Approximate counting with m counters: a detailed analysis

Helmut Prodinger

Department of Mathematics, University of Stellenbosch 7602, Stellenbosch, South Africa, hproding@sun.ac.za

Abstract

The classical algorithm *approximate counting* was recently modified by Cichon and Macyna: Instead of one counter, m counters are used, and the assignment of an incoming item to one of the counters is random. The parameter of interest is the sum of the values of all the counters. We analyse expectation and variance, getting explicit and asymptotic formulæ.

Keywords: Approximate counting, Rice's method, cancellations, q-analysis.

1. Introduction

Approximate counting is a technique that was first analysed by Flajolet [1]; some subsequent papers [2, 3, 4, 5] added to the analysis.

A counter C is kept, and each time an item arrives and needs to be counted, a random experiment is performed; if the current value of the counter is i, then with probability 2^{-i} the counter is increased by 1, otherwise it keeps its value; at the beginning, the counter value is $C = 1$. After n random increments, the value of the counter is typically close to $\log_2 n$, and the cited papers contain exact and asymptotic values for average and variance.

Very recently, Cichon and Macyna [6] used this idea as follows: Instead of 1 counter, they keep m counters, where $m \geq 1$ is an integer. For our subsequent analysis we will assume that m is fixed. When a new element arrives (and needs to be counted), it is randomly (with probability $\frac{1}{m}$) assigned to one of the m counters, and then the random experiment is performed as usual. The parameter that Cichon and Macyna are interested in is the total number of changes of any counter. In other words, if we (for convenience) assume that the initial setting of a counter to the value 1 counts as a change, Cichon is interested in the sum of the values of the m counters. This is the parameter that we will study in this paper.

We find an expression for $P_{N,l}$, the probability that the combined counter values are l, after N random experiments, in terms of the classical values $p_{n,l}$ for just one counter. Furthermore, we find exact and asymptotic expressions for expectation and variance. As in the classical case, the expectation is of the form $A_1 \log N + A_2 + \delta(\log_2 N) + o(1)$, with a periodic function $\delta(x)$ of period 1 and mean value 0. (The mean value is the zeroth term in its Fourier series expansion.) The variance is small: $A_3 + \delta_V(\log_2 N) + o(1)$. We will not compute the Fourier expansion of this periodic function explicitly, although it can be done.

Here are a few words about what to expect from our analysis: Intuitively, we expect that roughly N/m elements will go to each counter, so we expect a result like $\mathbb{E}(\mathsf{C}_N^{(m)}) \sim m\mathbb{E}(\mathsf{C}_{N/m}^{(m)})$, where $\mathsf{C}_N^{(m)}$ is the random variable describing the combined values of m counters after N random increments, and a similar formula for the variance. Indeed, our asymptotic analysis confirms these claims; differences would only occur for lower order terms that we do not compute here.

The paper clearly consists of two parts: The *explicit* computation of the first two moments (Section 2) and the asymptotic evaluation of them (Section 3). While it is quite an achievement to obtain these explicit evaluations, there is a prize to be paid: The computations to achieve all this are not simple. We also borrow from the paper [7] in terms of methods. Where this paper was particularly brief (because of length restrictions) we try here to re-explain things and provide more help for the interested reader. The path that we follow here is not surprising for the specialists. However, the present case offers several challenges, whence we decided to provide the calculations in full, since we feel that even reasonably experienced readers might not be able to fill them in, would they not been given.

Our approach is to use generating functions throughout. The explicit expressions that we obtain are alternating functions for which we use the convenient technique called *Rice's method* [8].

We will need a few abbreviations that we collect here: $L = \log 2$, $Q_n := (1-\frac{1}{2})\dots(1-\frac{1}{2^n})$, $Q(x) := \prod_{k\ge1}(1-\frac{x}{2^k})$, so that $Q_n = Q(1)/Q(2^{-n})$; the latter expression is used to define Q_n for arbitrary n, not just nonnegative integers.

We will prove the following asymptotic result, where $\delta(x)$ and $\delta_V(x)$ are periodic functions of period 1 and mean zero.

Theorem 1. *The average and the variance of Cichon-Macyna's m-approximate counter scheme admit the following asymptotic expansions as $N \to \infty$:*

$$E_N \sim m\Big[\log_2 N - \log_2 m + \frac{1}{2} - \alpha + \frac{\gamma}{L} + \delta(\log_2 N)\Big],$$
$$V_N \sim m\Big[1 - \alpha - \beta + \frac{2\tau}{L} + \delta_V(\log_2 N)\Big].$$

The constants (also independent of m) are

$$\alpha = \sum_{j\ge1}\frac{1}{2^j-1}, \qquad \beta = \sum_{j\ge1}\frac{1}{(2^j-1)^2} \quad \textit{and} \quad \tau = \sum_{j\ge1}\frac{(-1)^{j-1}}{j(2^j-1)}.$$

2. Generating functions

If we think what has happened after N random experiments, then we notice that $N = n_1 + \dots + n_m$ and n_i elements were assigned to counter C_i. The probability for such a split is a multinomial;

$$\binom{N}{n_1,\dots,n_m}\frac{1}{m^N} = \frac{N!}{n_1!\dots n_m!m^N}.$$

Then we get the probability that after N experiments the sum of the m counters is l:

$$P_{N,l} = m^{-N}\sum_{n_1+\dots+n_m=N}\binom{N}{n_1,\dots,n_m}\sum_{l_1+\dots+l_m=l} p_{n_1,l_1}\dots p_{n_m,l_m}.$$

Let

$$G_n(u) := \sum_{l\ge0} p_{n,l}u^l.$$

This generating function (studied in the classical papers) is not very nice itself, but the first and second (factorial) moments $G'_n(1)$ and $G''_n(1)$ are explicitly known:

$$G'_n(1) = 1 - \sum_{k=1}^{n} \binom{n}{k} (-1)^k 2^{-k} Q_{k-1},$$

$$G''_n(1) = \sum_{k=1}^{n} \binom{n}{k} (-1)^k 2^{1-k} Q_{k-1}(T_{k-1} - 1), \quad \text{with} \quad T_k = \sum_{j=1}^{k} \frac{1}{2^k - 1}.$$

