Analysis of the Expected Number of Bit Comparisons Required by Quickselect

James Allen Fill[1]

Department of Applied Mathematics and Statistics

The Johns Hopkins University

jimfill@jhu.edu and http://www.ams.jhu.edu/~fill/

and

Takéhiko Nakama

Department of Applied Mathematics and Statistics

The Johns Hopkins University

nakama@jhu.edu and http://www.ams.jhu.edu/~nakama/

ABSTRACT

When algorithms for sorting and searching are applied to keys that are represented as bit strings, we can quantify the performance of the algorithms not only in terms of the number of key comparisons required by the algorithms but also in terms of the number of bit comparisons. Some of the standard sorting and searching algorithms have been analyzed with respect to key comparisons but not with respect to bit comparisons. In this paper, we investigate the expected number of bit comparisons required by `Quickselect` (also known as `Find`). We develop exact and asymptotic formulae for the expected number of bit comparisons required to find the smallest or largest key by `Quickselect` and show that the expectation is asymptotically linear with respect to the number of keys. Similar results are obtained for the average case. For finding keys of arbitrary rank, we derive an exact formula for the expected number of bit comparisons that (using rational arithmetic) requires only finite summation (rather than such operations as numerical integration) and use it to compute the expectation for each target rank.

AMS 2000 *subject classifications.* Primary 68W40; secondary 68P10, 60C05.

Key words and phrases. `Quickselect`, `Find`, searching algorithms, asymptotics, average-case analysis, key comparisons, bit comparisons.

Date. March 2, 2009.

[1]Research for both authors supported by NSF grant DMS–0406104, and by The Johns Hopkins University's Acheson J. Duncan Fund for the Advancement of Research in Statistics.

1 Introduction and Summary

When an algorithm for sorting or searching is analyzed, the algorithm is usually regarded either as comparing keys pairwise irrespective of the keys' internal structure or as operating on representations (such as bit strings) of keys. In the former case, analyses often quantify the performance of the algorithm in terms of the number of key comparisons required to accomplish the task; `Quickselect` (also known as `Find`) is an example of those algorithms that have been studied from this point of view. In the latter case, if keys are represented as bit strings, then analyses quantify the performance of the algorithm in terms of the number of bits compared until it completes its task. Digital search trees, for example, have been examined from this perspective.

In order to fully quantify the performance of a sorting or searching algorithm and enable comparison between key-based and digital algorithms, it is ideal to analyze the algorithm from both points of view. However, to date, only `Quicksort` has been analyzed with both approaches; see Fill and Janson [3]. Before their study, `Quicksort` had been extensively examined with regard to the number of key comparisons performed by the algorithm (e.g., Knuth [13], Régnier [19], Rösler [20], Knessl and Szpankowski [10], Fill and Janson [2], Neininger and Rüschendorf [17]), but it had not been examined with regard to the number of bit comparisons in sorting keys represented as bit strings. In their study, Fill and Janson assumed that keys are independently and uniformly distributed over (0,1) and that the keys are represented as bit strings. [They also conducted the analysis for a general absolutely continuous distribution over (0,1).] They showed that the expected number of bit comparisons required to sort n keys is asymptotically equivalent to $n(\ln n)(\lg n)$ as compared to the lead-order term of the expected number of *key* comparisons, which is asymptotically $2n \ln n$. We use ln and lg to denote natural and binary logarithms, respectively, and use log when the base does not matter (for example, in remainder estimates).

In this paper, we investigate the expected number of bit comparisons required by `Quickselect`. Hoare [8] introduced this search algorithm, which is treated in most textbooks on algorithms and data structures. `Quickselect` selects the m-th smallest key (we call it the rank-m key) from a set of n distinct keys. (The keys are typically assumed to be distinct, but the algorithm still works—with a minor adjustment—even if they are not distinct.) The algorithm finds the target key in a recursive and random fashion. First, it selects a pivot uniformly at random from n keys. Let k denote the rank of the pivot. If $k = m$, then the algorithm returns the pivot. If $k > m$, then the algorithm recursively operates on the set of keys smaller than the pivot and returns the rank-m key. Similarly, if $k < m$, then the algorithm recursively operates on the set of keys larger than the pivot and returns the $(k - m)$-th smallest key from the subset. Although previous studies (e.g., Knuth [11], Mahmoud *et al.* [15], Prodiner [18], Grübel and U. Rösler [7], Lent and Mahmoud [14], Mahmoud and Smythe [16], Devroye [1], Hwang and Tsai [9]) examined `Quickselect` with regard to key comparisons, this study is the first to analyze the bit complexity of the algorithm.

We suppose that the algorithm is applied to n distinct keys that are represented as bit strings and that the algorithm operates on individual bits in order to find a target key. We also assume that the n keys are uniformly and independently distributed in $(0, 1)$. For instance, consider applying `Quickselect` to find the smallest key among three keys k_1, k_2, and k_3 whose binary representations are .01001100..., .00110101..., and .00101010..., respectively. If the algorithm selects k_3 as a pivot, then it compares each of k_1 and k_2 to k_3 in order to determine the rank of k_3. When k_1 and k_3 are compared, the algorithm requires 2 bit comparisons to determine that k_3 is smaller than k_1 because the two keys have the same first digit and differ at the second digit. Similarly, when k_2 and k_3 are compared, the algorithm requires 4 bit comparisons to determine that k_3 is smaller than k_2. After these comparisons, key k_3 has been identified as smallest. Hence the search for the smallest key requires a total of 6 bit comparisons (resulting from the two key comparisons).

We let $\mu(m,n)$ denote the expected number of bit comparisons required to find the rank-m key in a file of n keys by `Quickselect`. By symmetry, $\mu(m,n) = \mu(n+1-m,n)$. First, we develop exact and asymptotic formulae for $\mu(1,n) = \mu(n,n)$, the expected number of bit comparisons required to find the smallest key by `Quickselect`, as summarized in the following theorem.

Theorem 1.1. *The expected number $\mu(1,n)$ of bit comparisons required by* `Quickselect` *to find the smallest key in a file of n keys that are independently and uniformly distributed in $(0,1)$ has the following exact and asymptotic expressions:*

$$\begin{aligned}\mu(1,n) &= 2n(H_n-1) + 2\sum_{j=2}^{n-1} B_j \frac{n-j+1-\binom{n}{j}}{j(j-1)(1-2^{-j})} && (1.1)\\ &= cn - \frac{1}{\ln 2}(\ln n)^2 - \left(\frac{2}{\ln 2}+1\right)\ln n + O(1), && (1.2)\end{aligned}$$

where H_n and B_j denote harmonic and Bernoulli numbers, respectively, and

$$c := 2\sum_{k=0}^{\infty}\left(1 + 2^{-k}\sum_{j=1}^{2^k}\ln\frac{j}{2^k}\right) \doteq 5.27938. \tag{1.3}$$

With $\chi_k := \frac{2\pi i k}{\ln 2}$ and $\gamma :=$ Euler's constant $\doteq 0.57722$, the constant c can alternatively be expressed as

$$c = \frac{28}{9} + \frac{17-6\gamma}{9\ln 2} - \frac{4}{\ln 2}\sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\zeta(1-\chi_k)\Gamma(1-\chi_k)}{\Gamma(4-\chi_k)(1-\chi_k)}. \tag{1.4}$$

The asymptotic formula shows that the expected number of bit comparisons is asymptotically linear in n with lead-order coefficient approximately equal to 5.27938. Hence the expected number of bit comparisons is asymptotically different from that of *key* comparisons required to find the smallest key only by a constant factor (the expectation for key comparisons is asymptotically $2n$). Details of the derivations of the formulae are described in Section 3.

Complex-analytical methods are utilized to obtain the asymptotic formula (1.2) with c in the form (1.4) and seem to be indispensable for obtaining asymptotics beyond the lead term. [We remark that, although it involves the imaginary numbers χ_k, the expression (1.4) is real because the terms with indices k and $-k$ are complex conjugates.] In Section 3.2 we again use complex-analytical methods to reexpress (1.4) in the form (1.3). Having done all this we suspected that there must be a purely real-analytical way to obtain directly the lead-order asymptotics $\mu(1,n) \sim cn$ with c in the form (1.3). Indeed, there is: See Remark 3.3.

In Sections 4 and 5 we move on to derive exact and asymptotic expressions for the expected number of bit comparisons for the average case. We denote this expectation by $\mu(\bar{m}, n)$. In the average case, the parameter m in $\mu(m, n)$ is considered a discrete uniform random variable; hence $\mu(\bar{m}, n) = \frac{1}{n}\sum_{m=1}^{n} \mu(m, n)$. The derived asymptotic formula shows that $\mu(\bar{m}, n)$ is also asymptotically linear in n; see (4.48). More detailed results for $\mu(\bar{m}, n)$ are described in Section 4.

Lastly, in Section 5, we derive an exact expression of $\mu(m, n)$ for each fixed m that is suited for computations. Our preliminary exact formula for $\mu(m, n)$ [shown in (2.8)] entails infinite summation and integration. As a result, it is not a desirable form for numerically computing the expected number of bit comparisons. Hence we establish another exact formula that only requires finite summation and use it to compute $\mu(m, n)$ for $m = 1, \ldots, n$, $n = 2, \ldots, 25$. The computation leads to the following conjectures: (i) for fixed n, $\mu(m, n)$ increases in m for $m \le \frac{n+1}{2}$ and is symmetric about $\frac{n+1}{2}$; and (ii) for fixed m, $\mu(m, n)$ increases in n (asymptotically linearly).

2 Preliminaries

To investigate the bit complexity of `Quickselect`, we follow the general approach developed by Fill and Janson [3]. Let $U_1, \ldots, U_n$ denote the n keys uniformly and independently distributed on (0, 1), and let $U_{(i)}$ denote the rank-i key. Then, for $1 \le i < j \le n$ (assume $n \ge 2$),

$$P\{U_{(i)} \text{ and } U_{(j)} \text{ are compared}\} = \begin{cases} \dfrac{2}{j-m+1} & \text{if } m \le i \\ \dfrac{2}{j-i+1} & \text{if } i < m < j \\ \dfrac{2}{m-i+1} & \text{if } j \le m. \end{cases} \tag{2.1}$$

To determine the first probability in (2.1), note that $U_{(m)}, \ldots, U_{(j)}$ remain in the same subset until the first time that one of them is chosen as a pivot. Therefore, $U_{(i)}$ and $U_{(j)}$ are compared if and only if the first pivot chosen from $U_{(m)}, \ldots, U_{(j)}$ is either $U_{(i)}$ or $U_{(j)}$. Analogous arguments establish the other two cases.

For $0 < s < t < 1$, it is well known that the joint density function of $U_{(i)}$ and $U_{(j)}$ is given by

$$f_{U_{(i)},U_{(j)}}(s,t) \;:=\; \binom{n}{i-1, 1, j-i-1, 1, n-j} s^{i-1}(t-s)^{j-i-1}(1-t)^{n-j}. \tag{2.2}$$

Clearly, the event that $U_{(i)}$ and $U_{(j)}$ are compared is independent of the random variables $U_{(i)}$ and $U_{(j)}$. Hence, defining

$$\begin{aligned}
P_1(s, t, m, n) &= \sum_{m \le i < j \le n} \frac{2}{j - m + 1} f_{U_{(i)}, U_{(j)}}(s, t), & (2.3) \\
P_2(s, t, m, n) &= \sum_{1 \le i < m < j \le n} \frac{2}{j - i + 1} f_{U_{(i)}, U_{(j)}}(s, t), & (2.4) \\
P_3(s, t, m, n) &= \sum_{1 \le i < j \le m} \frac{2}{m - i + 1} f_{U_{(i)}, U_{(j)}}(s, t), & (2.5) \\
P(s, t, m, n) &= P_1(s, t, m, n) + P_2(s, t, m, n) + P_3(s, t, m, n) & (2.6)
\end{aligned}$$

[the sums in (2.3)–(2.5) are double sums over i and j], and letting $\beta(s, t)$ denote the index of the first bit at which the keys s and t differ, we can write the expectation $\mu(m, n)$ of the number of bit comparisons required to find the rank-m key in a file of n keys as

$$\begin{aligned}
\mu(m, n) &= \int_0^1 \int_s^1 \beta(s, t) P(s, t, m, n) \, dt \, ds & (2.7) \\
&= \sum_{k=0}^{\infty} \sum_{l=1}^{2^k} \int_{(l-1)2^{-k}}^{(l-\frac{1}{2})2^{-k}} \int_{(l-\frac{1}{2})2^{-k}}^{l 2^{-k}} (k + 1) P(s, t, m, n) \, dt \, ds; & (2.8)
\end{aligned}$$

in this expression, note that k represents the last bit at which s and t agree.

3 Analysis of $\mu(1, n)$

In Section 3.1, we derive the exact expression for $\mu(1, n)$ shown in Theorem 1.1. In Section 3.2, we prove the asymptotic result stated in Theorem 1.1.

