf{E}\sum_{j,k\geqslant 1} G_jG_k - \left(\sum_{k\geqslant 1}\mathsf{Pr}(G_k)\right)^2 \\ &= \sum_{j,k\geqslant 1}\mathsf{Pr}(G_j\cap G_k) - \left(\sum_{k\geqslant 1}\mathsf{Pr}(G_k)\right)^2\end{aligned} \tag{8}$$

and we need to find $\mathsf{Pr}(G_j \cap G_k)$. As we will now show

$$\mathsf{Pr}(G_j\cap G_k) = \begin{cases}\mathsf{Pr}(G_j) & \text{if } j=k, \\ 0 & \text{if } |j-k|=1, \\ (1-pq^j-pq^k)^n - (1-pq^{j-1}-pq^j-pq^k)^n & \text{if } |j-k|\geqslant 2. \\ \quad -(1-pq^j-pq^{k-1})^n + (1-pq^{j-1}-pq^j-pq^{k-1}-pq^k)^n & \end{cases}$$

Since the first two assertions are clear, consider $|j-k|\geqslant 2$. Then

$$\begin{aligned}
\mathsf{Pr}(G_j\cap G_k) &= \mathsf{Pr}\left(\bigcup_{\ell=1}^{n}\{\Gamma_\ell=j\}\cap\bigcup_{m=1}^{n}\{\Gamma_m=k\}\cap\bigcap_{r=1}^{n}\{\Gamma_r\neq j+1,k+1\}\right)\\
&= \mathsf{Pr}\left(\bigcup_{\ell=1}^{n}\{\Gamma_\ell=j\}\cap\bigcap_{r=1}^{n}\{\Gamma_r\neq j+1,k+1\}\right)\\
&\quad - \mathsf{Pr}\left(\bigcup_{\ell=1}^{n}\{\Gamma_\ell=j\}\cap\bigcap_{r=1}^{n}\{\Gamma_r\neq j+1,k,k+1\}\right)\\
&= \mathsf{Pr}\left(\bigcap_{r=1}^{n}\{\Gamma_r\neq j+1,k+1\}\right)-\mathsf{Pr}\left(\bigcap_{r=1}^{n}\{\Gamma_r\neq j,j+1,k+1\}\right)\\
&\quad - \mathsf{Pr}\left(\bigcap_{r=1}^{n}\{\Gamma_r\neq j+1,k,k+1\}\right)+\mathsf{Pr}\left(\bigcap_{r=1}^{n}\{\Gamma_r\neq j,j+1,k,k+1\}\right)\\
&= \left(1-pq^j-pq^k\right)^n-\left(1-pq^{j-1}-pq^j-pq^k\right)^n\\
&\quad -\left(1-pq^j-pq^{k-1}-pq^k\right)^n+\left(1-pq^{j-1}-pq^j-pq^{k-1}-pq^k\right)^n.
\end{aligned}$$

This, combined with (6) and Lemma 3(ii) gives, for $|j-k|\geqslant 2$

$$\begin{aligned}
|\mathsf{Pr}(G_j\cap G_k)-\mathsf{Pr}(G_j)\mathsf{Pr}(G_k)| &\leqslant |(1-pq^j)^n(1-pq^k)^n-(1-pq^j-pq^k)^n|\\
&\quad +|(1-pq^j)^n(1-pq^k-pq^{k-1})^n-(1-pq^j-pq^k-pq^{k-1})^n|\\
&\quad +|(1-pq^k)^n(1-pq^j-pq^{j-1})^n-(1-pq^k-pq^j-pq^{j-1})^n|\\
&\quad +|(1-pq^j-pq^{j-1})^n(1-pq^k-pq^{k-1})^n\\
&\quad -(1-pq^j-pq^{j-1}-pq^k-pq^{k-1})^n|\\
&= O\left(\frac{1}{n}\right)(npq^j\mathrm{e}^{-npq^j})(npq^k\mathrm{e}^{-npq^k}),
\end{aligned}$$

which implies that

$$\sum_{|j-k|\geqslant 2}\mathsf{Pr}(G_j\cap G_k)\sim\sum_{|j-k|\geqslant 2}\mathsf{Pr}(G_j)\mathsf{Pr}(G_k).$$

Since

$$\begin{aligned}
\sum_{|j-k|\geqslant 2}\mathsf{Pr}(G_j)\mathsf{Pr}(G_k) &= \sum_{j,k\geqslant 1}\mathsf{Pr}(G_j)\mathsf{Pr}(G_k)-\sum_{j\geqslant 1}(\mathsf{Pr}(G_j))^2-2\mathsf{Pr}(G_j)\mathsf{Pr}(G_{j+1})\\
&= \left(\sum_{j\geqslant 1}(\mathsf{Pr}(G_j)\right)^2-\sum_{j\geqslant 1}(\mathsf{Pr}(G_j))^2-2\mathsf{Pr}(G_j)\mathsf{Pr}(G_{j+1}),
\end{aligned}$$

(8) implies that

$$\mathsf{var}(Y_n)\sim\sum_{j\geqslant 1}\mathsf{Pr}(G_j)-\sum_{j\geqslant 1}(\mathsf{Pr}(G_j))^2-2\sum_{j\geqslant 1}\mathsf{Pr}(G_j)\mathsf{Pr}(G_{j+1}). \tag{9}$$

The first sum was already evaluated and it is the matter of the same routine to analyze the remaining two. We first appeal to Lemma 3(i) to replace the exact expressions for $\mathsf{Pr}(G_j)$ by approximations by exponential functions. This gives,

$$(\mathsf{Pr}(G_j))^2 \sim (\mathrm{e}^{-npq^j}(1-\mathrm{e}^{-npq^{j-1}}))^2 = \mathrm{e}^{-2npq^j} + \mathrm{e}^{-2np(1+q)q^{j-1}} - 2\mathrm{e}^{np(1+2q)q^{j-1}},$$

and

$$\begin{aligned}\mathsf{Pr}(G_j)\mathsf{Pr}(G_{j+1}) &\sim \mathrm{e}^{-npq^j}\mathrm{e}^{-npq^{j+1}}(1-\mathrm{e}^{-npq^{j-1}})(1-\mathrm{e}^{-npq^j})\\ &= \mathrm{e}^{-np(1+q)q^j} - \mathrm{e}^{-np(1+q(1+q))q^{j-1}} - \mathrm{e}^{-np(2+q)q^j} + \mathrm{e}^{-np(1+q)^2q^{j-1}}.\end{aligned}$$

Mellin transforms of the sums of these asymptotic expressions for $\sum(\mathsf{Pr}(G_j))^2$ and $\sum\mathsf{Pr}(G_j)\mathsf{Pr}(G_{j+1})$ are, respectively,

$$\Gamma(s)\frac{q^s}{p^s}\left(\frac{1}{2^sq^s} + \frac{1}{2^s(1+q)^s} - \frac{2}{(1+2q)^s}\right)\frac{1}{q^s-1}$$

and

$$\Gamma(s)\frac{q^s}{p^s}\left(\frac{1}{q^s(1+q)^s} - \frac{1}{(1+q(1+q))^s} - \frac{1}{q^s(2+q)^s} + \frac{1}{(1+q)^{2s}}\right)\frac{1}{q^s-1}.$$

