THE ABSENT-MINDED PASSENGERS PROBLEM: A MOTIVATING CHALLENGE SOLVED BY COMPUTER ALGEBRA

CARSTEN SCHNEIDER

ABSTRACT. In [6] an exciting case study has been initiated in which experimental mathematics and symbolic computation are utilized to discover new properties concerning the so-called Absent-Minded Passengers Problem. Based on these results, Doron Zeilberger raised some challenging tasks to gain further probabilistic insight. In this note we report on this enterprise. In particular, we demonstrate how the computer algebra packages of RISC can be used to carry out the underlying heavy calculations.

1. A CHALLENGING EMAIL

On January 22, 2020 I received the following email by Doron Zeilberger:

```
Dear Carsten,

I (and Shalosh) just posted a paper

https://arxiv.org/abs/2001.06839

with a challenge to you (see the middle of page 4)

Can you (and Sigma) extend theorem 5 of that paper
to the general case with k absent-minded passengers?

These expressions should be expressions in n,k,
the harmonic numbers H_{n-1}, H_{k-1} and their
generalizations (the partial sum , to n-1, k-1, respectively of
zeta(n) and zeta(k)).

If you and Sigma can do the fourth moment, and derive
the asymptotic in n (with a fixed but arbitrary k), I will
donate $100$ to the OEIS in your honor.

...

Best wishes,
Doron
```

When I received this email, I was thrilled: First, if one gets such an email – in particular from Doron Zeilberger, one is automatically eager to solve it. And second, I highly appreciate "The On-Line Encyclopedia of Integer Sequences" (OEIS) at `http://www.oeis.org` and supporting it by a donation of Doron Zeilberger gave an extra strong motivation.

Summarizing, the above email provoked various heavy calculations by means of computer algebra that will be introduced in the following.

2. The underlying problem and symbolic summation

The combinatorial problem of absent-minded passengers can be introduced as follows.

Definition 2.1. Consider a plane with $n \geq 2$ seats and suppose that n passengers enter the plane step-wise taking their seats. In addition, suppose that the first $k \geq 1$ passengers are absent-minded, i.e., they lost their seat ticket and take a seat (uniformally) at random. The remaining $n - k$ not absent-minded (but shy) passengers take their dedicated seats (as given in the plane ticket) if it is still free; otherwise, they choose (uniformally) at random one of the still available free seats. For $0 \leq r \leq n$, let $p_{n,k,r}$ be the probability that exactly r passengers sit in the wrong seat and let X_n be the random variable for "the number of passengers sitting in the wrong seat". Then the expected value for the passengers sitting in the wrong seat is

$$E(X_n) = \sum_{r=1}^{n} r \, p_{n,k,r}$$

and its variance is

$$V(X_n) = E(X_n^2) - E(X_n)^2 = \sum_{r=1}^{n} r^2 \, p_{n,k,r} - E(X_n)^2.$$

The situation of one absent-minded passenger ($k = 1$) has been considered in [17, 5] and has been explored further in [7] for the general case $k \geq 1$. Among other fascinating results, closed forms for $E(X_n)$ and $V(X_n)$ have been obtained in [7] by skillful combinatorial arguments. More precisely, the definite sum representations

$$E(X_n) = \frac{k(-1+n)}{n} + \sum_{i=1}^{-k+n} \frac{k}{1-i+n} \tag{2.1}$$

and

$$V(X_n) = \sum_{i=1}^{-k+n} \frac{(1-i-k+n)\left(1 - \frac{1-i-k+n}{1-i+n}\right)}{1-i+n} \tag{2.2}$$
$$] \quad + 2\left(\frac{(-1+k)k}{2(-1+n)n^2} + \sum_{i=1}^{k} \sum_{j=1}^{-k+n} \frac{\frac{1-j-k+n}{-j+n} - \frac{1-j-k+n}{1-j+n}}{n}\right)$$

have been derived and simplified to

$$E(X_n) = \frac{k(-1+n)}{n} - kS_1(k) + kS_1(n) \tag{2.3}$$

and

$$V(X_n) = \frac{2k - k^2 - 2n - 2kn + 2k^2n + 2n^2 - kn^2}{(-1+n)n^2} \tag{2.4}$$

$$- \frac{k(2+n)S_1(k)}{n} + \frac{k(2+n)S_1(n)}{n} + k^2 S_2(k) - k^2 S_2(n)$$

in terms of the harmonic numbers

$$S_1(n) = \sum_{i=1}^{n} \frac{1}{i^o}$$

of order $o \geq 1$ (often they are also denoted by $H_n^{(o)} = S_o(n)$ with the special case $H_n = S_1(n)$).
In the following we will prefer to write such expressions in terms of the modified harmonic numbers

$$\bar{S}_o(n) = S_o(n-1) = \sum_{i=1}^{n-1} \frac{1}{i^o}$$

yielding, e.g., the more compact expressions

$$E(X_n) = -1 + k - k\bar{S}_1(k) + k\bar{S}_1(n) \tag{2.5}$$

and

$$V(X_n) = \frac{(-1+k)k}{(-1+n)n} - \frac{k(2+n)\bar{S}_1(k)}{n} + \frac{k(2+n)\bar{S}_1(n)}{n} + k^2\bar{S}_2(k) - k^2\bar{S}_2(n). \tag{2.6}$$