We can now form a generating function for the case of m counters:

$$\sum_l P_{N,l} u^l = m^{-N} \sum_{n_1+\cdots+n_m=N} \binom{N}{n_1, \ldots, n_m} G_{n_1}(u) \ldots G_{n_m}(u)$$

or

$$\sum_l P_{N,l} u^l \frac{m^N}{N!} = \sum_{n_1+\cdots+n_m=N} \frac{1}{n_1! \ldots n_m!} G_{n_1}(u) \ldots G_{n_m}(u).$$

Multiplying this by z^N and summing, we find the fundamental relationship

$$\sum_{l,N} P_{N,l} u^l \frac{(mz)^N}{N!} = \sum_N z^N \sum_{n_1+\cdots+n_m=N} \frac{1}{n_1! \ldots n_m!} G_{n_1}(u) \ldots G_{n_m}(u) = \Big(\sum_n \frac{z^n}{n!} G_n(u) \Big)^m. \tag{1}$$

Differentiating this with respect to u once or twice, followed by setting $u := 1$, produces generating functions for the first and second factorial moments.

Taking one derivative we get:

$$\sum_{l,N} P_{N,l}\, l\, \frac{(mz)^N}{N!} = m \Big(\sum_n \frac{z^n}{n!} G_n(1) \Big)^{m-1} \sum_n \frac{z^n}{n!} G'_n(1) = m e^{z(m-1)} \sum_n \frac{z^n}{n!} G'_n(1).$$

Reading off the coefficient of z^N and multiplying it by $N!/m^N$ leads to an expression for the expected value:

$$\begin{aligned}
E_N = \mathbb{E}(\mathsf{C}_N^{(m)}) &= m^{1-N} N! [z^N] e^{z(m-1)} \sum_n \frac{z^n}{n!} G'_n(1) = m^{1-N} \sum_{n=0}^{N} \binom{N}{n} (m-1)^{N-n} G'_n(1) \\
&= m^{1-N} \sum_{n=0}^{N} \binom{N}{n} (m-1)^{N-n} \Big[1 - \sum_{k=1}^{n} \binom{n}{k} (-1)^k 2^{-k} Q_{k-1} \Big] \\
&= m - m^{1-N} \sum_{k=1}^{N} \sum_{n=k}^{N} \binom{N}{n} \binom{n}{k} (m-1)^{N-n} (-1)^k 2^{-k} Q_{k-1} \\
&= m - m^{1-N} \sum_{k=1}^{N} \sum_{n=k}^{N} \binom{N}{k} \binom{N-k}{n-k} (m-1)^{N-n} (-1)^k 2^{-k} Q_{k-1} \\
&= m - m^{1-N} \sum_{k=1}^{N} \binom{N}{k} m^{N-k} (-1)^k 2^{-k} Q_{k-1} = m - m \sum_{k=1}^{N} \binom{N}{k} (-1)^k (2m)^{-k} Q_{k-1}.
\end{aligned}$$

Note that these simplifications have resulted in a form that does not look too different from the classical instance $m = 1$. The corresponding simplifications for the second (factorial) moment, which are going to follow, are significantly harder. We decided to present them in full since it would cost an interested reader a few hours to fill them in herself.

Performing two differentiations, we get

$$\sum_{l,N} P_{N,l}\, l(l-1)\,\frac{(mz)^N}{N!} = m(m-1)e^{z(m-2)}\Big(\sum_n \frac{z^n}{n!}G_n'(1)\Big)^2 + me^{z(m-1)}\Big(\sum_n \frac{z^n}{n!}G_n''(1)\Big).$$

Now we need to perform some auxiliary computations: In these, we will use

$$\sum_n \binom{N}{n}\binom{n}{k}\binom{N-n}{j} = \sum_n \binom{N}{k+j}\binom{k+j}{k}\binom{N-k-j}{n-k} = 2^{N-k-j}\binom{N}{k+j}\binom{k+j}{k}.$$

Firstly,

$$\begin{aligned}
N![z^N]\Big(\sum_n \frac{z^n}{n!}G_n'(1)\Big)^2 &= \sum_n \binom{N}{n}G_n'(1)G_{N-n}'(1)\\
&= \sum_n \binom{N}{n}\Big[1-\sum_{j=1}^{n}\binom{n}{j}(-1)^j2^{-j}Q_{j-1}\Big]\Big[1-\sum_{k=1}^{N-n}\binom{N-n}{k}(-1)^k2^{-k}Q_{k-1}\Big]\\
&= 2^N - 2\sum_n \binom{N}{n}\sum_{j=1}^{n}\binom{n}{j}(-1)^j2^{-j}Q_{j-1}\\
&\quad + \sum_n \binom{N}{n}\sum_{j=1}^{n}\binom{n}{j}(-1)^j2^{-j}Q_{j-1}\sum_{k=1}^{N-n}\binom{N-n}{k}(-1)^k2^{-k}Q_{k-1}\\
&= 2^N - 2\sum_n \sum_{j=1}^{n}\binom{N}{j}\binom{N-j}{n-j}(-1)^j2^{-j}Q_{j-1}\\
&\quad + \sum_{j=1}^{N}\sum_{k=1}^{N}(-1)^j2^{-j}Q_{j-1}2^{N-k-j}\binom{N}{k+j}\binom{k+j}{k}(-1)^k2^{-k}Q_{k-1}\\
&= 2^N - 2^{N+1}\sum_j \binom{N}{j}(-1)^j4^{-j}Q_{j-1}\\
&\quad + \sum_{j=1}^{N}\sum_{k=1}^{N}(-1)^j2^{-j}Q_{j-1}2^{N-k-j}\binom{N}{k+j}\binom{k+j}{k}(-1)^k2^{-k}Q_{k-1}.
\end{aligned}$$