It follows that the integrand in the Mellin inversion has simple poles at exactly the same χ_k's as before and the rest of the argument can be completed by the same residue calculations as for the expected value. In particular, the residua at $\chi_0 = 0$ of the above expressions are, respectively,

$$\frac{-\ln 2q - \ln 2(1+q) + 2\ln(1+2q)}{\ln q}$$

and

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

Combining the above two expressions with (9) and (7) we obtain that the main non-oscillatory term in the variance is

$$\frac{2}{\ln q}\ln\left(\frac{(1+2q)(1+q(1+q))(2+q)}{2(1+q)^4}\right),$$

as specified in (4). Residua at other poles contribute to oscillatory terms and the explicit expressions for them could be obtained without difficulty, but we skip further details. □

Remarks. (i) For $q = \frac{1}{2}$ formula (3) becomes

$$\mathsf{E}Y_n \sim \frac{\ln 3}{\ln 2} - 1 + \frac{2}{\ln 2}\sum_{k=1}^{\infty}\Re\left(\frac{\mathrm{e}^{2k\pi\mathrm{i}\ln 3/\ln 2}-1}{\mathrm{e}^{2k\pi\mathrm{i}\ln 3/\ln 2}}\Gamma\left(\frac{2k\pi\mathrm{i}}{\ln 2}\right)\exp\left(\frac{2k\pi\mathrm{i}}{\ln 2}\ln n\right)\right).$$

In particular, the oscillatory component does not vanish. Its plot is given in Fig. 1. This contrasts with the situation considered by Louchard and Prodinger in [5] where there were oscillations for $q \neq \frac{1}{2}$ but not for $q = \frac{1}{2}$ (see (1)). The same applies to the variance.

(ii) The simple minded approach presented above can be used to recover some of the results obtained by Louchard and Prodinger. For example, the random variable X_n referred to in Theorem 1 satisfies

$$X_n = \sum_{j\geqslant 1} V_j,$$

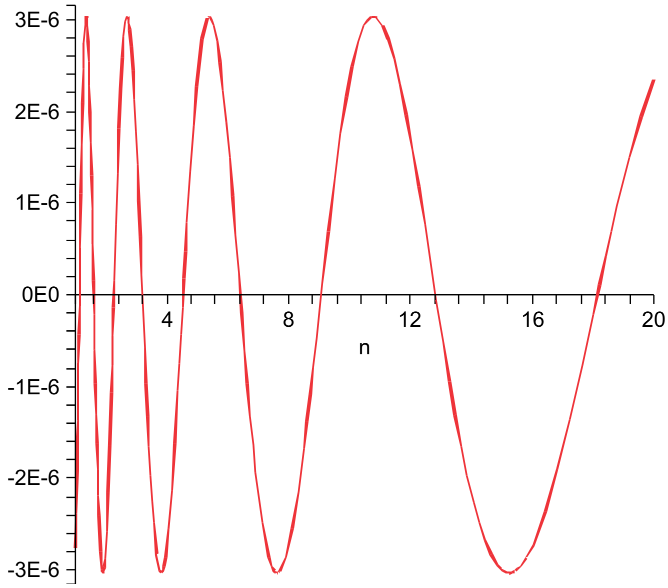

Fig. 1. Oscillatory contribution to $\mathsf{E}Y_n$ for $p = 1/2$.

where

$$V_j = \bigcap_{k=1}^{n} \{\Gamma_k \neq j\} \cap \bigcup_{\ell=1}^{n} \{\Gamma_\ell > j\}.$$

Since,

$$\bigcap_{k=1}^{n} \{\Gamma_k \neq j\} = V_j \cup \bigcap_{k=1}^{n} (\{\Gamma_k \neq j\} \cap \{\Gamma_k \leqslant j\}) = V_j \cup \bigcap_{k=1}^{n} \{\Gamma_k \leqslant j-1\},$$

we have

$$V_j = \bigcap_{k=1}^{n} \{\Gamma_k \neq j\} \backslash \bigcap_{k=1}^{n} \{\Gamma_k \leqslant j-1\}.$$

Hence,

$$\mathsf{E}X_n = \sum_{j=1}^{\infty} \left\{ \left(1 - pq^{j-1}\right)^n - \left(1 - q^{j-1}\right)^n \right\}, \tag{10}$$

and exactly the same argument as earlier could be used to obtain the asymptotic expression for the expectation in (10). In particular, the non-oscillatory term in the asymptotic expression for $\mathsf{E}X_n$ is $\ln\, p/\ln\, q$, and the explicit expression for the oscillations can be obtained by taking into account contributions of the residua at χ_k for $k \neq 0$. We omit further details, but we would like to mention that Louchard and Prodinger used different approach which required an extra arguments to show that the oscillations vanish for $p = q = \frac{1}{2}$. With our approach this is clear, since for $p = q$ (10) becomes

$$\mathsf{E}X_n = \sum_{j=1}^{\infty} \left\{ \left(1 - \frac{1}{2^j}\right)^n - \left(1 - \frac{1}{2^{j-1}}\right)^n \right\} = 1,$$

which gives the first part of (1).

The computation for the second part is not much more involved. Because of the last expression, the variance of X_n is equal to

$$2\sum_{j<k}\Pr(V_j\cap V_k),$$

and since for $j<k$

$$V_j\cap V_k=\bigcap_{\ell=1}^{n}\{\Gamma_\ell\neq j,k\}\setminus\bigcap_{\ell=1}^{n}\{\Gamma_\ell\neq j,\Gamma_\ell\leqslant k-1\},$$

this is further equal to

$$\begin{aligned}
&2\sum_{j<k}\left\{\left(1-\frac{1}{2^j}-\frac{1}{2^k}\right)^n-\left(1-\frac{1}{2^j}-\frac{1}{2^{k-1}}\right)^n\right\}\\
&\quad=2\sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty}\left\{\left(1-\frac{1}{2^j}-\frac{1}{2^k}\right)^n-\left(1-\frac{1}{2^j}-\frac{1}{2^{k-1}}\right)^n\right\}\\
&\quad=2\sum_{j=1}^{\infty}\lim_{M\to\infty}\left\{\left(1-\frac{1}{2^j}-\frac{1}{2^M}\right)^n-\left(1-\frac{1}{2^j}-\frac{1}{2^j}\right)^n\right\}\\
&\quad=2\sum_{j=1}^{\infty}\left\{\left(1-\frac{1}{2^j}\right)^n-\left(1-\frac{1}{2^{j-1}}\right)^n\right\}=2,
\end{aligned}$$

as specified in the second part of (1). It should be added, however, that this method does not seem to be useful as far as computation of higher moments of X_n or Y_n are concerned as one would have to deal with increasingly more complicated sums of the products of indicator functions.