While the simplification from (2.1) to (2.3) is straightforward, more work has to be carried out to derive (2.4) from (2.2). In [7] further details are suppressed how these simplifications have been carried out. Here I want to point out that such (usually painful) classical manipulations are meanwhile obsolete. For instance, by loading in the computer algebra packages

In[1]:= **<< Sigma.m**

> Sigma - A summation package by Carsten Schneider © RISC-JKU

In[2]:= **<< HarmonicSums.m**

> HarmonicSums by Jakob Ablinger © RISC-JKU

In[3]:= **<< EvaluateMultiSums.m**

> EvaluateMultiSums by Carsten Scneider © RISC-JKU

into Mathematica the multi-sum expression for the variance

In[4]:= $\mathbf{V} = \dfrac{\mathbf{k(-1+n)}}{\mathbf{n^2}} + \displaystyle\sum_{i=1}^{-k+n} \dfrac{(1-i-k+n)(1-\frac{1-i-k+n}{1-i+n})}{1-i+n}$
$+ \, 2 \left(\dfrac{(-1+k)k}{2(-1+n)n^2} + \displaystyle\sum_{i=1}^{k} \displaystyle\sum_{j=1}^{-k+n} \dfrac{\frac{1-j-k+n}{-j+n} - \frac{1-j-k+n}{1-j+n}}{n} \right);$

can be simplified to a closed form within seconds by executing the command[1]

In[5]:= **EvaluateMultiSum[V, {}, {k, n}, {1, 2}, {n, Infinity}]**

Out[5]= $\dfrac{2k - k^2 - 2n - 2kn + 2k^2n + 2n^2 - kn^2}{(-1+n)n^2} - \dfrac{k(2+n)S[1,k]}{n} + \dfrac{k(2+n)S[1,n]}{n} + k^2S[2,k] - k^2S[2,n]$

Remark. Internally, the command `EvaluateMultiSums` of the package `EvaluateMultiSums.m` [14] uses the summation paradigms of telescoping, creative telescoping and recurrence solving of the summation package `Sigma.m` [12]; the underlying algorithms generalize the hypergeometric case [9] to the class of indefinite nested sums and products in the setting of difference fields and rings [8, 15, 16]. In addition, the calculations are supported by special function

---

[1]Within the `Sigma` package (and later the `HarmonicSums` package) $S_o(n)$ is denoted by `S[o,n]`.

algorithms of the package `HarmonicSums.m` [2, 3]. We note that the user is completely dispensed from carrying out the summation tools explicitly. However, all the calculation steps are accomplished by proof certificates (based on the creative telescoping paradigm introduced in [18]). Thus if necessary, a rigorous correctness proof can be extracted.

3. The generating function approach for large moments

As elaborated in details in [6] the expectation, the variance and higher moments can be elegantly described with the generating function approach. Namely, consider the generating function

$$f_n^{(k)}(w) = \sum_{r=0}^{n} p_{n,k,r}\, w^r$$

of the probability $p_{n,k,r}$ introduced in Definition 2.1. Then the expectation and variance (see also Definition 2.1) can be straightforwardly connected to the generating function via

$$E(X_n) = \frac{d}{dw} f_n^{(k)}(w)|_{w=1}$$

and

$$V(X_n) = \left(w\frac{d}{dw}\right)^2 f_n^{(k)}(w)|_{w=1}.$$

More generally, define

$$M_0(n,k) = 1$$

and

$$M_l(n,k) = \left(w\frac{d}{dw}\right)^l f_n^{(k)}(w)|_{w=1} \tag{3.1}$$

for $l \geq 1$; note that $M_1(n,k) = E(X_n)$. Then one can define the lth moment by

$$m_l(n,k) = \sum_{i=0}^{l} \binom{l}{i}(-1)^{l-i} M_i(n,k) M_1^{l-i}(n,k). \tag{3.2}$$

In the above email but also in [6] the task was assigned to compute (besides the known moments $m_1(n,k) = 0$ and $m_2(n,k) = V(X_n)$) further moments $m_l(n,k)$ (at least for $l = 3, 4$).