Then

$$\begin{aligned}
&N![z^N]e^{z(m-2)}\Big(\sum_n \frac{z^n}{n!}G_n'(1)\Big)^2\\
&= \sum_s \binom{N}{s}(m-2)^{N-s}2^s - \sum_s \binom{N}{s}(m-2)^{N-s}2^{s+1}\sum_j \binom{s}{j}(-1)^j4^{-j}Q_{j-1}\\
&\quad + \sum_s \binom{N}{s}(m-2)^{N-s}\sum_{j=1}^{s}\sum_{k=1}^{s}(-1)^j2^{-j}Q_{j-1}2^{s-k-j}\binom{s}{k+j}\binom{k+j}{k}(-1)^k2^{-k}Q_{k-1}
\end{aligned}$$

$$= m^N - \sum_s (m-2)^{N-s} 2^{s+1} \sum_j \binom{N}{j}\binom{N-j}{s-j}(-1)^j 4^{-j} Q_{j-1}$$

$$+ \sum_s (m-2)^{N-s} \sum_{j=1}^{s} \sum_{k=1}^{s} (-1)^j 2^{-j} Q_{j-1} 2^{s-k-j} \binom{N}{k+j}\binom{N-k-j}{s-k-j}\binom{k+j}{k}(-1)^k 2^{-k} Q_{k-1}$$

$$= m^N + \sum_{j=1}^{N} \binom{N}{j} 2^{j+1} m^{N-j} (-1)^{j-1} 4^{-j} Q_{j-1}$$

$$+ \sum_{j=1}^{N} \sum_{k=1}^{N} (-1)^j 2^{-j} Q_{j-1} \binom{N}{k+j} m^{N-k-j} \binom{k+j}{k} (-1)^k 2^{-k} Q_{k-1}$$

$$= m^N - 2m^N \sum_{j=1}^{N} \binom{N}{j} (-1)^j (2m)^{-j} Q_{j-1} + m^N \sum_{j=1}^{N} \sum_{k=1}^{N} \binom{N}{k+j}\binom{k+j}{k} (-1)^{k+j} (2m)^{-k-j} Q_{j-1} Q_{k-1}.$$

Let us summarize what we just found:

$$N![z^N] e^{z(m-2)} \Big(\sum_n \frac{z^n}{n!} G_n'(1) \Big)^2 = m^N - 2m^N \sum_{j=1}^{N} \binom{N}{j} (-1)^j (2m)^{-j} Q_{j-1}$$
$$+ m^N \sum_{j=1}^{N} \sum_{k=1}^{N} \binom{N}{k+j}\binom{k+j}{k} (-1)^{k+j} (2m)^{-k-j} Q_{j-1} Q_{k-1}.$$

This must be multiplied by $m(m-1)$ and divided by m^N, for the first component of the second factorial moment:

$$m(m-1) - 2m(m-1) \sum_{j=1}^{N} \binom{N}{j} (-1)^j (2m)^{-j} Q_{j-1} + m(m-1) \sum_{k=1}^{N} \binom{N}{k} (-1)^k (2m)^{-k} \sum_{j=1}^{k-1} \binom{k}{j} Q_{j-1} Q_{k-1-j}.$$

The second contribution comes from here:

$$m e^{z(m-1)} \Big(\sum_n \frac{z^n}{n!} G_n''(1) \Big).$$

We read off coefficients:

$$N![z^N] e^{z(m-1)} \Big(\sum_n \frac{z^n}{n!} G_n''(1) \Big) = \sum_n \binom{N}{n} (m-1)^{N-n} \sum_{k=1}^{n} \binom{n}{k} (-1)^k 2^{1-k} Q_{k-1} (T_{k-1} - 1)$$
$$= \sum_n (m-1)^{N-n} \sum_{k=1}^{n} \binom{N}{k}\binom{N-k}{n-k} (-1)^k 2^{1-k} Q_{k-1} (T_{k-1} - 1)$$
$$= \sum_{k=1}^{N} \binom{N}{k} m^{N-k} (-1)^k 2^{1-k} Q_{k-1} (T_{k-1} - 1)$$
$$= 2m^N \sum_{k=1}^{N} \binom{N}{k} (-1)^k (2m)^{-k} Q_{k-1} (T_{k-1} - 1).$$

Combining everything, we find the second factorial moment:

$$m(m-1) - 2m(m-1)\sum_{j=1}^{N}\binom{N}{j}(-1)^j(2m)^{-j}Q_{j-1}$$

$$+\, m(m-1)\sum_{k=1}^{N}\binom{N}{k}(-1)^k(2m)^{-k}\sum_{j=1}^{k-1}\binom{k}{j}Q_{j-1}Q_{k-1-j} + 2m\sum_{k=1}^{N}\binom{N}{k}(-1)^k(2m)^{-k}Q_{k-1}(T_{k-1}-1).$$

Theorem 2. *The first and second factorial moments have the following explicit expressions:*

$$\begin{aligned}
E_N &= m - m\sum_{k=1}^{N}\binom{N}{k}(-1)^k(2m)^{-k}Q_{k-1},\\
E_N^{(2)} &= m(m-1) - 2m^2\sum_{j=1}^{N}\binom{N}{j}(-1)^j(2m)^{-j}Q_{j-1}\\
&\quad + m(m-1)\sum_{k=1}^{N}\binom{N}{k}(-1)^k(2m)^{-k}\sum_{j=1}^{k-1}\binom{k}{j}Q_{j-1}Q_{k-1-j} + 2m\sum_{k=1}^{N}\binom{N}{k}(-1)^k(2m)^{-k}Q_{k-1}T_{k-1}.
\end{aligned}$$

3. Asymptotics

Our goal (explained in more detail a bit later) is to rewrite an alternating sum as a contour integral:

$$\sum_{k=1}^{N}\binom{N}{k}(-1)^k f(k) = \frac{1}{2\pi\mathbf{i}}\int_{\mathcal{C}}\frac{(-1)^N N!}{z(z-1)\dots(z-N)}f(z)dz.$$

The asymptotic evaluation of the integral can then be achieved via residues. The challenge is here (and also in similar problems of the past) to extend the (discrete) sequence $f(k)$ to an analytic function $f(z)$. This will be done first.