3. An analog of (2)

In this section we focus on a particular case $p=\frac{1}{2}$. Let $R_{n,r}$ be the event that the sample of n i.i.d. GEOM($\frac{1}{2}$) random variables has r gaps, and let $p_n(r)$ be its probability. In other words, $p_n(r)=\Pr(Y_n=r)$ where Y_n is defined in Proposition 2. Then

Theorem 4. *For $r=0$ we have*

$$p_n(0)=\frac{2^{n-1}}{2^n-1},\tag{11}$$

and for $r\geqslant 1$

$$p_n(r)=p_*(r)-\eta_P(\ln n)+O(n^{\beta-1}),\tag{12}$$

where

$$p_*(r)=-\frac{1}{\ln 2}\sum_{m\geqslant 1}\frac{p_m(r)-p_m(r-1)}{m4^m}(2^m-1),\tag{13}$$

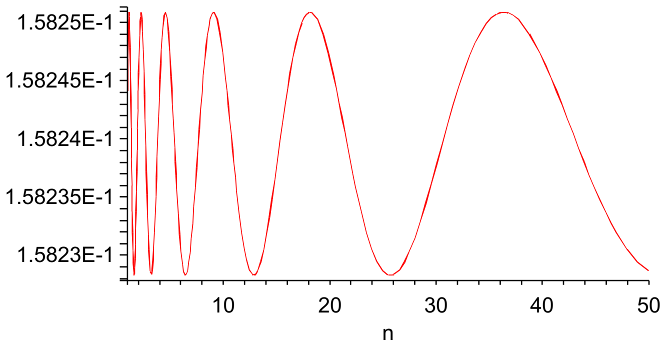

Fig. 2. The values of $p_n(1)$ for $1 \leqslant n \leqslant 50$.

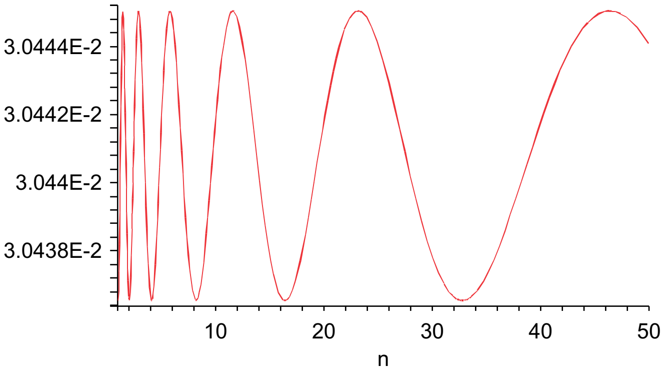

Fig. 3. The values of $p_n(2)$ for $1 \leqslant n \leqslant 50$.

and

$$\eta_P(x) = \frac{2}{\ln 2} \Re \left(\sum_{k \geqslant 1} \sum_{m \geqslant 1} \frac{p_m(r) - p_m(r-1)}{m! 4^m} (2^m - 1) \Gamma\left(m + \frac{2k\pi \mathrm{i}}{\ln 2}\right) \mathrm{e}^{-(2k\pi \mathrm{i}/\ln 2)x} \right),$$

and $0 < \beta < 1$.

Remark. While statement (12) seems circular, in sense that it represents $p_n(r)$ in terms of an infinite series involving that same quantity, the point to make is that the series converges exponentially fast, and thus is much better amenable to various numerical evaluations (including bounds on the approximation error) than the recurrence (15) (see e.g. [4,2] for similar instances). The function η_P has the usual features; in particular the magnitude of its oscillations is a fraction of 10^{-5} of the first term in (12). Plots of the asymptotic expressions for $p_n(r)$ for $1 \leqslant r \leqslant 5$ are given in Figs. (2)–(6).

Proof of Theorem 4. The proof is based on what has become a well established, 4-step program (see [3] for a good exposition): we will first derive a recurrence relation for the $p_n(r)$, then convert it into a functional equation by poissonization. We will then analyze the asymptotic behavior of the solution of this functional equation (by means of the Mellin transform and residue calculation). Finally, we come back to the asymptotics of $p_n(r)$ by analytical depoissonization.

To begin, for a set A, let $p_n(r, A)$ and $p_n(r|A)$ denote $\Pr(R_{n,r} \cap A)$ and $\Pr(R_{n,r}|A)$, respectively. We let

$$m_n := \min\{\Gamma_k : 1 \leqslant k \leqslant n\},$$

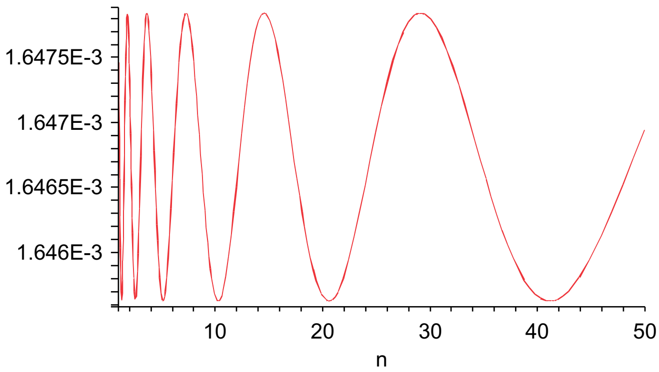

Fig. 4. The values of $p_n(3)$ for $1 \leqslant n \leqslant 50$.

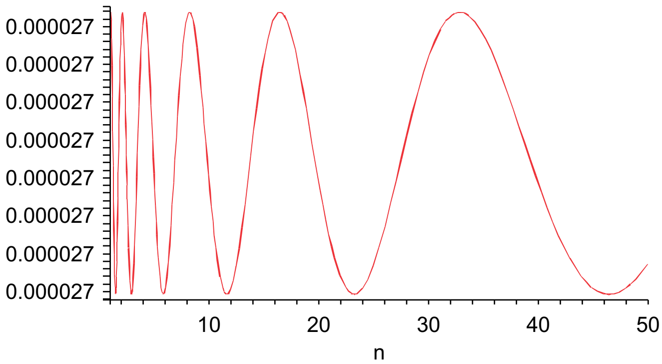

Fig. 5. The values of $p_n(4)$ for $1 \leqslant n \leqslant 50$.

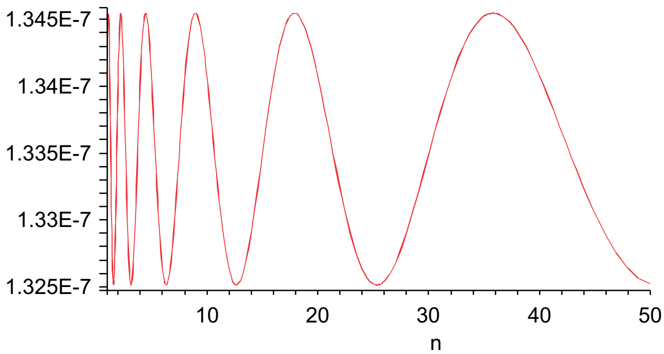

Fig. 6. The values of $p_n(5)$ for $1 \leqslant n \leqslant 50$.