3.1 Exact Computation of $\mu(1, n)$

Since the contribution of $P_2(s, t, m, n)$ or $P_3(s, t, m, n)$ to $P(s, t, m, n)$ is zero for $m = 1$, we have $P(s, t, 1, n) = P_1(s, t, 1, n)$ [see (2.4) through (2.6)]. Let $x := s,\ y := t - s,\ z := 1 - t$. Then

$$\begin{aligned}
P_1(s, t, 1, n) &= z^n \sum_{1 \le i < j \le n} \frac{2}{j} \binom{n}{i - 1, 1, j - i - 1, 1, n - j} x^{i-1} y^{j-i-1} z^{-j} \\
&= 2 z^n \int_z^{\infty} \eta^{-n-1} \sum_{1 \le i < j \le n} \binom{n}{i - 1, 1, j - i - 1, 1, n - j} x^{i-1} y^{j-i-1} \eta^{n-j} \, d\eta \\
&= 2 z^n \int_z^{\infty} \eta^{-n-1} n(n - 1)(x + y + \eta)^{n-2} \, d\eta \\
&= 2 z^n n(n - 1) \int_z^{\infty} \eta^{-3} \left(\frac{t}{\eta} + 1 \right)^{n-2} \, d\eta. & (3.1)
\end{aligned}$$

Making the change of variables $v = \frac{t}{\eta} + 1$ and integrating, and recalling $z = 1 - t$, we find, after some calculation,

$$P_1(s,t,1,n) \quad = \quad 2\sum_{j=2}^{n}(-1)^j \binom{n}{j} t^{j-2}. \tag{3.2}$$

From (2.8) and (3.2),

$$\begin{aligned}
\mu(1,n) &= \sum_{k=0}^{\infty}(k+1)\sum_{l=1}^{2^k}\int_{(l-1)2^{-k}}^{(l-\frac{1}{2})2^{-k}}\int_{(l-\frac{1}{2})2^{-k}}^{l2^{-k}} P_1(s,t,1,n)\,dt\,ds \\
&= 2\sum_{k=0}^{\infty}(k+1)\sum_{l=1}^{2^k}\int_{(l-1)2^{-k}}^{(l-\frac{1}{2})2^{-k}}\int_{(l-\frac{1}{2})2^{-k}}^{l2^{-k}}\sum_{j=2}^{n}(-1)^j\binom{n}{j}t^{j-2}\,dt\,ds \\
&= 2\sum_{k=0}^{\infty}(k+1)\sum_{l=1}^{2^k}\sum_{j=2}^{n}(-1)^j\binom{n}{j}\int_{(l-\frac{1}{2})2^{-k}}^{l2^{-k}} t^{j-2}[(l-\tfrac{1}{2})2^{-k}-(l-1)2^{-k}]\,dt \\
&= \sum_{k=0}^{\infty}(k+1)\sum_{l=1}^{2^k}\sum_{j=2}^{n}\frac{(-1)^j\binom{n}{j}}{j-1}2^{-k}\{(l2^{-k})^{j-1}-[(l-\tfrac{1}{2})2^{-k}]^{j-1}\} \\
&= \sum_{j=2}^{n}\frac{(-1)^j\binom{n}{j}}{j-1}\sum_{k=0}^{\infty}(k+1)2^{-kj}\sum_{l=1}^{2^k}[l^{j-1}-(l-\tfrac{1}{2})^{j-1}].
\end{aligned} \tag{3.3}$$

To further transform (3.3), define

$$a_{j,r} = \begin{cases} \dfrac{B_r}{r}\dbinom{j-1}{r-1} & \text{if } r \geq 2 \\ \frac{1}{2} & \text{if } r = 1 \\ \frac{1}{j} & \text{if } r = 0, \end{cases} \tag{3.4}$$

where B_r denotes the r-th Bernoulli number. Let $S_{n,j} := \sum_{l=1}^{n} l^{j-1}$. Then $S_{n,j} = \sum_{r=0}^{j-1} a_{j,r} n^{j-r}$ (see Knuth [13]), and

$$\begin{aligned}
&\sum_{l=1}^{2^k}[l^{j-1}-(l-\tfrac{1}{2})^{j-1}] = S_{2^k,j} - 2^{-(j-1)}\sum_{l=1}^{2^k}(2l-1)^{j-1} \\
&= S_{2^k,j} - 2^{-(j-1)}(S_{2^{k+1},j} - 2^{j-1}S_{2^k,j}) = 2S_{2^k,j} - 2^{-(j-1)}S_{2^{k+1},j} \\
&= 2\sum_{r=0}^{j-1} a_{j,r}2^{k(j-r)} - 2^{-(j-1)}\sum_{r=0}^{j-1} a_{j,r}2^{(k+1)(j-r)} = 2\sum_{r=1}^{j-1} a_{j,r}2^{k(j-r)}(1-2^{-r}).
\end{aligned} \tag{3.5}$$

From (3.3) and (3.5),

$$\mu(1,n) \quad = \quad 2\sum_{j=2}^{n}\frac{(-1)^j\binom{n}{j}}{j-1}\sum_{k=0}^{\infty}(k+1)2^{-kj}\sum_{r=1}^{j-1}a_{j,r}2^{k(j-r)}(1-2^{-r}).$$

Here

$$\begin{aligned}\sum_{k=0}^{\infty}(k+1)2^{-kj}\sum_{r=1}^{j-1}a_{j,r}2^{k(j-r)}(1-2^{-r}) &= \sum_{k=0}^{\infty}(k+1)\sum_{r=1}^{j-1}a_{j,r}2^{-kr}(1-2^{-r})\\ &= \sum_{r=1}^{j-1}a_{j,r}(1-2^{-r})\sum_{k=0}^{\infty}(k+1)2^{-kr} = \sum_{r=1}^{j-1}a_{j,r}(1-2^{-r})^{-1}.\end{aligned}$$

Hence

$$\begin{aligned}\mu(1,n) &= 2\sum_{j=2}^{n}\frac{(-1)^j\binom{n}{j}}{j-1}\sum_{r=1}^{j-1}a_{j,r}(1-2^{-r})^{-1} = 2\sum_{r=1}^{n-1}(1-2^{-r})^{-1}\sum_{j=r+1}^{n}\frac{(-1)^j\binom{n}{j}}{j-1}a_{j,r}\\ &= 2\sum_{j=2}^{n}\frac{(-1)^j\binom{n}{j}}{j-1}+2\sum_{r=2}^{n-1}(1-2^{-r})^{-1}\frac{B_r}{r}\sum_{j=r+1}^{n}\frac{(-1)^j\binom{n}{j}\binom{j-1}{r-1}}{j-1}\\ &= 2\sum_{j=2}^{n}\frac{(-1)^j\binom{n}{j}}{j-1}+2\sum_{r=2}^{n-1}(1-2^{-r})^{-1}\frac{B_r}{r}\left[\sum_{j=r}^{n}\frac{(-1)^j\binom{n}{j}\binom{j-1}{r-1}}{j-1}-\frac{(-1)^r\binom{n}{r}}{r-1}\right].\end{aligned} \tag{3.6}$$

The sum $\sum_{j=r}^{n}\frac{(-1)^j\binom{n}{j}\binom{j-1}{r-1}}{j-1}$ can be simplified as follows:

$$\begin{aligned}\sum_{j=r}^{n}\frac{(-1)^j}{j-1}\binom{n}{j}\binom{j-1}{r-1} &= \frac{1}{r-1}\sum_{j=r}^{n}(-1)^j\binom{n}{j}\binom{j-2}{r-2} = \frac{1}{r-1}\sum_{j=r}^{n}(-1)^j\binom{n}{j}\binom{j-2}{j-r}\\ &= \frac{(-1)^r}{r-1}\sum_{j=0}^{n}\binom{n}{n-j}\binom{1-r}{j-r} = \frac{(-1)^r}{r-1}\binom{n+1-r}{n-r}\\ &= (-1)^r\frac{n+1-r}{r-1}.\end{aligned} \tag{3.7}$$

Plugging (3.7) into (3.6) and recalling $B_{2k+1}=0$ for $k\geq 1$, we finally obtain

$$\begin{aligned}\mu(1,n) &= 2\sum_{j=2}^{n}\frac{(-1)^j\binom{n}{j}}{j-1}+2\sum_{r=2}^{n-1}(1-2^{-r})^{-1}\frac{B_r}{r}\left[\frac{(-1)^r(n-r+1)}{r-1}-\frac{(-1)^r\binom{n}{r}}{r-1}\right]\\ &= 2\sum_{j=2}^{n}\frac{(-1)^j\binom{n}{j}}{j-1}+2\sum_{j=2}^{n-1}B_j\frac{n-j+1-\binom{n}{j}}{j(j-1)(1-2^{-j})}\\ &= 2n(H_n-1)+2t_n,\end{aligned} \tag{3.8}$$

where H_n denotes the n-th harmonic number and

$$t_n := \sum_{j=2}^{n-1}\frac{B_j}{j(1-2^{-j})}\left[\frac{n-\binom{n}{j}}{j-1}-1\right]. \tag{3.9}$$

The last equality in (3.8) follows from the easy identity

$$\sum_{k=1}^{n}\frac{(-1)^{k-1}\binom{n}{k}}{k} = H_n.$$

3.2 Asymptotic Analysis of $\mu(1,n)$

In order to obtain an asymptotic expression for $\mu(1,n)$, we analyze t_n in (3.8)–(3.9). The following lemma provides an exact expression for t_n that easily leads to an asymptotic expression for $\mu(1,n)$:

Lemma 3.1. *For $n \geq 2$, let $u_n := t_{n+1} - t_n$ (with $t_2 = 0$) and $v_n := u_{n+1} - u_n$. Let γ denote Euler's constant ($\doteq 0.57722$), and define $\chi_k := \frac{2\pi ik}{\ln 2}$. Then*

(i)

$$v_n = \frac{1}{n+1} + \frac{\frac{H_{n+2}}{\ln 2} - (\frac{\gamma}{\ln 2} - \frac{1}{2})}{(n+1)(n+2)} - \Sigma_n,$$

where

$$\Sigma_n := \sum_{k \in \mathbb{Z}\setminus\{0\}} \frac{\zeta(1-\chi_k)\Gamma(n+1)\Gamma(1-\chi_k)}{(\ln 2)\Gamma(n+3-\chi_k)};$$

(ii)

$$u_n = -H_n + a - \frac{H_{n+1}}{(\ln 2)(n+1)} + \left(\frac{\gamma - 1}{\ln 2} - \frac{1}{2}\right)\frac{1}{n+1} + \tilde{\Sigma}_n,$$

where

$$\begin{aligned} a &:= \frac{14}{9} + \frac{17 - 6\gamma}{18 \ln 2} - \frac{2}{\ln 2} \sum_{k \in \mathbb{Z}\setminus\{0\}} \frac{\zeta(1-\chi_k)\Gamma(1-\chi_k)}{\Gamma(4-\chi_k)(1-\chi_k)}, \\ \tilde{\Sigma}_n &:= \sum_{k \in Z\setminus\{0\}} \frac{\zeta(1-\chi_k)\Gamma(1-\chi_k)}{(\ln 2)(1-\chi_k)} \frac{\Gamma(n+1)}{\Gamma(n+2-\chi_k)}; \end{aligned}$$

(iii)

$$\begin{aligned} t_n &= -(nH_n - n - 1) + a(n-2) - \frac{1}{2\ln 2}\left[H_n^2 + H_n^{(2)} - \frac{7}{2}\right] \\ &\quad + \left(\frac{\gamma-1}{\ln 2} - \frac{1}{2}\right)\left(H_n - \frac{3}{2}\right) + b - \tilde{\tilde{\Sigma}}_n, \end{aligned}$$

where

$$\begin{aligned} b &:= \sum_{k \in \mathbb{Z}\setminus\{0\}} \frac{2\zeta(1-\chi_k)\Gamma(-\chi_k)}{(\ln 2)(1-\chi_k)\Gamma(3-\chi_k)}, \\ \tilde{\tilde{\Sigma}}_n &:= \sum_{k \in \mathbb{Z}\setminus\{0\}} \frac{\zeta(1-\chi_k)\Gamma(-\chi_k)\Gamma(n+1)}{(\ln 2)(1-\chi_k)\Gamma(n+1-\chi_k)}, \end{aligned}$$

and $H_n^{(2)}$ denotes the n-th Harmonic number of order 2, i.e., $H_n^{(2)} := \sum_{i=1}^{n} \frac{1}{i^2}$.

In this lemma, u_n and v_n are derived in order to obtain the exact expression for t_n in (iii). From (3.8), the exact expression for t_n also provides an alternative exact expression for $\mu(1,n)$.

Before proving Lemma 3.1, we complete the proof of Theorem 1.1 using part (iii). We know

$$H_n = \ln n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + O(n^{-4}), \tag{3.10}$$

$$H_n^{(2)} = \frac{\pi^2}{6} - \frac{1}{n} + \frac{1}{2n^2} + O(n^{-3}). \tag{3.11}$$

Combining (3.10)–(3.11) with (3.8) and Lemma 3.1(iii), we obtain an asymptotic expression for $\mu(1,n)$:

$$\mu(1,n) = 2an - \frac{1}{\ln 2}(\ln n)^2 - \left(\frac{2}{\ln 2} + 1\right)\ln n + O(1). \tag{3.12}$$

The term $O(1)$ in (3.12) has fluctuations of small magnitude due to Σ_n, which is periodic in $\log n$ with amplitude smaller than 0.00110. Thus, as shown in Theorem 1.1, the asymptotic slope in (3.12) is

$$c = 2a = \frac{28}{9} + \frac{17-6\gamma}{9\ln 2} - \frac{4}{\ln 2}\sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\zeta(1-\chi_k)\Gamma(1-\chi_k)}{\Gamma(4-\chi_k)(1-\chi_k)}. \tag{3.13}$$

Let S denote the sum in c:

$$S := \sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\zeta(1-\chi_k)\Gamma(1-\chi_k)}{\Gamma(4-\chi_k)(1-\chi_k)} = \sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\zeta(1-\chi_k)}{(3-\chi_k)(2-\chi_k)(1-\chi_k)^2}, \tag{3.14}$$

where the formula $\Gamma(1+x) = x\Gamma(x)$ is used to derive the second expression. Both expressions involve the imaginary numbers χ_k, but S is a real number. We investigate S and express it using only real functions. We have the following result:

Theorem 3.2. *Let S be the sum defined at* (3.14). *Then*

$$\frac{S}{\ln 2} = \tilde{S} - \rho, \tag{3.15}$$

where

$$\tilde{S} := \sum_{k=0}^{\infty} 2^{-k} h(2^k), \ \ h(m) := \frac{1}{2}(m\ln m - \ln m! - m) + \frac{3}{8} - \frac{1}{24}m^{-1}, \tag{3.16}$$

and

$$\rho := -\frac{17-6\gamma}{36\ln 2} - \frac{1}{12}.$$

Proof of Theorem 3.2. Choose and fix $0<\theta<1$. We show that the integral

$$J := \int_{\theta-i\infty}^{\theta+i\infty} \frac{\zeta(1-s)}{(1-2^{-s})(3-s)(2-s)(1-s)^2}\, ds$$

equals $2\pi i\tilde{S}$ on the one hand and equals $2\pi i[\rho+(S/\ln 2)]$ on the other hand. Equating these two expressions gives the desired result.

