

# Samples of geometric random variables with multiplicity constraints

Margaret Archibald and Arnold Knopfmacher<sup>†</sup>

*The John Knopfmacher Centre for Applicable Analysis and Number Theory, Department of Mathematics, University of the Witwatersrand, P. O. Wits, 2050, Johannesburg, South Africa*  
*marchibald@maths.wits.ac.za, arnoldk@cam.wits.ac.za*

We investigate the probability that a sample  $\Gamma = (\Gamma_1, \Gamma_2, \dots, \Gamma_n)$  of independent, identically distributed random variables with a geometric distribution has no elements occurring exactly  $j$  times, where  $j$  belongs to a specified finite ‘forbidden set’  $A$  of multiplicities. Specific choices of the set  $A$  enable one to determine the asymptotic probabilities that such a sample has no variable occurring with multiplicity  $b$ , or which has all multiplicities greater than  $b$ , for any fixed integer  $b \geq 1$ .

**Keywords:** Geometric random variable, Mellin transform, Poisson transform, multiplicity

## 1 Introduction

We study samples of independent, identically distributed (i.i.d.) random variables with a geometric distribution. Specifically, let  $\Gamma_1, \Gamma_2, \Gamma_3, \dots$  be i.i.d. geometric random variables with parameter  $p$ , that is,  $\mathbb{P}(\Gamma_1 = j) = pq^{j-1}$ ,  $j \in \mathbb{N}$ , with  $p + q = 1$ . There is now an extensive literature on the combinatorics of geometric random variables and its applications in Computer Science. We are interested in the probability that a random sample of  $n$  such variables consists of elements whose multiplicities belong to specified sets. That is, we place restrictions on the number of times any element/letter can occur in the sample.

As a simple example, we may wish to consider a sample where none of the  $n$  elements occur exactly  $b$  times. In this case  $A = \{b\}$ . Another example of such a forbidden set is when a letter can occur only  $b$  times or more (or not at all), i.e.,  $A = \{1, 2, \dots, b-1\}$ , where  $b \geq 2$ . Note that we do not allow 0 in the forbidden set. Previously in (HK05; LP05), certain geometric samples with 0 in the forbidden set were studied under the names of ‘complete’ and ‘gap-free’ samples.

**Theorem 1** *Let  $A$  be any finite set of positive integers. The probability  $p_n$  that a geometric sample of length  $n$  has no letter appearing with multiplicity  $j$ , for any  $j \in A$  is (asymptotically as  $n \rightarrow \infty$ ),*

$$p_n = 1 - \frac{T^*(0)}{\ln(1/q)} - \frac{2}{\ln(1/q)} \Re \left( \sum_{k=1}^{\infty} \exp\{\chi_k \ln(q/n)\} T^*(\chi_k) \right) + O(n^{-1}),$$

where we set  $\chi_k := \frac{2k\pi i}{\ln(1/q)}$ , and where

$$T^*(0) = \sum_{j \in A} p^j \sum_{n \geq 0} p_n q^n \frac{1}{n+j} \binom{n+j}{j} \quad (1.1)$$

and

$$T^*(\chi_k) = \sum_{j \in A} \frac{p^j}{j!} \sum_{n \geq 0} p_n \frac{q^n}{n!} \Gamma(n+j+\chi_k), \quad \text{for } k \in \mathbb{Z} \setminus \{0\}. \quad (1.2)$$

<sup>†</sup>This material is based upon work supported by the National Research Foundation under grant number 2053740

## 2 Outline of the proof of Theorem 1

We want a recursion on the variable  $p_n$ , the probability that the sample does not have letters which appear exactly  $j$  times if  $j$  is an element of the forbidden set  $A \subset \mathbb{N}$ . Let  $\mathcal{B}$  represent the set of all permitted samples. If we let  $p_n = \mathbb{P}(\Gamma \in \mathcal{B})$  (the probability that the sample  $\Gamma = (\Gamma_1, \Gamma_2, \dots, \Gamma_n)$  has no letter occurring exactly  $j$  times for  $j \in A$ ), then we can write

$$p_n = \sum_{\substack{j \geq 0 \\ j \notin A}} \mathbb{P}\left(\{\Gamma \in \mathcal{B}\} \cap \left\{\sum_{\ell=1}^n I_{\Gamma_\ell=1} = j\right\}\right) = \sum_{\substack{j \geq 0 \\ j \notin A}} \mathbb{P}\left(\Gamma \in \mathcal{B} \mid \sum_{\ell=1}^n I_{\Gamma_\ell=1} = j\right) \cdot \mathbb{P}\left(\sum_{\ell=1}^n I_{\Gamma_\ell=1} = j\right),$$

where the indicator function  $I$  takes values 1 for true and 0 for false. Using the law of total probability, and the memoryless property of geometric random variables, we obtain the recursion:

$$p_n = \sum_{j=0}^n p_{n-j} \binom{n}{j} p^j q^{n-j} - \sum_{j \in A} p_{n-j} \binom{n}{j} p^j q^{n-j}. \tag{2.1}$$

We would like to see how  $p_n$  behaves asymptotically as  $n \rightarrow \infty$ . The Poissonisation technique we use can be seen in (JS98; Szp01), but we follow more specifically the process used in (HK05; JS97). Namely, we consider the Poisson transform of the sequence  $(p_n)$ , analyse its asymptotics with Mellin transforms, then de-Poissonise to recover the asymptotics of  $(p_n)$ . To do this, we make use of the exponential generating function  $P(z)$ , which is the Poisson transform of  $(p_n)$ , given by  $P(z) := \sum_{n \geq 0} p_n \frac{z^n}{n!} e^{-z}$ . We use (2.1) to show that

$$P(z) - e^{-z} = \sum_{n \geq 1} p_n \frac{z^n}{n!} e^{-z} = P(qz) - e^{-pz} \sum_{j \in A} \frac{(pz)^j}{j!} P(qz) - e^{-z}.$$