be the minimum of a sample of length n. Then we have

$$\begin{aligned} p_n(r) &= p_n(r, m_n = 1) + p_n(r, m_n > 1) \\ &= p_n(r, m_n = 1) + p_n(r | m_n > 1)\Pr(m_n > 1) \\ &= p_n(r, m_n = 1) + p_n(r)/2^n, \end{aligned}$$

where in the last step we used memoryless property of geometric random variables. In particular,

$$p_n(r, m_n = 1) = \left(1 - \frac{1}{2^n}\right) p_n(r), \quad r \geqslant 0, \tag{14}$$

and since it was shown in [2] that $p_n(0, m_n = 1) = \frac{1}{2}$, (11) follows. For $r \geqslant 1$ by conditioning on the number of Γ_k's that are equal to 1 we obtain

$$\begin{aligned}
p_n(r, m_n = 1) &= \sum_{k=1}^{n} \binom{n}{k} \frac{1}{2^n} p_n(r | \sum_{\ell} I_{\Gamma_\ell = 1} = k) \\
&= \sum_{k=1}^{n-1} \binom{n}{k} \frac{1}{2^n} p_n(r | \sum_{\ell} I_{\Gamma_\ell = 1} = k) + \frac{1}{2^n} p_n(r | \sum_{\ell} I_{\Gamma_\ell = 1} = n) \\
&= \sum_{k=1}^{n-1} \binom{n}{k} \frac{1}{2^n} \{p_{n-k}(r, m_{n-k} = 1) + p_{n-k}(r, m_{n-k} \geqslant 2)\} + \frac{1}{2^n} I_{r=0} \\
&= \sum_{k=1}^{n-1} \binom{n}{k} \frac{1}{2^n} \left\{p_{n-k}(r, m_{n-k} = 1) + \frac{1}{2^{n-k}} p_{n-k}(r-1)\right\} + \frac{1}{2^n} I_{r=0} \\
&= \sum_{k=r+1}^{n-1} \binom{n}{k} \frac{1}{2^n} p_k(r, m_k = 1) + \sum_{k=r}^{n-1} \binom{n}{k} \frac{1}{2^{n+k}} p_k(r-1) + \frac{1}{2^n} I_{r=0}.
\end{aligned}$$

Utilizing $p_n(0, m_n = 1) = \frac{1}{2}$ and (14) the above can be written as: for $r \geqslant 1$ and $n \geqslant r + 1$,

$$\begin{aligned}
\left(1 - \frac{1}{2^n}\right) p_n(r) &= \frac{1}{2^n} \sum_{k=r+1}^{n-1} \binom{n}{k} \left(1 - \frac{1}{2^k}\right) p_k(r) + \frac{1}{2^n} \sum_{k=r}^{n-1} \binom{n}{k} \frac{p_k(r-1)}{2^k}, \\
p_n(0) &= \frac{2^{n-1}}{2^n - 1}.
\end{aligned} \tag{15}$$

We now form a two-variable functional equation by first letting for $n \geqslant 1$

$$B_n(u) := \sum_{r \geqslant 0} p_n(r) u^r = \sum_{r=0}^{n-1} p_n(r) u^r,$$

and then

$$\mathscr{P}(z, u) := \sum_{n \geqslant 1} B_n(u) \mathrm{e}^{-z} \frac{z^n}{n!}.$$

We then have

$$\sum_{n \geqslant 1} \sum_{r \geqslant 0} u^r \mathrm{e}^{-z} \frac{z^n}{n!} \left(1 - \frac{1}{2^n}\right) p_n(r) = \mathscr{P}(z, u) - \mathrm{e}^{-z/2} \mathscr{P}(z/2, u),$$

and, on the other hand,

$$
\begin{aligned}
&\sum_{n\geqslant 1}\sum_{r\geqslant 0} u^r \mathrm{e}^{-z}\frac{z^n}{n!}\left(1-\frac{1}{2^n}\right)p_n(r)\\
&\quad=\sum_{n\geqslant 2}\sum_{r\geqslant 1} u^r \mathrm{e}^{-z}\frac{z^n}{n!}\left(1-\frac{1}{2^n}\right)p_n(r)+\sum_{n\geqslant 1}\mathrm{e}^{-z}\frac{z^n}{n!}\left(1-\frac{1}{2^n}\right)p_n(0)\\
&\quad=\sum_{n\geqslant 2}\sum_{r\geqslant 1} u^r \mathrm{e}^{-z}\frac{z^n}{n!}\left\{\frac{1}{2^n}\sum_{k=r+1}^{n-1}\binom{n}{k}\left(1-\frac{1}{2^k}\right)p_k(r)+\frac{1}{2^n}\sum_{k=r}^{n-1}\binom{n}{k}\frac{p_k(r-1)}{2^k}\right\}\\
&\qquad+\sum_{n\geqslant 1}\mathrm{e}^{-z}\frac{z^n}{n!}\left(1-\frac{1}{2^n}\right)\frac{2^{n-1}}{2^n-1}\\
&\quad=\sum_{n\geqslant 2}\sum_{r\geqslant 1} u^r \mathrm{e}^{-z}\left\{\sum_{k=r+1}^{n-1}\frac{(z/2)^k}{k!}\left(1-\frac{1}{2^k}\right)p_k(r)\frac{(z/2)^{n-k}}{(n-k)!}\right.\\
&\qquad\left.+\sum_{k=r}^{n-1}\frac{(z/4)^k}{k!}p_k(r-1)\frac{(z/2)^{n-k}}{(n-k)!}\right\}+\frac{1}{2}\mathrm{e}^{-z}(\mathrm{e}^{z}-1)\\
&\quad=\sum_{r\geqslant 1}\left\{\mathrm{e}^{-z/2}\sum_{k\geqslant r+1} u^r \mathrm{e}^{-z/2}\frac{(z/2)^k}{k!}\left(1-\frac{1}{2^k}\right)p_k(r)\sum_{n=k+1}^{\infty}\frac{(z/2)^{n-k}}{(n-k)!}\right.\\
&\qquad\left.+\mathrm{e}^{-3z/4}\sum_{k\geqslant r} u^r \mathrm{e}^{-z/4}\frac{(z/4)^k}{k!}p_k(r-1)\sum_{n=k+1}^{\infty}\frac{(z/2)^{n-k}}{(n-k)!}\right\}+\frac{1-\mathrm{e}^{-z}}{2}\\
&\quad=(1-\mathrm{e}^{-z/2})\sum_{k\geqslant 2}\sum_{r\geqslant 1} u^r \mathrm{e}^{-z/2}\frac{(z/2)^k}{k!}\left(1-\frac{1}{2^k}\right)p_k(r)\\
&\qquad+u(\mathrm{e}^{z/2}-1)\mathrm{e}^{-3z/4}\sum_{k\geqslant 1}\sum_{r\geqslant 0} u^r \mathrm{e}^{-z/4}\frac{(z/4)^k}{k!}p_k(r)+\frac{1-\mathrm{e}^{-z}}{2}\\
&\quad=(1-\mathrm{e}^{-z/2})\left\{\sum_{k\geqslant 1}\sum_{r\geqslant 0} u^r \mathrm{e}^{-z/2}\frac{(z/2)^k}{k!}\left(1-\frac{1}{2^k}\right)p_k(r)\right.\\
&\qquad\left.-\sum_{k\geqslant 1}\mathrm{e}^{-z/2}\frac{(z/2)^k}{k!}\left(1-\frac{1}{2^k}\right)p_k(0)\right\}\\
&\qquad+u(\mathrm{e}^{z/2}-1)\mathrm{e}^{-3z/4}\mathscr{P}(z/4,u)+\frac{1-\mathrm{e}^{-z}}{2}\\
&\quad=(1-\mathrm{e}^{-z/2})\left\{\mathscr{P}(z/2,u)-\mathrm{e}^{-z/4}\mathscr{P}(z/4,u)-\frac{1-\mathrm{e}^{-z/2}}{2}\right\}\\
&\qquad+u\mathrm{e}^{-z/4}(1-\mathrm{e}^{-z/2})\mathscr{P}(z/4,u)+\frac{1-\mathrm{e}^{-z}}{2}.
\end{aligned}
$$

This leads to the following functional equation:

$$\mathscr{P}(z,u)=\mathscr{P}(z/2,u)-T(z,u)+f(z), \tag{16}$$

where we have set

$$T(z,u):=(1-u)(1-\mathrm{e}^{-z/2})\mathrm{e}^{-z/4}\mathscr{P}(z/4,u) \quad \text{and} \quad f(z):=\mathrm{e}^{-z/2}(1-\mathrm{e}^{-z/2}).$$

For the remainder of our discussion u is a complex variable that is restricted to a compact domain, say $|u|\leqslant\frac{1}{2}$. The following fact is self-evident, we omit the proof.