To get the first expression for J, we calculate

$$J=\sum_{k=0}^{\infty}\int_{\theta-i\infty}^{\theta+i\infty}\frac{\zeta(1-s)2^{-ks}}{(3-s)(2-s)(1-s)^2}\,ds=\sum_{k=0}^{\infty}2^{-k}\int_{1-\theta-i\infty}^{1-\theta+i\infty}\frac{\zeta(t)2^{kt}}{t^2(1+t)(2+t)}\,dt.$$

But, for any positive integer m and any $\alpha>1$,

$$\int_{1-\theta-i\infty}^{1-\theta+i\infty}\frac{\zeta(t)m^t}{t^2(1+t)(2+t)}\,dt=\int_{\alpha-i\infty}^{\alpha+i\infty}\frac{\zeta(t)m^t}{t^2(1+t)(2+t)}\,dt-2\pi i\mathrm{Res}_{t=1}\left[\frac{\zeta(t)m^t}{t^2(1+t)(2+t)}\right],$$

which follows from residue calculus, taking into account the contribution of the simple pole of the integrand at 1. Here

$$\mathrm{Res}_{t=1}\left[\frac{\zeta(t)m^t}{t^2(1+t)(2+t)}\right]=\frac{1}{6}m$$

and

$$\int_{\alpha-i\infty}^{\alpha+i\infty}\frac{\zeta(t)m^t}{t^2(1+t)(2+t)}\,dt=\sum_{j=1}^{\infty}\int_{\alpha-i\infty}^{\alpha+i\infty}\frac{(j/m)^{-t}}{t^2(1+t)(2+t)}\,dt.$$

Further, since

$$\frac{1}{t^2(t+1)(t+2)}=\frac{-3/4}{t}+\frac{1/2}{t^2}+\frac{1}{t+1}-\frac{1/4}{t+2},$$

we have by Mellin inversion that

$$\frac{1}{2\pi i}\int_{\alpha-i\infty}^{\alpha+i\infty}\frac{x^{-t}}{t^2(1+t)(2+t)}\,dt$$

equals

$$f(x):=-\frac{3}{4}-\frac{1}{2}\ln x+x-\frac{1}{4}x^2$$

for $0\le x\le 1$ and equals 0 for $x\ge 1$. (Note that this requires only $\alpha>0$.) So

$$\begin{aligned}\int_{\alpha-i\infty}^{\alpha+i\infty}\frac{\zeta(t)m^t}{t^2(1+t)(2+t)}\,dt &= 2\pi i\sum f\left(\frac{j}{m}\right)=2\pi i\left[\frac{1}{2}(m\ln m-\ln m!)-\frac{1}{3}m+\frac{3}{8}-\frac{1}{24}m^{-1}\right]\\ &= 2\pi i\left[h(m)+\frac{1}{6}m\right],\end{aligned}$$

where the sum is over $1\le j\le m$ (or $1\le j\le m-1$), and therefore

$$\int_{1-\theta-i\infty}^{1-\theta+i\infty}\frac{\zeta(t)m^t}{t^2(1+t)(2+t)}\,dt=2\pi i h(m).$$

Thus we obtain our first expression for J. We remark that the series $\tilde{S}$ converges geometrically rapidly.

To obtain the second expression for J we move the horizontal (i.e., real) coordinate of the vertical line of integration over from θ to $-C$ where C is large positive number ($C \to \infty$). By residue calculus, we find

$$J = 2\pi i \left\{ \mathrm{Res}_{s=0} \left[\frac{\zeta(1-s)}{(1-2^{-s})(3-s)(2-s)(1-s)^2} \right] + \frac{S}{\ln 2} \right\} = 2\pi i \left(\rho + \frac{S}{\ln 2} \right),$$

as desired. □

Using Theorem 3.2 it is straightforward to derive the alternative expression

$$c = 2 \sum_{k=0}^{\infty} \left(1 + 2^{-k} \sum_{j=1}^{2^k} \ln \frac{j}{2^k} \right) \tag{3.17}$$

for the linear coefficient c in (3.13). Grabner and Prodinger [6] obtained an earlier draft of this manuscript and independently conducted a similar analysis of S leading to (3.17). They also showed how to compute c efficiently to high precision and in particular computed c to 50 decimal places.

Remark 3.3. The lead-order asymptotics $\mu(1,n) \sim cn$ with c in the form (3.17) can also be obtained simply using real-analytical arguments. Start with (2.7) with $m = 1$ and recall that $P(s,t,1,n) = P_1(s,t,1,n)$ is given by (3.2) to see that

$$\mu(1,n) = 2 \int_0^1 \int_0^t \beta(s,t) t^{-2} [(1-t)^n - 1 + nt] \, ds \, dt.$$

An easy dominated-convergence argument then shows that $\mu(1,n) \sim cn$ with c given in the integral form

$$c = 2 \int_0^1 \int_0^t \beta(s,t) t^{-1} \, ds \, dt.$$

Writing

$$\beta(s,t) = \sum_{k=0}^{\infty} \mathbf{1}(s \text{ and } t \text{ agree in their first } k \text{ bits})$$

and breaking up the double integral according to the first k bits of t leads to the summation form (3.17) of c. We omit the details. We do not know how to obtain asymptotics for $\mu(1,n)$ beyond the lead term by this sort of approach.

Now we prove Lemma 3.1:

Proof of Lemma 3.1. (i) Since

$$\begin{aligned} u_n &= t_{n+1} - t_n = \sum_{j=2}^{n} \frac{B_j}{j(1-2^{-j})} \left[\frac{(n+1) - \binom{n+1}{j}}{j-1} - 1 \right] - \sum_{j=2}^{n-1} \frac{B_j}{j(1-2^{-j})} \left[\frac{n - \binom{n}{j}}{j-1} - 1 \right] \\ &= -\sum_{j=2}^{n} \frac{B_j}{j(j-1)(1-2^{-j})} \left[\binom{n}{j-1} - 1 \right], \end{aligned}$$

it follows that

$$\begin{aligned}
v_n &= u_{n+1}-u_n = -\sum_{j=2}^{n+1}\frac{B_j}{j(j-1)(1-2^{-j})}\left[\binom{n+1}{j-1}-1\right]+\sum_{j=2}^{n+1}\frac{B_j}{j(j-1)(1-2^{-j})}\left[\binom{n}{j-1}-1\right] \\
&= -\sum_{k=0}^{n-1}\binom{n}{k}\frac{B_{k+2}}{(k+2)(k+1)[1-2^{-(k+2)}]} \\
&= \sum_{k=0}^{n-1}(-1)^k\binom{n}{k}\frac{\zeta(-1-k)}{(k+1)[1-2^{-(k+2)}]} \qquad (3.18)\\
&= \frac{(-1)^n}{2\pi i}\int_{\mathcal{C}}\frac{\zeta(-1-s)}{(s+1)[1-2^{-(s+2)}]}\frac{n!}{s(s-1)\cdots(s-n)}\,ds, \qquad (3.19)
\end{aligned}$$

where $\mathcal{C}$ is a positively oriented closed curve that encircles the integers $0,\ldots,n-1$ and does not include or encircle any of the following points: $-2+\chi_k$ (where $\chi_k := \frac{2\pi ik}{\ln 2}$), $k\in\mathbb{Z}$; -1; and n. Equality (3.18) follows from the fact that the Bernoulli numbers are extrapolated by the Riemann zeta function taken at nonnegative integers: $B_k = -k\zeta(1-k)$. [The coefficients $(-1)^k$ do not concern us since the Bernoulli numbers of odd index greater than 1 vanish.] Equality (3.19) follows from a direct application of residue calculus, taking into account contributions of the simple poles at the integers $0,\ldots,n-1$.

Let $\phi(s)$ denote the integrand in (3.19):

$$\phi(s)=\frac{\zeta(-1-s)}{(s+1)[1-2^{-(s+2)}]}\frac{n!}{s(s-1)\cdots(s-n)}.$$

We consider a positively oriented rectangular contour $\mathcal{C}_l$ with horizontal sides $\mathrm{Im}(s)=\lambda_l$ and $\mathrm{Im}(s)=-\lambda_l$, where $\lambda_l := \frac{(2l+1)\pi}{\ln 2}$, $l\in\mathbb{Z}^+$, and vertical sides $\mathrm{Re}(s)=n-\theta$ and $\mathrm{Re}(s)=-\lambda_l$, where $0<\theta<1$. By elementary bounds on $\phi(s)$ along $\mathcal{C}_l$ and the fact that

$$\int_{n-\theta-i\infty}^{n-\theta+i\infty}\phi(s)\,ds = 0 \qquad (3.20)$$

(this is implicit on page 113 of Flajolet and Sedgewick [5] and explicitly proved in the Appendix), one can show that

$$\lim_{l\to\infty}\int_{\mathcal{C}_l}\phi(s)\,ds = 0.$$

Accounting for residues due to the poles encircled by $\mathcal{C}_l$, we obtain

$$\begin{aligned}
v_n &= (-1)^{n+1}\left\{\mathrm{Res}_{s=-1}[\phi(s)]+\mathrm{Res}_{s=-2}[\phi(s)]+\sum_{k\in\mathbb{Z}\setminus\{0\}}\mathrm{Res}_{s=-2+\chi_k}[\phi(s)]\right\}\\
&= -\frac{1}{n+1}+\frac{\frac{H_{n+2}}{\ln 2}-\left(\frac{\gamma}{\ln 2}-\frac{1}{2}\right)}{(n+1)(n+2)}-\Sigma_n, \qquad (3.21)
\end{aligned}$$

where

$$\Sigma_n := \sum_{k\in\mathbb{Z}\setminus\{0\}}\frac{\zeta(1-\chi_k)\Gamma(n+1)\Gamma(1-\chi_k)}{(\ln 2)\Gamma(n+3-\chi_k)}. \qquad (3.22)$$

$\square$

(ii) We have $u_2 = t_3 - t_2 = t_3 = -\frac{1}{9}$. Hence, from (i),

$$\begin{aligned}
u_n &= u_2 + \sum_{j=2}^{n-1} v_j = -\frac{1}{9} + \sum_{j=2}^{n-1} v_j \\
&= -\frac{1}{9} - \sum_{j=2}^{n-1} \frac{1}{j+1} + \frac{1}{\ln 2} \sum_{j=2}^{n-1} \frac{H_{j+2}}{(j+1)(j+2)} - \left(\frac{\gamma}{\ln 2} - \frac{1}{2}\right) \sum_{j=2}^{n-1} \frac{1}{(j+1)(j+2)} - \sum_{j=2}^{n-1} \Sigma_j \\
&= -\frac{1}{9} - (H_n - H_2) + \frac{1}{\ln 2} \sum_{j=2}^{n-1} \frac{H_{j+2}}{(j+1)(j+2)} - \left(\frac{\gamma}{\ln 2} - \frac{1}{2}\right) \left(\frac{1}{3} - \frac{1}{n+1}\right) - \sum_{j=2}^{n-1} \Sigma_j \\
&= \frac{14}{9} - \frac{\gamma}{3 \ln 2} - H_n + \left(\frac{\gamma}{\ln 2} - \frac{1}{2}\right) \frac{1}{n+1} + \frac{1}{\ln 2} \sum_{j=2}^{n-1} \frac{H_{j+2}}{(j+1)(j+2)} - \sum_{j=2}^{n-1} \Sigma_j. \qquad (3.23)
\end{aligned}$$

Here

$$\begin{aligned}
\sum_{j=2}^{n-1} \frac{H_{j+2}}{(j+1)(j+2)} &= \sum_{j=3}^{n} \frac{H_{j+1}}{j} - \sum_{j=4}^{n+1} \frac{H_j}{j} \\
&= \frac{H_4}{3} + \sum_{j=4}^{n} \frac{H_{j+1} - H_j}{j} - \frac{H_{n+1}}{n+1} \qquad (3.24) \\
&= \frac{17}{18} - \frac{H_n + 1}{n+1}, \qquad (3.25)
\end{aligned}$$

where we assume $n \geq 3$ for (3.24), but (3.25) holds also for $n = 2$. In regard to $\sum_{j=2}^{n-1} \Sigma_j$, note that

$$\Sigma_n = - \sum_{k \in Z \setminus \{0\}} \frac{\zeta(1-\chi_k)\Gamma(1-\chi_k)}{(\ln 2)(1-\chi_k)} \left[\frac{\Gamma(n+2)}{\Gamma(n+3-\chi_k)} - \frac{\Gamma(n+1)}{\Gamma(n+2-\chi_k)}\right],$$

so that

$$\sum_{j=2}^{n-1} \Sigma_j = - \sum_{k \in Z \setminus \{0\}} \frac{\zeta(1-\chi_k)\Gamma(1-\chi_k)}{(\ln 2)(1-\chi_k)} \left[\frac{\Gamma(n+1)}{\Gamma(n+2-\chi_k)} - \frac{\Gamma(3)}{\Gamma(4-\chi_k)}\right]. \qquad (3.26)$$

Define

$$\tilde{\Sigma}_n := \sum_{k \in Z \setminus \{0\}} \frac{\zeta(1-\chi_k)\Gamma(1-\chi_k)}{(\ln 2)(1-\chi_k)} \frac{\Gamma(n+1)}{\Gamma(n+2-\chi_k)}. \qquad (3.27)$$

Then, combining (3.23), (3.25), and (3.26), we obtain

$$u_n = -H_n + a - \frac{H_{n+1}}{(\ln 2)(n+1)} + \left(\frac{\gamma - 1}{\ln 2} - \frac{1}{2}\right) \frac{1}{n+1} + \tilde{\Sigma}_n,$$

where

$$a := \frac{14}{9} + \frac{17-6\gamma}{18\ln 2} - \frac{2}{\ln 2}\sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\zeta(1-\chi_k)\Gamma(1-\chi_k)}{\Gamma(4-\chi_k)(1-\chi_k)}. \tag{3.28}$$

