ON ASYMPTOTICS OF CERTAIN SUMS ARISING IN CODING THEORY

Wojciech Szpankowski*
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.

Abstract

Recently, T. Kløve [IEEE IT, 41, 1995] analyzed the average worst case probability of undetected error for linear \([n, k; q]\) codes of length \(n\) and dimension \(k\) over an alphabet of size \(q\). The following sum \(S_n = \sum_{i=1}^{n} \binom{n}{i} (i/n)^i (1 - i/n)^{n-i}\) arose, which has also some other applications in coding theory, average case analysis of algorithms, and combinatorics. T. Kløve conjectured an asymptotic expansion of this sum, and we prove its enhanced version in this note. Furthermore, we consider a more challenging sum arising in the upper bound of the average worst case probability of undetected error over systematic codes derived by Massey. Namely:

\[
S_{n,k} = \sum_{i=1}^{n} \binom{n-k}{i} (i/n)^i (1 - i/n)^{n-i} \quad \text{for } k \geq 0.
\]

We obtain an asymptotic expansion of \(S_{n,k}\), and this leads to a conclusion that Massey’s bound on the average worst case probability over all systematic codes is better for every \(k\) than the corresponding Kløve’s bound over all codes \([n, k; q]\). The technique used in this note belongs to the analytical analysis of algorithms and is based on some enumeration of trees, singularity analysis, Lagrange’s inversion formula, and Ramanujan’s identities. In fact, \(S_n\) turns out to be related to the so called Ramanujan’s \(Q\)-function which finds plenty applications (e.g., hashing with linear probing, the birthday paradox problem, random mappings, caching, memory conflicts, etc.)

Index Terms: Error detections, linear codes, enumeration of trees, singularity analysis, asymptotic expansions, Lagrange’s inversion formula, Ramanujan’s \(Q\)-function.

*This research was supported in part by NSF Grants NCR-9206315 and CCR-9201078.
1. Formulation of the Problem and Main Results

Consider linear \([n, k; q]\) codes of length \(n\) and dimension \(k\) over \(q\)-ary symmetric channels (cf. [14]). Kløve [9] proved that the average over all codes \([n, k; q]\) of the worst case probability of undetected error \(P(n, k, q)\) is given by

\[
P(n,1,q) = \frac{q-1}{q^n-1} S_n, \tag{1}
\]
\[
P(n,k,q) \leq \frac{q^k-1}{q^n-1} S_n, \tag{2}
\]

where \(S_n\) is defined as

\[
S_n = \sum_{i=1}^{n} \binom{n}{i} \left( \frac{i}{n} \right)^i \left( 1 - \frac{i}{n} \right)^{n-i}. \tag{3}
\]

Kløve used some old results of Riordan and Sloane, and Watson to establish an upper bound on the sum \(S_n\). Also, based on extensive numerical computations, the author of [9] conjectured that

\[
S_n = \sqrt{\frac{n\pi}{2}} - \frac{1}{3} + \frac{c}{\sqrt{n}} + O(1/n)
\]

where \(c \approx 0.1044\). We prove below an enhanced version of this conjecture. The technique used to establish this result is of its own interest since it can be applied to many other problems arising in coding, pattern matching, analysis of algorithms, data compression, and so forth (cf. [7, 15, 18]). It also turns out (cf. Remark (iii) below) that \(S_n\) is related to the so called Ramanujan’s \(Q\)-function (cf. [6, 11]) which finds applications to hashing with linear probing, the birthday paradox problem, random mappings, caching, memory conflicts, etc. (cf. [3, 6, 5, 11, 12, 13]).

Let us consider a slightly more general sum, namely:

\[
S_n(y) = \sum_{i=1}^{n} \binom{n}{i} \left( \frac{i + y}{n} \right)^i y \left( 1 - \frac{i + y}{n} \right)^{n-i-y} \tag{4}
\]

where \(y\) is real (in fact, \(y\) can be a complex number, too). The following result proves Kløve’s conjecture.