We must first rewrite a function that appears in the second factorial moment. The strategy for that was laid down in [7]. The main challenge is to define

$$\sum_{j=1}^{k-1}\binom{k}{j}Q_{j-1}Q_{k-1-j}$$

when k is allowed to be a complex number.

The computations will use the abbreviation

$$a_{n+1} = (-1)^n 2^{-\binom{n+1}{2}}/Q_n,$$

and the following relations due to Euler [9]:

$$\frac{1}{Q(t)} = \sum_{n\ge0}\frac{t^n}{2^n Q_n}, \qquad Q(t) = \sum_{n\ge0}a_{n+1}t^n.$$

Using these (partition) identities, we will be able to apply the binomial theorem

$$\sum_{j=1}^{k-1}\binom{k}{j}X^j = (1+X)^k - 1 - X^k,$$

where X is complicated but explicit. And in this form, the extension $(1+X)^z - 1 - X^z$ is feasible.

Following this strategy, we get

$$\begin{aligned}
\psi(N) &:= \sum_{j=1}^{N-1}\binom{N}{j}Q_{j-1}Q_{N-j-1}\\
&= \sum_{j=1}^{N-1}\binom{N}{j}\frac{Q_\infty}{Q(2^{1-j})}\frac{Q_\infty}{Q(2^{1-(N-j)})}\\
&= Q_\infty^2\sum_{j=1}^{N-1}\binom{N}{j}\sum_{s,t\ge0}\frac{2^{(1-j)s}}{2^sQ_s}\frac{2^{(1-(N-j))t}}{2^tQ_t}\\
&= Q_\infty^2\sum_{j=1}^{N-1}\binom{N}{j}\sum_{s,t\ge0}\frac{2^{-js}}{Q_s}\frac{2^{-(N-j)t}}{Q_t}\\
&= Q_\infty^2\sum_{s,t\ge0}\frac{1}{Q_sQ_t}\Big((2^{-s}+2^{-t})^N - 2^{-Ns} - 2^{-Nt}\Big)\\
&= 2Q_\infty^2\sum_{0\le s\le t}\frac{1}{Q_sQ_t}\Big((2^{-s}+2^{-t})^N - 2^{-Ns} - 2^{-Nt}\Big) - Q_\infty^2\sum_{0\le s}\frac{1}{Q_s^2}\Big((2^{1-s})^N - 2^{1-Ns}\Big)\\
&= 2Q_\infty^2\sum_{0\le s\le t}\frac{2^{-sN}}{Q_sQ_t}\Big((1+2^{-(t-s)})^N - 1 - 2^{-N(t-s)}\Big) - (2^N-2)Q_\infty^2\sum_{0\le s}\frac{1}{Q_s^2}(2^{-s})^N\\
&= 2Q_\infty^2\sum_{s,h\ge0}\frac{2^{-sN}}{Q_sQ_{s+h}}\Big((1+2^{-h})^N - 1 - 2^{-hN}\Big) - (2^N-2)Q_\infty^2\sum_{s\ge0}\frac{1}{Q_s^2}2^{-Ns}.
\end{aligned}$$

Here, we can replace N by an arbitrary value:

$$\psi(z) = 2Q_\infty^2\sum_{s,h\ge0}\frac{2^{-sz}}{Q_sQ_{s+h}}\Big((1+2^{-h})^z - 1 - 2^{-hz}\Big) - (2^z-2)Q_\infty^2\sum_{s\ge0}\frac{1}{Q_s^2}2^{-zs}.$$

However, we will rework this expression, so that it fits our needs. In particular, we need the expansion of $\psi(z)$ around $z=0$, which is not obvious from the present representation, as a naive approach to get it would result in divergent series. The following computations belong to the realm of q-series (basic hypergeometric series) manipulations. The goal is to make the series in question converge very fast. Again, we decided to present the computations in full since even the present author needed quite some time to do them, and not all steps are completely straightforward.

$$\begin{aligned}
Q_\infty^2\sum_{s\ge0}\frac{1}{Q_s^2}2^{-zs} &= Q_\infty\sum_{s\ge0}\frac{Q(2^{-s})}{Q_s}2^{-zs} = Q_\infty\sum_{s\ge0}\frac{1}{Q_s}2^{-zs}\sum_{n\ge0}a_{n+1}2^{-sn}\\
&= Q_\infty\sum_{n\ge0}a_{n+1}\sum_{s\ge0}\frac{2^{-(z+n)s}}{Q_s} = Q_\infty\sum_{n\ge0}a_{n+1}\frac{1}{Q(2^{1-z-n})} = \sum_{n\ge0}a_{n+1}Q_{z+n-1}.
\end{aligned}$$

Further,

$$
\begin{aligned}
Q_\infty^2 \sum_{s,h\ge 0} \frac{2^{-sz}}{Q_s Q_{s+h}}\Big((1+2^{-h})^z - 1\Big) &= Q_\infty^2 \sum_{s,h\ge 0} \frac{2^{-sz}}{Q_s Q_{s+h}} \sum_{\lambda\ge 1} \binom{z}{\lambda} 2^{-h\lambda} \\
&= Q_\infty \sum_{s,h\ge 0} \frac{2^{-sz}}{Q_s} \sum_{\lambda\ge 1} \binom{z}{\lambda} 2^{-h\lambda} Q(2^{-s-h}) \\
&= Q_\infty \sum_{s,h\ge 0} \frac{2^{-sz}}{Q_s} \sum_{\lambda\ge 1} \binom{z}{\lambda} 2^{-h\lambda} \sum_{n\ge 0} a_{n+1} 2^{-n(s+h)} \\
&= Q_\infty \sum_{s\ge 0} \frac{2^{-sz}}{Q_s} \sum_{\lambda\ge 1} \binom{z}{\lambda} \sum_{n\ge 0} a_{n+1} 2^{-ns} \sum_{h\ge 0} 2^{-h\lambda} 2^{-nh} \\
&= Q_\infty \sum_{\lambda\ge 1} \binom{z}{\lambda} \sum_{n\ge 0} a_{n+1} \sum_{s\ge 0} \frac{2^{-s(n+z)}}{Q_s} \frac{1}{1-2^{-\lambda-n}} \\
&= Q_\infty \sum_{\lambda\ge 1} \binom{z}{\lambda} \sum_{n\ge 0} a_{n+1} \frac{1}{Q(2^{1-n-z})} \frac{1}{1-2^{-\lambda-n}} \\
&= \sum_{n\ge 0} a_{n+1} Q_{z+n-1} \sum_{\lambda\ge 1} \binom{z}{\lambda} \frac{1}{1-2^{-\lambda-n}}.
\end{aligned}
$$

