DIFFERENCE OF SUMS CONTAINING PRODUCTS OF BINOMIAL COEFFICIENTS AND THEIR LOGARITHMS

ALLEN R. MILLER* AND IRA S. MOSKOWITZ†

Abstract. Properties of the difference of two sums containing products of binomial coefficients and their logarithms which arise in the application of Shannon's information theory to a certain class of covert channels are deduced. Some allied consequences of the latter are also recorded.

Key words. sums of binomial coefficients and their logarithms, covert channels, anonymity, traffic analysis

AMS subject classifications. 05A10; 94A40

1. Introduction. Traffic analysis in anonymizing network configurations can be used to open a covert communication channel. Recently, Moskowitz et al. [2] have developed a mathematical model (whose salient feature is given by equation (1.1) below) that describes the situation of oneway messages passing from private Enclave_1 to private Enclave_2 in which each communication is encrypted and passes through exit and entry mix firewalls. Furthermore, there are $n+1$ senders in Enclave_1, and one of them called Alice is malicious. The other n clueless transmitters are benign. Every sender may send at most one message per unit time t to Enclave_2. All messages from Enclave_1 to Enclave_2 pass through a public line that is subject to eavesdropping by an eavesdropper called Eve who knows the value of n. The only action Eve can take is to count the number of messages per t going from Enclave_1 to Enclave_2. The n clueless transmitters all send their messages per t as independently and identically distributed Bernoulli random variables with parameter q. Alice acts independently (through ignorance of the n clueless senders) when deciding to send a message. Thus, Alice by sending or not sending a message affects the number of messages that Eve counts. This in brief is the covert channel.

A normalizing noise term $S(n)$ is defined corresponding to the degree of anonymity afforded by the n clueless senders transmitting as fair coins ($q = 1/2$) which is the situation when maximal anonymity occurs. For $n = 1, 2, 3, \ldots$ $S(n)$ is shown in [2] to be given by

$$S(n) = \frac{1}{2^n} \sum_{k=0}^{n} \left[\frac{1}{2} \binom{n+1}{k} \ln \binom{n+1}{k} - \binom{n}{k} \ln \binom{n}{k} \right], \qquad (1.1)$$

and by employing Shannon's information theory the capacity is given by $1 - S(n)/\ln 2$. Moreover, an important property of this model is that the capacity should decrease monotonically to zero with increasing n.

For brevity in the sequel we define $T(n) \equiv 2^n S(n)$. The expression for $T(n)$ is of interest in its own right, since it is the difference of two divergent sums that contain products of binomial coefficients and their logarithms. Moreover, these types of sums appear not to be readily found in the mathematical literature.

In the present investigation we shall obtain in section 2 a simpler algebraic representation for $T(n)$ which contains the sum $\sum_{k=0}^{n} \binom{n}{k} \ln(k+1)$ that is similar, for

*1616 Eighteenth Street NW, Washington, DC 20009.

†Center for High Assurance Computer Systems, Naval Research Laboratory, Washington, DC 20375.

example, to a particular sum involved in a representation for $\ln \Gamma(x+1)$ deduced by Stirling for $x > -1$ (see [1, equation (5.47)]). We shall also derive two integral representations and two representations involving integrals for $S(n)$. Then in section 3 we give a detailed proof that $S(n)$ is monotonically increasing with increasing n. This property of $S(n)$ alluded to above was verified numerically in [2] for $n \leq 7750$. Finally, in section 4, we show that $S(n) \to \ln 2$ as $n \to \infty$. We record some consequences of the latter as well.

