#A7 INTEGERS 24 (2024)

EVALUATING THE GENERALIZED BUCHSTAB FUNCTION AND REVISITING THE VARIANCE OF THE DISTRIBUTION OF THE SMALLEST COMPONENTS OF COMBINATORIAL OBJECTS

Claude Gravel
Department of Computer Science, Toronto Metropolitan University, Canada
gravel@torontomu.ca

Daniel Panario
School of Mathematics and Statistics, Carleton University, Canada
daniel@math.carleton.ca

Received: 12/30/22, Revised: 8/13/23, Accepted: 12/20/23, Published: 1/2/24

Abstract

Let $n \geq 1$ and let X_n be the random variable representing the size of the smallest component of a random combinatorial object made of n elements. Combinatorial objects belong to parametric classes. This article focuses on the exp-log class with parameter $K = 1$ (permutations, derangements, polynomials over a finite field, etc.) and $K = 1/2$ (surjective maps, 2-regular graphs, etc.). The generalized Buchstab function Ω_K is essential in evaluating probabilistic and statistical quantities. For $K = 1$, it is known that $\mathrm{Var}(X_n) = C(n + O(n^{-\epsilon}))$ for some $\epsilon > 0$ and sufficiently large n. We revisit the evaluation of $C = 1.3070\ldots$ using different methods: analytic estimation using tools from complex analysis, numerical integration using Taylor expansions, and computation of the exact distributions for $n \leq 4000$ using the recursive nature of the counting problem. In general, for any K, the quantity $1/\Omega_K(x)$ for $x \geq 1$ is related to the asymptotic proportion of n-objects with large smallest components. We show how the coefficients of the Taylor expansion of $\Omega_K(x)$ for $\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1$ depend on those for $\lfloor x \rfloor - 1 \leq x - 1 < \lfloor x \rfloor$. We use this family of coefficients to evaluate $\Omega_K(x)$.

1. Introduction

Let $n \geq 1$ and let X_n be the random variable representing the size of the smallest component of a random combinatorial object made of n elements. By a random combinatorial object, we mean a combinatorial object chosen uniformly at random among all possible combinatorial objects of size n. The cardinality of the support of X_n is, in principle, $n + 1$. Since the length of the smallest component obviously

DOI: 10.5281/zenodo.10450967

cannot be between $\lfloor n/2 \rfloor + 1$ and $n-1$ inclusively, the range of X_n is $1, 2, \ldots, \lfloor n/2 \rfloor$ together with n. For some reasons that will become clear hereafter, we add zero probabilities to extend the range of X_n over all integers between 1 and n inclusively.

The theory about combinatorial objects and analytical methods required to understand many of the references in this paper is in [6]. Our results in Section 2 are valid for the class that contains permutations, derangements, and monic polynomials over finite fields, to name a few. Our result in Section 3 applies to all combinatorial objects in the exp-log class. We let readers consult [6] for the proper definitions of the exp-log class of combinatorial objects.

We can take the typical permutations or monic polynomials over finite fields. The latter receives special treatment in [9]. References [12] and [13] give local results about the probability distribution of X_n and asymptotic results about the k-th moment of X_n. One of our goals in this paper is to revisit some results concerning the second moment to compute the variance of X_n, denoted by $\mathrm{Var}(X_n)$. We recall that, by definition,

$$\mathrm{Var}(X_n) = \sum_{k=1}^{n} \big(k - \mathbf{E}(X_n)\big)^2 \, \mathbf{P}\{X_n = k\} = \mathbf{E}(X_n^2) - (\mathbf{E}(X_n))^2, \tag{1}$$

where $\mathbf{P}\{X_n = k\}$ is the probability that X_n equals k, and $\mathbf{E}(X_n)$ is the expectation of X_n.

The k-th moment of X_n, $\mathbf{E}(X_n^k)$, is expressed as an integral involving the ordinary Buchstab function ω, which is defined over the real interval $[1, \infty)$ by

$$\omega(x) = \frac{1}{x} \quad \text{for } 1 \le x \le 2, \quad \text{and} \quad \frac{\mathrm{d}(x\omega(x))}{\mathrm{d}x} = \omega(x-1) \quad \text{for } x \ge 2. \tag{2}$$

As mentioned in [12], the k-th moment of X_n involves the quantity $\int_1^\infty t^{-k}\omega(t)\mathrm{d}t$. Besides the original paper by Buchstab [3] in which the function is defined and analyzed, numerous other papers discuss its various properties and applications, such as [2]. The book [15] contains many proofs of the properties of the Buchstab function.

Theorem 5 from [13] stipulates that

$$\mathrm{Var}(X_n) = C\big(n + O(n^{-\epsilon})\big) \quad \text{for some } \epsilon > 0. \tag{3}$$

The constant C from (3) is given by

$$C = 2\int_1^\infty \frac{\omega(t)}{t^2}\mathrm{d}t. \tag{4}$$

Remark 1. In [1], [11], [12], and [13], the integration interval for (4) starts at 2. When computing the variance, the authors inadvertently forgot to add $3/4$, resulting from the integration over the interval $[1, 2)$. This mistake leads to confusion for some researchers; see [5].

Let S_n be the set of permutations on n elements, and let $S_{k,n} \subsetneq S_n$ be those permutations with smallest cycles of length k for $1 \le k \le n$. Denote the cardinality of $S_{k,n}$ by $s_{k,n}$. Let $c_k = (k-1)!$ for $k \ge 1$, and let $[n/k] = 1$ if and only if $k|n$; otherwise $[n/k] = 0$. Then, in [12] it is proved that

$$s_{k,n} = \sum_{i=1}^{\lfloor n/k \rfloor} \frac{c_k^i}{i!} \frac{n!}{(k!)^i (n-ki)!} \sum_{j=k+1}^{n-ki} s_{j,n-ki} + [n/k] \frac{c_k^{n/k}}{(n/k)!} \frac{n!}{(k!)^{n/k}} \tag{5}$$

$$= \sum_{i=1}^{\lfloor n/k \rfloor} \frac{n!}{k^i i! (n-ki)!} \sum_{j=k+1}^{n-ki} s_{j,n-ki} + [n/k] \frac{n!}{(n/k)! k^{n/k}}. \tag{6}$$

In order to simplify the notation from [12] to fit our purpose here, we change slightly the notation from $L^s_{k,n}$ to $s_{k,n}$.