Lemma 5. *We have*

(a) $|B_n(u)|\leqslant|1/(1-u)|$.
(b) $\mathscr{P}(z,u)$ *is an entire function of* z.
(c) *For* $z\geqslant 0$, $|\mathscr{P}(z,u)|\leqslant|1/(1-u)|$ *is a bounded function of* z *and* $\mathscr{P}(0,u)=0$.

It follows in particular, that the Mellin transform $\mathscr{M}(\mathscr{P})(s,u)$ of $\mathscr{P}(z,u)$ is well-defined and is analytic in the vertical strip $-1<\Re(s)<0$. Applying the Mellin transform to (16) we get

$$\mathscr{M}(\mathscr{P})(s,u)=2^s\mathscr{M}(\mathscr{P})(s,u)-\mathscr{M}(T)(s,u)+\mathscr{M}(f)(s,u). \tag{17}$$

For $\Re(s)>-1$

$$\mathscr{M}(f)(s,u)=\int_0^\infty \mathrm{e}^{-z/2}(1-\mathrm{e}^{-z/2})z^{s-1}\,\mathrm{d}z=(2^s-1)\Gamma(s),$$

so we turn our attention to $\mathscr{M}(T)(s,u)$. By Lemma 5 $\mathscr{P}(z/4,u)$ is bounded for $\geqslant 0$, hence its Mellin transform

$$\mathscr{M}(T)(s,u)=(1-u)\int_0^\infty(1-\mathrm{e}^{-z/2})\mathrm{e}^{-z/4}\mathscr{P}(z/4,u)z^{s-1}\,\mathrm{d}z$$

is analytic in the half plane $\Re(s)>-1$. We first derive an expression valid for $\Re(s)>0$ and then extend it by analytic continuation. Write

$$\mathscr{M}(T)(s,u)=(1-u)(I_1-I_2), \tag{18}$$

where

$$I_1=\int_0^\infty \mathrm{e}^{-z/4}\mathscr{P}(z/4,u)z^{s-1}\,\mathrm{d}z \quad \text{and} \quad I_2=\int_0^\infty \mathrm{e}^{-3z/4}\mathscr{P}(z/4,u)z^{s-1}\,\mathrm{d}z.$$

Now for $\Re(s)>0$

$$\begin{aligned} I_1&=\int_0^\infty \mathrm{e}^{-z/4}\mathscr{P}(z/4,u)z^{s-1}\,\mathrm{d}z=\int_0^\infty \mathrm{e}^{-z/4}\left(\sum_{n\geqslant 1}B_n(u)\frac{(z/4)^n}{n!}\right)\mathrm{e}^{-z/4}z^{s-1}\,\mathrm{d}z\\ &=\sum_{n\geqslant 1}\frac{B_n(u)}{n!4^n}\int_0^\infty \mathrm{e}^{-z/2}z^{n+s-1}\,\mathrm{d}z=\sum_{n\geqslant 1}\frac{B_n(u)2^{n+s}}{n!4^n}\Gamma(n+s)\end{aligned}$$

(interchange of integration and summation is justified since the expression in question is absolutely convergent). Likewise, we have

$$I_2=\sum_{n\geqslant 1}\frac{B_n(u)}{n!4^n}\Gamma(n+s).$$

Inserting these two equations to (18) we obtain

$$\mathscr{M}(T)(s,u) = (1-u)\left(\sum_{n\geqslant 1} \frac{B_n(u)}{n!4^n}\Gamma(n+s)(2^{n+s}-1)\right), \tag{19}$$

which is still valid for $\Re(s) > -1$ by analytic continuation. It follows from (17) that

$$\mathscr{M}(\mathscr{P})(s,u) = \frac{\mathscr{M}(T)(s,u)}{2^s-1} - \Gamma(s)$$

and by the Mellin inversion for $-1 < \sigma < 0$

$$\mathscr{P}(z,u) = \frac{1}{2\pi \mathrm{i}}\int_{\sigma-\mathrm{i}\infty}^{\sigma+\mathrm{i}\infty}\left(\frac{\mathscr{M}(T)(s,u)}{2^s-1} - \Gamma(s)\right) z^{-s}\,\mathrm{d}s. \tag{20}$$

We now shift the integration path to the right side of zero. Since the function $\mathscr{M}(T)(s,u)$ is implicit, a rigorous justification of that step is required. The following lemma whose proof is postponed until the appendix will suffice for that purpose:

Lemma 6. *Let $s = \sigma + \mathrm{i}t$, where σ, t are real numbers. Assume that σ is in a bounded set A contained in $\sigma > -1$. Then as $|t| \to \infty$*

$$|\mathscr{M}(T)(\sigma + \mathrm{i}ft, u)| = O_A(\mathrm{e}^{-0.34|t|}),$$

where the big 'O' constant holds uniformly for $\sigma \in A$ and $|u| \leqslant \frac{1}{2}$.

We now consider the integral around a rectangular contour with vertices, say, $-1/2 \pm \mathrm{i}t$, $1 \pm \mathrm{i}t$, where t eventually approaches ∞. By Lemma 6 and the assumption $z > 0$, the integrals along the vertical sides converge absolutely and the integrals along the horizontal sides go to zero as $t \to \infty$. We get from (20)

$$\begin{aligned}\mathscr{P}(z,u) = &-\sum_k \operatorname{Res}\left(\frac{\mathscr{M}(T)(s,u)}{2^s-1} - \Gamma(s)z^{-s}\right)_{s=\chi_k} \\ &+ \frac{1}{2\pi i}\int_{1-\mathrm{i}\infty}^{1+\mathrm{i}\infty}\left(\frac{\mathscr{M}(T)(s,u)}{2^s-1} - \Gamma(s)\right) z^{-s}\,\mathrm{d}s.\end{aligned}$$