Lastly,

$$
\begin{aligned}
-Q_\infty^2 \sum_{s,h\ge 0} \frac{2^{-sz}}{Q_s Q_{s+h}} 2^{-hz} = -Q_\infty^2 \sum_{s,h\ge 0} \frac{2^{-(s+h)z}}{Q_s Q_{s+h}} &= -Q_\infty \sum_{s,h\ge 0} \frac{2^{-(s+h)z}}{Q_s} Q(2^{-s-h}) \\
&= -Q_\infty \sum_{s,h\ge 0} \frac{2^{-(s+h)z}}{Q_s} \sum_{n\ge 0} a_{n+1} 2^{-(s+h)n} \\
&= -Q_\infty \sum_{s\ge 0} \frac{2^{-sz}}{Q_s} \sum_{n\ge 0} a_{n+1} 2^{-sn} \sum_{h\ge 0} 2^{-h(z+n)} \\
&= -Q_\infty \sum_{s\ge 0} \sum_{n\ge 0} \frac{2^{-s(z+n)}}{Q_s} a_{n+1} \frac{1}{1-2^{-z-n}} \\
&= -Q_\infty \sum_{n\ge 0} \frac{1}{Q(2^{1-z-n})} a_{n+1} \frac{1}{1-2^{-z-n}} \\
&= -\sum_{n\ge 0} a_{n+1} Q_{z+n-1} \frac{1}{1-2^{-z-n}}.
\end{aligned}
$$

Putting everything together, we find

$$
\begin{aligned}
\psi(z) &= 2 \sum_{n\ge 0} a_{n+1} Q_{z+n-1} \sum_{\lambda\ge 1} \binom{z}{\lambda} \frac{1}{1-2^{-\lambda-n}} \\
&\quad - 2 \sum_{n\ge 0} a_{n+1} Q_{z+n-1} \frac{1}{1-2^{-z-n}} - (2^z - 2) \sum_{n\ge 0} a_{n+1} Q_{z+n-1} \\
&= 2 \sum_{n\ge 0} a_{n+1} Q_{z+n-1} \sum_{\lambda\ge 1} \binom{z}{\lambda} \Big[1 + \frac{1}{2^{\lambda+n}-1}\Big] - 2 \sum_{n\ge 0} a_{n+1} Q_{z+n-1} \frac{1}{1-2^{-z-n}} - (2^z - 2) \sum_{n\ge 0} a_{n+1} Q_{z+n-1}
\end{aligned}
$$

$$
\begin{aligned}
&= 2(2^z-1)\sum_{n\ge0} a_{n+1}Q_{z+n-1} + 2\sum_{n\ge0} a_{n+1}Q_{z+n-1}\sum_{\lambda\ge1}\binom{z}{\lambda}\frac{1}{2^{\lambda+n}-1}\\
&\quad - 2\sum_{n\ge0} a_{n+1}Q_{z+n-1}\frac{1}{1-2^{-z-n}} - (2^z-2)\sum_{n\ge0} a_{n+1}Q_{z+n-1}\\
&= 2\sum_{n\ge0} a_{n+1}Q_{z+n-1}\sum_{\lambda\ge1}\binom{z}{\lambda}\frac{1}{2^{\lambda+n}-1} - 2\sum_{n\ge0} a_{n+1}Q_{z+n-1}\frac{1}{1-2^{-z-n}} + 2^z\sum_{n\ge0} a_{n+1}Q_{z+n-1}.
\end{aligned}
$$

It was already computed in [7] that

$$
\sum_{n\ge1} a_{n+1}Q_{n-1} - 2\sum_{n\ge1} a_{n+1}Q_{n-1}\frac{1}{1-2^{-n}} = \alpha+\beta = \sum_{j\ge1}\frac{2^j}{(2^j-1)^2},
$$

with

$$
\alpha = \sum_{j\ge1}\frac{1}{2^j-1} \quad \text{and} \quad \beta = \sum_{j\ge1}\frac{1}{(2^j-1)^2}.
$$

However, for the convenience of the reader, we will reproduce this here, also, because the presentation in [7] is very brief. We start with