For a fixed n, we have the following properties:

$$s_{n,n} = (n-1)!, \quad s_{k,n} = 0 \text{ for } \lfloor n/2 \rfloor + 1 \le k \le n-1, \quad \text{and } \sum_{k=1}^{n} s_{k,n} = n!.$$

We have for a fixed $n \ge 1$ that

$$\mathbf{P}\{X_n = k\} = \frac{s_{k,n}}{n!} \quad \text{for } 1 \le k \le n.$$

In Section 2, we evaluate C from (3) using different approaches. Another of our goals in Section 3, is to evaluate the generalized Buchstab[1] function with parameter $K > 0$ defined by

$$\Omega_K(x) = \begin{cases} 1 & \text{for } 1 \le x < 2, \\ 1 + K \int_2^x \frac{\Omega_K(u-1)}{u-1} \mathrm{d}u & \text{for } x \ge 2. \end{cases} \tag{7}$$

The quantity $1/\Omega_K(x)$ gives the fraction of n-objects with large smallest components; more precisely, Theorem 1.1 from [1] stipulates that

$$\lim_{n \to \infty} \frac{s_{\lfloor xn \rfloor, \lfloor xn \rfloor}}{\sum_{i=n}^{\lfloor xn \rfloor} s_{\lfloor xn \rfloor, i}} = \frac{1}{\Omega_K(x)} \quad \text{for } x > 1.$$

For the sake of completeness and to gain insight into how the Buchstab function connects to combinatorial analysis, we end this introduction by briefly recalling how Buchstab introduced his function ω when studying the factorization of natural numbers into primes. The primes are like the irreducible factors of a polynomial, or the cycles of a permutation, etc. Let $\xi \in \{1, \dots, n\}$ with its decomposition into

[1]We thank an anonymous referee for bringing to our attention that the function considered here is not a possible generalization of the original Buchstab because there is no K such that Ω_K coincides with ω on the interval $[1, 2)$.

primes given as $p_1(\xi)\cdots p_k(\xi)=\xi$ such that $p_1(\xi)\le p_2(\xi)\le\ldots\le p_k(\xi)$. We count the number of ξ's with their smallest prime factor less than m; in other words, set

$$\Psi(n,m)=\operatorname{card}\{\xi\in\{1,\ldots,n\}\colon p_1(\xi)\le m\}.$$

Then in [3] it is shown that

$$\Psi(n,m)=1+\sum_{p\le m}\Psi\left(\frac{n}{p},p\right)\quad\text{for all } 1<m\le n.$$

The previous summation is over all the primes p less than or equal to m. The functional equation given by Ψ is related to the Dickman function, which we do not discuss here; see [15] for a detailed analysis of the Dickman function and the Buchstab function.

2. Approaches

2.1. Analytic Estimation

This section mostly recalls results from [11] and [13]. The approach from [13] to obtain the limiting quantities for $\mathbf{P}\{X_n\ge k\}$ and $\mathbf{E}(X_n^\ell)$ as $k,n\to\infty$ and $\ell\ge 1$ uses singularity analysis of exponential generating functions for combinatorial objects. For an in-depth coverage of singularity analysis applied to combinatorics, see [6].

Permutations form a typical class of combinatorial objects that we choose here for our discussion, but the results are not limited only to permutations. The cycles are the irreducible components of a permutation. Let $C(z)=\sum_{i=0}^{\infty}C_i z^i/i!$ be the exponential generating function for counting cycles of given lengths. Then, the exponential generating function for counting permutations of given sizes is

$$L(z)=\exp(C(z))=\sum_{i=0}^{\infty}L_i\frac{z^i}{i!}.$$

For a fixed $n>0$, we are interested in counting permutations with the smallest cycles of length at least k for $1\le k\le n$. Let $S(z)$ be the generating function for counting permutations with the smallest cycles of length at least k for $1\le k\le n$. Then we have

$$S(z)=\exp\left(\sum_{i=k}^{\infty}C_i\frac{z^i}{i!}\right)-1=\sum_{i=0}^{\infty}S_i\frac{z^i}{i!}.$$

Therefore, the tail of the probability distribution of X_n is given by

$$\mathbf{P}\{X_n\ge k\}=\frac{S_n}{L_n}.$$

Using singularity analysis, in [13] it is shown that if $k, n \to \infty$, then

$$\mathbf{P}\{X_n \geq k\} = \frac{1}{k}\omega\left(\frac{n}{k}\right) + O\left(\frac{1}{k^{1+\epsilon}}\right) \quad \text{for some } \epsilon > 0. \tag{8}$$

Theorem 1 states the asymptotic behavior of the moments.

Theorem 1. *For some function $h(n)$ that tends to infinity slower than $\log(n)$ and for some $\epsilon > 0$ independent of n, we have that*

$$\mathbf{E}(X_n) = e^{-\gamma}\log(n)\left(1 + O\left(\frac{h(n)}{\log(n)}\right)\right),$$
$$\mathbf{E}(X_n^\ell) = \ell n^{\ell-1}\left(\int_1^\infty \frac{\omega(x)}{x^\ell}\mathrm{d}x\right)\left(1 + O\left(\frac{1}{n^\epsilon}\right)\right) \quad \text{for integers } \ell \geq 2.$$

Proof. We consider the case when $\ell \geq 2$. We give the main steps for proving Theorem 1. By definition, we have

$$\mathbf{E}(X_n^\ell) = \sum_{k=1}^{\infty}\left(k^\ell - (k-1)^\ell\right)\mathrm{P}\{X_n \geq k\}. \tag{9}$$