The residue computation is straightforward. We get

$$\begin{aligned}\mathscr{P}(z,u) = 1 &- \frac{1}{\ln 2}\mathscr{M}(T)(0,u) - \frac{1}{\ln 2}\sum_{k\neq 0}\mathscr{M}(T)\left(\frac{2k\pi \mathrm{i}}{\ln 2}, u\right) z^{-(2k\pi\mathrm{i}/\ln 2)} \\ &+ \frac{1}{2\pi \mathrm{i}}\int_{1-\mathrm{i}\infty}^{1+\mathrm{i}\infty}\left(\frac{\mathscr{M}(T)(s,u)}{2^s-1} - \Gamma(s)\right) z^{-s}\,\mathrm{d}s \\ = 1 &- \frac{1}{\ln 2}\sum_{m\geqslant 1}\frac{(1-u)B_m(u)(2^m-1)}{m4^m} \\ &- \frac{1}{\ln 2}\sum_{k\neq 0}\sum_{m\geqslant 1}\frac{(1-u)B_m(u)(2^m-1)}{m!4^m}\Gamma\left(m + \frac{2k\pi \mathrm{i}}{\ln 2}\right) z^{-(2k\pi i/\ln 2)} \\ &+ \frac{1}{2\pi i}\int_{1-\mathrm{i}\infty}^{1+\mathrm{i}\infty}\left(\frac{\mathscr{M}(T)(s,u)}{2^s-1} - \Gamma(s)\right) z^{-s}\,\mathrm{d}s.\end{aligned}$$

We remark that because of Lemma 6 the double sum above and the integral along $\sigma = 1$ are absolutely convergent.

We will verify in Appendix the applicability of the depoissonization process. This will imply in particular that it will suffice to focus on $z \to \infty$ through the positive reals. But then the integral above is o(1) so that as $z \to \infty$ through reals,

$$\mathscr{P}(z,u) = 1 - \frac{1}{\ln 2}\sum_{m\geqslant 1}\frac{(1-u)B_m(u)(2^m-1)}{m4^m} + (1-u)\eta_P(\ln z) + o(z^{-1}), \tag{21}$$

where the little 'o' constant is uniform for $|u| \leqslant \frac{1}{2}$ and where

$$\eta_P(x) := \frac{1}{\ln 2}\sum_{k\neq 0}\sum_{m\geqslant 1}\frac{B_m(u)(2^m-1)}{m!4^m}\Gamma\left(m + \frac{2k\pi \mathrm{i}}{\ln 2}\right)\mathrm{e}^{-(2k\pi\mathrm{i}/\ln 2)x}.$$

After depoissonization we will get for any β, $0 < \beta < 1$,

$$B_n(u) = P(n,u) + O(n^{\beta-1}),$$

where the big 'O' constant holds uniformly for $|u| \leqslant \frac{1}{2}$. Making explicit what we have from Eq. (21)

$$\begin{aligned} B_n(u) &= 1 - \frac{1}{\ln 2}\sum_{m\geqslant 1}\frac{(1-u)B_m(u)(2^m-1)}{m4^m} + (1-u)\eta_P(\ln n) + \mathrm{o}(n^{-1}) + O(n^{\beta-1}) \\ &= 1 - \frac{1}{\ln 2}\sum_{m\geqslant 1}\frac{(1-u)B_m(u)(2^m-1)}{m4^m} + (1-u)\eta_P(\ln n) + O(n^{\beta-1}), \end{aligned}$$

where both the big 'O' and the little 'o' constants hold uniformly for $|u| \leqslant \frac{1}{2}$. Because of uniformity we can multiply the above equation by $\frac{1}{u^{r+1}}$, $r \geqslant 1$, and integrate the resulting equation with respect to u along the circle $|u| = \frac{1}{2}$. This proves the result. $\square$

Appendix

Proof of Lemma 6. Recall (19) and use the estimate of $B_n(u)$ in Lemma 5(a) to get

$$\begin{aligned} |\mathscr{M}(T)(\sigma + \mathrm{i}t, u)| &\leqslant \left|(1-u)\left(\sum_{n\geqslant 1}\frac{B_n(u)}{n!4^n}\Gamma(n+s)(2^{n+s}-1)\right)\right| \\ &\leqslant \sum_{n\geqslant 1}\frac{1}{n!4^n}\,|\Gamma(n+\sigma+\mathrm{i}t)|\,(2^{n+\sigma}+1). \end{aligned}$$

Since the situation is symmetric with respect to t it suffices to consider $t \to \infty$. Note that $n \geqslant 1$ and $\sigma > -1$ so that $n + \sigma > 0$. This implies that $2^{n+\sigma} > 1$ and thus the last expression is bounded by

$$2\sum_{n\geqslant 1}\frac{1}{n!4^n}|\Gamma(n+\sigma+\mathrm{i}t)|2^{n+\sigma} = 2^{\sigma+1}\sum_{n\geqslant 1}\frac{|\Gamma(n+\sigma+\mathrm{i}t)|}{\Gamma(n+\sigma)}\cdot\frac{\Gamma(n+\sigma)}{n!2^n}. \tag{22}$$

Now we focus on $|\Gamma(n+\sigma+\mathrm{i}t)|/\Gamma(n+\sigma)$. By Stirling's formula we have

$$\exp\left\{\left(n+\sigma-\frac{1}{2}\right)\ln|n+\sigma+\mathrm{i}t| - t\tan^{-1}\frac{t}{n+\sigma}\right.$$
$$\left. -(n+\sigma-\frac{1}{2})\ln(n+\sigma)\right\}\left(1+O_A\left(\frac{1}{n}\right)\right)$$
$$=\exp\left(\frac{(n+\sigma)}{2}\ln\left(1+\frac{t^2}{(n+\sigma)^2}\right) - t\tan^{-1}\frac{t}{n+\sigma}\right)\frac{(n+\sigma)^{1/2}}{((n+\sigma)^2+t^2)^{1/4}}\left(1+O_A\left(\frac{1}{n}\right)\right)$$
$$\leqslant \exp\left\{(n+\sigma)\left(\frac{1}{2}\ln\left(1+\frac{t^2}{(n+\sigma)^2}\right) - \frac{t}{n+\sigma}\tan^{-1}\frac{t}{n+\sigma}\right)\right\}\left(1+O_A\left(\frac{1}{n}\right)\right)$$
$$=\exp\left\{(n+\sigma)h\left(\frac{t}{n+\sigma}\right)\right\}\left(1+O_A\left(\frac{1}{n}\right)\right),$$

where the function $h(x)$ is defined by

$$h(x) := \tfrac{1}{2}\ln(1+x^2) - x\tan^{-1}x, \quad x\geqslant 0.$$

Note that $h(0)=h'(0)=0$ and $h'(x)=-\tan^{-1}x\leqslant 0$. Hence $h(x)$ is negative for $x\geqslant 0$. Let

$$g(x) := \begin{cases} -\frac{x^2}{3} & \text{if } 0\leqslant x\leqslant 2,\\ -x+\frac{2}{3} & \text{if } x\geqslant 2.\end{cases}$$

Then $h(x)\leqslant g(x)$ for $x\geqslant 0$. Replacing $h(x)$ by $g(x)$ we see that the left-hand side of (22) is bounded by

$$O_A\left(\sum_{n\geqslant 1}\exp\left\{(n+\sigma)g\left(\frac{t}{n+\sigma}\right)\right\}\frac{\Gamma(n+\sigma)}{n!2^n}\right).$$