**Theorem 1.** *For large \(n\) and any real \(y\) the sum \(S_n\) admits the following asymptotic expansion*

\[
S_n(y) = e^{2y} \left( \sqrt{\frac{n\pi}{2}} - \left( \frac{1}{3} - 2y \right) + \frac{\sqrt{2\pi}}{24} \frac{1}{\sqrt{n}} - \frac{4}{135} (1 + y) \frac{1}{n} + O(1/n^{3/2}) \right). \tag{5}
\]

We observe that \(\sqrt{2\pi}/24 = 0.1044428448\) as predicted by Kløve.

Furthermore, Kløve discussed in [9] Massey’s bound [16] for the average worst case probability \(P(n, k, q)\) of undetected error over all systematic binary \([n, k]\) codes. More generally,
for any $q$-ary systematic codes Massey proved that \[ P(n, k, q) \leq q^{k-n}(S_n - S_{n,k}) , \] where for given $k \geq 0$

\[ S_{n,k} = \sum_{i=1}^{n} \binom{n-k}{i} \left( \frac{j}{n} \right)^i \left( 1 - \frac{i}{n} \right)^{n-i} . \] (7)

Note that $S_n = S_{n,0}$.

Massey’s bound (6) is more subtle to deal with since one can expect that the leading terms of the asymptotic expansions of $S_n$ and $S_{n,k}$ may be of the same order. In addition, for $k = O(n)$ the second term in the asymptotic expansion plays a dominant role. In such cases, the asymptotic expansion of the sum $S_{n,k}$ is of prime interest. In this note we only consider $k = O(1)$.

In the next section, using similar technique to the one applied to prove Theorem 1, we establish an asymptotic expansion of $S_{n,k}$ for all $k \geq 0$.

**Theorem 2.** (i) For large $n$, the sum $S_{n,k}$ admits the following asymptotic expansion for $k \leq 2$

\[ S_{n,1} = \frac{1}{2} \sqrt{\frac{n \pi}{2}} - \frac{2}{3} + \frac{\sqrt{2} \pi}{48} \frac{1}{\sqrt{n}} + O \left( \frac{1}{n} \right) , \] (8)

\[ S_{n,2} = \frac{3}{8} \sqrt{\frac{n \pi}{2}} - \frac{2}{3} - \frac{\sqrt{2} \pi}{192} \frac{1}{\sqrt{n}} + O \left( \frac{1}{n} \right) . \] (9)

(ii) In general, for fixed $k \geq 1$

\[ S_{n,k} = \frac{1}{2^k \pi \binom{2k}{k}} \sqrt{\frac{n \pi}{2}} - \frac{2}{3} + O \left( \frac{1}{\sqrt{n}} \right) . \] (10)

for large $n$.

**Remarks.** (i) Define $\tilde{S}_{n,k} = S_{n,k} + 1$. T. Kløve in a private communication [10] observed that $\tilde{S}_n = 2\tilde{S}_{n,1}$. His derivation follows: First, note that by changing the order of summation we obtain

\[ \tilde{S}_{n,1} = \frac{1}{n^n} \sum_{i=0}^{n} \binom{n-1}{i} i^i (n-i)^{n-i} = \frac{1}{n^n} \sum_{i=0}^{n} \binom{n-1}{n-i} (n-i)^{n-i} i^i . \]

Hence

\[ 2\tilde{S}_{n,1} = \frac{1}{n^n} \sum_{i=0}^{n} \left\{ \binom{n-1}{i} + \binom{n-1}{n-i} \right\} i^i (n-i)^{n-i} = \tilde{S}_n . \]
(ii) The coefficient $\alpha_k = 2^{-2k} \binom{2k}{k}$ in front of $\sqrt{n\pi}/2$ in $S_{n,k}$ was conjectured by T. Kløve [10] based on author’s preliminary asymptotics of $S_{n,k}$ for $k \leq 6$. This guess was instrumental for us to prove the formula on $\alpha_k$. We thank T. Kløve for his help.