Let $\nu(n) = \lfloor n^{\epsilon'} \rfloor$ such that $0 < \epsilon' < \epsilon$ where ϵ is given from (8). Then $\nu(n) \to \infty$ as $n \to \infty$, so we split the sum from (9) using ν, and we obtain

$$\begin{aligned}\mathbf{E}(X_n^\ell) &= \sum_{k=1}^{\nu(n)-1}\left(k^\ell - (k-1)^\ell\right)\mathrm{P}\{X_n \geq k\} + \sum_{k=\nu(n)}^{\infty}\left(k^\ell - (k-1)^\ell\right)\mathrm{P}\{X_n \geq k\}\\ &\overset{\text{def}}{=} S_1 + S_2.\end{aligned}$$

Using (8), and the fact that $\mathbf{P}\{X_n \geq n+1\} = 0$, we have $S_1 = O\big((\nu(n))^{\ell-1}\big)$ because $(k^\ell - (k-1)^\ell) \in O(k^{\ell-1})$ and $\mathrm{P}\{X_n \geq k\} \in O(1/k^{1+\epsilon})$ in the range $1 \leq k < \nu(n)$. In the range $\nu(n) \leq k \leq n$, we have $(k^\ell - (k-1)^\ell) \in O(\ell k^{\ell-1})$, and therefore

$$\begin{aligned}S_2 &= \sum_{k=\nu(n)}^{\infty}\left(k^\ell - (k-1)^\ell\right)\mathrm{P}\{X_n \geq k\}\\ &= \ell\left(\sum_{k=\nu(n)}^{n} k^{\ell-2}\omega\left(\frac{n}{k}\right)\right)\big(1 + O(\nu(n)^{-\epsilon})\big).\end{aligned} \tag{10}$$

The sum within (10) is a Riemann sum estimated by its corresponding integral

$$\begin{aligned}\sum_{k=\nu(n)}^{n} k^{\ell-2}\omega\left(\frac{n}{k}\right) &= \int_0^n t^{\ell-2}\omega\left(\frac{n}{t}\right)\mathrm{d}t + O\left(\frac{1}{n}\right)\\ &= n^{\ell-1}\int_1^\infty \frac{\omega(x)}{x^\ell}\mathrm{d}x + O\left(\frac{1}{n}\right) \quad \text{with } \frac{n}{t} = x.\end{aligned}$$

The proof for the case $\ell = 1$ is quite similar, and the range $\nu(n) \le k \le n$ is divided further into two ranges $\nu(n) \le k < n\mu(u)$ and $n\mu(n) \le k \le n$ where $\mu(n)$, for some well-chosen function μ, is defined as in [11]. □

Remark 2. The sum in (10) goes up to n inclusively and not $n/2$; thus, the range of integration starts at 1 and not 2. Because $\mathbf{P}\{X_n = k\} = 0$ for $\lfloor n/2 \rfloor + 1 \le k \le n-1$, we also point out that

$$\mathbf{P}\{X_n \ge k\} = \sum_{i=k}^{n} \mathbf{P}\{X_n = i\} = \mathbf{P}\{X_n = n\} \quad \text{for } \lfloor n/2 \rfloor + 1 \le k \le n.$$

Going back to the variance of X_n, we have the following theorem that ends our section on the analytical estimation for $\operatorname{Var}(X_n)/n$ as $n \to \infty$.

Theorem 2. *For some $\epsilon > 0$ independent of n, we have that*

$$\operatorname{Var}(X_n) = nC\left(1 + O\left(\frac{1}{n^{\epsilon}}\right)\right) \quad \textit{with} \quad C = 2\int_1^{\infty} \frac{\omega(x)}{x^2}\mathrm{d}x.$$

Proof. We have by definition that $\operatorname{Var}(X_n) = \mathbf{E}(X_n^2) - \big(\mathbf{E}(X_n)\big)^2$. We use (8) and consider the second moment. Hence we have

$$\begin{aligned}
\mathbf{E}(X_n^2) &= \sum_{k=1}^{\infty} \big(k^2 - (k-1)^2\big)\mathbf{P}\{X_n \ge k\} = \sum_{k=1}^{\infty} \big(2k-1\big)\mathbf{P}\{X_n \ge k\} \\
&= \sum_{k=1}^{n} \big(2k-1\big)\mathbf{P}\{X_n \ge k\} \\
&= \sum_{k=1}^{n} \Big(2k-1\Big)\left(\frac{1}{k}\omega\left(\frac{n}{k}\right) + O\left(\frac{1}{k^{1+\epsilon}}\right)\right) \quad \text{for some } \epsilon > 0 \\
&\sim 2\sum_{k=1}^{n} \omega\left(\frac{n}{k}\right). \qquad (11)
\end{aligned}$$

The expression (11) is a Riemann sum estimated in a way similar to that of Theorem 1. The quantity $\big(\mathbf{E}(X_n)\big)^2$ is negligible compared to $\mathbf{E}(X_n^2)$ as $n \to \infty$. Hence, we have that

$$\operatorname{Var}(X_n) \sim 2n\int_1^{\infty} \frac{\omega(x)}{x^2}\mathrm{d}x \quad \text{as } n \to \infty.$$

Reference [14] proves that $\omega(x) \to e^{-\gamma}$ where γ is the Euler-Mascheroni constant. More specifically, it proves that $|\omega(x) - e^{-\gamma}| < 10^{-4}$ for $x > 4$. Therefore, we have that

$$C = 2\int_1^{\infty} \frac{\omega(x)}{x^2}\mathrm{d}x = 2\int_1^{4} \frac{\omega(x)}{x^2}\mathrm{d}x + 2\int_4^{\infty} \frac{e^{-\gamma}}{x^2}\mathrm{d}x + 2\int_4^{\infty} \frac{\omega(x) - e^{-\gamma}}{x^2}\mathrm{d}x.$$

Using the quantities from [11] for

$$2\int_2^\infty \frac{\omega(x)}{x^2}\mathrm{d}x = 0.5586\ldots,$$

and, this time, taking into account the evaluation of the integral over $[1,2]$ that yields exactly $3/4$, we obtain up to four significant figures that $C = 1.3068\ldots$, and thus

$$\frac{\mathrm{Var}(X_n)}{n} \to 1.3068\ldots \quad \text{as } n \to \infty.$$

The proof is now complete. □

2.2. Numerical Integration

We use an idea from [8] in Theorem 3 to evaluate $\omega(x)$, for any $x \geq 1$, with arbitrary finite precision. We use Theorem 3 to evaluate C. The quantity n in this section is not the same as previously stated, which stood for the number of elements considered in our combinatorial object, while n here stands for the integral part of a real number, as is standard in numerical approximations.

