Journal of Integer Sequences, Vol. 16 (2013),
Article 13.3.7

Distribution of Values of the Binomial Coefficients and the Stern Sequence

Jennifer Lansing[1]
Department of Mathematics
University of Illinois at Urbana-Champaign
1409 W. Green St.
Urbana, IL 61801
USA
jlweber@illinois.edu

Abstract

The distribution of values in a sequence, as seen in counting the number of terms larger than a scalar times the average value is applied to the the Stern sequence and the binomial coefficients. We also show that the suitably scaled logarithms of $\{\binom{n}{k} : 0 \leq k \leq n\}$ converges to a distribution, and prove an implicit formula for the convergent function. We leave an open problem regarding the distribution of the Stern sequence.

1 Overview

In Section 2, we give a brief motivation and overview of problems on the distribution of values in sequences, as well as the distribution of spacings in sequences. Then in Section 3, we investigate the distribution of values of the binomial coefficients in comparison to the scaled average value, and we also present some data on this question. Then in Section 4, we make a slight variation in the problem, prove that a limiting distribution exists, and prove an implicit formula for the limiting distribution. Finally, in Section 5, we look at the distribution of values of the Stern sequence.

[1]The author acknowledges support from National Science Foundation grant DMS 08-38434"EMSW21-MCTP: Research Experience for Graduate Students".

2 Background and motivation

Historically there have been two stages to understanding the distribution of sequences, from a number theoretical standpoint. The first and more classical stage involves looking at the distribution of values in a sequence. Hermann Weyl [16] made many advances in this area by proving certain sequences are uniformly distributed. For example, Weyl proved that if α is an irrational number, then for any positive integer d, the sequence $\{\alpha n^d\}$ is uniformly distributed.

Once the distribution of values for a sequence is well understood, then a more modern approach is to look at the distribution of spacings between consecutive terms. Some famous results in this area include work by Hooley [4, 5] and Gallagher [2] on the distribution of gaps between consecutive primes. The limiting distribution for average gaps between primes is Poisson, and so are the spacings of fractional parts of lacunary sequences (see [10]). Another famous result in this area is the Steinhaus conjecture, also known as the "three-gap theorem". For example, for any irrational α, the gaps between consecutive terms after ordering the sequence $\{\alpha n\}$ up to a certain N, will only take 3 values, one of which is the sum of the other two (see [12, 15]). For a more complete background on the distribution of spacings, see the first few pages of [1].

The distribution of values is well understood for many sequences, but what about the Stern sequence? With the goal in mind of studying the distribution of spacing of the Stern sequence, we first need to understand the distribution of values. The Stern sequence [14], which originated in a triangular array similar to Pascal's triangle, can also be defined recursively. The Stern sequence, denoted by $s(n)$, satisfies the recurrences

$$s(2n) = s(n), \quad s(2n+1) = s(n+1) + s(n) \quad \text{with } s(0) = 0 \text{ and } s(1) = 1.$$

Understanding the distribution of values for a row of the Stern sequence in the triangular array (called the diatomic array), is not an easy problem. As a means of trying to understand the distribution of the Stern sequence, we look at a potentially similar sequence, the binomial coefficients. The initial idea that these two sequences might be comparable in behavior of distribution came from the sum of values in a row. For example, the sum of values in the n-th row of the diatomic array is 3^n, and the sum of the binomial coefficients $\binom{n}{k}$ for a fixed n is 2^n. With this motivation of understanding distribution of the Stern sequence better, the author investigated the distribution of values of the binomial coefficients.

3 First glance

As a means of understanding the distribution of the values of the binomial coefficients, we compare them to the average value. First note that the average value of the binomial coefficients is $2^n/(n+1)$.

For a fixed n, we define

$$F(\lambda, n) := \frac{\#\left\{0 \le k \le n : \binom{n}{k} \ge \lambda \frac{2^n}{n+1}\right\}}{n+1}$$

to be the counting function for the number of binomial coefficients that are larger than λ times the average value.

Remark 1. A probabilistic interpretation of this would be finding at how many points k, with $k = 0, 1, \ldots, n$, does the probability mass function $f_k(n) := \binom{n}{k} 2^{-n}$ lie above $\lambda/(n+1)$. However, we will proceed from a number theoretic standpoint.

Then for various values of n, we compute $F(\lambda, n)$ to see if there is a limiting function as n goes to infinity. Figure 1 compares the values of $F(\lambda, 2^8)$, $F(\lambda, 2^9)$, and $F(\lambda, 2^{10})$. The curve

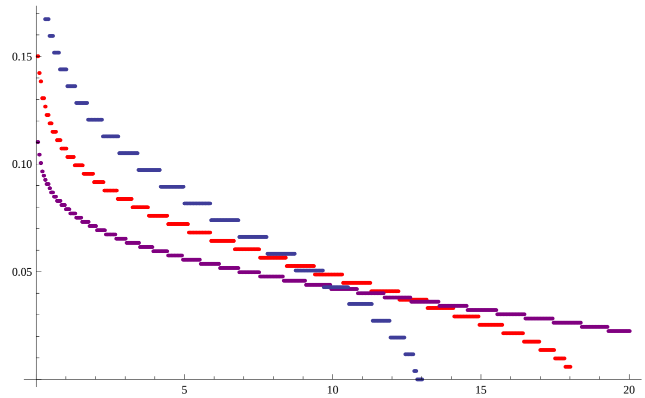


Figure 1: Comparing $F(\lambda, 2^8)$, $F(\lambda, 2^9)$, and $F(\lambda, 2^{10})$

of $F(\lambda, 2^8)$ is given in blue and is on top for λ close to 0. The curve of $F(\lambda, 2^9)$ is given in red and is the middle curve. The curve for $F(\lambda, 2^{10})$ is given in purple and is on bottom for λ close to 0. The data suggests that a nontrivial limiting function does not exist, possibly because the comparing function grows too fast as compared to the binomial coefficients. It is likely that $F(\lambda, n)$ converges to $F(\lambda) = 0$ for $\lambda > 0$, with $F(0) = 1$.