□

(iii) Closely following the derivation of u_n described above, we obtain (for $n \geq 2$)

$$\begin{aligned} t_n &= t_2 + \sum_{j=2}^{n-1} u_j = \sum_{j=2}^{n-1} u_j \\ &= -\sum_{j=2}^{n-1} H_j + a(n-2) - \frac{1}{\ln 2}\sum_{j=3}^{n}\frac{H_j}{j} + \left(\frac{\gamma-1}{\ln 2} - \frac{1}{2}\right)\left(H_n - \frac{3}{2}\right) + \sum_{j=2}^{n-1}\tilde{\Sigma}_j \\ &= -(nH_n - n - 1) + a(n-2) - \frac{1}{2\ln 2}\left[H_n^2 + H_n^{(2)} - \frac{7}{2}\right] \\ &\quad + \left(\frac{\gamma-1}{\ln 2} - \frac{1}{2}\right)\left(H_n - \frac{3}{2}\right) + b - \tilde{\tilde{\Sigma}}_n, \end{aligned} \tag{3.29}$$

where

$$b := \sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{2\zeta(1-\chi_k)\Gamma(-\chi_k)}{(\ln 2)(1-\chi_k)\Gamma(3-\chi_k)}, \tag{3.30}$$

$$\tilde{\tilde{\Sigma}}_n := \sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\zeta(1-\chi_k)\Gamma(-\chi_k)\Gamma(n+1)}{(\ln 2)(1-\chi_k)\Gamma(n+1-\chi_k)}. \tag{3.31}$$

□

4 Analysis of the Average Case: $\mu(\bar{m},n)$

4.1 Exact Computation of $\mu(\bar{m},n)$

Here we consider the parameter m in $\mu(m,n)$ as a discrete random variable with probability mass function $P\{m=i\} = \frac{1}{n}$, $i = 1,2,\ldots,n$, and average over m while the parameter n is fixed. Thus, using the notation defined in (2.3) through (2.7),

$$\begin{aligned} \mu(\bar{m},n) &= \frac{1}{n}\sum_{m=1}^{n}\mu(m,n) = \frac{1}{n}\sum_{m=1}^{n}\int_0^1\int_s^1 \beta(s,t)P(s,t,m,n)\,dt\,ds \\ &= \int_0^1\int_s^1 \beta(s,t)\ \frac{1}{n}\sum_{m=1}^{n} P(s,t,m,n)\,dt\,ds = \mu_1(\bar{m},n) + \mu_2(\bar{m},n) + \mu_3(\bar{m},n), \end{aligned}$$

where, for $l = 1,2,3$,

$$\mu_l(\bar{m},n) = \int_0^1\int_s^1 \beta(s,t)\ \frac{1}{n}\sum_{m=1}^{n} P_l(s,t,m,n)\ dt\,ds. \tag{4.1}$$

Here $\mu_1(\bar{m}, n) = \mu_3(\bar{m}, n)$, since

$$P_3(1-t', 1-s', n-m'+1, n) = P_1(s', t', m', n)$$

by an easy symmetry argument we omit, and so

$$\begin{aligned}
\mu_3(\bar{m}, n) &= \int_0^1 \int_s^1 \beta(s,t)\, \frac{1}{n} \sum_{m=1}^{n} P_3(s,t,m,n)\; dt\, ds \\
&= \int_0^1 \int_{s'}^1 \beta(1-t', 1-s')\, \frac{1}{n} \sum_{m'=1}^{n} P_3(1-t', 1-s', n-m'+1, n)\; dt'\, ds' \\
&= \int_0^1 \int_{s'}^1 \beta(s', t')\, \frac{1}{n} \sum_{m'=1}^{n} P_1(s', t', m', n)\; dt'\, ds' \\
&= \mu_1(\bar{m}, n).
\end{aligned}$$

Therefore

$$\mu(\bar{m}, n) = 2\mu_1(\bar{m}, n) + \mu_2(\bar{m}, n), \tag{4.2}$$

and we will compute $\mu_1(\bar{m}, n)$ and $\mu_2(\bar{m}, n)$ exactly in Sections 4.1.1-2.

4.1.1 Exact Computation of $\mu_1(\bar{m}, n)$

We use the following lemma in order to compute $\mu_1(\bar{m}, n)$ exactly:

Lemma 4.1.

$$\begin{aligned}
&\int_0^1 \int_s^1 \beta(s,t)\, \frac{1}{n} \sum_{m=2}^{n} P_1(s,t,m,n)\, dt\, ds \\
&= 2\sum_{j=2}^{n-1} \frac{(-1)^j \binom{n-1}{j}}{j(j-1)} + \frac{2}{9} \sum_{j=2}^{n-1} \frac{(-1)^j \binom{n-1}{j}}{j-1} - 2\sum_{j=3}^{n-1} B_j \frac{n-j+1-\binom{n-1}{j-1}}{j(j-1)(j-2)(1-2^{-j})} \\
&\quad -2\sum_{j=2}^{n-1} \frac{(-1)^j \binom{n-1}{j}}{(j+1)j(j-1)(1-2^{-j})}.
\end{aligned}$$

Before proving the lemma, we complete the computation of $\mu_1(\bar{m}, n)$. Note that

$$\begin{aligned}
\mu_1(\bar{m}, n) &= \int_0^1 \int_s^1 \beta(s,t)\, \frac{1}{n} \sum_{m=1}^{n} P_1(s,t,m,n)\, dt\, ds \\
&= \frac{1}{n} \int_0^1 \int_s^1 \beta(s,t)\, P_1(s,t,1,n)\, dt\, ds + \int_0^1 \int_s^1 \beta(s,t)\, \frac{1}{n} \sum_{m=2}^{n} P_1(s,t,m,n)\, dt\, ds \\
&= \frac{1}{n}\mu(1,n) + \int_0^1 \int_s^1 \beta(s,t)\, \frac{1}{n} \sum_{m=2}^{n} P_1(s,t,m,n)\, dt\, ds.
\end{aligned}$$

Therefore, by (3.8) and Lemma 4.1, we obtain

$$\begin{aligned}
\mu_1(\bar{m}, n) &= \frac{2}{n}\sum_{j=2}^{n}\frac{(-1)^j\binom{n}{j}}{j-1} + \frac{2}{n}\sum_{j=2}^{n-1} B_j \frac{n-j+1-\binom{n}{j}}{j(j-1)(1-2^{-j})} \\
&\quad + 2\sum_{j=2}^{n-1}\frac{(-1)^j\binom{n-1}{j}}{j(j-1)} + \frac{2}{9}\sum_{j=2}^{n-1}\frac{(-1)^j\binom{n-1}{j}}{j-1} - 2\sum_{j=3}^{n-1} B_j \frac{n-j+1-\binom{n-1}{j-1}}{j(j-1)(j-2)(1-2^{-j})} \\
&\quad - 2\sum_{j=2}^{n-1}\frac{(-1)^j\binom{n-1}{j}}{(j+1)j(j-1)(1-2^{-j})} \\
&= n-1-4\sum_{j=3}^{n}\frac{(-1)^j\binom{n-1}{j-1}}{j(j-1)(j-2)} + \frac{2}{n}\sum_{j=2}^{n-1} B_j \frac{n-j+1-\binom{n}{j}}{j(j-1)(1-2^{-j})} \\
&\quad + \frac{2}{9}\sum_{j=2}^{n-1}\frac{(-1)^j\binom{n-1}{j}}{j-1} - 2\sum_{j=3}^{n-1} B_j \frac{n-j+1-\binom{n-1}{j-1}}{j(j-1)(j-2)(1-2^{-j})} \\
&\quad - 2\sum_{j=2}^{n-1}\frac{(-1)^j\binom{n-1}{j}}{(j+1)j(j-1)(1-2^{-j})},
\end{aligned} \tag{4.3}$$

where the second equality holds since

$$\begin{aligned}
&\frac{2}{n}\sum_{j=2}^{n}\frac{(-1)^j\binom{n}{j}}{j-1} + 2\sum_{j=2}^{n-1}\frac{(-1)^j\binom{n-1}{j}}{j(j-1)} \\
&= 2\sum_{j=2}^{n}\frac{(-1)^j(n-1)!}{j!(n-j)!(j-1)} - 2\sum_{j=3}^{n}\frac{(-1)^j(n-1)!}{(j-1)!(n-j)!(j-1)(j-2)} \\
&= n-1+2\sum_{j=3}^{n}\frac{(-1)^j(n-1)!}{(j-1)!(n-j)!(j-1)}\left[\frac{1}{j}-\frac{1}{j-2}\right] \\
&= n-1-4\sum_{j=3}^{n}\frac{(-1)^j\binom{n-1}{j-1}}{j(j-1)(j-2)}.
\end{aligned}$$

In Section 4.1.2 we combine the expression for $\mu_1(\bar{m}, n)$ in (4.3) with a similar expression for $\mu_2(\bar{m}, n)$ to obtain an exact expression for $\mu(\bar{m}, n)$. The remainder of this section is devoted to proving Lemma 4.1. For this, the following expression for $P_1(s,t,m,n)$ will prove useful:

Lemma 4.2. *Let $m \geq 2$ and let $x := s$, $y := t-s$, $z := 1-t$. Then the quantity $P_1(s,t,m,n)$ defined at* (2.3) *satisfies*

$$\begin{aligned}
&P_1(s,t,m,n) \\
&= 2n\int_0^x \frac{1}{(\xi+y)^2}\left[\Upsilon_1(m,n,\xi,x,y,z) - \Upsilon_2(m,n,\xi,x,y,z) + \Upsilon_3(m,n,\xi,x,y,z)\right] d\xi,
\end{aligned} \tag{4.4}$$

where

$$\begin{aligned}
\Upsilon_1(m,n,\xi,x,y,z) &:= \binom{n-1}{m-2}(x-\xi)^{m-2}(n-m)(\xi+y+z)^{n-m+1},\\
\Upsilon_2(m,n,\xi,x,y,z) &:= \binom{n-1}{m-2}(x-\xi)^{m-2}(n-m+1)z(\xi+y+z)^{n-m},\\
\Upsilon_3(m,n,\xi,x,y,z) &:= \binom{n-1}{m-2}(x-\xi)^{m-2}z^{n-m+1}.
\end{aligned}$$

Proof of Lemma 4.2. By (2.2)–(2.3),

$$\begin{aligned}
P_1(s,t,m,n) &= \sum_{m\le i<j\le n} \frac{2}{j-m+1}\frac{n!}{(i-1)!(j-i-1)!(n-j)!}\; x^{i-1}y^{j-i-1}z^{n-j}\\
&= \sum_{m\le i<j\le n} \frac{2}{j-m+1}\frac{n!}{(n-m-1)!}\binom{n-m-1}{i-m,\, j-i-1,\, n-j}\frac{(i-m)!}{(i-1)!}\; x^{i-1}y^{j-i-1}z^{n-j}\\
&= \frac{2\,n!}{(n-m-1)!}\sum_{m\le i<j\le n} \frac{1}{j-m+1}\binom{n-m-1}{i-m,\, j-i-1,\, n-j}\frac{(i-m)!}{(i-1)!}\; x^{i-1}y^{j-i-1}z^{n-j}.
\end{aligned} \tag{4.5}$$

In order to compactly describe the derivation of (4.4), we define the following indefinite integration operator T:

$$T(f(x)) := \int_0^x f(\xi)\, d\xi.$$

We really should write $(Tf)(x)$ rather than $T(f(x))$, but we would like to use shorthand such as $T(x^j) = \frac{x^{j+1}}{j+1}$ when $j > -1$. The operator T treats its argument f as a function of x; the other variables involved in f (namely, y and z) are treated as constants. The notation T^l will denote the l-th iterate of T. In this notation, for $m < i$,

$$\frac{(i-m)!}{(i-1)!}\; x^{i-1} = T^{m-1}(x^{i-m}),$$

and the sum in (4.5) equals

$$T^{m-1}\left(\sum_{m\le i<j\le n} \frac{1}{j-m+1}\binom{n-m-1}{i-m,\, j-i-1,\, n-j}\; x^{i-m}y^{j-i-1}z^{n-j}\right).$$

Here

$$\frac{1}{j-m+1}\; z^{n-j} = z^{n-m+1}\int_z^\infty \eta^{-(j-m+1)-1} d\eta,$$

so

$$\begin{aligned}
&T^{m-1}\left(\sum_{m\leq i<j\leq n}\frac{1}{j-m+1}\binom{n-m-1}{i-m,j-i-1,n-j}\,x^{i-m}y^{j-i-1}z^{n-j}\right)\\
&= z^{n-m+1}\,T^{m-1}\left(\int_z^{\infty}\left[\sum_{m\leq i<j\leq n}\binom{n-m-1}{i-m,j-i-1,n-j}\,x^{i-m}y^{j-i-1}\eta^{-j+m-2}\right]d\eta\right)\\
&= z^{n-m+1}\,T^{m-1}\left(\int_z^{\infty}\eta^{-n+m-2}(x+y+\eta)^{n-m-1}d\eta\right)\\
&= z^{n-m+1}\,T^{m-1}\left(\int_z^{\infty}\eta^{-3}\left(\frac{t}{\eta}+1\right)^{n-m-1}d\eta\right)
\end{aligned}\tag{4.6}$$

(note that $x+y=t$). Making the change of variables $v=\frac{t}{\eta}+1$ and integrating, we obtain, after some computation,

$$\begin{aligned}
&\int_z^{\infty}\eta^{-3}\left(\frac{t}{\eta}+1\right)^{n-m-1}d\eta\\
&= \frac{1}{t^2(n-m+1)(n-m)}\left[(n-m)\left(1+\frac{t}{z}\right)^{n-m+1}-(n-m+1)\left(1+\frac{t}{z}\right)^{n-m}+1\right].
\end{aligned}\tag{4.7}$$