We recall that we need to evaluate

$$C = 2\int_1^\infty \frac{\omega(t)}{t^2}\mathrm{d}t = \lim_{n\to\infty} \frac{\mathrm{Var}(X_n)}{n}. \tag{12}$$

For notational simplicity, we use $f : [1,\infty) \to [0,1]$ to denote the function $x \mapsto \omega(x)/x^2$. As mentioned previously, $|\omega(x) - e^{-\gamma}| < 10^{-4}$ for $x > 4$, and f is therefore bounded. The function f is also continuous because it is the composition of two continuous functions on $[1,\infty)$. We have that $f(x) \to 0$ as $x \to \infty$. Hence, the Riemann sum of f is convergent. We can approximate numerically its Riemann sum, that is, $\int_1^\infty f(t)\mathrm{d}t$, up to a desired accuracy by truncating the integral; this is because $f(x) \to 0$.

A popular method to approximate an integral is the trapezoidal method with a regular grid of points. Consider the interval $[1, n^*]$ where $n^* \in \mathbb{N}$ shall be determined later. Given the nature of ω (and so f), we consider for now an interval of the form $[n, n+1]$ where $n \in \mathbb{N}$. A point from a regular grid on $[n, n+1]$ can be put conveniently into the form $x_i = n + i\delta$ for $0 \leq i \leq \ell$ where $\delta = 2^{-\ell}$. We therefore have that

$$\sum_{i=0}^{2^\ell - 1} \delta \frac{\big(f(n+i\delta) + f(n+(i+1)\delta)\big)}{2} \to \int_n^{n+1} f(t)\mathrm{d}t \quad \text{as } \ell \to \infty. \tag{13}$$

To evaluate C with four significant digits, we can select $n^* = 10000$ and $\ell = 14$ so that $\delta < 10^{-4}$ using, for instance, the sharp bounds on numerical integration from [4]. Now, it remains to know how to compute numerically $\omega(x)$ for $x \geq 1$, which we do using the Taylor series as given by Theorem 3.

Theorem 3. *Consider the Taylor expansions of ω with respect to the z variable for each unit length interval of the form $[n, n+1)$. More precisely let*

$$\omega\left(n+\frac{1+z}{2}\right)=\sum_{i=0}^{\infty} c_{n,i} z^i \quad \textit{for } n \geq 1 \textit{ and for } -1 \leq z < 1.$$

Let $c_{n,i}$ the i-th term for the n-th sequence $\mathbf{c}_n$ for $n \geq 1$ and $i \geq 0$. Then we have

$$\begin{aligned}
c_{1,i} &= \frac{2}{3}\left(\frac{-1}{3}\right)^i, \\
c_{n+1,0} &= \frac{1}{2n+3}\sum_{i=0}^{\infty} c_{n,i}\left(2(n+1)+\frac{(-1)^i}{i+1}\right) \quad \textit{for } n > 1, \\
c_{n+1,i} &= \frac{1}{2n+3}\left(\frac{c_{n,i}}{n} - c_{n+1,i-1}\right) \quad \textit{for } n > 1 \textit{ and } i \geq 1.
\end{aligned}$$

Proof. Let $n \geq 1$ and let $x = n + t \geq 1$ with $n = \lfloor x \rfloor$ and $0 \leq t < 1$. If ω has a Taylor expansion in $[n, n+1)$, that is, it has the coefficients $c_{n,i}$, then we obtain the coefficients $c_{n+1,i}$ of the Taylor expansion in $[n+1, n+2)$ as follows. We integrate the difference-differential equation (2) and have that

$$\begin{aligned}
\int_{u=n+1}^{u=n+1+t} \mathrm{d}(u\omega(u)) &= (n+1+t)\omega(n+1+t) - (n+1)\omega(n+1) \\
&= \int_{u=n+1}^{u=n+1+t} \omega(u-1)\mathrm{d}u \\
&= \int_{x=0}^{x=t} \omega(n+x)\mathrm{d}x, \quad \text{with } u = n+1+x.
\end{aligned}$$

The affine transformation $t = z = 2t + 1$ transforms the fractional part $t \in [0, 1)$ into a centered-around-0 value $z \in [-1, 1)$. Equivalently, $t = (z+1)/2$, and we have that

$$\left(n+1+\frac{z+1}{2}\right)\omega\left(n+1+\frac{z+1}{2}\right) - (n+1)\omega(n+1) \tag{14}$$

$$= \int_{v=0}^{v=(z+1)/2} \omega(n+1+v)\mathrm{d}v$$

$$= \frac{1}{2}\int_{u=-1}^{u=z} \omega\left(n+\frac{u+1}{2}\right)\mathrm{d}u \quad \text{with } v = \frac{u+1}{2}. \tag{15}$$

Using the Taylor expansion around $u = 0$ of ω in the interval $[n, n+1)$ in terms of the dummy variable of integration, we have

$$\omega\left(n+\frac{u+1}{2}\right) = \sum_{i=0}^{\infty} c_{n,i} u^i \quad \text{for } -1 \leq u \leq z < 1. \tag{16}$$

Hence by substituting (16) into (15):

$$\int_{u=-1}^{u=z} \omega\Big(n + \frac{u+1}{2}\Big)\mathrm{d}u = \int_{-1}^{z} \sum_{i=0}^{\infty} c_{n,i} u^i \mathrm{d}u = \sum_{i=0}^{\infty} c_{n,i} \frac{\big(z^{i+1} - (-1)^{i+1}\big)}{i+1}. \qquad (17)$$

By continuity of ω, we have also that

$$\lim_{z\to 1} \omega\Big(n + \frac{z+1}{2}\Big) = \omega(n+1) = \lim_{z\to 1} \sum_{i=0}^{\infty} c_{n,i} z^i = \sum_{i=0}^{\infty} c_{n,i}. \qquad (18)$$