(iii) P. Kirschenhofer and H. Prodinger observed that $S_n$ is related to Ramanujan’s $Q$-function defined as

$$Q(n) = \sum_{k=1}^{n-1} \frac{n!}{(n-k)!(n^k)}. \quad (11)$$

Actually, P. Kirschenhofer in a private communication [8] proved that $S_n = Q(n)$, and his derivation is presented at the end of Section 2.1 for the completeness of this analysis.

Asymptotics of $Q(n)$ were discussed in [6, 11]. Theorems 1 and 2 can be used to assess the quality of the Massey and Kløve upper bounds for the probability of undetected errors. The following is an easy conclusion from what we have told so far.

**Corollary** For all $[n,k; q]$ codes and all systematic codes $[n,k; q]$ the following bounds respectively hold for the probability of undetected errors

$$P(n,k,q) \leq (q^k - 1)q^{-n} \left( \frac{n\pi}{2} + O(1) \right), \quad (12)$$

$$P(n,k,q) \leq q^k (1 - \alpha_k)q^{-n} \left( \frac{n\pi}{2} + O(1) \right), \quad (13)$$

where $\alpha_k = 2^{-2k} \binom{2k}{k}$. Thus for large $n$ and every $k$, Massey’s bound on the average worst case probability over all systematic codes $[n,k; q]$ is better than corresponding Kløve’s bound over all codes $[n,k; q]$.

In the next section we prove the above theorems using a combination of enumeration of trees [19], Lagrange’s inversion formula [18, 19], and singularity analysis [4]. The proof might be of its own interest even if it applies standard tools from the toolkit of the analytical analysis of algorithms (cf. [11, 18, 19]) due to several applications mentioned above.

2. Analysis: Trees Enumerations and Singularity Analysis

We discuss separately proofs of Theorem 1 and Theorem 2 in the next two subsections.

2.1 Basic Analysis: Proof of Theorem 1

We first concentrate on proving Theorem 1 for $y = 0$. The extension to any $y$ is simple, and we deal with it at the end of this subsection. It turns out that it is easier to work with

$$S_n = \sum_{i=0}^{n} \binom{n}{i} \left( \frac{i}{n} \right)^i \left( 1 - \frac{i}{n} \right)^{n-i}. \quad (14)$$
instead of $S_n$. Note that $S_n = \tilde{S}_n - 1$. Let $s_n = n^n \tilde{S}_n$, and we denote by $s(z)$ the exponential generating function of $s_n$, that is, $s(z) = \sum_{n \geq 0} s_n z^n / n!$. From (3) and elementary properties of convolutions we observe that $s(z) = (B(z))^2$ where

$$B(z) = \sum_{k=0}^{\infty} \frac{k^k}{k!} z^k,$$

where $|z| < e^{-1}$ to assure convergence of the above series. (To see the above, one should take convolution of the sequence $a_n = n^n / n!$ with itself.)

We need a formula for $B(z)$ which turns out to be related to a well known result on a tree enumeration due to A. Cayley (cf. [18, 19]). Let us consider the number $t_n$ of rooted labeled trees on $n$ vertices. It can be proved that $t_n = n^{n-1}$ (cf. [19]), and its exponential generating function $T(z)$ satisfies the following functional equation [18, 19]

$$T(z) = ze^{T(z)}.$$

where

$$T(z) = \sum_{n=1}^{\infty} \frac{n^{n-1}}{n!} z^n$$

for $|z| < e^{-1}$. The series expansion (17) follows from (16) and Lagrange’s inversion formula (cf. [19]).

