The expected value and the variance of a multiplicity of a given part size in a random composition of an integer

Boris L. Kheyfets *
Department of Mathematics
Drexel University
Philadelphia, PA 19104-2875
blk24@drexel.edu

Abstract

The expected value and the variance of a multiplicity of a given part size in a random composition of an integer is obtained. This result was used in [1] to analyze algorithms for computing the Walsh-Hadamard transform.

The expected multiplicity of a part size can be viewed as one aspect of a measure of the "m-distinctness" [2] of a random composition. Because of its simplicity the formula has already found a nice application to theoretical computer science. In the paper [1], which is devoted to the analysis of a class of divide and conquer recurrences arising from the computation of the WHT, the expected multiplicity of a part size was used (Lemma 2, Section 5) to calculate the mean value of times of execution loops in WHT algorithm.

We start presenting the result of this paper with necessary definitions. We consider a k-tuple $\kappa_n = (\gamma_1, \ldots, \gamma_k)$ where all γ_j's are positive integers, called *parts*, and the number k is called a "*number of parts*". If $\sum_{j=1}^{k} \gamma_j = n$, then we call κ_n - a *composition* of an integer n. The values of γ_j's are called "*part sizes*". The *multiplicity* of a part size is the number of parts with that size. There are exactly 2^{n-1} different compositions of n. This fact can be found e.g. in Andrews [3, Example 3, p. 63]. We denote the set of all compositions of n by Ω_n. A "random composition" means a

*Supported in part by NSA grant MSPF-02G-043

composition chosen accordingly to the uniform probability measure on Ω_n. The following theorem is true.

Theorem 1. *Let $X = X_{n,j}$ denote the multiplicity of part size j in a random composition of n. For $n \geq j+1$,*

$$E(X) = \frac{n+3-j}{2^{j+1}},$$

and for $n \geq 2j+1$,

$$Var(X) = \frac{3n - 2nj + 3j^2 - 12j + 5}{2^{2(j+1)}} + \frac{n+3-j}{2^{j+1}}.$$

Proof. Let $g_j(n,m)$ be the number of compositions of n in which part size j has multiplicity m. In terms of generating function [4]:

$$G_j(z,u) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} g_j(n,m) z^n u^m = \frac{1-z}{1-2z-(u-1)z^j(1-z)},$$

the expected value of X is [5, p. 108]:

$$E(X) = \frac{[z^n]\left(\frac{\partial G_j(z,u)}{\partial u}\right)_{u=1}}{2^{n-1}}.$$

Let $H(z) = 1-2z$ and $K(z) = z^j(1-z)$. Expending

$$\begin{aligned} G_j(z,u) &= \frac{1-z}{H(z)}\frac{1}{1-(u-1)\frac{K(z)}{H(z)}} \\ &= \frac{1-z}{H(z)}\left(1+(u-1)\frac{K(z)}{H(z)} + (u-1)^2\frac{K^2(z)}{H^2(z)} + \ldots\right) \qquad (1) \end{aligned}$$

we see that

$$E(X) = 2^{-(n-1)}[z^n]\left(\frac{(1-z)^2 z^j}{(1-2z)^2}\right).$$

In view of the expansion

$$(1-2z)^{-2} = \sum_{k=0}^{\infty}(k+1)2^k z^k,$$

we have

$$\begin{aligned} E(X) &= 2^{-(n-1)}[z^n]\sum_{k=0}^{\infty}(k+1)2^k(1-2z+z2)z^{k+j} \\ &= 2^{-(j+1)}(4(n-j+1) - 4(n-j) + (n-j-1)) \\ &= \frac{n+3-j}{2^{j+1}} \qquad (n \geq j+1). \end{aligned}$$

Note. If $n = j$ then $E(X) = 2^{-(j-1)}$ and if $n < j$ then $E(X) = 0$. To determine the variance, we begin by computing the second factorial moment. From

$$E(X(X-1)) = \frac{[z^n]\left(\frac{\partial^2 G_j(z,u)}{\partial u^2}\right)_{u=1}}{2^{n-1}}$$

and the expansion (1), we have

$$E(X(X-1)) = 2\cdot 2^{-(n-1)}\,[z^n]\left(\frac{(1-z)^3 z^{2j}}{(1-2z)^3}\right).$$

Thus

$$\begin{aligned}
E(X(X-1)) &= 2\cdot 2^{-(n-1)}\,[z^n]\sum_{k=0}^{\infty}\binom{k+2}{2}2^k(1-3z+3z^2-z^3)z^{k+2j} \\
&= 2^{-2(j+1)}\left(16\binom{n-2j+2}{2}-24\binom{n-2j+1}{2}\right. \\
&\quad \left.+12\binom{n-2j}{2}-2\binom{n-2j-1}{2}\right) \\
&= \frac{n^2-4nj+4j^2+9n-18j+14}{2^{2(j+1)}} \qquad (n\geq 2j+1).
\end{aligned}$$

Hence

$$\begin{aligned}
Var(X) &= E(X(X-1))+E(X)-E^2(X) \\
&= \frac{n^2-4nj+4j^2+9n-18j+14}{2^{2(j+1)}}+\frac{n+3-j}{2^{j+1}}-\left(\frac{n+3-j}{2^{j+1}}\right)^2 \\
&= \frac{3n-2nj+j^2-12j+5}{2^{2(j+1)}}+\frac{n+3-j}{2^{j+1}} \qquad (n\geq 2j+1).
\end{aligned}$$

This completes the proof.

Acknowledgement. We wish to thank the anonymous referee, whose suggestion allowed a significant improvement of the manuscript.

References

[1] P. Hitczenko, J. Johnson, H.-J. Huang, Distribution of a Class of Divide and Conquer Recurrences Arising from the Computation of the Walsh-Hadamard Transform, (preprint, http://www.mcs.drexel.edu/~phitczen/whtpaper.pdf), 2004.

[2] G. Louchard, The number of distinct part sizes of some multiplicity in compositions of an Integer, *A Probabilistic Analysis*, 2003.

[3] G.E. Andrews, *The Theory of Partitions*, Addison - Wesley, Reading, MA, 1976.

[4] P. Hitczenko, C. Rousseau, C.D. Savage, A Generatingfunctionology Approach to a Problem of Wilf, *Journal of Combinatorial and Applied Mathematics* **142** (2002), 107–114.

[5] H.S. Wilf, *Generatingfunctionology*, 2nd ed., Academic Press, San Diego, 1994.