Using the Taylor expansion around $z = 0$ of ω in the interval $[n+1, n+2)$, we obtain

$$\omega\Big(n + 1 + \frac{z+1}{2}\Big) = \sum_{i=0}^{\infty} c_{n+1,i} z^i \quad \text{for } -1 \le z < 1.$$

Then substituting (18) into (14) and (17) into (15) yields:

$$(2n+3+z) \sum_{i=0}^{\infty} c_{n+1,i} z^i = 2(n+1) \sum_{i=0}^{\infty} c_{n,i} + \sum_{i=0}^{\infty} c_{n,i} \frac{\big(z^{i+1} - (-1)^{i+1}\big)}{i+1}. \qquad (19)$$

Substituting $z = 0$ in (19), we get

$$c_{n+1,0} = \frac{1}{2n+3} \sum_{i=0}^{\infty} c_{n,i} \Big(2(n+1) + \frac{(-1)^i}{i+1}\Big). \qquad (20)$$

By using (20) and equating coefficients with the same power of z, we find $c_{n+1,i}$ for $i \ge 1$:

$$\begin{aligned}(2n+3+z)c_{n+1,0} + (2n+3+z) \sum_{i=1}^{\infty} c_{n+1,i} z^i \\ = 2(n+1) \sum_{i=0}^{\infty} c_{n,i} + \sum_{i=0}^{\infty} c_{n,i} \frac{\big(z^{i+1} + (-1)^{i}\big)}{i+1},\end{aligned}$$

$$c_{n+1,0} z + (2n+3+z) \sum_{i=1}^{\infty} c_{n+1,i} z^i = c_{n,0} z + \sum_{i=1}^{\infty} c_{n,i} \frac{z^{i+1}}{i+1},$$

and

$$\begin{aligned}(2n+3+z) \sum_{i=1}^{\infty} c_{n+1,i} z^i \\ = (2n+3)c_{n+1,1} z + (2n+3) \sum_{i=2}^{\infty} c_{n+1,i} z^i + \sum_{i=1}^{\infty} c_{n+1,i} z^{i+1}.\end{aligned}$$

The previous equation holds if and only if

$$\big((2n+3)c_{n+1,i} + c_{n+1,i-1}\big)z^i = \frac{c_{n,i-1}z^i}{i} \quad \text{for all } i \geq 1.$$

We finally find the Taylor expansion of $1/x$ around $x = 1$ with $1 \leq x = 1 + t \leq 2$ and $t = (1+z)/2$ for $-1 \leq z < 1$, and have

$$\omega\left(1 + \frac{1+z}{2}\right) = \frac{2}{3}\frac{1}{(1+(z/3))} = \frac{2}{3}\sum_{i=0}^{\infty}\left(\frac{-1}{3}\right)^i z^i = \sum_{i=0}^{\infty} c_{1,i}z^i.$$

The proof is now complete. □

We point out that the centered-around-0 flavor of the Taylor expansion with coefficients $\mathbf{c}_n$ allows faster convergence around the endpoints n and $n+1$; see [8]. We compute the first n^* sequences with their first J terms, provided we have a library that does real arithmetic with finite and arbitrary precision.

Algorithm 1 Trapezoidal rule by using Taylor coefficient of the Buchstab function on the interval $[n, n+1)$ for $n \in \mathbb{N}$

Input: ℓ, n, $\{c_{n,j}\}_{j=0}^{J}$
Output: s, the sum from (13).

$\delta \leftarrow 2^{-\ell}$
$s \leftarrow 0$
for $i = 0$ **to** $2^\ell - 1$ **do**
    $y_0 \leftarrow 0$
    $y_1 \leftarrow 1$
    $t_0 \leftarrow i\delta$
    $t_1 \leftarrow (i+1)\delta$
    $z_0 \leftarrow 1$
    $z_1 \leftarrow 1$
    for $j = 0$ **to** J **do**
        $y_0 \leftarrow y_0 + c_{n,j}z_0$
        $y_1 \leftarrow y_1 + c_{n,j}z_1$
        $z_0 \leftarrow z_0(2t_0 - 1)$
        $z_1 \leftarrow z_1(2t_1 - 1)$
    end for
    $s \leftarrow s + \frac{y_0}{(n+t_0)^2} + \frac{y_1}{(n+t_1)^2}$
end for
$s \leftarrow \frac{s\delta}{2}$

To obtain C, we iteratively call Algorithm 1 for values of $n = 1, 2, \ldots, n^*$ with the coefficients for the Taylor expansion of ω on the interval $[n, n+1)$. We add the result of all iterations together and obtain $C = 1.3070\ldots$, which confirms comfortably the estimation from Section 2.1.

We end this section with comments about Algorithm 1. We have from line (7) that $t_1 = t_0 + \delta$. The loop at line (10) computes the Taylor polynomial of degree J of the Buchstab function $\omega(n + (1+z)/2)$ for the specific values of $z = z_0$ and $z = z_1$. During the j-th iteration at lines (11) and (12), we have that $y_b = \sum_{k=0}^{j} c_{n,k} z_b^k$ for $b = 0$ and $b = 1$, respectively. Lines (13) and (14) are for updating z_0 and z_1, respectively, for the next iteration, that is, the $(j+1)$-th iteration. We recall the meaning of the left side of the limiting expression in Equation (13) is that the height of a rectangle is $\big(f(n + i\delta) + f(n + (i+1)\delta\big)/2$ with $f(x) = \omega(x)/x^2$ in our case, and its length is δ; therefore, line (16) sums over the heights of all the rectangles. Averaging two consecutive heights by 2 is carried out only once at line (18) to save a few operations. Similarly, we take into account the length δ, which is identical for each rectangle, only once at line (18).