2. Reduction of $S(n)$ to a simpler sum. Recall that $T(n) = 2^n S(n)$ where $S(n)$ is given by equation (1.1). We shall show below that

$$T(n) = 2^n \ln(n+1) - \sum_{k=0}^{n} \binom{n}{k} \ln(k+1) \tag{2.1}$$

and that $S(n)$ may be written elegantly as

$$S(n) = \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} \ln\left(\frac{n+1}{k+1}\right) . \tag{2.2}$$

To prove the latter we define

$$\alpha(n) \equiv \sum_{k=0}^{n} \binom{n}{k} \ln \binom{n}{k} ,$$

where the binomial coefficient $\binom{n}{k} = n!/k!(n-k)!$ which we agree vanishes if $k > n$ or if $k < 0$. Thus

$$\alpha(n) = \ln n! \sum_{k=0}^{n} \binom{n}{k} - \sum_{k=0}^{n} \binom{n}{k} \ln k! - \sum_{k=0}^{n} \binom{n}{k} \ln(n-k)! . \tag{2.3}$$

The first sum equals 2^n. By noting that

$$\binom{n}{k} = \binom{n}{n-k} \tag{2.4}$$

and reversing the order of summation (i.e. by letting $k \mapsto n-k$) in the third sum, we see that the second and third sums in equation (2.3) are equal so that

$$\alpha(n) = 2^n \ln n! - 2 \sum_{k=0}^{n} \binom{n}{k} \ln k! \tag{2.5a}$$

and so

$$\alpha(n+1) = 2^{n+1} \ln(n+1)! - 2 \sum_{k=0}^{n+1} \binom{n+1}{k} \ln k! . \tag{2.5b}$$

From equation (1.1) and the definitions of $T(n)$ and $\alpha(n)$ it is evident that

$$T(n) = \frac{1}{2}\alpha(n+1) - \alpha(n)$$

and so by using equations (2.5) we have

$$T(n) = 2^n \ln(n+1) + \sum_{k=0}^{n+1} \left[2 \left(\begin{array}{c} n \\ k \end{array} \right) - \left(\begin{array}{c} n+1 \\ k \end{array} \right) \right] \ln k! \; . \tag{2.6}$$

Now observing that

$$\left(\begin{array}{c} n+1 \\ k \end{array} \right) = \left(\begin{array}{c} n \\ k \end{array} \right) + \left(\begin{array}{c} n \\ k-1 \end{array} \right) \tag{2.7}$$

equation (2.6) gives

$$T(n) = 2^n \ln(n+1) + \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \ln k! - \sum_{k=1}^{n+1} \left(\begin{array}{c} n \\ k-1 \end{array} \right) \ln k! \; .$$

Finally, since

$$\sum_{k=1}^{n+1} \left(\begin{array}{c} n \\ k-1 \end{array} \right) \ln k! = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \ln(k+1)!$$

we obtain equation (2.1).

Thus also

$$S(n) = \ln(n+1) - \frac{1}{2^n} \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \ln(k+1) \tag{2.8}$$

which shows that $-S(n)$ is the difference of a divergent weighted sum of logarithms $\ln(k+1)$ (since $\sum_{k=0}^{n} 2^{-n} \left(\begin{array}{c} n \\ k \end{array} \right) = 1$) and the montonically increasing logarithm $\ln(n+1)$. This observation brings to mind and is analogous to one of the many formulations for Euler's constant γ (cf. e.g. [1, equation (6.65)]) which is the limiting value as $n \to \infty$ of the much simpler difference of the divergent harmonic series $\sum_{k=1}^{n} 1/k$ and $\ln(n+1)$. (However, we show in section 4 that $S(n) \to \ln 2$ as $n \to \infty$.) We have already noted that series of the type appearing in equation (2.1) appear in the literature in other contexts.

We conclude this section by deriving four integral representations for $S(n)$. We start with (see [3, section 2.6.17., equation (1)])

$$\ln \left(\frac{n+1}{k+1} \right) = \int_0^1 (x^n - x^k) \frac{dx}{\ln x} \; , \tag{2.9}$$

where $n > -1$ and $k > -1$. Multiplying both sides of equation (2.9) by the binomial coefficient $\left(\begin{array}{c} n \\ k \end{array} \right)$ and summing over the index k gives

$$\sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \ln \left(\frac{n+1}{k+1} \right) = \int_0^1 [2^n x^n - (1+x)^n] \frac{dx}{\ln x} \; .$$