$$
\begin{aligned}
\sum_{n\ge1} a_{n+1}Q_{n-1} &= Q(1)\sum_{n\ge1} a_{n+1}\frac{1}{Q(2^{1-n})} = Q(1)\sum_{n\ge1} a_{n+1}\sum_{m\ge0}\frac{2^{-nm}}{Q_m}\\
&= \sum_{m\ge0} Q(2^{-m})\sum_{n\ge1} a_{n+1}2^{-nm} = \sum_{m\ge0} Q(2^{-m})(Q(2^{-m})-1).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\sum_{n\ge1} a_{n+1}Q_{n-1}\frac{1}{1-2^{-n}} &= Q(1)\sum_{n\ge1} a_{n+1}\frac{1}{Q(2^{1-n})}\frac{1}{1-2^{-n}} = Q(1)\sum_{n\ge1} a_{n+1}\sum_{m\ge0}\frac{2^{-nm}}{Q_m}\sum_{j\ge0}2^{-nj}\\
&= \sum_{j\ge0}\sum_{m\ge0} Q(2^{-m})\sum_{n\ge1} a_{n+1}2^{-n(m+j)} = \sum_{j\ge0}\sum_{m\ge0} Q(2^{-m})(Q(2^{-m-j})-1).
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\sum_{n\ge1} a_{n+1}Q_{n-1} - 2\sum_{n\ge1} a_{n+1}Q_{n-1}\frac{1}{1-2^{-n}}\\
&\quad= \sum_{m\ge0} Q(2^{-m})(Q(2^{-m})-1) - 2\sum_{j\ge0}\sum_{m\ge0} Q(2^{-m})(Q(2^{-m-j})-1)\\
&\quad= \sum_{m\ge0}(Q(2^{-m})-1)(Q(2^{-m})-1) - 2\sum_{j\ge0}\sum_{m\ge0}(Q(2^{-m})-1)(Q(2^{-m-j})-1)\\
&\qquad+ \sum_{m\ge0}(Q(2^{-m})-1) - 2\sum_{j\ge0}\sum_{m\ge0}(Q(2^{-m-j})-1)\\
&\quad= -\Big(\sum_{m\ge0}(Q(2^{-m})-1)\Big)^2 - \sum_{m\ge0}(2m+1)(Q(2^{-m})-1)\\
&\quad= -\Big(\sum_{n\ge1} a_{n+1}\sum_{m\ge0}2^{-mn}\Big)^2 - \sum_{n\ge1} a_{n+1}\sum_{m\ge0}2^{-mn} - 2\sum_{n\ge1} a_{n+1}\sum_{m\ge0} m2^{-mn}
\end{aligned}
$$

$$= -\Big(\sum_{n\ge1} a_{n+1}\frac{1}{1-2^{-n}}\Big)^2 - \sum_{n\ge1} a_{n+1}\frac{1}{1-2^{-n}} - 2\sum_{n\ge1} a_{n+1}\frac{2^{-n}}{(1-2^{-n})^2}.$$

The following formula is obtained by partial fraction decomposition:

$$\frac{Q(1)}{Q(t)} = \sum_{n\ge1} a_n \frac{1}{1-\frac{t}{2^n}} = \sum_{n\ge1} a_{n+1}\frac{1}{1-\frac{t}{2^{n+1}}} + \frac{1}{1-\frac{t}{2}}.$$

Therefore

$$\begin{aligned}\sum_{n\ge1} a_{n+1}\frac{1}{1-2^{-n}} &= \lim_{t\to2}\Big(\frac{Q(1)}{Q(t)} - \frac{1}{1-\frac{t}{2}}\Big) = \lim_{t\to2}\frac{1}{1-\frac{t}{2}}\Big(\frac{Q(1)}{Q(\frac{t}{2})} - 1\Big)\\ &= \lim_{t\to1}\frac{1}{1-t}\Big(\frac{Q(1)}{Q(t)} - 1\Big) = -\Big(\frac{Q(1)}{Q(t)}\Big)'\Big|_{t=1} = -\alpha.\end{aligned}$$

By differentiation,

$$\Big(\frac{Q(1)}{Q(t)}\Big)' = \sum_{n\ge1} a_n \frac{\frac{1}{2^n}}{(1-\frac{t}{2^n})^2} = \sum_{n\ge1} a_{n+1}\frac{\frac{1}{2^{n+1}}}{(1-\frac{t}{2^{n+1}})^2} + \frac{\frac12}{(1-\frac{t}{2})^2}.$$

Therefore

$$2\Big(\frac{Q(1)}{Q(t)}\Big)' - \frac{1}{(1-\frac{t}{2})^2} = \sum_{n\ge1} a_{n+1}\frac{\frac{1}{2^n}}{(1-\frac{t}{2^{n+1}})^2}.$$

So

$$\begin{aligned}\sum_{n\ge1} a_{n+1}\frac{2^{-n}}{(1-2^{-n})^2} &= \lim_{t\to2}\Big[2\Big(\frac{Q(1)}{Q(t)}\Big)' - \frac{1}{(1-\frac{t}{2})^2}\Big]\\ &= 2\lim_{t\to2}\Big[\frac{Q(1)}{Q(t)} - \frac{1}{1-\frac{t}{2}}\Big]' = -\frac12\Big(\frac{Q(1)}{Q(t)}\Big)''\Big|_{t=1} = -\frac{\alpha^2+\beta}{2}.\end{aligned}$$

Putting things together,

$$\begin{aligned}\sum_{n\ge1} a_{n+1}Q_{n-1} - 2\sum_{n\ge1} a_{n+1}Q_{n-1}\frac{1}{1-2^{-n}} &= -\Big(\sum_{n\ge1} a_{n+1}\frac{1}{1-2^{-n}}\Big)^2 - \sum_{n\ge1} a_{n+1}\frac{1}{1-2^{-n}} - 2\sum_{n\ge1} a_{n+1}\frac{2^{-n}}{(1-2^{-n})^2}\\ &= -\alpha^2 + \alpha + \alpha^2 + \beta = \alpha + \beta.\end{aligned}$$

Now we need a few expansions that are elementary:

$$Q(x) \sim Q(1)\Big[1 - \alpha(x-1) + \frac{\alpha^2-\beta}{2}(x-1)^2\Big], \quad x \to 1,$$

and thus

$$Q(2^{-z}) \sim Q(1)\Big[1 - \alpha(2^{-z}-1) + \frac{\alpha^2-\beta}{2}(2^{-z}-1)^2\Big] \sim Q(1)\Big[1 + L\alpha z + \frac{L^2}{2}(\alpha^2 - \alpha - \beta)z^2\Big].$$

Therefore

$$Q_z \sim 1 - L\alpha z + \frac{L^2}{2}(\alpha^2 + \alpha + \beta)z^2$$

and

$$Q_{z-1} = \frac{Q_z}{1-2^{-z}} \sim \Big[1 - L\alpha z + \frac{L^2}{2}(\alpha^2+\alpha+\beta)z^2\Big]\frac{1}{1-2^{-z}} \sim \frac{1}{Lz} + \frac{1}{2} - \alpha + \frac{L}{2}\Big(\alpha^2+\beta+\frac{1}{6}\Big)z.$$