The combinatorial approach of [7] to derive multi-sum expressions of $m_l(X_n)$ for larger l seems hopeless. Even though it would have been pure fun to challenge `Sigma.m`, similarly as in in [10], with more complicated sums than (2.3). Another exciting approach has been carried out in [6, Theorem 5] by guessing the moments $m_l(n,k)$ for the special case $k = 1$ and $l = 2, \ldots, 8$. As it turns out (and proposed by Doron Zeilberger in his email), this attempt can be pushed forward by the following powerful formula given in [6, Theorem 3]:

$$\begin{aligned}
f_n^{(k)}(w) &= \frac{1}{n}\sum_{r=0}^{k} r!\binom{k}{r} w^r(1-w)^{k-r}\prod_{i=0}^{n-k-1}(r\,w+i+1) \\
&= \frac{1+r\,w}{n}\sum_{r=0}^{k} r!\binom{k}{r} w^r(1-w)^{k-r}(2+r\,w)_{n-k-1};
\end{aligned} \tag{3.3}$$

here $(x)_k = x(x-1)\ldots(x-k+1)$ denotes the Pochhammer symbol. As a consequence, one can calculate straightforwardly any value of $M_l(n,k)$ and thus of $m_l(n,k)$ for particularly chosen n and k. E.g., we can compute

In[6]:= $\mathbf{F = \dfrac{1 + r\,w}{n} \sum\limits_{r=0}^{k} r! \dbinom{k}{r} w^r (1-w)^{k-r} (2 + r\,w)_{n-k-1}}$/.$\{\mathbf{n \to 10, k \to 2}\}$;

In[7]:= $\mathbf{MSpec0 = F}$/.$\mathbf{w \to 1}$

Out[7]= 1

In[8]:= $\mathbf{MSpec1 = w * D[F, w]}$/.$\mathbf{w \to 1}$

Out[8]= $\dfrac{5869}{1260}$

In[9]:= $\mathbf{MSpec2 = w * D[w * D[F, w], w]}$/.$\mathbf{w \to 1}$

Out[9]= $\dfrac{50293}{2100}$

In[10]:= $\mathbf{MSpec3 = w * D[w * D[w * D[F, w], w], w]}$/.$\mathbf{w \to 1}$

Out[10]= $\dfrac{9966821}{75600}$

and activating the formula (3.2) for $n = 10$, $k = 2$ and $l = 3$ yields the third moment

In[11]:= $\mathbf{mSpec3 = -MSpec0 * MSpec1^3 + 3 * MSpec1^2 - 3 * MSpec2 * MSpec1 + MSpec3}$

Out[11]= $-\dfrac{702653939}{1000188000}$

But even more is possible with the hypergeometric sum representation (3.3). As already promoted in [6] one can derive easily the closed form expressions of $M_l(X_n)$ for symbolic n and k. For this task we observe that

$$M_l(n,k) = \left(w \tfrac{d}{dw}\right)^l f_n^{(k)}(w)\big|_{w=1}$$

$$= \left(w \tfrac{d}{dw}\right)^l \left[\frac{1 + r\,w}{n} \sum_{r=0}^{k} r! \binom{k}{r} w^r (1-w)^{k-r} (2 + r\,w)_{n-k-1}\right]\Bigg|_{w=1}$$

$$= \left(w \tfrac{d}{dw}\right)^l \left[\frac{1 + r\,w}{n} \sum_{r=\max(0,k-l)}^{k} r! \binom{k}{r} w^r (1-w)^{k-r} (2 + r\,w)_{n-k-1}\right]\Bigg|_{w=1}, \quad (3.4)$$

i.e., for any $l \in \mathbb{Z}_{\geq 0}$ at most $l+1$ summands contribute and the remaining summands vanish with the evaluation $w = 1$. Moreover the differentiation of the arising building blocks in (3.4) can be carried out easily. For instance if we differentiate $(2 + r\,w)_{n-k-1}$ twice w.r.t. w by using the Mathematica-command D and the Sigma-commands ToSigma and CollectProdSum we get

In[12]:= $\mathbf{D[D[(2 + r\,w)_{n-k-1}, w], w]}$//$\mathbf{ToSigma}$//$\mathbf{CollectProdSum}$

Out[12]= $\left(\dfrac{2r^2}{(1+rw)^2} + \dfrac{2r^2 S[1, rw]}{1+rw} + r^2 S[1, rw]^2 - \dfrac{2r^2 S[1, -k+n+rw]}{1+rw} - 2r^2 S[1, rw] S[1, -k+n+rw] + r^2 S[1, -k+n+rw]^2\right.$