We use the asymptotics

$$\frac{\Gamma(n+\sigma)}{\Gamma(n+1)} = n^{\sigma-1}\left(1+O_A\left(\frac{1}{n}\right)\right),$$

where the big 'O' constant holds uniformly over the bounded set A, not only for an individual σ. The sum is reduced to

$$O_A\left(\sum_{n\geqslant 1}\exp\left\{(n+\sigma)g\left(\frac{t}{n+\sigma}\right)\right\}\frac{1}{(2-\varepsilon)^n}\right),$$

for an arbitrarily small $\varepsilon>0$. We split the above sum into S_1+S_2 according to the definition of $g(x)$:

$$S_1 = \sum_{n+\sigma\leqslant t/2} \mathrm{e}^{(n+\sigma)(-t/(n+\sigma)+(2/3))-n\ln(2-\varepsilon)},$$
$$S_2 = \sum_{n+\sigma\geqslant t/2} \mathrm{e}^{(n+\sigma)(-(1/3)(t^2/(n+\sigma)^2))-n\ln(2-\varepsilon)}.$$

Now

$$S_1 = \mathrm{e}^{-t}\sum_{n+\sigma\leqslant t/2}\mathrm{e}^{2(n+\sigma)/3-n\ln(2-\varepsilon)} = \mathrm{e}^{2\sigma/3}\mathrm{e}^{-t}\sum_{n+\sigma\leqslant t/2}\mathrm{e}^{n((2/3)-\ln(2-\varepsilon))}.$$

Since $\frac{2}{3}-\ln 2\approx -0.02648<0$ the sum converges and as $t\to\infty$ we have

$$S_1 = O_A(\mathrm{e}^{-t}). \tag{23}$$

Next,

$$S_2 = \sum_{n+\sigma \geqslant t/2} \mathrm{e}^{(n+\sigma)(-(1/3)(t^2/(n+\sigma)^2))-n\ \ln(2-\varepsilon)} = \sum_{n+\sigma \geqslant t/2} \mathrm{e}^{-(1/3)(t^2/(n+\sigma))-n\ \ln(2-\varepsilon)}.$$

Note that $n+\sigma \geqslant t/2$ implies $t/n \leqslant 2(1+\sigma/n)$. Since σ is bounded, we have $t/n = O_A(1)$ and

$$\frac{t^2}{n+\sigma} = \frac{t^2}{n} + O_A\left(\frac{t^2}{n^2}\right) = \frac{t^2}{n} + O_A(1).$$

Insert this in S_2 and omit the possibly finite many terms whose summation index is about $t/2$ to obtain

$$S_2 = O_A(\mathrm{e}^{-(t\ \ln(2-\varepsilon)/2)}) + O_A\left(\sum_{n \geqslant t/2} \mathrm{e}^{-(1/3)(t^2/n)-n\ \ln(2-\varepsilon)}\right). \tag{24}$$

Consider n as a *continuous* variable and let

$$j(n) := -\frac{1}{3}\frac{t^2}{n} - n\ln(2-\varepsilon).$$

It is easy to see that $j(n)$ attains the unique maximum at

$$n_{\max} = \frac{t}{\sqrt{3\ \ln(2-\varepsilon)}} \simeq (0.693)t,$$

with the maximum value

$$j(n_{\max}) = \frac{-2\sqrt{\ln(2-\varepsilon)}}{\sqrt{3}}t.$$

Decompose the sum in (24) as

$$\sum_{n \geqslant t/2} \mathrm{e}^{-(1/3)(t^2/n)-n\ \ln(2-\varepsilon)} = \sum_{0.8t \geqslant n \geqslant t/2} \mathrm{e}^{-(1/3)(t^2/n)-n\ \ln(2-\varepsilon)} + \sum_{n \geqslant 0.8t} \mathrm{e}^{-(1/3)(t^2/n)-n\ \ln(2-\varepsilon)}.$$

Since the summands in the second summation are decreasing, replacing the summation by the corresponding integral and estimating the first sum trivially we get

$$\begin{aligned}
&\left(0.8t - \frac{t}{2}\right)\mathrm{e}^{j(n_{\max})} + \int_{0.79t}^{\infty} \mathrm{e}^{-(1/3)(t^2/n)-n\ \ln(2-\varepsilon)}\,\mathrm{d}n \\
&\quad = \left(0.8t - \frac{t}{2}\right)\mathrm{e}^{(-2\sqrt{\ln(2-\varepsilon)}/\sqrt{3})t} + t\int_{0.79}^{\infty} \mathrm{e}^{-(1/3)(t/n)-nt\ \ln(2-\varepsilon)}\,\mathrm{d}n \\
&\quad = 0.3t\mathrm{e}^{(-2\sqrt{\ln(2-\varepsilon)})/(\sqrt{3})t} + t\int_{0.79}^{\infty} \mathrm{e}^{-t((1/3n)+n\ln(2-\varepsilon))}\,\mathrm{d}n.
\end{aligned}$$

The function $(1/3n) + n\ \ln(2-\varepsilon)$ is decreasing, so the asymptotics for the integral as $t \to \infty$ is straightforward. We have

$$\int_{0.79}^{\infty} \mathrm{e}^{-t((1/3n)+n\ \ln(2-\varepsilon))}\,\mathrm{d}n = O(t^{-1}\mathrm{e}^{-t((1/3(0.79))+(0.79)\ \ln(2-\varepsilon))}) = O(\mathrm{e}^{-0.969t}),$$

knowing that $1/3(0.79) + (0.79)\ \ln 2 \approx .969\,527\,200\,9$. We now use the value $-2\sqrt{\ln 2}/\sqrt{3} = -.961\,351\,257\,7$ to infer that

$$\sum_{n \geqslant t/2} \mathrm{e}^{-(1/3)(t^2/n)-n\ \ln(2-\varepsilon)} = O(\mathrm{e}^{-.961t}).$$

Inserting this in (24), we obtain $S_2 = O_A(\mathrm{e}^{-0.34t})$, which combined with (23) completes the proof. $\square$

Justification of depoissonization process. According to [3, Corollary 1, p.34] we need to construct a linear cone $\mathscr{L}(\theta)$ such that for some α, β, $0<\alpha, \beta<1$, there are positive constants A, B, R such that if $|z|>R$, $z\in\mathscr{L}(\theta)$ then

$$|\mathscr{P}(z,u)|\leqslant B|z|^{\beta}, \tag{25}$$

and if $|z|\geqslant R$, $z\notin\mathscr{L}(\theta)$ then

$$|\mathscr{P}(z,u)\mathrm{e}^{z}|\leqslant A\mathrm{e}^{\alpha|z|}. \tag{26}$$

hold uniformly for $|u|\leqslant 1/2$.