Dividing each side of the latter equation by 2^n and noting equation (2.2) we deduce for $n = 1, 2, 3 \ldots$

$$S(n) = \int_0^1 \left[x^n - \left(\frac{1+x}{2} \right)^n \right] \frac{dx}{\ln x} \, . \tag{2.10a}$$

Now making the transformation $x = e^{-t}$ in the latter integral yields

$$S(n) = \int_0^\infty e^{-(n+1)t} \left[\left(\frac{1+e^t}{2} \right)^n - 1 \right] \frac{dt}{t} \, . \tag{2.10b}$$

However, letting $n = 0$ in equation (2.9) we may start with

$$\ln(k+1) = \int_0^1 (x^k - 1) \frac{dx}{\ln x} \, .$$

Multiplying the latter equation by $\binom{n}{k}$ and summing over k now gives

$$\sum_{k=0}^{n} \binom{n}{k} \ln(k+1) = \int_0^1 [(1+x)^n - 2^n] \frac{dx}{\ln x}$$

which upon dividing by 2^n and noting equation (2.8) yields

$$S(n) = \ln(n+1) + \int_0^1 \left[1 - \left(\frac{1+x}{2} \right)^n \right] \frac{dx}{\ln x} \, . \tag{2.11a}$$

Upon making the transformation $x = e^{-t}$ in the integral this result can also be written as

$$S(n) = \ln(n+1) + \int_0^\infty \left[\left(\frac{1+e^{-t}}{2} \right)^n - 1 \right] \frac{e^{-t}}{t} dt \, . \tag{2.11b}$$

3. Monotonicity and boundedness of $S(n)$. We now prove that $S(n)$ is increasing with increasing $n \geq 1$, i.e. $S(n+1) > S(n)$ for $n = 1, 2, 3, \ldots$.

First we observe that when the positive integer $n \geq 1$ is replaced by the real variable $t > 0$ in equations (2.10) and (2.11), then the latter provide continuous and differentiable analogues of the discrete sum $S(n)$ given by equation (1.1). Thus, for example, we have from equation (2.11a) for $t > 0$

$$S(t) = \ln(1+t) + \int_0^1 \left[1 - \left(\frac{1+x}{2} \right)^t \right] \frac{dx}{\ln x} \, .$$

Now differentiating $S(t)$ with respect to t gives

$$\frac{dS}{dt} = \frac{1}{1+t} - \int_0^1 \left(\frac{1+x}{2} \right)^t \left[\frac{\ln(1+x) - \ln 2}{\ln x} \right] dx \, , \tag{3.1}$$

where it is easy to see that on the closed interval $0 \leq x \leq 1$

$$0 \;\leq\; \frac{\ln(1+x) - \ln 2}{\ln x} \;\leq\; \frac{1}{2} \, .$$

Thus obviously

$$\int_0^1 \left(\frac{1+x}{2}\right)^t \left[\frac{\ln(1+x)-\ln 2}{\ln x}\right] dx \;<\; \frac{1}{2^{1+t}} \int_0^1 (1+x)^t dx \;.$$

Since

$$\frac{1}{2^{1+t}} \int_0^1 (1+x)^t dt \;=\; \frac{1-2^{-(1+t)}}{1+t} \;<\; \frac{1}{1+t}$$

it is evident that

$$\int_0^1 \left(\frac{1+x}{2}\right)^t \left[\frac{\ln(1+x)-\ln 2}{\ln x}\right] dx \;<\; \frac{1}{1+t} \;.$$

Finally, from equation (3.1) we have $dS/dt > 0$ and so $S(t)$ is a strictly increasing function of $t > 0$ thus proving our assertion for $S(n)$.

We conclude this section by showing for $n \geq 1$ that

$$S(n) < 2(1-2^{-n-1}) < 2 \;. \tag{3.2}$$

Assuming the latter, since $S(n)$ is monotonically increasing and bounded we may define

$$\mu \equiv \lim_{n\to\infty} S(n) \;,$$

where from inequalities (3.2) we must have $0 < \mu < 2$.