$\left. + r^2 S[2, rw] - r^2 S[2, -k+n+rw]\right)(2+rw)_{-1-k+n}$

In this way, one can compute $M_1(n,k)(= E(X_n))$ and rediscovers (2.5). Similarly, one gets, e.g.,

$$M_2(n,k) = (-1+k)^2 + \frac{\left(-2k + kn - 2k^2 n\right)\bar{S}_1(k)}{n} + k^2\left(\bar{S}_1(k)\right)^2$$

$$+ \frac{\left(2k - kn + 2k^2 n\right)\bar{S}_1(n)}{n} - 2k^2 \bar{S}_1(k)\bar{S}_1(n) + k^2\left(\bar{S}_1(n)\right)^2 + k^2 \bar{S}_2(k) - k^2 \bar{S}_2(n)$$

$$+ \frac{(-1+k)k}{(-1+n)n}\delta(-2+k)$$

$$(3.5)$$

with

$$\delta(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

Note that in the special case $k = 1$ the factor $\frac{(-1+k)k}{(-1+n)n}$ does not contribute; this comes from the fact that in the summation of (3.4) the lower bound $k - l$ should be not negative.

In a nutshell, one can calculate sufficiently many $M_l(n, k)$ and using the formula (3.2) one gets the desired moments $m_l(n, k)$. E.g., with $M_1(n, k) = E(X_n)$ given in (2.5) and (3.5) one can reproduce $m_2(n, k)$ as given in (2.6). Similarly, one obtains, e.g.,

$$
\begin{aligned}
m_3(n, k) = &-\frac{(-2+k)(-1+k)k}{(-2+n)(-1+n)n} + 2k^3 \bar{S}_3(n) + \frac{(6k^2 - 5kn - kn^2)\bar{S}_1(k)}{(-1+n)n} - \frac{3k(\bar{S}_1(k))^2}{n} \\
&+ \frac{(-6k^2 + 5kn + kn^2)\bar{S}_1(n)}{(-1+n)n} + \frac{6k\bar{S}_1(k)\bar{S}_1(n)}{n} - \frac{3k(\bar{S}_1(n))^2}{n} + \frac{3(-k + 2k^2 + k^2n)\bar{S}_2(k)}{n} \\
&- 2k^3 \bar{S}_3(k) - \frac{3(-k + 2k^2 + k^2n)\bar{S}_2(n)}{n} + \frac{3(2k^2 - k^3 - 2kn + k^2n)}{(-1+n)^2 n}\delta(-2+k)
\end{aligned}
$$

and

$$
\begin{aligned}
m_4(n, k) = &\frac{(-3+k)(-2+k)(-1+k)k}{(-3+n)(-2+n)(-1+n)n} + \frac{6k - 12k^2 + 30k^3 - 6k^4 + 18kn - 29k^2n - 7k^2n^2}{(-1+n)n}\bar{S}_2(n) \\
&+ \frac{\bar{S}_2(k)}{(-1+n)n}\left(-6k + 12k^2 - 30k^3 + 6k^4 - 18kn + 29k^2n + 7k^2n^2\right) \\
&+ \frac{\bar{S}_1(k)}{(-2+n)(-1+n)n}\left(-52k + 36k^2 - 12k^3 + 40kn - 11kn^2 - kn^3\right) \\
&+ \frac{\bar{S}_1(n)}{(-2+n)(-1+n)n}\left(52k - 36k^2 + 12k^3 - 40kn + 11kn^2 + kn^3\right) \\
&+ \frac{3(-2k + 4k^2 - 6kn + 3k^2n + k^2n^2)(\bar{S}_1(k))^2}{(-1+n)n} - \frac{4k(\bar{S}_1(k))^3}{n} + 3k^4(\bar{S}_2(n))^2 \\
&- \frac{6(-2k + 4k^2 - 6kn + 3k^2n + k^2n^2)\bar{S}_1(k)\bar{S}_1(n)}{(-1+n)n} + \frac{12k(\bar{S}_1(k))^2\bar{S}_1(n)}{n} \\
&+ \frac{3(-2k + 4k^2 - 6kn + 3k^2n + k^2n^2)(\bar{S}_1(n))^2}{(-1+n)n} - \frac{12k\bar{S}_1(k)(\bar{S}_1(n))^2}{n} + \frac{4k(\bar{S}_1(n))^3}{n} \\
&- \frac{6(2k - 4k^2 + 2k^3 + k^3n)\bar{S}_1(k)\bar{S}_2(k)}{n} + \frac{6(2k - 4k^2 + 2k^3 + k^3n)\bar{S}_1(n)\bar{S}_2(k)}{n} \\
&- \frac{4(2k - 6k^2 + 6k^3 + 3k^3n)\bar{S}_3(k)}{n} + 6k^4\bar{S}_4(k) + \frac{6(2k - 4k^2 + 2k^3 + k^3n)\bar{S}_1(k)\bar{S}_2(n)}{n} \\
&- \frac{6(2k - 4k^2 + 2k^3 + k^3n)\bar{S}_1(n)\bar{S}_2(n)}{n} - 6k^4\bar{S}_2(k)\bar{S}_2(n) \\
&+ \frac{4(2k - 6k^2 + 6k^3 + 3k^3n)\bar{S}_3(n)}{n} - 6k^4\bar{S}_4(n) - \frac{2}{(-2+n)^2(-1+n)^2 n}\left(-4k + 36k^2\right. \\
&- 30k^3 + 6k^4 - 24kn + k^2n + 15k^3n - 4k^4n + 16kn^2 - 15k^2n^2 + 3k^3n^2\big)\delta(-3+k) \\
&+ \left[-\frac{6\bar{S}_1(k)}{(-1+n)^2 n}\left(-2k - 5k^2 + 3k^3 + 10kn - 7k^2n + k^3n\right)\right.
\end{aligned}
$$