In order to find $B(z)$, we differentiate once $T(z)$, and observe that $s(z) = zT'(z) + 1$. Furthermore, differentiating the functional equation (16), and after some simple algebra, we finally show that

$$B(z) = \frac{1}{1 - T(z)},$$

thus

$$s(z) = \frac{1}{(1 - T(z))^2}.$$

We have to stress again that this equation is true only for $|z| < e^{-1}$, that is, $z = e^{-1}$ is a singularity of $B(z)$ which can be used to obtain asymptotics of $s_n$. This can be seen by viewing (16) as a definition of $z(T) = Te^{-T}$ function that achieves its maximum value $z = e^{-1}$ at $T = 1$.

To apply the singularity analysis of Flajolet and Odlyzko [4] we need to expand $T(z)$ around $z = e^{-1}$. We first consider equation (16), write $z = Te^{-T}$, and then treat $T$ as an independent variable. Expanding the last equation around $e^{-1}$ one gets

$$Te^{-T} = e^{-1} + \frac{e^{-1}}{2}(T - 1)^2 + \frac{e^{-1}}{3}(T - 1)^3 + O((T - 1)^4).$$
Solving for $T$ we finally arrive at (we have used Maple)

$$
T(z) - 1 = \sqrt{2(1 - \epsilon z)} + \frac{2}{3}(1 - \epsilon z) + \frac{11\sqrt{2}}{36}(1 - \epsilon z)^{3/2} + \frac{43}{135}(1 - \epsilon z)^{2} + O((1 - \epsilon z)^{5/2}).
$$

Write now $h(z) = (1 - \epsilon z)$, and then

$$
s(z) = \frac{1}{2h(z) \left(1 + \frac{2\sqrt{2}}{3}h(z) + \frac{11}{36}h(z) + O(h^{3/2}(z))\right)^{2}}
$$

$$
= \frac{1}{2(1 - \epsilon z)} + \frac{\sqrt{2}}{3\sqrt{1 - \epsilon z}} + \frac{1}{36} + \frac{\sqrt{2}}{540}\sqrt{1 - \epsilon z} + O(1 - \epsilon z),
$$

and finally

$$
\bar{S}_n = \frac{n!}{n^n} [z^n] s(z)
$$

where $[z^n]f(z)$ denotes the coefficient at $z^n$ in $f(z)$.

To extract the coefficients at $z^n$ from (21), we use the singularity analysis and transfer theorems of Flajolet and Odlyzko [4] which allow us to compute separately the coefficients for every function involved in the asymptotic expansion. We find:

$$
[z^n] \left( \frac{1}{\sqrt{1 - \epsilon z}} \right) = \frac{e^n}{\sqrt{\pi} n} \left( 1 - \frac{1}{8n} + O(1/n^2) \right),
$$

$$
[z^n] \left( \sqrt{1 - \epsilon z} \right) = -\frac{e^n}{\sqrt{\pi} n^3} \left( \frac{1}{2} + \frac{3}{16n} \right),
$$

$$
[z^n] \left( \frac{1}{1 - \epsilon z} \right) = e^n,
$$

$$
\frac{n!}{n^n} = e^{-n}\sqrt{2\pi} n \left( 1 + \frac{1}{12n} + O(1/n^2) \right).
$$

After some simply algebra, the theorem for $y = 0$ is finally proved.

To extend it to any $y$, we only need a new formula for the generating function $B(z)$. Entry 13 of Ramanujan (cf. [2], pp. 80, Eq. (16.8)) gives the following

$$
B(z) = \sum_{k=0}^{\infty} \frac{(y + k)^k}{k!} z^k = \frac{e^{yT(z)}}{1 - T(z)} \quad |z| < e^{-1}
$$

where $T(z)$ is defined in (16). Using this, and our previous arguments we easily prove Theorem 1 for any $y$. This completes the proof of Theorem 1.