From (4.5) and (4.6)–(4.7),

$$\begin{aligned}
&P_1(s,t,m,n)\\
&= \frac{2\,n!}{(n-m+1)!}\,T^{m-1}\left(t^{-2}[(n-m)(z+t)^{n-m+1}-(n-m+1)z(z+t)^{n-m}+z^{n-m+1}]\right).
\end{aligned}\tag{4.8}$$

Here

$$t^{-2}[(n-m)(z+t)^{n-m+1}-(n-m+1)z(z+t)^{n-m}+z^{n-m+1}] = \sum_{r=2}^{n-m+1}t^{r-2}\Upsilon(m,n,r,z),\tag{4.9}$$

where

$$\Upsilon(m,n,r,z) := (n-m)\binom{n-m+1}{r}z^{n-m+1-r}-(n-m+1)\binom{n-m}{r}z^{n-m+1-r}.\tag{4.10}$$

Then, since $t=x+y$,

$$\sum_{r=2}^{n-m+1}t^{r-2}\Upsilon(m,n,r,z)=\sum_{r=2}^{n-m+1}\Upsilon(m,n,r,z)\sum_{j=0}^{r-2}\binom{r-2}{j}x^jy^{r-2-j}.\tag{4.11}$$

From (4.8)–(4.11),

$$\begin{aligned}
&P_1(s,t,m,n) \\
&= \frac{2\,n!}{(n-m+1)!} T^{m-1}\left(\sum_{r=2}^{n-m+1} \Upsilon(m,n,r,z) \sum_{j=0}^{r-2} \binom{r-2}{j} x^j y^{r-2-j}\right) \\
&= \frac{2\,n!}{(n-m+1)!} \sum_{r=2}^{n-m+1} \Upsilon(m,n,r,z) \sum_{j=0}^{r-2} \binom{r-2}{j} y^{r-2-j} T^{m-1}(x^j) \\
&= \frac{2\,n!}{(n-m+1)!} \sum_{r=2}^{n-m+1} \Upsilon(m,n,r,z) \sum_{j=0}^{r-2} \binom{r-2}{j} y^{r-2-j} \frac{x^{j+m-1}}{(j+1)\cdots(j+m-1)}.
\end{aligned} \tag{4.12}$$

Because of the partial fraction expansion

$$\frac{1}{(j+1)\cdots(j+m-1)} = \frac{1}{(m-2)!}\sum_{l=0}^{m-2} \frac{(-1)^l\binom{m-2}{l}}{j+l+1},$$

it follows that

$$\begin{aligned}
&\sum_{j=0}^{r-2} \binom{r-2}{j} y^{r-2-j} \frac{x^{j+m-1}}{(j+1)\cdots(j+m-1)} \\
&= \sum_{j=0}^{r-2} \binom{r-2}{j} y^{r-2-j} \frac{x^{j+m-1}}{(m-2)!} \sum_{l=0}^{m-2} \frac{(-1)^l\binom{m-2}{l}}{j+l+1} \\
&= \frac{1}{(m-2)!} \sum_{l=0}^{m-2} (-1)^l \binom{m-2}{l} x^{m-2-l} \int_0^x \xi^l \sum_{j=0}^{r-2} \binom{r-2}{j} y^{r-2-j} \xi^j d\xi \\
&= \frac{1}{(m-2)!} \sum_{l=0}^{m-2} (-1)^l \binom{m-2}{l} x^{m-2-l} \int_0^x \xi^l (\xi+y)^{r-2} d\xi \\
&= \frac{1}{(m-2)!} \int_0^x (x-\xi)^{m-2} (\xi+y)^{r-2} d\xi.
\end{aligned} \tag{4.13}$$

From (4.12)–(4.13),

$$\begin{aligned}
P_1(s,t,m,n) &= \frac{2\,n!}{(n-m+1)!(m-2)!} \sum_{r=2}^{n-m+1} \Upsilon(m,n,r,z) \int_0^x (x-\xi)^{m-2}(\xi+y)^{r-2} d\xi \\
&= 2n\binom{n-1}{m-2} \int_0^x \sum_{r=2}^{n-m+1} \Upsilon(m,n,r,z)(x-\xi)^{m-2}(\xi+y)^{r-2} d\xi \\
&= 2n\binom{n-1}{m-2} \int_0^x \frac{(x-\xi)^{m-2}}{(\xi+y)^2} \sum_{r=2}^{n-m+1} \Upsilon(m,n,r,z)(\xi+y)^{r} d\xi.
\end{aligned} \tag{4.14}$$

Here, by (4.10),

$$\begin{aligned}
&\sum_{r=2}^{n-m+1} \Upsilon(m,n,r,z)(\xi+y)^r \\
&= \sum_{r=2}^{n-m+1} \left[(n-m)\binom{n-m+1}{r} z^{n-m+1-r} - (n-m+1)\binom{n-m}{r} z^{n-m+1-r}\right](\xi+y)^r \\
&= (n-m)\sum_{r=2}^{n-m+1}\binom{n-m+1}{r}(\xi+y)^r z^{n-m+1-r} - (n-m+1)\sum_{r=2}^{n-m+1}\binom{n-m}{r}(\xi+y)^r z^{n-m+1-r} \\
&= (n-m)[(\xi+y+z)^{n-m+1} - z^{n-m+1} - (n-m+1)(\xi+y)z^{n-m}] \\
&\quad -(n-m+1)z[(\xi+y+z)^{n-m} - z^{n-m} - (n-m)(\xi+y)z^{n-m-1}] \\
&= (n-m)(\xi+y+z)^{n-m+1} - (n-m+1)z(\xi+y+z)^{n-m} + z^{n-m+1}. \qquad (4.15)
\end{aligned}$$

Substitution of (4.15) into (4.14) gives the desired (4.4). □

Proof of Lemma 4.1. From Lemma 4.2, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{m=2}^{n} P_1(s,t,m,n) \\
&= 2\int_0^x \frac{1}{(\xi+y)^2}\sum_{m=2}^{n}[\Upsilon_1(m,n,\xi,x,y,z) - \Upsilon_2(m,n,\xi,x,y,z) + \Upsilon_3(m,n,\xi,x,y,z)]\,d\xi. \qquad (4.16)
\end{aligned}$$

Here

$$\begin{aligned}
&\sum_{m=2}^{n}\Upsilon_1(m,n,\xi,x,y,z) = (\xi+y+z)^2\,\frac{d}{dw}\left[\sum_{m=2}^{n}\binom{n-1}{m-2}(x-\xi)^{m-2}w^{n-m}\right]\Bigg|_{w=\xi+y+z} \\
&= (\xi+y+z)^2\,\frac{d}{dw}\left\{w^{-1}[(x-\xi+w)^{n-1} - (x-\xi)^{n-1}]\right\}\Big|_{w=\xi+y+z} \\
&= (\xi+y+z)^2\left\{-w^{-2}[(x-\xi+w)^{n-1} - (x-\xi)^{n-1}] + w^{-1}(n-1)(x-\xi+w)^{n-2}\right\}\Big|_{w=\xi+y+z} \\
&= (x-\xi)^{n-1} - 1 + (n-1)(\xi+y+z) \qquad (4.17)
\end{aligned}$$

(note that $x+y+z=1$). Similarly,

$$\begin{aligned}
&\sum_{m=2}^{n}\Upsilon_2(m,n,\xi,x,y,z) = z\,\frac{d}{dw}\left[\sum_{m=2}^{n}\binom{n-1}{m-2}(x-\xi)^{m-2}w^{n-m+1}\right]\Bigg|_{w=\xi+y+z} \\
&= z\,\frac{d}{dw}\left[(x-\xi+w)^{n-1} - (x-\xi)^{n-1}\right]\Big|_{w=\xi+y+z} = z\left[(n-1)(x-\xi+w)^{n-2}\right]\Big|_{w=\xi+y+z} \\
&= z(n-1), \qquad (4.18)
\end{aligned}$$

and

$$\sum_{m=2}^{n}\Upsilon_3(m,n,\xi,x,y,z) = \sum_{m=2}^{n}\binom{n-1}{m-2}(x-\xi)^{m-2}z^{n-m+1} = (x-\xi+z)^{n-1} - (x-\xi)^{n-1}.$$

Hence

$$\sum_{m=2}^{n}[\Upsilon_1(m,n,\xi,x,y,z) - \Upsilon_2(m,n,\xi,x,y,z) + \Upsilon_3(m,n,\xi,x,y,z)]$$
$$= (n-1)(\xi+y) - 1 + (x-\xi+z)^{n-1}. \tag{4.19}$$

Therefore, from (4.16) and (4.19), we obtain

$$\begin{aligned}
\frac{1}{n}\sum_{m=2}^{n} P_1(s,t,m,n) &= 2\int_0^x \frac{1}{(\xi+y)^2}[(n-1)(\xi+y) - 1 + (x-\xi+z)^{n-1}]\, d\xi \\
&= 2\int_0^x \frac{1}{(\xi+y)^2}\{(n-1)(\xi+y) - 1 + [1-(\xi+y)]^{n-1}\}\, d\xi \\
&= 2\int_0^x \frac{1}{(\xi+y)^2}\sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}(\xi+y)^j\, d\xi = 2\sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}\int_0^x(\xi+y)^{j-2}\, d\xi \\
&= 2\sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}\frac{(x+y)^{j-1} - y^{j-1}}{j-1} = 2\sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}\frac{t^{j-1} - (t-s)^{j-1}}{j-1}. \qquad (4.20)
\end{aligned}$$

We complete the proof by using (4.20) to compute $\int_0^1\int_s^1 \beta(s,t)\,\frac{1}{n}\sum_{m=2}^{n} P_1(s,t,m,n)\, ds\, dt$. We have

$$\begin{aligned}
&\int_0^1\int_s^1 \beta(s,t)\,\frac{1}{n}\sum_{m=2}^{n} P_1(s,t,m,n)\, ds\, dt \\
&= 2\int_0^1\int_s^1 \beta(s,t)\ \sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}\frac{t^{j-1}-(t-s)^{j-1}}{j-1}\, dt\, ds \\
&= 2\int_0^1\int_s^1 \beta(s,t)\ \sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}\frac{t^{j-1}}{j-1}\, dt\, ds \\
&\quad -2\int_0^1\int_s^1 \beta(s,t)\ \sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}\frac{(t-s)^{j-1}}{j-1}\, dt\, ds. \qquad (4.21)
\end{aligned}$$

Closely following the derivations shown in (3.3)–(3.8), one can show that

$$\begin{aligned}
&\int_0^1\int_s^1 \beta(s,t)\ \sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}\frac{t^{j-1}}{j-1}\, dt\, ds \\
&= \sum_{j=2}^{n-1}\frac{(-1)^j\binom{n-1}{j}}{j(j-1)} + \frac{1}{9}\sum_{j=2}^{n-1}\frac{(-1)^j\binom{n-1}{j}}{j-1} - \sum_{j=3}^{n-1} B_j\frac{n-j+1-\binom{n-1}{j-1}}{j(j-1)(j-2)(1-2^{-j})}. \qquad (4.22)
\end{aligned}$$

Thus, in order to complete the proof, it remains to show that

$$\int_0^1\int_s^1 \beta(s,t)\ \sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}\frac{(t-s)^{j-1}}{j-1}\, dt\, ds = \sum_{j=2}^{n-1}\frac{(-1)^j\binom{n-1}{j}}{(j+1)j(j-1)(1-2^{-j})}. \tag{4.23}$$

Indeed, we have

$$
\begin{aligned}
&\int_0^1 \int_s^1 \beta(s,t) \ \sum_{j=2}^{n-1} (-1)^j \binom{n-1}{j} \frac{(t-s)^{j-1}}{j-1}\, dt\, ds \\
&= \sum_{k=0}^{\infty} (k+1) 2^k \int_0^{2^{-(k+1)}} \int_{2^{-(k+1)}}^{2^{-k}} \sum_{j=2}^{n-1} (-1)^j \binom{n-1}{j} \frac{(t-s)^{j-1}}{j-1}\, dt\, ds \\
&= \sum_{k=0}^{\infty} (k+1) 2^k \int_0^{2^{-(k+1)}} \int_{2^{-(k+1)}-s}^{2^{-k}-s} \sum_{j=2}^{n-1} (-1)^j \binom{n-1}{j} \frac{v^{j-1}}{j-1}\, dv\, ds \\
&= \sum_{k=0}^{\infty} (k+1) 2^k \sum_{j=2}^{n-1} (-1)^j \binom{n-1}{j} \int_0^{2^{-k}} \frac{v^{j-1}}{j-1} \int_{[2^{-(k+1)}-v] \bigvee 0}^{(2^{-k}-v) \bigwedge 2^{-(k+1)}} ds\, dv. \qquad (4.24)
\end{aligned}
$$

Here

$$
\int_{[2^{-(k+1)}-v] \bigvee 0}^{(2^{-k}-v) \bigwedge 2^{-(k+1)}} ds = \begin{cases} v & \text{if } 0 \le v \le 2^{-(k+1)} \\ 2^{-k} - v & \text{if } 2^{-(k+1)} < v \le 2^{-k}. \end{cases} \qquad (4.25)
$$

Thus

$$
\begin{aligned}
\int_0^{2^{-k}} \frac{v^{j-1}}{j-1} \int_{[2^{-(k+1)}-v] \bigvee 0}^{(2^{-k}-v) \bigwedge 2^{-(k+1)}} ds\, dv &= \frac{1}{j-1} \left[\int_0^{2^{-(k+1)}} v^j\, dv + \int_{2^{-(k+1)}}^{2^{-k}} v^{j-1} (2^{-k} - v)\, dv \right] \\
&= \frac{2^{-k(j+1)}(1-2^{-j})}{(j+1)j(j-1)}. \qquad (4.26)
\end{aligned}
$$