Verification of (25): By Lemma 6 we choose $\theta=0.3$ (actually any positive number less that 0.34 will work) and we will use the Mellin inversion (20) with $\sigma=-\beta$. Since

$$|\Gamma(-\beta-\mathrm{i}t)|=O(\mathrm{e}^{-((\pi/2)-\varepsilon)|t|})$$

we get

$$\begin{aligned}|\mathscr{P}(z,u)|&\leqslant\frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\frac{\mathscr{M}(T)(-\beta+\mathrm{i}t,u)}{2^{-\beta+\mathrm{i}t}-1}-\Gamma(-\beta+\mathrm{i}t))z^{-(-\beta+\mathrm{i}t)}\right|\,\mathrm{d}t\\&=O_{\beta}\left(\int_{-\infty}^{\infty}(\mathrm{e}^{-0.34|t|}+\mathrm{e}^{-((\pi/2)-\varepsilon)|t|})\left|z^{-(-\beta+\mathrm{i}t)}\right|\,\mathrm{d}t\right)\\&=O_{\beta}\left(|z|^{\beta}\int_{-\infty}^{\infty}(\mathrm{e}^{-0.34|t|}+\mathrm{e}^{-((\pi/2)-\varepsilon)|t|})\mathrm{e}^{t\ \arg(z)}\,\mathrm{d}t\right).\end{aligned}$$

Because $z\in\mathscr{L}(0.3)$, $|\arg(z)|\leqslant 0.3$. As a result, the above integral is absolutely convergent and (25) follows.

Verification of (26): We use the idea of increasing domains. To this end, we choose $\lambda=3/2$ so that $\lambda/2$ and $\lambda/4$ are less than 1. For some $\xi>1$, define the domain D_m by

$$D_m=\{z: z\notin\mathscr{L}(0.3),\ \ \xi/4\leqslant|z|\leqslant\xi\lambda^{m+1}\},\quad m=0,1,2,\ldots.$$

We use induction to verify that (26) holds in all D_m. Let the α in question be any number such that $1>\alpha>\cos(0.3)$. We will see how to choose R, ξ in the sequel.

Now rewrite the functional equation (16) in the form

$$\mathscr{P}(z,u)=\mathscr{P}(z/2,u)-(1-u)(1-\mathrm{e}^{-z/2})\,\mathrm{e}^{-z/4}\mathscr{P}(z/4,u)+\mathrm{e}^{-z/2}(1-\mathrm{e}^{-z/2}).$$

Multiply the above by e^{z} and further reduce the resulting equation to

$$\mathscr{P}(z,u)\mathrm{e}^{z}=(\mathscr{P}(z/2,u)\mathrm{e}^{z/2})\mathrm{e}^{z/2}-(1-u)(\mathrm{e}^{z/2}-1)(\mathscr{P}(z/4,u)\mathrm{e}^{z/4})+(\mathrm{e}^{z/2}-1).$$

For $z\notin\mathscr{L}(0.3)$, $\Re(z)\leqslant|z|\ \cos 0.3$. Hence

$$\begin{aligned}|\mathscr{P}(z,u)\mathrm{e}^{z}|\leqslant{}&\left|\mathscr{P}(z/2,u)\mathrm{e}^{z/2}\right|\mathrm{e}^{(1/2)\Re(z)}+|1-u|(\mathrm{e}^{(1/2)\Re(z)}+1)\left|\mathscr{P}(z/4,u)\mathrm{e}^{z/4}\right|\\&+(\mathrm{e}^{(1/2)\Re(z)}+1)\\\leqslant{}&|\mathscr{P}(z/2,u)\mathrm{e}^{z/2}|\mathrm{e}^{(1/2)|z|\cos 0.3}+|1-u|(\mathrm{e}^{(1/2)|z|\cos 0.3}+1)\left|\mathscr{P}(z/4,u)\mathrm{e}^{z/4}\right|\\&+(\mathrm{e}^{(1/2)|z|\cos 0.3}+1).\end{aligned}\tag{27}$$

Choose ξ so large and fix it so that when $|z|\geqslant\xi$, we have simultaneously

$$\begin{aligned}&\mathrm{e}^{(1/2)|z|(\cos 0.3-\alpha)}\leqslant\frac{1}{3},\\&\mathrm{e}^{-|z|(3/4)\alpha}(\mathrm{e}^{(1/2)|z|\ \cos 0.3}+1)|1-u|\leqslant\frac{1}{3},\\&\mathrm{e}^{-|z|\alpha}(\mathrm{e}^{(1/2)|z|\cos 0.3}+1)\leqslant\frac{1}{3}.\end{aligned}\tag{28}$$

Such choice for ξ is realizable since $\cos 0.3 < \alpha$. With ξ chosen, choose $A > 1$ so that (26) is satisfied in D_0. This is feasible due to the compactness of D_0. We will show by induction that the (26) holds in all domains D_m with respect to the same parameters. Assume that it holds in D_m and let $z \in D_{m+1}$. If $z \in D_m$, we are done by the induction hypothesis. If $z \in D_{m+1} - D_m$, then by the definition of D_m we have

$$\xi\left(\frac{3}{2}\right)^{m+1} < |z| \leqslant \xi\left(\frac{3}{2}\right)^{m+2}.$$

Multiplying by $\frac{1}{2}$ gives

$$\frac{\xi}{4} \leqslant \frac{3\xi}{4} \leqslant \xi\left(\frac{3}{2}\right)^{m+1} \cdot \frac{1}{2} < |z/2| \leqslant \xi\left(\frac{3}{2}\right)^{m+1} \frac{3}{4} \leqslant \xi\left(\frac{3}{2}\right)^{m+1}.$$

That is,

$$\frac{\xi}{4} \leqslant |z/2| \leqslant \xi\left(\frac{3}{2}\right)^{m+1}.$$

Hence $z/2 \in D_m$. Similarly, $z/4 \in D_m$. By the induction hypothesis and (27), we have

$$\begin{aligned}
&|\mathscr{P}(z,u)\mathrm{e}^z| \\
&\quad \leqslant B\mathrm{e}^{\alpha(|z|/2)}\mathrm{e}^{(1/2)|z|\cos 0.3} + |1-u|(\mathrm{e}^{(1/2)|z|\cos 0.3} + 1)B\mathrm{e}^{\alpha(|z|/4)} + (\mathrm{e}^{(1/2)|z|\cos 0.3} + 1) \\
&\quad = \left(\mathrm{e}^{(1/2)|z|(\cos 0.3 - \alpha)} + |1-u|\mathrm{e}^{-|z|(3/4)\alpha}(\mathrm{e}^{(1/2)|z|\cos 0.3} + 1) + \frac{\mathrm{e}^{-\alpha|z|}}{B}(\mathrm{e}^{(1/2)|z|\cos 0.3} + 1)\right) B\mathrm{e}^{\alpha|z|}.
\end{aligned}$$

By (28) the expression in parenthesis is less than 1 which gives (26) and completes the induction (we may take $R = \xi$).

References

[1] W.M.Y. Goh, E. Schmutz, Gap-free set partitions, Random Struct. Algorithms 3 (1992) 9–18.

[2] P. Hitczenko, A. Knopfmacher, Gap-free compositions and gap-free samples of geometric random variable, Discrete Math. 294 (2005) 225–239.

[3] P. Jacquet, W. Szpankowski, Analytical depoissonization and its applications, Theoret. Comput. Sci. 201 (1998) 1–62.

[4] S. Janson, W. Szpankowski, Analysis of the asymmetric leader election algorithm, Electron. J. Combin. 4 (1997) 16.

[5] G. Lourchard, H. Prodinger, The number of gaps in sequences of geometrically distributed random variables, Preprint available at ⟨http://www.ulb.ac.be/di/mcs/louchard/⟩ (number 81 on the list) or at ⟨http://math.sun.ac.za/∼prodinger/pdffiles/gapsAPRIL27.pdf⟩.