In fact, using our derivation we can obtain as many terms in the asymptotic expansion of $S_n$ as we wish. It is easy to see that the full asymptotic expansion is as follows:

$$
S_n = \sum_{k=-1}^{\infty} \frac{c_k}{n^{k/2}}.
$$
In Theorem 1 we proved that $e_{-1} = \sqrt{\pi}/2$, $a_0 = -1/3$, $c_1 = \sqrt{2\pi}/24$, $c_2 = -4/135$, and $c_3 = -71/1152 \cdot \sqrt{\pi}/2$ (cf. [11] Sec. 1.2.11.3).

In passing, we should point out that since all terms in $S_n$ are positive, other approaches are possible (cf. [11, 17]). However, a detailed and subtle analysis is required for such methods.

Finally, as promised in Remark (iii), we now present Kirschenhofer’s proof of $S_n = Q(n)$ where $Q(n)$ is defined in (11) (cf. [8]). As we have already observed

$$S_n = -1 + \frac{n!}{n^n [z^n]} \frac{1}{(1 - T(z))^2}.$$  

Differentiating (16), after simple algebra, we show that

$$\frac{1}{(1 - T)^2} = z(T + \log \frac{1}{1 - T})',$$

thus

$$S_n = -1 + \frac{n!}{n^n [z^n]} (T + \log \frac{1}{1 - T}) = \frac{n!}{n^n [z^n]} \log \frac{1}{1 - T}.$$  

But, it is well known that (cf. [13])

$$\log \frac{1}{1 - T} = \sum \frac{n^{n-1} Q(n) z^n}{n!}.$$  

This, and the above, proves that $S_n = Q(n)$. Another derivation can be obtained by a careful application of the Lagrange inversion formula.

### 2.2 Enhanced Analysis: Proof of Theorem 2

Now, we can wrestle with the proof of Theorem 2. As we shall see below, our approach from Section 2.1 pays off when dealing the the more challenging sum $S_{n,k}$. The analysis of $S_{n,k}$ follows the same line of arguments as above, so we only sketch it. Let $s_{n,k} = n^n \bar{S}_{n,k}$ and $s_k(z) = \sum_{n \geq 0} s_{n,k} z^n/(n-k)!$ for all $k \geq 0$. Using (15) one easily see that

$$s_k(z) = z^k B(z) B^{(k)}(z)$$

where $B(z) = 1/(1 - T(z))$ (cf. (18)), and $B^{(k)}(z)$ denotes the $k$th derivative of $B(z)$. This can be verified by taking the convolution of the coefficients of $B(z)$ and $B^{(k)}(z)$. Thus,

$$\bar{S}_{n,k} = \frac{(n-k)!}{n^n [z^n]} \left(z^k B(z) B^{(k)}(z)\right).$$

Using Stirling’s formula for fixed $k$ we arrive at

$$\frac{(n-k)!}{n^n} = e^{-n} \sqrt{2\pi n} \frac{1}{n^k} \left(1 - \frac{k^2}{2n}\right) \left(1 - \frac{k(1-2k)}{2n}\right) \left(1 + \frac{1}{12n} + O\left(\frac{1}{n^2}\right)\right),$$

7
thus it suffices to get an explicit formula for $B^{(k)}(z)$, which is not that trivial computationally.

Let us first start with an explicit formula on $s_k(z)$ for some small values of $k$. By (18), one can find (we used Maple)

\[
\begin{align*}
  s_1(z) &= \frac{T(z)}{(1 - T(z))}, \\
  s_2(z) &= \frac{T(z)(4 - T(z))}{(1 - T(z))^2}, \\
  s_3(z) &= \frac{T^3(z)(27 - 14T(z) + 2T^2(z))}{(1 - T(z))^8}, \\
  s_4(z) &= \frac{T^4(z)(256 - 203T(z) + 58T^2(z) - 6T^3(z))}{(1 - T(z))^{10}}. 
\end{align*}
\]