$$+ \frac{6\bar{S}_1(n)}{(-1+n)^2 n}\big(-2k - 5k^2 + 3k^3 + 10kn - 7k^2 n + k^3 n\big) + \frac{1}{(-1+n)^3 n}\big(-k + 37k^2$$
$$- 42k^3 + 12k^4 - 34kn + 52k^2 n - 18k^3 n - 13kn^2 + 7k^2 n^2\big)\Big]\delta(-2+k)$$

For instance, specializing $m_4(n,k)$ to $k = 1$ gives

$$m_4(n,1) = \big(\frac{14+n}{n} - 6\bar{S}_2(n)\big)\bar{S}_1(n) + \frac{3(-2+n)\bar{S}_1(n)^2}{n} + \frac{4\bar{S}_1(n)^3}{n}$$
$$- \frac{(18+7n)\bar{S}_2(n)}{n} + 3\bar{S}_2(n)^2 + \frac{4(2+3n)\bar{S}_3(n)}{n} - 6\bar{S}_4(n)$$

for one absent-minded passenger (as derived in [6, Theorem 5]) and

$$m_4(n,2) = \big(\frac{2\big(-74+13n+13n^2\big)}{(-1+n)n} - \frac{24(1+2n)\bar{S}_2(n)}{n}\big)\bar{S}_1(n) + 48\bar{S}_2(n)^2$$
$$+ \frac{12\big(5-2n+n^2\big)\bar{S}_1(n)^2}{(-1+n)n} + \frac{8\bar{S}_1(n)^3}{n} - \frac{4(21+19n)\bar{S}_2(n)}{n}$$
$$+ \frac{16(7+6n)\bar{S}_3(n)}{n} - 96\bar{S}_4(n) + \frac{2\big(51-45n+19n^2\big)}{(-1+n)n}$$

for two absent-minded passengers.

Following this strategy we succeeded in computing the moments $m_l(n,k)$ up to $l \le 15$ on a machine with 12 cores and 1.5TB memory, but failed to proceed due to time and space limitations. Luckily, another great trick from [19] enabled us to compute even more moments. Namely, the calculation of

$$\bar{M}_l(X_n) = \big(\tfrac{d}{dw}\big)^l f_n^{(k)}(w)|_{w=1}$$
$$= \big(\tfrac{d}{dw}\big)^l \left[\frac{1+r\,w}{n} \sum_{r=\max(0,k-l)}^{k} r!\binom{k}{r} w^r (1-w)^{k-r}(2+r\,w)_{n-k-1} \right]\Big|_{w=1} \tag{3.6}$$

turns out to be much faster (and less memory consuming) than the calculation of (3.4). Given the moments of this so-called exponential moments, one gets back the ordinary moments $M_l(n,k)$ by using the following formula from [19]:

$$M_l(X_n) = \sum_{r=1}^{l} S(l,r)\bar{M}_l(X_n) \tag{3.7}$$

where $S(l,r)$ denote the Stirling Numbers of the Second kind. In summary, we carried out the following steps:

Step 1: Calculate the exponential moments $\bar{M}_l(n,k)$ with the formula (3.6).
Step 2: Calculate the ordinary moments $M_l(n,k)$ with the formula (3.7).
Step 3: Finally, calculate the desired moments $m_l(n,k)$ with the formula (3.2).