We thus have a functional equation of the form:

$$P(z) = P(qz) - e^{-pz} \sum_{j \in A} \frac{(pz)^j}{j!} P(qz). \tag{2.2}$$

The technique we now use is the Mellin transform. A standard reference on Mellin transforms is (FGD95). We define the function (see (2.2))

$$T(z) := e^{-pz} \sum_{j \in A} \frac{(pz)^j}{j!} P(qz) \quad \left( = \sum_{n \geq 0} p_n \frac{q^n}{n!} \sum_{j \in A} \frac{p^j}{j!} z^{n+j} e^{-z} \right). \tag{2.3}$$

We note that the Mellin transform of  $T(z)$  has a fundamental strip of at least  $\langle -1, \infty \rangle$ . Now we find the Mellin transform of (2.3) to be

$$T^*(s) = \sum_{n \geq 0} p_n \frac{q^n}{n!} \sum_{j \in A} \frac{p^j}{j!} \mathcal{M}(z^{n+j} e^{-z}) = \sum_{n \geq 0} p_n \frac{q^n}{n!} \sum_{j \in A} \frac{p^j}{j!} \Gamma(n+j+s).$$

In particular we will make use of the values  $T^*(0)$  and  $T^*(\chi_k)$ , as given in (1.1) and (1.2). After iterating (2.2) we get

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

for any  $m \geq 0$  and thus in the limit as  $m \rightarrow \infty$ ,  $P(z) = 1 - \sum_{j=0}^{\infty} T(q^j z)$ . We define  $Q(z) := P(z) - 1 = -\sum_{j=0}^{\infty} T(q^j z)$ . Then the corresponding Mellin transform is

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

where  $Q^*(s)$  exists in the fundamental strip  $\langle -1, 0 \rangle$ . The inverse Mellin transform is

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

for any  $-1 < c < 0$ . The residue theorem can be used to evaluate this. This process (compare to (HK05)) results in

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

Now,  $x^{-s}Q^*(s) = x^{-s}T^*(s)/(q^{-s} - 1)$  has simple poles at  $q^{-s} = 1$ , i.e. at  $\chi_k = 2k\pi i/\ln(1/q)$ ,  $k \in \mathbb{Z}$ , and the corresponding residues are

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

Of these quantities, all but the  $k = 0$  term contribute oscillations of small amplitude. We need to use asymptotic de-Poissonisation to deduce that  $p_n \sim P(n) = Q(n) + 1$ . We consider the theorem given in Szpankowski, (Szp01, page 463), whose five conditions are met by choosing  $\gamma_1(z) = 0$ ,  $\gamma_2(z) = 1$ , and  $t(z) = -T(z)$ . We can now deduce that

$$p_n = P(n) + O(n^{-1}) = Q(n) + 1 + O(n^{-1})$$

where

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

with  $T^*(0)$  and  $T^*(\chi_k)$  given in (1.1) and (1.2) respectively. This concludes the proof of Theorem 1.

### 3 The complementary set $\mathbb{N} \setminus A$ is finite

We consider now the complementary problem where the *permitted* set  $B = \mathbb{N} \setminus A$  is finite. The de-Poissonisation method cannot be used in this case, but we can bound the probabilities by elementary means to show that  $p_n \rightarrow 0$  for all finite sets  $B$ . (Consequently fluctuations are also absent in these cases.) Suppose the permitted set of multiplicities  $B = \mathbb{N} \setminus A$  is finite with largest element  $k$ . The probability that all multiplicities belong to such a set  $B$  is bounded above by the case when  $B = \{0, 1, 2, \dots, k\}$ . Now samples where all multiplicities are at most  $k$  are themselves a subset of the set of samples with the weaker restriction that there are at most  $k$  ones.

The probability of exactly  $j$  ones in a sample of length  $n$  is  $\binom{n}{j} p^j (1-p)^{n-j}$ . Hence the probability of at most  $k$  ones is

$$\sum_{j=0}^k \binom{n}{j} p^j (1-p)^{n-j} \leq (k+1) \binom{n}{k} (1-p)^{n-k} = O(n^k q^n),$$

which is exponentially small.

### 4 Further work

We aim to continue this work by considering other examples of forbidden sets  $A$ . In addition, we would like to reconsider the results of this paper from an urn model standpoint, as was used to obtain the results in (LP05).

### Acknowledgements

We would like to thank Helmut Prodinger for the helpful advice he gave at the start of this project, relating to finding asymptotics from the recursion (2.1).

### References

- [FGD95] P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, 144:3–58, 1995.
- [HK05] P. Hitczenko and A. Knopfmacher. Gap-free compositions and gap-free samples of geometric random variables. *Discrete Mathematics*, 294:225–239, 2005.
- [JS97] S. Janson and W. Szpankowski. Analysis of the asymmetric leader election algorithm. *Electr. J. Combin.*, 4(1, research paper 17):16pp, 1997.

- [JS98] P. Jacquet and W. Szpankowski. Analytical de-poissonization and its applications. *Theoret. Comput. Sci.*, 201:1–62, 1998.
- [LP05] G. Louchard and H. Prodinger. The number of gaps in sequences of geometrically distributed random variables. *Preprint*, 304, 2005.
- [Pro93] H. Prodinger. Combinatorial problems of geometrically distributed random variables and applications in computer science. *Publications de l'IRMA (Strasbourg)*, 30:87–95, 1993.
- [Szp01] W. Szpankowski. *Average Case Analysis of Algorithms on Sequences*. John Wiley and Sons, New York, 2001.