Now we can expand the function $\psi(z)$ around $z=0$:

$$\begin{aligned}
\psi(z) &= 2\sum_{n\ge0} a_{n+1}Q_{z+n-1}\sum_{\lambda\ge1}\binom{z}{\lambda}\frac{1}{2^{\lambda+n}-1} - 2\sum_{n\ge0} a_{n+1}Q_{z+n-1}\frac{1}{1-2^{-z-n}} + 2^z\sum_{n\ge0} a_{n+1}Q_{z+n-1}\\
&= 2Q_{z-1}\sum_{\lambda\ge1}\frac{z}{\lambda}\binom{z-1}{\lambda-1}\frac{1}{2^{\lambda+n}-1} + 2\sum_{n\ge1} a_{n+1}Q_{z+n-1}\sum_{\lambda\ge1}\binom{z}{\lambda}\frac{1}{2^{\lambda+n}-1}\\
&\quad - 2Q_{z-1}\frac{1}{1-2^{-z}} - 2\sum_{n\ge1} a_{n+1}Q_{z+n-1}\frac{1}{1-2^{-z-n}} + 2^zQ_{z-1} + 2^z\sum_{n\ge1} a_{n+1}Q_{z+n-1}\\
&\sim 2\frac{1}{Lz}\sum_{\lambda\ge1}\frac{z}{\lambda}\binom{z-1}{\lambda-1}\frac{1}{2^{\lambda}-1} + O(z) - 2\Big[\frac{1}{Lz}+\frac{1}{2}-\alpha+\frac{L}{2}\Big(\alpha^2+\beta+\frac{1}{6}\Big)z\Big]\Big[\frac{1}{Lz}+\frac{1}{2}+\frac{Lz}{12}\Big]\\
&\quad - 2\sum_{n\ge1} a_{n+1}Q_{n-1}\frac{1}{1-2^{-n}} + (1+Lz)\Big[\frac{1}{Lz}+\frac{1}{2}-\alpha\Big] + \sum_{n\ge1} a_{n+1}Q_{n-1}\\
&\sim \frac{2}{L}\sum_{\lambda\ge1}\frac{1}{\lambda}(-1)^{\lambda-1}\frac{1}{2^{\lambda}-1} - 2\Big[\frac{1}{Lz}+\frac{1}{2}-\alpha+\frac{L}{2}\Big(\alpha^2+\beta+\frac{1}{6}\Big)z\Big]\Big[\frac{1}{Lz}+\frac{1}{2}+\frac{Lz}{12}\Big]\\
&\quad + (1+Lz)\Big[\frac{1}{Lz}+\frac{1}{2}-\alpha\Big] + \alpha + \beta\\
&\sim -\frac{2}{L^2z^2} + \frac{2\alpha-1}{Lz} - \alpha^2 + \alpha + \frac{2}{3} + \frac{2\tau}{L},
\end{aligned}$$

where we use

$$\tau := \sum_{j\ge1}\frac{(-1)^{j-1}}{j(2^j-1)}.$$

After these preparations, we can engage into the asymptotics, using Rice's method. This method has been described in [8]. We briefly summarize what we need:

An alternating sum can be written as a contour integral:

$$\sum_{k=1}^{N}\binom{N}{k}(-1)^k f(k) = \frac{1}{2\pi\mathbf{i}}\int_{\mathcal{C}}\frac{(-1)^N N!}{z(z-1)\dots(z-N)}f(z)dz.$$

Here, the positively oriented curve $\mathcal{C}$ enclosed the poles $1,2,\dots,N$, and no others. This formula follows from simple residue calculations. Note also that

$$\frac{(-1)^N N!}{z(z-1)\dots(z-N)} = -\frac{\Gamma(N+1)\Gamma(-z)}{\Gamma(N+1-z)}.$$

Extending the curve of integration, we encounter extra residues; in order to keep the formula correct, these residues must be subtracted. They give us the terms of the asymptotic expansion of interest. There is in all our examples a pole at $z=0$, and it will give us the dominant contribution. From a term 2^z-1 in the denominator, we will also have poles at $\chi_k := \frac{2\pi\mathbf{i}k}{L}$, and the contributions establish the (Fourier series of the) periodic oscillation in the asymptotic expansion.

Neglecting these (usually tiny) oscillations, we can write in a suggestive way:

$$\sum_{k=1}^{N}\binom{N}{k}(-1)^k f(k) \sim \mathrm{Res}_{z=0}\frac{\Gamma(N+1)\Gamma(-z)}{\Gamma(N+1-z)}f(z).$$

Now we want to apply this to the expected value

$$m - m\sum_{k=1}^{N}\binom{N}{k}(-1)^k(2m)^{-k}Q_{k-1}.$$

Here, we need this function:

$$f(z) = -(2m)^{-z}Q_{z-1}.$$

The residue calculations can, after our preliminaries, be done by a computer. In papers written in the eighties, such computations would have been done by humans, but once all the relevant functions are expanded to a sufficient number of terms, there isn't anything that we can learn. These expansion we provided, and the other ones Maple "knows". This is in sharp contrast to the (long) computations in earlier parts of this paper; they cannot be done by a computer, at least not to the knowledge of the present author.

The result is:

$$\log_2 N - \log_2 m - \frac{1}{2} - \alpha + \frac{\gamma}{L}.$$

Altogether:

$$E_N \sim m\Big[\log_2 N - \log_2 m + \frac{1}{2} - \alpha + \frac{\gamma}{L} - \frac{1}{L}\sum_{k\neq 0}\Gamma(\chi_k)e^{-2\pi i k\cdot\log_2 N}\Big].$$

We don't show the residue calculation at $z = \chi_k$ but it is much easier, since there is only a simple pole.

Now we turn to the second factorial moment.