To show inequalities (3.2), upon observing that

$$\ln\left(\frac{n+1}{k+1}\right) < \frac{n+1}{k+1} \qquad (k = 0, 1, \ldots, n) \;,$$

we have from equation (2.2)

$$S(n) < \frac{n+1}{2^n} \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \frac{1}{k+1} \;.$$

Since (see e.g.[3, section 4.2.2., equation (42)])

$$\sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \frac{1}{k+1} = \frac{1}{n+1}(2^{n+1}-1) \;,$$

it is evident that $S(n) < 2 - 2^{-n}$ and we are done.

4. The limiting value of $S(n)$ as $n \to \infty$. From equation (2.2) we may write

$$S(n) = -\frac{1}{2^n} \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \ln\left(1 - \frac{n-k}{n+1}\right) \;.$$

Letting $k \mapsto n-k$ and using equation (2.4) in this result then yields

$$S(n) = -\frac{1}{2^n} \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \ln\left(1 - \frac{k}{n+1}\right) \;. \tag{4.1}$$

Since for $0 \le z < 1$

$$\ln(1-z) = -\sum_{p=1}^{\infty} \frac{z^p}{p} \tag{4.2}$$

setting $z = k/(n+1)$ we have from equations (4.1) and (4.2)

$$S(n) = \frac{1}{2^n} \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \sum_{p=1}^{\infty} \frac{1}{p} \left(\frac{k}{n+1} \right)^p .$$

Formally interchanging the order of summation then gives

$$S(n) = \frac{1}{2^n} \sum_{p=1}^{\infty} \frac{1}{p} \; \frac{1}{(n+1)^p} \sum_{k=1}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) k^p \; , \tag{4.3}$$

where for integers $p > 0$

$$\sum_{k=1}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) k^p = 2^{n-p} \left(\begin{array}{c} n \\ p \end{array} \right) p! + 2^n \sum_{k=1}^{p-1} \left(\begin{array}{c} n \\ k \end{array} \right) \left(-\frac{1}{2} \right)^k \sum_{\ell=0}^{k} (-1)^\ell \left(\begin{array}{c} k \\ \ell \end{array} \right) \ell^p \; , \tag{4.4}$$

and it is understood that the latter second term vanishes when $p = 1$. Equation (4.4) is derived *ab initio* by Schwatt (cf. [4, p. 84, equation (34)]) where here a misprint is now corrected.

Since $\left(\begin{array}{c} n \\ p \end{array} \right) = n!/p!(n-p)!$ we have from equations (4.3) and (4.4)

$$S(n) = \sum_{p=1}^{\infty} \frac{(1/2)^p}{p} \; \frac{M_p(n)}{(n+1)^p} + \sum_{p=1}^{\infty} \frac{1}{p} \; \frac{Q_{p-1}(n)}{(n+1)^p} \; , \tag{4.5}$$

where we have defined

$$M_p(n) \equiv \frac{n!}{(n-p)!} \tag{4.6}$$

and

$$Q_{p-1}(n) \equiv \sum_{k=1}^{p-1} \left(\begin{array}{c} n \\ k \end{array} \right) \left(-\frac{1}{2} \right)^k \sum_{\ell=0}^{k} (-1)^\ell \left(\begin{array}{c} k \\ \ell \end{array} \right) \ell^p \; . \tag{4.7}$$

Observing that the first and second p-summations in equation (4.5) will terminate respectively when $p > n$ and $p - 1 > n$, this result can only be valid as a divergent asymptotic expansion for $S(n)$ about $n = \infty$.