Using this approach, we succeeded in calculating generously[2] the moments $m_l(n,k)$ up to $l \le 21$ in a recent amount of time and memory. The moments up to order 16 are online available and the link can be found in [11]. The corresponding timings[3] for the different steps, the total time and the size of the moments are summarized in the table of Figure 1.

---

[2]Originally only $l \le 4$ was required in the challenge.
[3]The calculations have been distributed (parallelized) neatly using about 10 Mathematica subkernels.

l	time for step 1	time for step 2	time for step 3	total time	size of $m_l(n,k)$
1	0.16 s	0.016 s	0.96 s	1.14 s	88 B
2	1.86 s	0.025 s	0.068 s	1.95 s	1536 B
3	3.06 s	0.067 s	0.182 s	3.307 s	6 KB
4	4.61 s	0.15 s	0.48 s	5.24 s	25 KB
5	6.66 s	0.33 s	1.18 s	8.18 s	73 KB
6	9.55 s	0.75 s	2.81 s	13.12 s	210 KB
7	13.81 s	1.70 s	6.51 s	22.02 s	531 KB
8	21.53 s	3.73 s	15.14 s	40.40 s	1.3 MB
9	35.79 s	8.03 s	34.50 s	78.33 s	2.9 MB
10	63 s	18 s	75 s	156 s	6.4 MB
11	115 s	37 s	160 s	313 s	13 MB
12	222 s	77 s	335 s	634 s	27 MB
13	454 s	233 s	678 s	1364 s	52 MB
14	1101 s	579 s	1344 s	3024 s	98 MB
15	2559 s	1063 s	2611 s	6233 s	180 MB
16	5249 s	2380 s	4995 s	12625 s	326 MB
17	11510 s	4238 s	9521 s	25270 s	573 MB
18	22357 s	8807 s	17669 s	48834 s	993 MB
19	48597 s	17843 s	32300 s	98740 s	1.7 GB
20	95457 s	30467 s	59384 s	185309 s	2.8 GB
21	180954s	57621s	126000s	364576s	4.7 GB

FIGURE 1. Timings and size of the calculated moments $m_l(n,k)$.

We note that for $l \in \mathbb{Z}_{\geq 0}$ the closed form expression $m_l(n,k)$ can be given in terms of the harmonic numbers $S_r(n)$ and $S_r(k)$ (resp. $\bar{S}_r(n)$ and $\bar{S}_r(k)$) with $1 \leq r \leq l$. We note further that the sequences generated by the harmonic numbers $\{S_r(n) \mid r \geq 1\}$ (resp. $\{\bar{S}_r(n) \mid r \geq 1\}$) are algebraically independent among the rational sequences, i.e., the representation of $m_l(n,k)$ in terms of the harmonic numbers is optimal. Interesting enough, the algebraic independence can be shown with the help of the summation paradigm of parameterized (creative) telescoping; see [13, Example 6.3]. For more general classes of harmonic sums we refer also to [4].

4. ASYMPTOTIC EXPANSIONS

Two days after Doron Zeilberger's first email, I got a slight update of his challenge:

```
Finally, to get the donation in your honor you have to complete
the challenge. Find asymptotic expressions for $m_l(n)$ for
at least l=7 (it would be nice to also have l=8), and
prove (hopefully automatically) that
```

$$\lim_{n \to \infty} \frac{m_1(n)}{m_2(n)^{1/2}}$$

```
equals 0 if l is odd and (2*l-1)(2*l-3)...1  if l is even.
This would be a partial elementary reproof of the
asymptotic normality proved for arbitrary r using "fancy" probability
of the asymptotic normality, in the Henze-Last paper.
```

Given the calculations of $m_l(n,k)$ we proceed as follows. For $l = 2$ we load in the computed closed form expression

$$\text{In[13]:= } \mathbf{m2} = \frac{\mathbf{2k - k^2 - 2n - 2kn + 2k^2n + 2n^2 - kn^2}}{\mathbf{(-1 + n)n^2}}$$
$$- \frac{\mathbf{k(2 + n)S[1, k]}}{\mathbf{n}} + \frac{\mathbf{k(2 + n)S[1, k]}}{\mathbf{n}} + \mathbf{k^2S[2, k] - k^2S[2, n];}$$

and use the `HarmonicSums` command to calculate the first terms of the expansion in n:

$$\text{In[14]:= } \mathbf{SExpansion[m2, n, 3]/.LG[n] \to Log[n] + \gamma}$$

$$\text{Out[14]= } \frac{\mathbf{7(-1+k)k}}{\mathbf{6n^3}} + \frac{\mathbf{k(-1+6k)}}{\mathbf{12n^2}} - \mathbf{k^2S_2(\infty)} + \frac{\frac{\mathbf{k}}{\mathbf{2}} + \mathbf{k^2 - 2kS_1(k)}}{\mathbf{n}} - \mathbf{kS[1,k] + k^2S[2,k] + \left(k + \frac{2k}{n}\right)(\gamma + Log[n])}$$

where $\gamma = 0.577216$ is Euler's constant, $S_2(\infty) = \zeta(2) = \frac{\pi^2}{6}$, and $\zeta(z) = \sum_{i=1}^{\infty} \frac{1}{i^z}$ denotes the Riemann-Zeta function. Summarizing, we have calculated

$$\begin{aligned}
m_2(n,k) = & \frac{7(-1+k)k}{6n^3} + \frac{k(-25+18k)}{12n^2} + \frac{k(-1+2k)}{2n} - \left(k + \frac{2k}{n}\right)S_1(k) \\
& + k^2 S_2(k) - k^2\zeta(2) + \left(k + \frac{2k}{n}\right)(\gamma + \log(n)) + O(\tfrac{1}{n^4}).
\end{aligned} \tag{4.1}$$

Analogously, we obtain, e.g., the expansion of $m_3(n,k)$:

$$\begin{aligned}
m_3(n,k) = & -\frac{k\left(31 - 38k + 8k^2\right)}{4n^3} - \frac{k\left(73 - 90k + 12k^2\right)}{12n^2} + \frac{\frac{1}{2}k(-1+6k) - 3k(-1+2k)\zeta(2)}{n} \\
& + (\gamma + \log(n))\left(k - \frac{k(-13+12k)}{2n^3} - \frac{3k(-3+2k)}{n^2} + \frac{6k}{n} + \frac{6kS_1(k)}{n}\right) \\
& + \left(-k + \frac{k(-13+12k)}{2n^3} + \frac{3k(-3+2k)}{n^2} - \frac{6k}{n}\right)S_1(k) \\
& - \frac{3kS_1(k)^2}{n} + \left(3k^2 + \frac{3k(-1+2k)}{n}\right)S_2(k) - 2k^3S_3(k) - 3k^2\zeta(2) + 2k^3\zeta(3) \\
& - \frac{3k(\gamma + \log(n))^2}{n} + \left(-\frac{3(-2+k)^2k}{n^3} + \frac{3(-2+k)k}{n^2}\right)\delta(-2+k) + O(\tfrac{1}{n^4}).
\end{aligned}$$

In total we calculated these expansions of $m_l(n,k)$ up to $l \leq 18$ with the following timings

l	5	6	7	8	9	10	11	12	13	14	15	16	17	18
time	$1.2s$	$3.6s$	$9.6s$	25	$65s$	$167s$	$405s$	$986s$	$2419s$	$1.6h$	$3.8h$	$9.8h$	$22.7h$	$53.4h.$

Here we parallelized the calculations on 10 Mathematica kernels. E.g., we needed 53.4 hours to obtain the expansion for $m_{18}(n,k)$; however summing up all the timings of the used kernels we needed in total 16 days of CPU time to tackle the case $l = 18$. The expansions up to order 16 are online available and the corresponding link can be found in [11].

To fully win Doron Zeilberger's challenge, we also dealt with the limit

$$c_l(k) = \lim_{n\to\infty} \frac{m_l(n,k)}{m_2(n,k)^{l/2}}, \qquad l \geq 2.$$

Looking at (4.1) the leading term in the expansion of $m_2(n,k)$ equals $\log(n)\,k$. In addition the leading term in $m_3(n,k)$ is also $\log(n)\,k$. Thus

$$c_3(k) = \lim_{n\to\infty} \frac{m_3(n,k)}{m_3(n,k)^{3/2}} = \lim_{n\to\infty} \frac{1}{\sqrt{k\log(n)}} = 0.$$

Similarly, given the other computed expansions we get

$$c_{2l+1} = 0, \quad 0 \leq l \leq 8$$

and

l	2	4	6	8	10	12	14	16	18
$c_l(k)$	1	3	15	105	945	10395	135135	2027025	34459425.

As a consequence we can confirm that the first values $c_{2l}(k)$ agree with the double factorial of odd numbers (sequence A001147 in OEIS), i.e.,

$$c_{2l}(k) = (2l-1)!! = \prod_{i=1}^{l}(2i-1) = \frac{(2l-1)!}{(l-1)!2^{l-1}}$$

holds for all $1 \le l \le 9$.