We need one more preparation:

$$T_{z-1} = \alpha - \frac{1}{2^z-1} - \sum_{j\geq 1}\frac{1}{2^{j+z}-1} \sim -\frac{1}{Lz} + \frac{1}{2} - \frac{Lz}{12} + L(\alpha+\beta)z.$$

Now the residue calculation at $z = 0$ can be done by a computer. There are 3 sums, that can be done individually. We don't display them, since they are quite long, each of type $A_1\log^2 n + A_2\log n + A_3$, and eventually there are many cancellations when one computes the variance. Recall that the variance is computed as second factorial moment plus expectation minus expectation squared. We find:

$$m - m\alpha - \frac{11}{12}m^2 - \frac{2m(m-1)\tau}{L} - m\beta + \frac{m^2\pi^2}{6L^2}.$$

This is the asymptotic main term of the variance, apart from the periodic component that we did not compute. Note carefully that the square of the expectation contains the square of a periodic function, which is again periodic, but does *not* have mean value zero. This mean value of $\delta^2(x)$

is a tiny quantity, but it is often very important, as explained for instance in [10]. Taking it into account, we find for the variance

$$m - m\alpha - \frac{11}{12}m^2 - \frac{2m(m-1)\tau}{L} - m\beta + \frac{m^2\pi^2}{6L^2} - m^2\frac{1}{L^2}\sum_{k\neq 0}\Gamma(-\chi_k)\Gamma(\chi_k) + \delta_V(\log_2 n).$$

The coefficient of m^2 is

$$\frac{\pi^2}{6L^2} - \frac{2\tau}{L} - \frac{11}{12} - \frac{1}{L^2}\sum_{k\neq 0}\Gamma(-\chi_k)\Gamma(\chi_k).$$

But this quantity is exactly equal to zero, as explained in the survey article [11]. To wet the reader's appetite, this surprising identity is related to the identity for *Dedekind's η-function*.

This reduces the variance to

$$m - m\alpha - m\beta + \frac{2m\tau}{L} + \delta_V(\log_2 n).$$

In [12] we find another handy formula: Set

$$h(x) = \sum_{k\geq 1}\frac{e^{kx}}{(e^{kx}-1)^2},$$

then

$$h(x) = \frac{\pi^2}{6x^2} - \frac{1}{2x} + \frac{1}{24} - \frac{2\pi^2}{x^2}h\Big(\frac{4\pi^2}{x}\Big).$$

For $x = L$ this gives

$$\alpha + \beta = \frac{\pi^2}{6L^2} - \frac{1}{2L} + \frac{1}{24} - \frac{2\pi^2}{L^2}h\Big(\frac{4\pi^2}{L}\Big).$$

Using this and the previous formula, we find that the coefficient of m is

$$\frac{1}{2L} + \frac{1}{24} + \frac{2\pi^2}{L^2}h\Big(\frac{4\pi^2}{L}\Big) - \frac{1}{L^2}\sum_{k\neq 0}\Gamma(-\chi_k)\Gamma(\chi_k),$$

which is numerically very close to

$$\frac{1}{2L} + \frac{1}{24}.$$

4. Conclusion

This was a first step towards a precise analysis of the model of m counters. It is possible that methods developed by Guy Louchard, as described in [13] might be suitable to derive *all* moments asymptotically.

Also, it might be possible that methods developed recently by Hwang and coauthors [14] provide a faster access to the variance.

The answer to these assumptions lies in the (near) future.

Added during the revision, January 2012. Such computations have indeed been started already and will hopefully appear soon. As expected, they lead to a faster asymptotic evaluation, but do not lead to explicit forms as obtained in this paper.

Acknowledgment. The insightful comments of two referees are gratefully acknowledged.

References

[1] P. Flajolet, Approximate counting: a detailed analysis, BIT 25 (1985) 113–134.

[2] P. Kirschenhofer, H. Prodinger, Approximate counting: An alternative approach, RAIRO Theoretical Informatics and Applications 25 (1991) 43–48.

[3] H. Prodinger, Approximate counting via Euler transform, Mathematica Slovaka 44 (1994) 569–574.

[4] H. Prodinger, Hypothetic analyses: Approximate counting in the style of Knuth, path length in the style of Flajolet, Theoretical Computer Science 100 (1992) 243–251.

[5] G. Louchard, H. Prodinger, Generalized approximate counting revisited, TCS 391 (2008) 109–125.

[6] J. Cichoń, W. Macyna, Approximate counters for flash memory, 17th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA), Toyama, Japan.

[7] P. Kirschenhofer, H. Prodinger, W. Szpankowski, Digital search trees again revisited: The internal path length perspective, SIAM Journal on Computing 23 (1994) 598–616.

[8] P. Flajolet, R. Sedgewick, Mellin transforms and asymptotics: Finite differences and Rice's integrals, TCS 144 (1995) 101–124.

[9] G. Andrews, The Theory of Partitions, Vol. 2 of Encyclopedia of Mathematics and its Applications, Addison-Wesley, 1976.

[10] H. Prodinger, Compositions and Patricia tries: no fluctuations in the variance!, Proceedings of the 6th Workshop on ALENEX and the 1st workshop onANALCO, L. Arge, G. Italiano, R. Sedgewick (editors) (2004) 211–215.

[11] H. Prodinger, Periodic oscillations in the analysis of algorithms—and their cancellations, Journal of the Iranian Statistical Society 3 (2004) 251–270.

[12] P. Kirschenhofer, H. Prodinger, J. Schoissengeier, Zur Auswertung gewisser numerischer Reihen mit Hilfe modularer Funktionen, Lecture Notes in Mathematics 1262 (1987) 108–110.

[13] G. Louchard, H. Prodinger, Asymptotics of the moments of extreme-value related distribution functions, Algorithmica 46 (2006) 431–467, long version: http://www.ulb.ac.be/di/mcs/louchard/moml.ps.

[14] H.-K. Hwang, M. Fuchs, V. Zacharovas, Asymptotic variance of random symmetric digital search trees, Discrete Mathematics and Theoretical Computer Science 12 (2010) 103–166.