However, it is easily seen from equation (4.6) that

$$M_p(n) = n(n-1) \cdots (n-(p-1))$$

and so $M_p(n)$ is a monic polynomial in n of degree p. Thus for each $p > 0$ it is evident that

$$\frac{M_p(n)}{(n+1)^p} = 1 + O\left(\frac{1}{n} \right) \qquad (n \to \infty) \; . \tag{4.8a}$$

Moreover, from equation (4.7) it is also easily seen that $Q_{p-1}(n)$ is a polynomial in n of degree $p-1$. Let c_{p-1} be the coefficient of n^{p-1} in $Q_{p-1}(n)$. Then also from equation (4.7) we have

$$c_{p-1} = \frac{(-1/2)^{p-1}}{(p-1)!} \sum_{\ell=0}^{p-1} (-1)^\ell \left(\begin{array}{c} p-1 \\ \ell \end{array} \right) \ell^p \ .$$

Schwatt [4, p. 101, equation (196)] has shown for $p \geq 1$ that

$$\sum_{\ell=0}^{p-1} (-1)^\ell \left(\begin{array}{c} p-1 \\ \ell \end{array} \right) \ell^p = (-1)^{p-1} \frac{1}{2} (p-1) p!$$

so that

$$c_{p-1} = \left(\frac{1}{2} \right)^p p(p-1) \ .$$

Thus we have

$$\frac{Q_{p-1}(n)}{(n+1)^p} = \left(\frac{1}{2} \right)^p p(p-1) O\left(\frac{1}{n} \right) \qquad (n \to \infty) \tag{4.8b}$$

and by using equations (4.8) we have from equation (4.5) the asymptotic result

$$S(n) = \left[1 + O\left(\frac{1}{n} \right) \right] \sum_{p=1}^{\infty} \frac{(1/2)^p}{p} + O\left(\frac{1}{n} \right) \sum_{p=1}^{\infty} \left(\frac{1}{2} \right)^p (p-1) \qquad (n \to \infty) \ . \tag{4.9}$$

Recalling equation (4.2) the first sum in equation (4.9) equals $\ln 2$ and it is easily shown that the second sum reduces to unity so that we can now write equation (4.9) as

$$S(n) = \left[1 + O\left(\frac{1}{n} \right) \right] \ln 2 + O\left(\frac{1}{n} \right) \qquad (n \to \infty) \ .$$

Finally, since we have already shown in section 3 that $\lim_{n \to \infty} S(n)$ exists as a positive number $\mu < 2$, it is evident from the latter result that $\mu = \ln 2$.

We end by recording two interesting corollaries that are obtained immediately from equations (2.10), namely

$$\lim_{n \to \infty} \int_0^1 \left[x^n - \left(\frac{1+x}{2} \right)^n \right] \frac{dx}{\ln x} = \ln 2$$

and

$$\lim_{n \to \infty} \int_0^1 e^{-(n+1)x} \left[\left(\frac{1+e^x}{2} \right)^n - 1 \right] \frac{dx}{x} = \ln 2 \ .$$

5. Conclusions. We have discussed how an important application in anonymity and covert channels leads to the problem of deducing properties of the difference of sums containing products of binomial coefficients and their logarithms. Intuitively, the capacity of the covert channel should decrease monotonically to zero as the number of transmitters increases, and the analysis provided herein proves this.

Acknowledgments. We thank the anonymous reviewers for spotting an error in an earlier version of this paper. In correcting the paper we were able to vastly simplify the proof, so we are doubly indebted to the reviewers.

REFERENCES

[1] R.L. Graham, D.E. Knuth and O. Patashnik, *Concrete Mathematics*, Addison-Wesley, Reading, 1989.

[2] I.S. Moskowitz, R.E. Newman, D.P. Crepeau and A.R. Miller, *Covert channels and anonymizing networks*, Proceedings 2003 Workshop on Privacy in the Electronic Society (eds. P. Samarati & P. Syverson), pp. 79-88, ACM Press, 2003.

[3] A.P. Prudnikov, Yu. A. Brychkov and O.I. Marichev, *Integral and Series, Vol. 1*, Gordon and Breach, New York, 1986.

[4] I.J. Schwatt, An Introduction to the Operations with Series, 2nd edition, Chelsea, New York, 1924.