From (4.24) and (4.26), we obtain

$$
\begin{aligned}
&\int_0^1 \int_s^1 \beta(s,t) \ \sum_{j=2}^{n-1} (-1)^j \binom{n-1}{j} \frac{(t-s)^{j-1}}{j-1}\, dt\, ds \\
&= \sum_{k=0}^{\infty} (k+1) 2^k \sum_{j=2}^{n-1} (-1)^j \binom{n-1}{j} \frac{2^{-k(j+1)}(1-2^{-j})}{(j+1)j(j-1)} \\
&= \sum_{j=2}^{n-1} (-1)^j \binom{n-1}{j} \frac{1-2^{-j}}{(j+1)j(j-1)} \sum_{k=0}^{\infty} (k+1) 2^{-kj} \\
&= \sum_{j=2}^{n-1} (-1)^j \binom{n-1}{j} \frac{1-2^{-j}}{(j+1)j(j-1)} \frac{1}{(1-2^{-j})^2} \\
&= \sum_{j=2}^{n-1} \frac{(-1)^j \binom{n-1}{j}}{(j+1)j(j-1)(1-2^{-j})}, \qquad (4.27)
\end{aligned}
$$

and (4.23) is proved. □

4.1.2 Exact Computation of $\mu_2(\bar{m}, n)$ and $\mu(\bar{m}, n)$

The derivations for obtaining a computationally preferable exact expression for $\mu_2(\bar{m}, n)$ are entirely analogous to those for $\mu_1(\bar{m}, n)$ described in the previous section (Section 4.1.1). Thus we omit details. As described in Section 3.1, $P_2(s, t, m, n)$ is zero for $m = 1$ and for $m = n$, so, from (4.1),

$$\mu_2(\bar{m}, n) = \int_0^1 \int_s^1 \beta(s, t) \, \frac{1}{n} \sum_{m=2}^{n-1} P_2(s, t, m, n) \, dt \, ds. \tag{4.28}$$

Therefore we first derive a computationally desirable expression for $\frac{1}{n} \sum_{m=2}^{n-1} P_2(s, t, m, n)$. Again, let $x := s,\ y := t - s,\ z := 1 - t$. Then

$$\begin{aligned}
&\frac{1}{n} \sum_{m=2}^{n-1} P_2(s, t, m, n) \\
&= \frac{1}{n} \sum_{m=2}^{n-1} \sum_{1 \le i \le m < j \le n} \frac{2}{j - i + 1} \binom{n}{i - 1, 1, j - i - 1, 1, n - j} x^{i-1} y^{j-i-1} z^{n-j} \\
&= \frac{1}{n} \sum_{m=2}^{n-1} S_1(m, n, x, y, z) - \frac{1}{n} \sum_{m=2}^{n-1} S_2(m, n, x, y, z) - \frac{1}{n} \sum_{m=2}^{n-1} S_3(m, n, x, y, z), \qquad (4.29)
\end{aligned}$$

where

$$\begin{aligned}
S_1(m, n, x, y, z) &:= \sum_{1 \le i < j \le n} \frac{2}{j - i + 1} \binom{n}{i - 1, 1, j - i - 1, 1, n - j} x^{i-1} y^{j-i-1} z^{n-j}, \\
S_2(m, n, x, y, z) &:= \sum_{m \le i < j \le n} \frac{2}{j - i + 1} \binom{n}{i - 1, 1, j - i - 1, 1, n - j} x^{i-1} y^{j-i-1} z^{n-j}, \\
S_3(m, n, x, y, z) &:= \sum_{1 \le i < j \le m} \frac{2}{j - i + 1} \binom{n}{i - 1, 1, j - i - 1, 1, n - j} x^{i-1} y^{j-i-1} z^{n-j}.
\end{aligned}$$

Fill and Janson [3] showed that $S_1(m, n, x, y, z) = 2 \sum_{j=2}^{n} (-1)^j \binom{n}{j} (t - s)^{j-2}$. Hence

$$\frac{1}{n} \sum_{m=2}^{n-1} S_1(m, n, x, y, z) = \frac{2(n - 2)}{n} \sum_{j=2}^{n} (-1)^j \binom{n}{j} (t - s)^{j-2}. \tag{4.30}$$

Following the derivations shown in (4.5) through (4.20), one can show that

$$\begin{aligned}
\frac{1}{n} \sum_{m=2}^{n-1} S_2(m, n, x, y, z) &= 2y^{-2} x [(x + z)^{n-1} - 1 + y(n - 1)] \qquad (4.31) \\
&= 2(t - s)^{-2} s \{[1 - (t - s)]^{n-1} - 1 + (t - s)(n - 1)\} \\
&= 2s \sum_{j=2}^{n-1} (-1)^j \binom{n - 1}{j} (t - s)^{j-2}. \qquad (4.32)
\end{aligned}$$

To obtain a similar expression for $\frac{1}{n}\sum_{m=2}^{n-1} S_3(m,n,x,y,z)$, we note that, letting $m' := n+1-m,\ i' := n+1-j,\ j' := n+1-i$,

$$
\begin{aligned}
S_3(m,n,x,y,z) &= \sum_{m'\le i'<j'\le n} \frac{2}{j'-i'+1}\binom{n}{n-j',1,j'-i'-1,1,i'-1} x^{n-j'} y^{j'-i'-1} z^{i'-1} \\
&= S_2(n+1-m,n,z,y,x).
\end{aligned}
$$

Thus

$$
\begin{aligned}
\frac{1}{n}\sum_{m=2}^{n-1} S_3(m,n,x,y,z) &= \frac{1}{n}\sum_{m=2}^{n-1} S_2(n+1-m,n,z,y,x) \\
&= \frac{1}{n}\sum_{m=2}^{n-1} S_2(m,n,z,y,x). \qquad (4.33)
\end{aligned}
$$

Inspecting (4.31)–(4.33), we find

$$
\frac{1}{n}\sum_{m=2}^{n-1} S_3(m,n,x,y,z) = 2(1-t)\sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}(t-s)^{j-2}. \qquad (4.34)
$$

From (4.29), (4.30), (4.32), and (4.34),

$$
\begin{aligned}
\frac{1}{n}\sum_{m=1}^{n-1} P_2(s,t,m,n) &= \frac{2(n-2)}{n}\sum_{j=2}^{n}(-1)^j\binom{n}{j}(t-s)^{j-2} - 2s\sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}(t-s)^{j-2} \\
&\quad -2(1-t)\sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}(t-s)^{j-2} \\
&= \frac{2(n-2)}{n}\sum_{j=2}^{n}(-1)^j\binom{n}{j}(t-s)^{j-2} - 2\sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}(t-s)^{j-2} \\
&\quad +2\sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}(t-s)^{j-1}. \qquad (4.35)
\end{aligned}
$$

Hence, from (4.28) and (4.35),

$$
\begin{aligned}
\mu_2(\bar{m},n) &= \frac{2(n-2)}{n}\int_0^1\int_s^1 \beta(s,t)\sum_{j=2}^{n}(-1)^j\binom{n}{j}(t-s)^{j-2}\,dt\,ds \\
&\quad -2\int_0^1\int_s^1 \beta(s,t)\sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}(t-s)^{j-2}\,dt\,ds \\
&\quad +2\int_0^1\int_s^1 \beta(s,t)\sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}(t-s)^{j-1}\,dt\,ds. \qquad (4.36)
\end{aligned}
$$

Fill and Janson [3] showed that

$$\int_0^1 \int_s^1 \beta(s,t) \sum_{j=2}^{n} (-1)^j \binom{n}{j} (t-s)^{j-2}\, dt\, ds = \sum_{j=2}^{n} \frac{(-1)^j \binom{n}{j}}{j(j-1)[1-2^{-(j-1)}]}. \tag{4.37}$$

A careful term-by-term inspection of the derivations shown in (4.24)–(4.27) reveals that

$$\begin{aligned}
\int_0^1 \int_s^1 \beta(s,t) \sum_{j=2}^{n-1} (-1)^j \binom{n-1}{j} (t-s)^{j-2}\, dt\, ds &= \sum_{j=2}^{n-1} \frac{(-1)^j \binom{n-1}{j}}{j(j-1)[1-2^{-(j-1)}]}, && (4.38)\\
\int_0^1 \int_s^1 \beta(s,t) \sum_{j=2}^{n-1} (-1)^j \binom{n-1}{j} (t-s)^{j-1}\, dt\, ds &= \sum_{j=2}^{n-1} \frac{(-1)^j \binom{n-1}{j}}{j(j+1)(1-2^{-j})}. && (4.39)
\end{aligned}$$

Combining (4.36)–(4.39), we obtain

$$\begin{aligned}
\mu_2(\bar{m}, n) &= \frac{2(n-2)}{n} \sum_{j=0}^{n-2} \frac{(-1)^j \binom{n}{j+2}}{(j+1)(j+2)[1-2^{-(j+1)}]} - 2 \sum_{j=2}^{n} \frac{(-1)^j \binom{n}{j}}{j(j-1)[1-2^{-(j-1)}]} + 2(n-1)\\
&= -\frac{4}{n} \sum_{j=2}^{n} \frac{(-1)^j \binom{n}{j}}{j(j-1)[1-2^{-(j-1)}]} + 2(n-1). && (4.40)
\end{aligned}$$

Finally, we complete the exact computation of $\mu(\bar{m}, n)$. From (4.2), (4.3), and (4.40), we have

$$\begin{aligned}
&\mu(\bar{m}, n) = 2\mu_1(\bar{m}, n) + \mu_2(\bar{m}, n)\\
&= 2(n-1) - 8 \sum_{j=3}^{n} \frac{(-1)^j \binom{n-1}{j-1}}{j(j-1)(j-2)} + \frac{4}{n} \sum_{j=2}^{n-1} B_j \frac{n-j+1-\binom{n}{j}}{j(j-1)(1-2^{-j})}\\
&\quad + \frac{4}{9} \sum_{j=2}^{n-1} \frac{(-1)^j \binom{n-1}{j}}{j-1} - 4 \sum_{j=3}^{n-1} B_j \frac{n-j+1-\binom{n-1}{j-1}}{j(j-1)(j-2)(1-2^{-j})} - 4 \sum_{j=2}^{n-1} \frac{(-1)^j \binom{n-1}{j}}{(j+1)j(j-1)(1-2^{-j})}\\
&\quad - \frac{4}{n} \sum_{j=2}^{n} \frac{(-1)^j \binom{n}{j}}{j(j-1)[1-2^{-(j-1)}]} + 2(n-1). && (4.41)
\end{aligned}$$

We rewrite or combine some of the terms in (4.41) for the asymptotic analysis of $\mu(\bar{m}, n)$

described in the next section. We define

$$\begin{aligned}
F_1(n) &:= \sum_{j=3}^{n} \frac{(-1)^j \binom{n}{j}}{(j-1)(j-2)}, \\
F_2(n) &:= \sum_{j=2}^{n-1} \frac{B_j}{j(1-2^{-j})} \left[\frac{n - \binom{n}{j}}{j-1} - 1 \right], \\
F_3(n) &:= \sum_{j=2}^{n-1} \frac{(-1)^j \binom{n-1}{j}}{j-1}, \\
F_4(n) &:= \sum_{j=3}^{n-1} \frac{B_j}{j(j-1)(1-2^{-j})} \left[\frac{n-1-\binom{n-1}{j-1}}{j-2} - 1 \right], \\
F_5(n) &:= \sum_{j=3}^{n} \frac{(-1)^j \binom{n}{j}}{j(j-1)(j-2)[1-2^{-(j-1)}]}.
\end{aligned}$$

The second, third, fourth, and fifth terms in (4.41) can be written as $-\frac{8}{n}F_1(n)$, $\frac{4}{n}F_2(n)$, $\frac{4}{9}F_3(n)$, and $-4F_4(n)$, respectively. The last three terms in (4.41) can be combined as follows:

$$\begin{aligned}
&-4 \sum_{j=2}^{n-1} \frac{(-1)^j \binom{n-1}{j}}{(j+1)j(j-1)(1-2^{-j})} - \frac{4}{n} \sum_{j=2}^{n} \frac{(-1)^j \binom{n}{j}}{j(j-1)[1-2^{-(j-1)}]} + 2(n-1) \\
&\qquad = \frac{4}{n} \sum_{j=3}^{n} \frac{(-1)^j \binom{n}{j}}{(j-1)(j-2)[1-2^{-(j-1)}]} - \frac{4}{n} \sum_{j=2}^{n} \frac{(-1)^j \binom{n}{j}}{j(j-1)[1-2^{-(j-1)}]} + 2(n-1) \\
&\qquad = \frac{8}{n} \sum_{j=3}^{n} \frac{(-1)^j \binom{n}{j}}{j(j-1)(j-2)[1-2^{-(j-1)}]} = \frac{8}{n} F_5(n).
\end{aligned}$$

Therefore

$$\mu(\bar{m}, n) = 2(n-1) - \tfrac{8}{n}F_1(n) + \tfrac{4}{n}F_2(n) + \tfrac{4}{9}F_3(n) - 4F_4(n) + \tfrac{8}{n}F_5(n). \tag{4.42}$$

4.2 Asymptotic Analysis of $\mu(\bar{m}, n)$

We derive an asymptotic expression for $\mu(\bar{m}, n)$ shown in (4.42). The computations described in this section are analogous to those in Section 3.2. Hence we merely sketch details to derive the asymptotic expression. First, we analyze $F_1(n)$. A routine complex-analytical argument similar to (but much easier than) the one described in Section 3.2 shows that

$$\begin{aligned}
F_1(n) &= (-1)^{n+1} \sum_{k=0}^{2} \mathrm{Res}_{s=k} \left[\frac{n!}{s(s-1)^2(s-2)^2(s-3)\cdots(s-n)} \right] \\
&= (-1)^{n+1} \left[\frac{(-1)^n}{2} + (-1)^n n H_{n-1} + \frac{(-1)^n}{2} n(n-1) \left(H_{n-2} - \frac{5}{2} \right) \right] \\
&= -\frac{1}{2} n(n-1) H_{n-2} + \frac{5}{4} n(n-1) - n H_{n-1} - \frac{1}{2} \\
&= -\frac{1}{2} n^2 \ln n + \left(\frac{5}{4} - \frac{\gamma}{2} \right) n^2 - n \ln n + \frac{n^2}{2(n-1)} - (\gamma+1)n + O(1).
\end{aligned} \tag{4.43}$$