5. CONCLUSION

When I received Doron Zeilberger's email, the first calculations were straightforward – except that I was (and I am still not) expert in probability theory and thus made some annoying errors in the beginning. However, pushing the calculations further to the 21th moment and computing its asymptotic expansions was a challenge. Here the computer algebra packages developed at RISC (`Sigma.m` and `EvaluateMultiSums.m` by myself and `HarmonicSums.m` by Jakob Ablinger) were of great help to calculate these huge expressions. I hope that this little note advertises how existing computer algebra tools in general and in particular those of the combinatorics group at RISC can push forward interesting (combinatorial) research topics. For instance, these packages have been heavily used in the last years to carry out large-scale QCD-calculations in the research field of elementary particle physics; see, e.g., [1] and references therein. Last but not least, I am very grateful to Doron Zeilberger who challenged me with these problems and initiated this fun project.

REFERENCES

[1] J. Ablinger, A. Behring, J. Blümlein, A. De Freitas, A. von Manteuffel, and C. Schneider. Calculating Three Loop Ladder and V-Topologies for Massive Operator Matrix Elements by Computer Algebra. *Comput. Phys. Comm.*, 202:33–112, 2016. arXiv:1509.08324 [hep-ph].

[2] J. Ablinger, J. Blümlein, and C. Schneider. Harmonic sums and polylogarithms generated by cyclotomic polynomials. *J. Math. Phys.*, 52(10):1–52, 2011. [arXiv:1007.0375 [hep-ph]].

[3] J. Ablinger, J. Blümlein, and C. Schneider. Analytic and algorithmic aspects of generalized harmonic sums and polylogarithms. *J. Math. Phys.*, 54(8):1–74, 2013. arXiv:1302.0378 [math-ph].

[4] J. Ablinger and C. Schneider. Algebraic independence of sequences generated by (cyclotomic) harmonic sums. *Annals of Combinatorics*, 22(2):213–244, 2018. arXiv:1510.03692 [cs.SC].

[5] B. Bolobás. *The Art of Mathematics: Coffee Time in Memphis*. Cambridge University Press, 2006.

[6] S.B. Ekhad and D. Zeilberger. The absent-minded passengers problem via computer algebra. *The Personal Journal of Shalosh B. Ekhad and Doron Zeilberger*, 2020. arXiv:2001.06839 [math.CO].

[7] N. Henze and G. Last. Absent-minded passengers. *Amer. Math. Monthly*, 126(10):867–875, 2019. arXiv:1809.10192 [math.PR].

[8] M. Karr. Summation in finite terms. *J. ACM*, 28:305–350, 1981.

[9] M. Petkovšek, H. S. Wilf, and D. Zeilberger. $A = B$. A K Peters, Wellesley, MA, 1996.

[10] H. Prodinger, C. Schneider, and S. Wagner. Unfair permutations. *Europ. J. Comb.*, 32:1282–1298, 2011.

[11] C. Schneider. Moments and their expansions up to order 16. March 3, 2020, online available at: https://www.risc.jku.at/people/cschneid/data/AMPassenger.tar.gz.

[12] C. Schneider. Symbolic summation assists combinatorics. *Sém. Lothar. Combin.*, 56:1–36, 2007. Article B56b.

[13] C. Schneider. Parameterized telescoping proves algebraic independence of sums. *Ann. Comb.*, 14:533–552, 2010. [arXiv:0808.2596]; for a preliminary version see FPSAC 2007.

[14] C. Schneider. Simplifying multiple sums in difference fields. In C. Schneider and J. Blümlein, editors, *Computer Algebra in Quantum Field Theory: Integration, Summation and Special Functions*, Texts and Monographs in Symbolic Computation, pages 325–360. Springer, 2013. arXiv:1304.4134 [cs.SC].

[15] C. Schneider. A difference ring theory for symbolic summation. *J. Symb. Comput.*, 72:82–127, 2016. arXiv:1408.2776 [cs.SC].

[16] C. Schneider. Summation Theory II: Characterizations of $R\Pi\Sigma$-extensions and algorithmic aspects. *J. Symb. Comput.*, 80(3):616–664, 2017. arXiv:1603.04285 [cs.SC].

[17] P. Winkler. *Mathematical Puzzles: A Connoisseur's Collection.* A K Peter, 2004.

[18] D. Zeilberger. The method of creative telescoping. *J. Symbolic Comput.*, 11:195–204, 1991.

[19] D. Zeilberger. The automatic central limit theorems generator (and much more!). In *Advances in combinatorial mathematics*, pages 165–174. Springer, Berlin, 2009.

(Carsten Schneider) Johannes Kepler University Linz, Research Institute for Symbolic Computation, A-4040 Linz, Altenberger Strasse 69, Austria

E-mail address: Carsten.Schneider@risc.jku.at