2.3. Recurrence Relation

We compute the probability distribution of X_n and then compute $\mathrm{Var}(X_n)$ for values of $n = 1, 2, \ldots, 4000$. Recalling (1), we have that

$$\mathrm{Var}(X_n) = \sum_{k=1}^{n} \big(k - \mathbf{E}(X_n)\big)^2 \, \mathbf{P}\{X_n = k\}.$$

Because

$$\mathbf{E}(X_n) = \sum_{k=1}^{n} k\mathbf{P}\{X_n = k\} \quad \text{and} \quad \mathbf{P}\{X_n = k\} = \frac{s_{k,n}}{n!},$$

the variance is a rational number, which is suitable to control the accuracy, as follows:

$$\frac{n! \sum_{k=1}^{n} k^2 s_{n,k} - \Big(\sum_{k=1}^{n} k s_{n,k}\Big)^2}{(n!)^2}.$$

We divide the quantity $\mathrm{Var}(X_n)$ by n in order to normalize it. We recall that $\mathrm{Var}(X_n) = C(n + O(n^{-\epsilon}))$ for some $\epsilon > 0$. When computing exactly $\mathrm{Var}(X_n)$ for a fixed n and comparing with the asymptotic formula, one would need the hidden factor of $n^{-\epsilon}$ and the value ϵ itself in order make a fair comparison; we nevertheless obtain numbers that are very close to the numbers from Sections 2.1 and 2.2:

$$\frac{\mathrm{Var}(X_{1000})}{1000} = 1.3004\ldots, \quad \frac{\mathrm{Var}(X_{2000})}{2000} = 1.3036\ldots,$$
$$\frac{\mathrm{Var}(X_{3000})}{3000} = 1.3047\ldots, \quad \frac{\mathrm{Var}(X_{4000})}{4000} = 1.3053\ldots.$$

The memory size on the machines available to us is the main limitation; however, it is enough to assert C up to two significant digits. A space of $12.7GB$ is needed to compute the triangular table for $n = 4000$. Storing the values in a triangular array allows us to compute the recurrence relation easily. Trimming the array of potentially unused cells is tough as n grows. Each array cell holds $s_{n,k}$ for a pair (n, k). The values $s_{n,k}$ are given by (6). We could compress the array slightly for $s_{n,k}$ when $\lfloor n/2 \rfloor + 1 \le k \le n - 1$ using methods described in [10] for instance, but we would not gain much for large values of n (like $n > 1000$) in terms of space and would yield a more complicated code. A possible algorithm for counting the $s_{n,k}$ is as in Algorithm 2.

Algorithm 2 Computing $s_{n,k}$

Input: N

Output: $s_{n,k}$ for $1 \le n \le N$ and $1 \le k \le n$

1: $s_{0,0} \leftarrow 1$
2: **for** $n = 1$ **to** N **do**
3:     $s_{n,0} \leftarrow 0$
4:     $s_{n,n} \leftarrow (n-1)!$
5: **end for**
6: **for** $n = 2$ **to** N **do**
7:     **for** $k = 1$ **to** $\lfloor n/2 \rfloor$ **do**
8:         $t_1 \leftarrow 0$
9:         **for** $i = 1$ **to** $\lfloor n/k \rfloor$ **do**
10:             $u_1 \leftarrow 0$
11:             **for** $j = k + 1$ **to** $n - ki$ **do**
12:                 $u_1 \leftarrow u_1 + s_{n-ki,j}$
13:             **end for**
14:             **if** $k + 1 \le n - ki$ **then**
15:                 $u_1 \leftarrow u_1 \frac{n!}{i!k^i(n-ki)!}$
16:             **end if**
17:             $t_1 \leftarrow t_1 + u_1$
18:         **end for**
19:         $t_2 \leftarrow 0$
20:         **if** k divides n **then**
21:             $t_2 \leftarrow \frac{n!}{(n/k)!k^{n/k}}$
22:         **end if**
23:         $s_{n,k} \leftarrow t_1 + t_2$
24:     **end for**
25: **end for**

We make just a few comments about Algorithm 2. From a data structure point of view, $n = 0$ and $k = 0$ are boundaries for the array, and lines (1) and (3) define the programming boundaries but are not part of the combinatorial objects and their related probability distributions a fortiori. The loop at line (7) runs up to $\lfloor n/2 \rfloor$ because we assume that $s_{n,k}$ are initialized to 0 by default for all valid n and k; this is usually the case in most advanced programming languages when declaring data structures.

We end this section with a small example. Table 1 shows $s_{n,k}$ for $1 \le n \le 10$.

	k									
n	1	2	3	4	5	6	7	8	9	10
10	2293839	525105	223200	151200	72576	0	0	0	0	362880
9	229384	52632	22400	18144	0	0	0	0	40320	
8	25487	5845	2688	1260	0	0	0	5040		
7	3186	714	420	0	0	0	720			
6	455	105	40	0	0	120				
5	76	20	0	0	24					
4	15	3	0	6						
3	4	0	2							
2	1	1								
1	1									

Table 1: Values of $s_{n,k}$ for $1 \le n \le 10$.

3. Generalized Buchstab Function

We recall the definition of the generalized Buchstab function with parameter $K > 0$, which is

$$\Omega_K(x) = \begin{cases} 1 & \text{for } 1 \le x < 2, \\ 1 + K \int_2^x \frac{\Omega_K(u-1)}{u-1} \mathrm{d}u & \text{for } x \ge 2. \end{cases} \tag{21}$$

The values of $1/\Omega_K(x)$ are asymptotic proportions of the large smallest components, as proved in [1]. More precisely, we recall that $s_{n,k}$, given as in (5) of Section 1, is the number of combinatorial n-objects with their smallest components having length k. For instance, classes of objects with parameter $K = 1/2$ include 2-regular graphs, surjective maps, etc. Classes of objects with parameter $K = 1$ include derangements, permutations, monic polynomials over a finite field, etc. The quantity $\sum_{i=k}^{n} s_{n,i}$ is the number of n-objects for which the smallest component has a size of at least k for $1 \le k \le n$. Let $x > 1$ and consider the ratio

$$\frac{s_{\lfloor xn \rfloor, \lfloor xn \rfloor}}{\sum_{i=n}^{\lfloor xn \rfloor} s_{\lfloor xn \rfloor, i}}. \tag{22}$$

Then it is shown in [1] that, for $x > 1$,

$$\lim_{n\to\infty} \frac{s_{\lfloor xn \rfloor, \lfloor xn \rfloor}}{\sum_{i=n}^{\lfloor xn \rfloor} s_{\lfloor xn \rfloor, i}} = \frac{1}{\Omega_K(x)}. \tag{23}$$

The limiting quantity (23) justifies our interest in evaluating the generalized Buchstab function.