Since $F_2(n)$ is equal to t_n, which is defined at (3.9) and analyzed in Section 3.2, we already have an asymptotic expression for $F_2(n)$. Next we derive an asymptotic expression for $F_3(n)$:

$$\begin{aligned} F_3(n) &= (-1)^n \sum_{k=0}^{1} \mathrm{Res}_{s=k} \left\{ \frac{(n-1)!}{s(s-1)^2(s-2)\cdots[s-(n-1)]} \right\} \\ &= nH_{n-2} - n - H_{n-2} + 2 \\ &= n\ln n + (\gamma - 1)n - \ln n + O(1). \end{aligned} \tag{4.44}$$

To obtain an asymptotic expression for $F_4(n)$, we closely follow the approach of Section 3.2. Let $\tilde{u}_n := F_4(n+1) - F_4(n)$. Then

$$\tilde{u}_n = -\sum_{j=3}^{n} \frac{B_j}{j(j-1)(j-2)(1-2^{-j})} \left[\binom{n-1}{j-2} - 1 \right].$$

Let $\tilde{v}_n := \tilde{u}_{n+1} - \tilde{u}_n$. Then, by computations similar to those performed for v_n in Section 3.2,

$$\begin{aligned} \tilde{v}_n &= -\sum_{k=0}^{n-2} (-1)^k \frac{\zeta(-2-k)}{(k+2)(k+1)[1-2^{-(k+3)}]} \binom{n-1}{k} \\ &= (-1)^{n+1} \sum_{k=1}^{3} \mathrm{Res}_{s=-k} \left\{ \frac{\zeta(-2-s)}{(s+2)(s+1)[1-2^{-(s+3)}]} \frac{(n-1)!}{s(s-1)\cdots[s-(n-1)]} \right\} \\ &\quad +(-1)^{n+1} \sum_{k\in\mathbb{Z}\setminus\{0\}} \mathrm{Res}_{s=-3+\chi_k} \left\{ \frac{\zeta(-2-s)}{(s+2)(s+1)[1-2^{-(s+3)}]} \frac{(n-1)!}{s(s-1)\cdots[s-(n-1)]} \right\} \\ &= \frac{1}{9n} - \frac{1}{n(n+1)} - \frac{1}{n(n+1)(n+2)} \left[\frac{\gamma}{\ln 2} - \frac{1}{2} - \frac{H_{n+2}}{n+2} \right] - \xi_n, \end{aligned}$$

where

$$\xi_n := \sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\zeta(1-\chi_k)\Gamma(1-\chi_k)\Gamma(n)}{(\ln 2)\Gamma(n+3-\chi_k)}.$$

Hence

$$\begin{aligned} \tilde{u}_n &= \tilde{u}_2 + \sum_{j=2}^{n-1} \tilde{v}_j \\ &= \frac{1}{9} H_{n-1} + \tilde{a} + \tilde{\xi}_n - \frac{1}{2\ln 2} \left(\frac{H_n}{n} - \frac{H_{n+1}}{n+1} \right) + \frac{1}{n} - \frac{3 + \ln 2 - 2\gamma}{4\ln 2} \frac{1}{n(n+1)}, \end{aligned}$$

where

$$\begin{aligned} \tilde{a} &:= \frac{7}{36\ln 2} - \frac{41}{72} - \frac{\gamma}{12\ln 2} - \sum_{k\in Z\setminus\{0\}} \frac{\zeta(1-\chi_k)\Gamma(1-\chi_k)}{(\ln 2)(2-\chi_k)\Gamma(4-\chi_k)}, \\ \tilde{\xi}_n &:= \sum_{k\in Z\setminus\{0\}} \frac{\zeta(1-\chi_k)\Gamma(1-\chi_k)\Gamma(n)}{(\ln 2)(2-\chi_k)\Gamma(n+2-\chi_k)}. \end{aligned}$$

Thus

$$\begin{aligned} F_4(n) &= F_4(2) + \sum_{j=2}^{n-1} \tilde{u}_j \\ &= \frac{1}{9} n H_{n-1} + \frac{8}{9} H_{n-1} + \left(\tilde{a} - \frac{1}{9}\right) n - \frac{8}{9} - \frac{3}{8 \ln 2} - \frac{3 + \ln 2 - 2\gamma}{8 \ln 2} - 2\tilde{a} + \tilde{b} - \tilde{\tilde{\xi}}_n \\ &\quad + \frac{1}{2 \ln 2} \frac{H_n}{n} + \frac{3 + \ln 2 - 2\gamma}{4 \ln 2} \frac{1}{n}, \end{aligned} \tag{4.45}$$

where

$$\begin{aligned} \tilde{b} &:= \sum_{k \in Z \setminus \{0\}} \frac{\zeta(1-\chi_k)\Gamma(1-\chi_k)}{(\ln 2)(2-\chi_k)(1-\chi_k)\Gamma(3-\chi_k)}, \\ \tilde{\tilde{\xi}}_n &:= \sum_{k \in Z \setminus \{0\}} \frac{\zeta(1-\chi_k)\Gamma(1-\chi_k)\Gamma(n)}{(\ln 2)(2-\chi_k)(1-\chi_k)\Gamma(n+1-\chi_k)}. \end{aligned}$$

Therefore

$$F_4(n) = \frac{1}{9} n \ln n + \left(\tilde{a} + \frac{1}{9}\gamma - \frac{1}{9}\right) n + \frac{8}{9} \ln n + O(1). \tag{4.46}$$

Finally, we analyze $F_5(n)$. By computations that are entirely analogous to those performed for $F_1(n)$, $F_2(n)$, and $F_4(n)$,

$$\begin{aligned} F_5(n) &= (-1)^{n+1} \sum_{k=0}^{2} \mathrm{Res}_{s=k} \left\{ \frac{n!}{[1 - 2^{-(s-1)}] s^2 (s-1)^2 (s-2)^2 (s-3) \cdots (s-n)} \right\} \\ &\quad + (-1)^{n+1} \sum_{k \in Z \setminus \{0\}} \mathrm{Res}_{s=1+\chi_k} \left\{ \frac{n!}{[1 - 2^{-(s-1)}] s^2 (s-1)^2 (s-2)^2 (s-3) \cdots (s-n)} \right\} \\ &= \frac{1}{4}(2H_n + 3 + 4 \ln 2) - \frac{n(n-1)}{2}(H_{n-2} - \ln 2 - 3) \\ &\quad - n \left[\frac{1}{2 \ln 2} (H_{n-1})^2 + \left(\frac{1}{2} - \frac{1}{\ln 2}\right) H_{n-1} + \frac{1}{2 \ln 2} H_{n-1}^{(2)} + \frac{2}{\ln 2} + \frac{\ln 2}{12} - \frac{1}{2} \right] \\ &\quad + \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{\Gamma(-1-\chi_k)\Gamma(n+1)}{(\ln 2)\chi_k(\chi_k^2 - 1)\Gamma(n - \chi_k)} \\ &= -\frac{1}{2} n^2 \ln n + \frac{3 + \ln 2 - \gamma}{2} n^2 - \frac{1}{2 \ln 2} n (\ln n)^2 + \left(\frac{1}{\ln 2} - \frac{1}{2}\right) n \ln n + O(n). \end{aligned} \tag{4.47}$$

Therefore, from (4.42)–(4.44) and (4.46)–(4.47), we obtain the following asymptotic formula for $\mu(\bar{m}, n)$:

$$\mu(\bar{m}, n) = 4(1 + \ln 2 - \tilde{a})n - \frac{4}{\ln 2}(\ln n)^2 + 4\left(\frac{2}{\ln 2} - 1\right) \ln n + O(1). \tag{4.48}$$

The asymptotic slope $4(1 + \ln 2 - \tilde{a})$ is approximately 8.20731.

5 Derivation of a Closed Formula for $\mu(m,n)$

The exact expression for $\mu(m,n)$ obtained in Section 2 [see (2.8)] involves infinite summation and integration. Hence it is not a preferable form for numerically computing the expectation. In this section, we establish another exact expression for $\mu(m,n)$ that only involves finite summation. We also use the formula to compute $\mu(m,n)$ for $m=1,\dots,n$, $n=2,\dots,20$.

As described in Section 2, it follows from equations (2.6)–(2.8) that

$$\mu(m,n) \quad = \quad \mu_1(m,n)+\mu_2(m,n)+\mu_3(m,n), \tag{5.1}$$

where, for $q=1,2,3$,

$$\mu_q(m,n) := \sum_{k=0}^{\infty}\sum_{l=1}^{2^k}\int_{(l-1)2^{-k}}^{(l-\frac{1}{2})2^{-k}}\int_{(l-\frac{1}{2})2^{-k}}^{l2^{-k}}(k+1)P_q(s,t,m,n)\ dt\ ds. \tag{5.2}$$

The same technique can be applied to eliminate the infinite summation and integration from each $\mu_q(m,n)$. We describe the technique for obtaining a closed expression of $\mu_1(m,n)$ in detail.

First, we transform $P_1(s,t,m,n)$ shown in (2.3) so that we can eliminate the integration in $\mu_1(m,n)$. Define

$$C_1(i,j) := \mathbf{1}(1\le m\le i<j\le n)\frac{2}{j-m+1}\binom{n}{i-1,1,j-i-1,1,n-j}, \tag{5.3}$$

where $\mathbf{1}(1\le m\le i<j\le n)$ is an indicator function that equals 1 if the event in braces holds and 0 otherwise. Since

$$\begin{aligned}&s^{i-1}(t-s)^{j-i-1}(1-t)^{n-j}\\ &= \quad s^{i-1}\sum_{u=0}^{j-i-1}\binom{j-i-1}{u}t^u(-1)^{j-i-1-u}s^{j-i-1-u}\sum_{v=0}^{n-j}\binom{n-j}{v}(-1)^{n-j-v}t^{n-j-v},\end{aligned}$$

it follows that

$$\begin{aligned}P_1(s,t,m,n) &= \sum_{m\le i<j\le n}C_1(i,j)\sum_{u=0}^{j-i-1}\sum_{v=0}^{n-j}\binom{j-i-1}{u}\binom{n-j}{v}s^{j-u-2}t^{n-j-v+u}(-1)^{n-i-u-v-1}\\ &= \sum_{m\le i<j\le n}C_1(i,j)\sum_{f=i-1}^{j-2}\sum_{h=j-f-2}^{n-f-2}s^ft^h\binom{j-i-1}{f-i+1}\binom{n-j}{h-j+f+2}(-1)^{h-i-j+1}\\ &= \sum_{f=m-1}^{n-2}\sum_{h=0}^{n-f-2}s^ft^hC_2(f,h),\end{aligned} \tag{5.4}$$

where

$$C_2(f,h) := \sum_{i=m}^{f+1}\sum_{j=f+2}^{f+h+2}C_1(i,j)\binom{j-i-1}{f-i+1}\binom{n-j}{h-j+f+2}(-1)^{h-i-j+1}.$$

Thus, from (5.2) and (5.4), we can eliminate the integration in $\mu_1(m,n)$ and express it using polynomials in l:

$$\mu_1(m,n) = \sum_{f=m-1}^{n-2}\sum_{h=0}^{n-f-2} C_3(f,h)\sum_{k=0}^{\infty}(k+1)\sum_{l=1}^{2^k} 2^{-k(f+h+2)}[l^{h+1}-(l-\tfrac{1}{2})^{h+1}][(l-\tfrac{1}{2})^{f+1}-(l-1)^{f+1}], \tag{5.5}$$

where

$$C_3(f,h) := \frac{1}{(n+1)(f+1)}C_2(f,h).$$

Note that

$$\begin{aligned} l^{h+1}-\left(l-\frac{1}{2}\right)^{h+1} &= -\sum_{j=0}^{h}\binom{h+1}{j}l^j\left(-\frac{1}{2}\right)^{h+1-j}, \\ \left(l-\frac{1}{2}\right)^{f+1}-(l-1)^{f+1} &= -\sum_{j'=0}^{f}\binom{f+1}{j'}l^{j'}(-1)^{f+1-j'}\left[1-\left(\frac{1}{2}\right)^{f+1-j'}\right]. \end{aligned}$$

Hence

$$\begin{aligned} &\left[l^{h+1}-\left(l-\frac{1}{2}\right)^{h+1}\right]\left[\left(l-\frac{1}{2}\right)^{f+1}-(l-1)^{f+1}\right] \\ &= \sum_{j'=0}^{f}\sum_{j=0}^{h}\binom{f+1}{j'}\binom{h+1}{j}(-1)^{f+h-j'-j}\left[1-\left(\frac{1}{2}\right)^{f+1-j'}\right]\left(\frac{1}{2}\right)^{h+1-j}l^{j'+j}, \end{aligned}$$

which can be rearranged to

$$\sum_{j=1}^{f+h+1} C_4(f,h,j)l^{j-1}, \tag{5.6}$$

where

$$C_4(f,h,j) := (-1)^{f+h-j+1}\left(\frac{1}{2}\right)^{h-j+2}\sum_{j'=0\vee(j-1-h)}^{(j-1)\wedge f}\binom{f+1}{j'}\binom{h+1}{j-1-j'}\left[1-\left(\frac{1}{2}\right)^{f+1-j'}\right]\left(\frac{1}{2}\right)^{j'}.$$

Therefore, from (5.5)–(5.6), we obtain

$$\begin{aligned} \mu_1(m,n) &= \sum_{f=m-1}^{n-2}\sum_{h=0}^{n-f-2} C_3(f,h)\sum_{k=0}^{\infty}(k+1)\sum_{l=1}^{2^k} 2^{-k(f+h+2)}\sum_{j=1}^{f+h+1} C_4(f,h,j)l^{j-1} \\ &= \sum_{f=m-1}^{n-2}\sum_{h=0}^{n-f-2}\sum_{j=1}^{f+h+1} C_5(f,h,j)\sum_{k=0}^{\infty}(k+1)2^{-k(f+h+2)}\sum_{l=1}^{2^k} l^{j-1}, \end{aligned}$$

where

$$C_5(f,h,j) := C_3(f,h) \cdot C_4(f,h,j).$$

Here, as described in Section 3.1,

$$\sum_{l=1}^{2^k} l^{j-1} = \sum_{r=0}^{j-1} a_{j,r} 2^{k(j-r)},$$

where $a_{j,r}$ is defined by (3.4). Now define

$$C_6(f,h,j,r) := a_{j,r}\; C_5(f,h,j).$$

Then

$$\begin{aligned} \mu_1(m,n) &= \sum_{f=m-1}^{n-2} \sum_{h=0}^{n-f-2} \sum_{j=1}^{f+h+1} \sum_{r=0}^{j-1} a_{j,r}\, C_5(f,h,j) \sum_{k=0}^{\infty} (k+1) 2^{-k(f+h+2+r-j)} \\ &= \sum_{f=m-1}^{n-2} \sum_{h=0}^{n-f-2} \sum_{j=1}^{f+h+1} \sum_{r=0}^{j-1} C_6(f,h,j,r)[1 - 2^{-(f+h+2+r-j)}]^{-2} \\ &= \sum_{a=1}^{n-1} C_7(a)(1-2^{-a})^{-2}, \end{aligned} \tag{5.7}$$

where

$$C_7(a) := \sum_{f=m-1}^{n-2} \sum_{h=\alpha}^{n-f-2} \sum_{j=\beta}^{f+h+1} C_6(f,h,j,a+j-(f+h+2)),$$

in which $\alpha := 0 \bigvee (a-f-1)$ and $\beta := 1 \bigvee (f+h+2-a)$.