Using now (20), and the asymptotic expansion of $(1 - e^{-b})^{-\delta}$, namely (cf. [4, 19])

\[
[z^n](1 - e^{-b})^{-\delta} = \begin{cases} 
  e^n \frac{n^{\delta-1}}{\Gamma(\delta)} \left(1 + \frac{b(\delta-1)}{2n} + O(1/n)\right) & b \notin \{0, -1, -2, \ldots\} \\
  e^n \left(\frac{n^{\delta-1}}{b!}\right) & b \in \{0, -1, -2, \ldots\}, 
\end{cases}
\]

(22)

where $\Gamma(z)$ is the Euler gamma function (cf. [11]), we immediately establish part (i) of Theorem 2 along the same lines as in the proof of Theorem 1. For example, using (20) we can expand $s_1(z)$ around $z = e^{-1}$ to obtain

\[
s_1(z) = \frac{1}{4(1 - e^{-1})} - \frac{\sqrt{2}}{12(1 - e^{-1})^{3/2}} - \frac{1}{4(1 - e^{-1})} + O\left(\frac{1}{\sqrt{1 - e^{-1}}}\right)
\]

which implies (8) of Theorem 2(i). In a similar fashion we obtain (9).

To prove part (ii) of Theorem 2, one needs an explicit formula for $s_k(z)$ for all $k \geq 1$. Based on the above formulas on $s_k(z)$ for small $k$, we observe

\[
s_k(z) = \frac{T^k(z) \left(\sum_{l=0}^{k-1} (-1)^l \beta_l(k) T^l(z)\right)}{(1 - T(z))^{2(k+1)}},
\]

(23)

where $\beta_l(k)$ $(l = 0, 1, \ldots, k - 1)$ are constants. Fortunately, to recover the first two terms of the asymptotic expansion of $S_{n,k}$ we need only $\alpha'_k = \sum_{l=0}^{k-1} (-1)^l \beta_l(k)$.

The coefficients $\alpha'_k$ satisfy a recurrence equation that we present next. Differentiating $s_k(z)$, and after a long and tedious algebra we obtain a relationship between $s_{k+1}(z)$ and $s_k(z)$ which further leads to the following

\[
\alpha'_{k+1} = (2k + 1)\alpha'_k
\]

(23)

with $\alpha'_1 = 1$. Solving the above recurrence, one gets

\[
\alpha'_k = 1 \cdot 3 \cdot 5 \cdots (2k - 1) = \frac{(2k)!}{2^kk!}.
\]
The rest is easy. Setting $T(z) = 1 + O(1 - \varepsilon z)$ and $1 - T(z) = \sqrt{2(1 - \varepsilon z)}(1 + O(\sqrt{1 - \varepsilon z})$ in (23), and using the above we finally show that for fixed $k$

$$s_k(z) = \frac{\alpha_k'}{2^{k+1}(1 - \varepsilon z)^{k+1}} + \frac{\alpha_k'\sqrt{2}}{2^{k+13}(1 - \varepsilon z)^{k+1/2}} + O((1 - \varepsilon z)^{-k}).$$

Then, by (22) we obtain

$$S_{n,k} = \frac{(n - k)!}{n^n} \left( \frac{\alpha_k' e^n}{2^{k+1}} \binom{n}{k} + \frac{\sqrt{2}\alpha_k' e^n}{2^{k+13}\Gamma(k + \frac{1}{2})} n^{-k+\frac{1}{2}}(1 + O(n^{-1}) + O(n^{k-1})) \right)$$

where (cf. [1])

$$\Gamma(k + \frac{1}{2}) = \frac{\alpha_k'}{2\pi \sqrt{k}}.$$

After some algebra, we finally prove part (ii) of Theorem 2.

ACKNOWLEDGEMENT

We thank Professors P. Kirschenhofer and H. Prodinger (Technical University of Vienna, Austria) for commenting on the first drafts of this note. We are particularly obliged to Professor T. Klove (University of Bergen, Norway) for finding some numerical slips in the first versions of this note, providing more numerical data and explanations, and for making a conjecture concerning the constant $\alpha_k$ that was instrumental in proving it.

References