We remark that from now on and up to Table 2 inclusively, the symbol n no longer refers to the size of a combinatorial object.

Following the ideas exposed in Section 2.2, let $n \geq 1$ be a natural number, and let $c_{n,i}$ be i-th coefficient of the Taylor expansion for $\Omega_K(z)$ in the interval $[n, n+1)$ with $-1 \leq z < 1$. More precisely, let

$$\Omega_K\Big(n + \frac{1+z}{2}\Big) = \sum_{i=0}^{\infty} c_{n,i} z^i \quad \text{for } -1 \leq z < 1. \tag{24}$$

As expected, the sequence $(c_{n,i})_{i\geq 0}$ depends on the previous sequence $(c_{n-1,i})_{i\geq 0}$ for $n > 2$. Our library can compute over $\mathbb{R}$ with arbitrary finite precision. The variable z in (24) is the fractional part of $x \in [n, n+1)$ centered around 0.

Theorem 4. *For $K > 0$, consider the Taylor expansions of Ω_K with respect to the z variable for each unit length interval of the form $[n, n+1)$. More precisely, let*

$$\Omega_K\Big(n + \frac{1+z}{2}\Big) = \sum_{i=0}^{\infty} c_{n,i} z^i \quad \textit{for } n \geq 1 \textit{ and for } -1 \leq z < 1.$$

For $n \geq 1$ and $i \geq 0$, and let α_i be defined by

$$\alpha_i = \sum_{j=0}^{i} \frac{(-1)^{i-j}}{(2n-1)^{i-j}} c_{n-1,j}.$$

Then we have

$$\begin{aligned}
c_{1,0} &= 1,\\
c_{1,i} &= 0 \quad \textit{for } i \geq 1,\\
c_{2,0} &= c_{2,0} = 1 + K \sum_{i=1}^{\infty} \frac{1}{i2^i},\\
c_{2,i} &= K \sum_{j=i}^{\infty} \frac{(-1)^{j-1}}{j2^j} \binom{j}{i} \quad \textit{for } i \geq 1,\\
c_{n,0} &= \sum_{i=0}^{\infty} c_{n-1,i} - \frac{K}{2n-1} \sum_{i=0}^{\infty} \frac{(-1)^{i+1}\alpha_i}{i+1} \quad \textit{for } n \geq 3,\\
c_{n,i} &= \frac{K\alpha_{i-1}}{(2n-1)i} \quad \textit{for } n \geq 3 \textit{ and } i \geq 1.
\end{aligned}$$

Proof. For $x \in [1,2)$, the function Ω_K is constant and so $c_{1,0} = 1$ and $c_{1,i} = 0$ for $i \geq 1$.

For $2 \leq x = 2 + ((1+z)/2) < 3$, the coefficients of the Taylor expansion are $1 + K\log(2 + (1+z)/2)$; hence the coefficients are given by

$$c_{2,0} = 1 + K\sum_{j=1}^{\infty} \frac{1}{j2^j} \quad \text{and} \quad c_{2,i} = K\sum_{j=i}^{\infty} \frac{(-1)^{j-1}}{j2^j}\binom{j}{i} \quad \text{for } i \geq 1. \tag{25}$$

Given $x \geq 3$ such that $x = n + ((z+1)/2)$ and $n \geq 3$, we assume that the sequence $(c_{n-1,i})_{i\geq 0}$ is known. We have

$$\begin{aligned}
\Omega_K\left(n + \left(\frac{1+z}{2}\right)\right) &= \sum_{i=0}^{\infty} c_{n,i} z^i \\
&= 1 + K\int_2^{n+(1+z)/2} \frac{\Omega_K(u-1)}{u-1}\mathrm{d}u \\
&= 1 + K\int_2^{n} \frac{\Omega_K(u-1)}{u-1}\mathrm{d}u + K\int_n^{n+(1+z)/2} \frac{\Omega_K(u-1)}{u-1}\mathrm{d}u \\
&= \Omega_K(n) + K\int_{u=n}^{u=n+(1+z)/2} \frac{\Omega_K(u-1)}{u-1}\mathrm{d}u \\
&= \Omega_K(n) + K\int_{v=-1}^{v=z} \frac{\Omega_K(n-1+(v+1)/2)}{2n-1+v}\mathrm{d}v \quad \text{with } v = 2u-2n-1 \\
&= \Omega_K(n) + \frac{K}{2n-1}\int_{u=-1}^{u=z} \left(\sum_{i=0}^{\infty} c_{n-1,i}u^i\right)\left(\sum_{i=0}^{\infty} \frac{(-1)^i u^i}{(2n-1)^i}\right)\mathrm{d}u \\
&= \Omega_K(n) + \frac{K}{2n-1}\int_{u=-1}^{u=z} \sum_{i=0}^{\infty}\left(\sum_{j=0}^{i} \frac{(-1)^{i-j}}{(2n-1)^{i-j}} c_{n-1,j}\right)u^i\mathrm{d}u \\
&= \Omega_K(n) + \frac{K}{2n-1}\int_{u=-1}^{u=z} \sum_{i=0}^{\infty} \alpha_i u^i\mathrm{d}u \\
&= \Omega_K(n) - \frac{K}{2n-1}\sum_{i=0}^{\infty} \frac{(-1)^{i+1}\alpha_i}{i+1} + \frac{K}{2n-1}\sum_{i=0}^{\infty} \frac{\alpha_i z^{i+1}}{i+1}.
\end{aligned} \tag{26}$$

The continuity of Ω_K implies that

$$\Omega_K(n) = \lim_{z\to 1} \Omega_K\left(n - 1 + \frac{1+z}{2}\right) = \lim_{z\to 1}\sum_{i=0}^{\infty} c_{n-1,i}z^i = \sum_{i=0}^{\infty} c_{n-1,i}.$$