The procedure described above can be applied to derive analogous exact formulae for $\mu_2(m,n)$ and $\mu_3(m,n)$. In order to derive the analogous exact formula for $\mu_2(m,n)$, one need only start the derivation by changing the indicator function in $C_1(i,j)$ [see (5.3)] to $\mathbf{1}(1 \le i < m < j \le n)$ and follow each step of the procedure; for $\mu_3(m,n)$, start the derivation by changing the indicator function to $\mathbf{1}(1 \le i < j \le m \le n)$.

Using the closed exact formulae of $\mu_1(m,n)$, $\mu_2(m,n)$, and $\mu_3(m,n)$, we computed $\mu(m,n)$ for $n = 2,3,\ldots,20$ and $m = 1,2,\ldots,n$. Figure 1 shows the results, which suggest the following: (i) for fixed n, $\mu(m,n)$ increases in m for $m \le \frac{n+1}{2}$ and is symmetric about $\frac{n+1}{2}$; (ii) for fixed m, $\mu(m,n)$ increases in n; (iii) $\max_m \mu(m,n)$ is asymptotically linear in n.

6 Discussion

Our investigation of the bit complexity of `Quickselect` revealed that the expected number of bit comparisons required by `Quickselect` to find the smallest or largest key from a set of n keys is asymptotically linear in n with the asymptotic slope approximately equal to 5.27938. Hence asymptotically it differs from the expected number of *key* comparisons to achieve the same task only by a constant factor. (The expectation for key comparisons is asymptotically $2n$; see Knuth [11] and Mahmoud *et al.* [15]). This result is rather contrastive to the

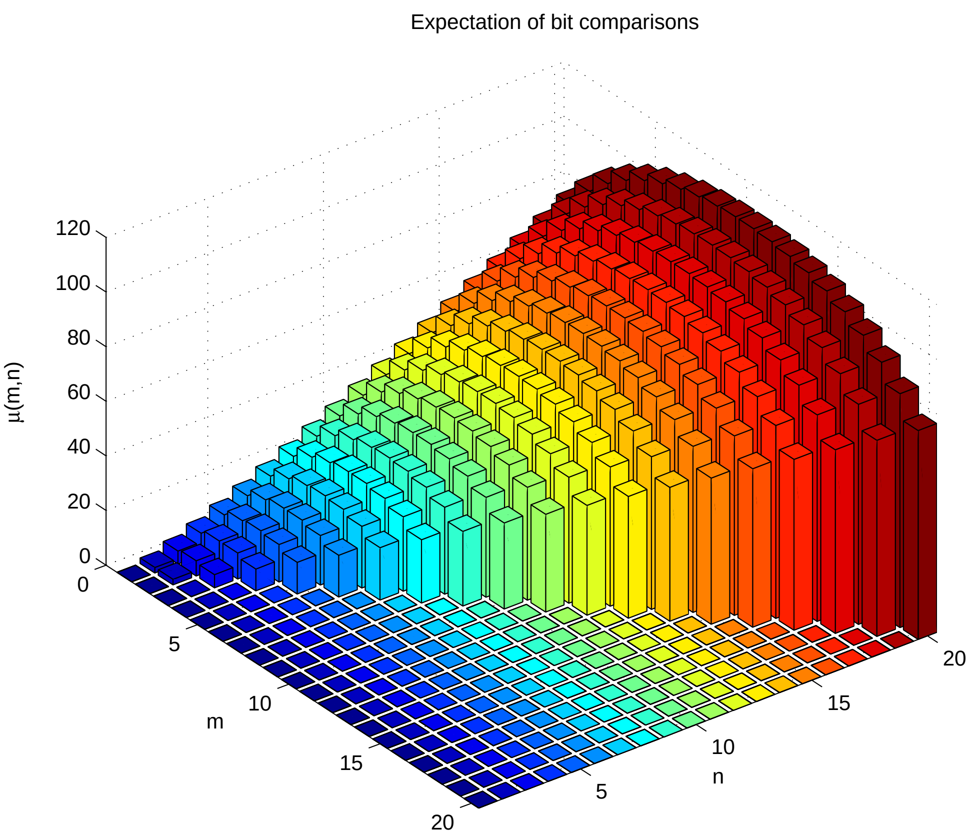

Figure 1: Expected number of bit comparisons for `Quickselect`. The closed formulae for $\mu_1(m,n)$, $\mu_2(m,n)$, and $\mu_3(m,n)$ were used to compute $\mu(m,n)$ for $n = 1, 2, \ldots, 20$ (n represents the number of keys) and $m = 1, 2, \ldots, n$ (m represents the rank of the target key).

`Quicksort` case in which (see Fill and Janson [3]) the expected number of bit comparisons is asymptotically $n(\ln n)(\lg n)$ whereas the expected number of key comparisons is asymptotically $2n \ln n$. Our analysis also showed that the expected number of bit comparisons for the average case remains asymptotically linear in n with the lead-order coefficient approximately equal to 8.20731. Again, the expected number is asymptotically different from that of key comparisons for the average case only by a constant factor. (The expected number of key comparisons for the average case is asymptotically $3n$; see Mahmoud *et al.* [15].)

Although we have yet to establish a formula analogous to (3.8) and (4.42) for the expected number of bit comparisons to find the m-th key for fixed m, we established an exact expression that only requires finite summation and used it to obtain the results shown in Figure 1. However, the formula remains complex. Written as a single expression, $\mu(m,n)$ is a seven-fold sum of rather elementary terms with each sum having order n terms (in the worst case); in this sense, the running time of the algorithm for computing $\mu(m,n)$ is of order n^7. The expression for $\mu(m,n)$ does not allow us to derive an asymptotic formula for it or to prove the three conjectures described at the end of Section 5. The situation is substantially better for the expected number of *key* comparisons to find the m-th key from a set of n keys; Knuth [11] showed that the expectation can be written as $2[n+3+(n+1)H_n-(m+2)H_m-(n+3-m)H_{n+1-m}]$.

In this paper, we considered independent and uniformly distributed keys in (0,1). In this case, each bit in bit strings is 1 with probability 0.5. Building on the present work and that of Fill and Janson [3], much more general key-distributions are treated by Vallée *et al.* [21]. Their generalization further elucidates the complexity of `Quickselect` and other algorithms.

Acknowledgment. We thank Philippe Flajolet, Svante Janson, and Helmut Prodinger for helpful discussions.

7 Appendix

In order to prove (3.20), it suffices to show that, for any positive integer m,

$$\int_{n-\theta-i\infty}^{n-\theta+i\infty} \zeta(-1-s)\, m^{-s} \frac{ds}{(s+1)s\cdots(s-n)} = 0$$

(note that $n \geq 2$ and $0 < \theta < 1$). Letting $t := -1-s$, it is thus sufficient to show that

$$J := \int_{-(n+1)+\theta-i\infty}^{-(n+1)+\theta+i\infty} \zeta(t)\, m^{t} \frac{dt}{t(t+1)\cdots[t+(n+1)]} = 0.$$

Using the residue theorem, we obtain

$$J = -2\pi i \left[\sum_{k=0}^{n}(-1)^k \frac{\zeta(-k)\, m^{-k}}{k!(n+1-k)!} + \frac{m}{(n+2)!}\right] + \int_{2-i\infty}^{2+i\infty} \zeta(t)\, m^{t} \frac{dt}{t(t+1)\cdots[t+(n+1)]}; \tag{7.1}$$

The "2" in the second term here could just as well be any real number exceeding 1. Here

$$\sum_{k=0}^{n}(-1)^k \frac{\zeta(-k)\, m^{-k}}{k!(n+1-k)!} = -\frac{1}{2\,(n+1)!} + \sum_{k=1}^{n} \frac{B_{k+1}\, m^{-k}}{(k+1)!(n+1-k)!} = \sum_{k=1}^{n+1} \frac{B_k\, m^{-(k-1)}}{k!(n+2-k)!}.$$

Therefore

$$\begin{aligned} &\sum_{k=0}^{n}(-1)^k \frac{\zeta(-k)\, m^{-k}}{k!(n+1-k)!} + \frac{m}{(n+2)!} \\ &= \frac{m^{-(n+1)}}{(n+1)!}\left[\sum_{k=1}^{n+1} \frac{B_k\,(n+1)!}{k!(n+2-k)!} m^{n+2-k} + \frac{m^{n+2}}{n+2}\right] \\ &= \frac{m^{-(n+1)}}{(n+1)!}\sum_{k=1}^{m-1} k^{n+1} = \frac{1}{(n+1)!}\sum_{k=1}^{m-1}\left(1-\frac{k}{m}\right)^{n+1}; \end{aligned} \tag{7.2}$$

for the second equality, see Knuth [12] (Exercise 1.2.11.2-4). On the other hand, Flajolet *et al.* [4] showed that

$$\int_{2-i\infty}^{2+i\infty} \zeta(t)\, m^t \frac{dt}{t(t+1)\cdots[t+(n+1)]} = \frac{2\pi i}{(n+1)!}\sum_{k=1}^{m-1}\left(1-\frac{k}{m}\right)^{n+1}. \tag{7.3}$$

Thus it follows from (7.1)–(7.3) that $J = 0$.

References

[1] L. Devroye. On the probablistic worst-case time of "Find". *Algorithmica*, 31:291–303, 2001.

[2] J. A. Fill and S. Janson. Quicksort asymptotics. *Journal of Algorithms*, 44:4–28, 2002.

[3] J. A. Fill and S. Janson. The number of bit comparisons used by Quicksort: An average-case analysis. *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, pages 293–300, 2004.

[4] P. Flajolet, P. Grabner, P. Kirschenhofer, H. Prodinger, and R. F. Tichy. Mellin transforms and asymptotics: digital sums. *Theoretical Computer Science*, 123:291–314, 1994.

[5] P. Flajolet and R. Sedgewick. Mellin transforms and asymptotics: Finite differences and Rice's integrals. *Theoretical Computer Science*, 144:101–124, 1995.

[6] P. J. Grabner and H. Prodinger. On a constant arising in the analysis of bit comparisons in Quickselect. *Quaestiones Mathematicae*, 31:303–306, 2008.

[7] R. Grübel and U. Rösler. Asymptotic distribution theory for Hoare's selection algorithm. *Advances in Applied Probability*, 28:252–269, 1996.

[8] C. R. Hoare. Find (algorithm 65). *Communications of the ACM*, 4:321–322, 1961.

[9] H. Hwang and T. Tsai. Quickselect and the Dickman function. *Combinatorics, Probability and Computing*, 11:353–371, 2002.

[10] C. Knessl and W. Szpankowski. Quicksort algorithm again revisited. *Discrete Mathematics and Theoretical Computer Science*, 3:43–64, 1999.

[11] D. E. Knuth. Mathematical analysis of algorithms. In *Information Processing 71 (Proceedings of IFIP Congress, Ljubljana, 1971)*, pages 19–27. North-Holland, Amsterdam, 1972.

[12] D. E. Knuth. *The Art of Computer Programming. Volume 1: Fundamental Algorithms*. Addison-Wesley, Reading, Massachusetts, 1998.

[13] D. E. Knuth. *The Art of Computer Programming. Volume 3: Sorting and Searching*. Addison-Wesley, Reading, Massachusetts, 1998.

[14] J. Lent and H. M. Mahmoud. Average-case analysis of multiple Quickselect: An algorithm for finding order statistics. *Statistics and Probability Letters*, 28:299–310, 1996.

[15] H. M. Mahmoud, R. Modarres, and R. T. Smythe. Analysis of Quickselect: An algorithm for order statistics. *RAIRO Informatique Théorique et Applications*, 29:255–276, 1995.

[16] H. M. Mahmoud and R. T. Smythe. Probabilistic analysis of multiple Quickselect. *Algorithmica*, 22:569–584, 1998.

[17] R. Neininger and L. Rüschendorf. Rates of convergence for Quickselect. *Journal of Algorithms*, 44:51–62, 2002.

[18] H. Prodinger. Multiple Quickselect—Hoare's Find algorithm for several elements. *Information Processing Letters*, 56:123–129, 1995.

[19] M. Régnier. A limiting distribution of Quicksort. *RAIRO Informatique Théorique et Applications*, 23:335–343, 1989.

[20] U. Rösler. A limit theorem for Quicksort. *RAIRO Informatique Théorique et Applications*, 25:85–100, 1991.

[21] B. Vallée, J. Clément, J. A. Fill, and P. Flajolet. The number of symbol comparisons in Quicksort and Quickselect. *Preprint*, 2009.