Hence (26) is rewritten as

$$\Omega_K\left(n + \frac{1+z}{2}\right) = \sum_{i=0}^{\infty} c_{n-1,i} - \frac{K}{2n-1}\sum_{i=0}^{\infty} \frac{\alpha_i(-1)^{i+1}}{i+1} + \frac{K}{2n-1}\sum_{i=0}^{\infty} \frac{\alpha_i z^{i+1}}{i+1}$$

$$= c_{n,0} + \sum_{i=1}^{\infty} \frac{K\alpha_{i-1}}{(2n-1)i} z^i = c_{n,0} + \sum_{i=1}^{\infty} c_{n,i} z^i.$$

The proof is complete. □

For instance, by reading $\Omega_1(2^{13})$ from the left half of Table 2 and recalling (22), the proportion of random permutations on at least 2^{14} elements, and with a cycle of smallest length at least 2^{13}, is close to $1/\Omega_1(2^{13}) \approx 0.000218$. We note that if the number of permuted elements is precisely 2^{14}, then there will be no smallest component of size at least 2^{13}; one can observe this from the recurrence relation in Section 2.3 as well.

Similarly, by reading $\Omega_{1/2}(2^{13})$ from the right half of Table 2 and recalling (22), the proportion of random 2-regular graphs with at least 2^{14} vertices, and with a large smallest component of at least 2^{13}, is close to $1/\Omega_{1/2}(2^{13}) \approx 0.0131$. We note that if the number of vertices is exactly 2^{14}, then there will be no smallest component of size at least 2^{13}.

$K=1$				$K=1/2$			
x	$\Omega_K(x)$	x	$\Omega_K(x)$	x	$\Omega_K(x)$	x	$\Omega_K(x)$
1	1	16	8.9874	1	1	16	3.3302
2	1	32	17.9749	2	1	32	4.7470
3	1.6941	64	35.9498	3	1.3470	64	6.7397
4	2.2468	128	71.8997	4	1.5866	128	9.5501
5	2.8085	256	143.7995	5	1.7971	256	13.5191
6	3.3703	512	287.5991	6	1.9856	512	19.1282
7	3.9320	1024	575.1983	7	2.1579	1024	27.0580
8	4.4937	2048	1150.3966	8	2.3175	2048	38.2705
9	5.0554	4096	2300.7932	9	2.4669	4096	54.1260
10	5.6171	8192	4567.8834	10	2.6077	8192	76.5480

Table 2: A few values of $\Omega_K(x)$ for $K = 1$ and $K = 1/2$.

We conclude this section by mentioning that [5] gives values for $1/\Omega_K(x)$ with $x = 2, 3, 4, 5$ and that, if we invert values from Table 2 for $x = 2, 3, 4, 5$, they agree with those from [5].

4. Conclusion

In this paper, we computed the normalization constant of the variance of the distribution of the smallest component of random combinatorial objects. We used different approaches: an analytic method based on the singularity analysis for generating functions, a numerical integration method using Taylor expansions for the Buchstab function, and the recurrence relation for counting the number of smallest components. All the methods yield $1.3070\ldots$. We also showed how to compute the value of the generalized Buchstab function by recursively building sequences of Taylor expansions for each unit interval of the form $[n, n+1)$ where $n \in \mathbb{N} \setminus \{0\}$. We can compute the asymptotic proportion of large smallest components for various random combinatorial objects by obtaining very accurate values of the generalized Buchstab function.

Acknowledgements. D. Panario is partially funded by the Natural Sciences and Engineering Research Council of Canada, reference number RPGIN-2018-05328. The authors thank an anonymous referee for suggestions and corrections that improved the paper.

References

[1] E. A. Bender, A. Mashatan, D. Panario, and B. Richmond, Asymptotics of combinatorial structures with large smallest component, *J. Combin. Theory Ser. A* **107**(1) (2004), 117-125.

[2] N. G. de Bruijn, On the number of uncancelled elements in the sieve of Eratosthenes, in *Proceedings of Koninklijke Nederlandse Akademie van Wetenschappen: Series A: Mathematical Sciences* **53**(5-6) (1950), 803-812.

[3] A. A. Buchstab, Asymptotic estimates of a general number-theoretic function, *Sb. Math.* **44** (1937), 1239-1246. (In Russian).

[4] D. Cruz-Uribe and C. J. Neugebauer, Sharp error bounds for the trapezoidal rule and Simpson's rule, *J. Inequal. Pure Appl.* **3** (2002), 1-22.

[5] S. Finch, Permute, graph, map, derange, preprint, `arXiv: 2111.05720`.

[6] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, New York, 2009.

[7] P. Flajolet, X. Gourdon and D. Panario, The complete analysis of a polynomial factorization algorithm over finite fields, *J. Algorithms* **40**(1) (2001), 37-81.

[8] G. Marsaglia, A. Zaman, and J. C. W. Marsaglia, Numerical solution of some classical differential-difference equations, *Math. Comp.* **53** (1989), 191-201.

[9] G. L. Mullen and D. Panario, *Handbook of Finite Fields*, Chapman & Hall/CRC, New York, 2013.

[10] G. Navarro, *Compact Data Structures: A Practical Approach*, Cambridge University Press, New York, 2016.

[11] D. Panario and B. Richmond, Analysis of Ben-Or's polynomial irreducibility test, *Random Structures Algorithms* **13**(3-4) (1998), 439-456.

[12] D. Panario and B. Richmond, Exact largest and smallest size of components, *Algorithmica* **31** (2001), 413-432.

[13] D. Panario and B. Richmond, Smallest components in decomposable structures: Exp-log class, *Algorithmica* **29**(1–2) (2001), 205–226.

[14] A. Selberg, The number of cancelled elements in the sieve of Eratosthenes, *Nordisk Matematisk Tidsskrift* **26** (1944), 79-84. (In Norwegian).

[15] G. Tenenbaum, *Introduction à la Théorie Analytique et Probabiliste des Nombres*, Belin, Paris, 2014.