4 Results

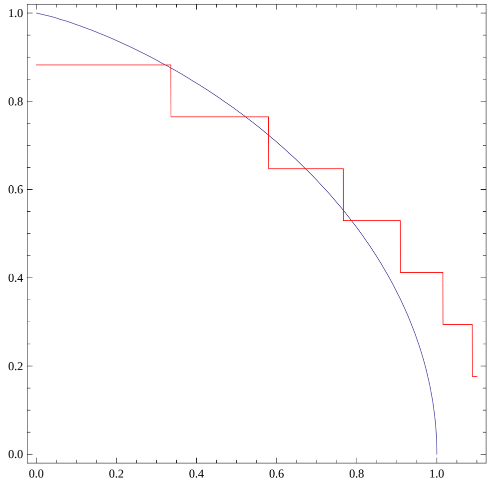
We now vary the problem in perhaps an unexpected way. Instead of allowing the variable λ to be a multiplier, we want the counting function to converge to a limiting function, so we raise the average value to λ instead. We now define the counting function to be

$$G(\lambda, n) := \frac{\#\left\{0 \le k \le n : \binom{n}{k} \ge \left(\frac{2^n}{n+1}\right)^\lambda\right\}}{n+1}.$$

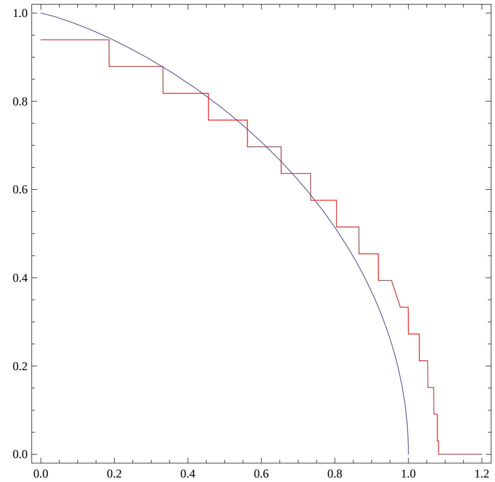
We will also denote the limit, whose existence we shall establish later, by

$$G(\lambda) := \lim_{n \to \infty} G(\lambda, n).$$

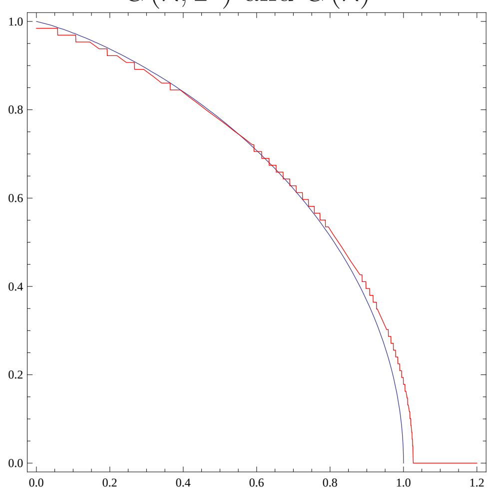
Computing $G(\lambda, n)$ for various n, as seen in Figure 2, we see that a limiting function does seem to exist. The sequence of functions seems to converge to $G(\lambda)$ fairly quickly; the error is roughly 0.0125 for $G(\lambda, 2^{10})$, 0.00556 for $G(\lambda, 2^{11})$, and 0.00312 for $G(\lambda, 2^{13})$. This data then leads us to the following theorem.



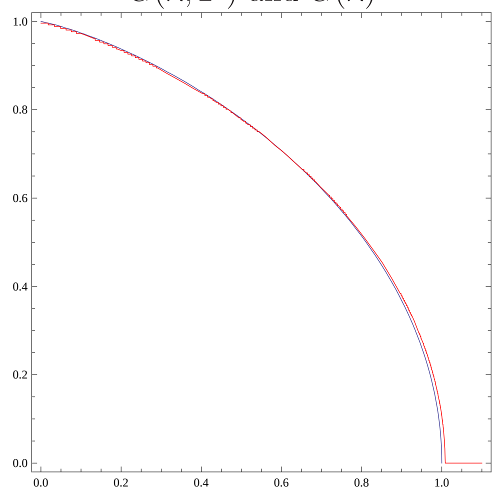
$G(\lambda, 2^4)$ and $G(\lambda)$



$G(\lambda, 2^5)$ and $G(\lambda)$



$G(\lambda, 2^7)$ and $G(\lambda)$



$G(\lambda, 2^9)$ and $G(\lambda)$

Figure 2: $G(\lambda, 2^4)$, $G(\lambda, 2^5)$, $G(\lambda, 2^7)$, $G(\lambda, 2^9)$ compared with $G(\lambda)$

Theorem 2. *The limit $G(\lambda)$ exists, and satisfies the relation*

$$1 - \frac{(1 + G(\lambda)) \ln(1 + G(\lambda)) + (1 - G(\lambda)) \ln(1 - G(\lambda))}{2 \ln 2} = \lambda. \tag{1}$$

While refinements of the asymptotics of the binomial coefficients can be found in [13], basic asymptotics and Stirling's Formula are sufficient to prove the result.

Proof. The larger values for the binomial coefficient occur in the middle, at approximately $n/2$. Our main term will come from $\sqrt{n} \leq k \leq n - \sqrt{n}$, but since the binomial coefficients are also symmetric, we will consider only the second half, or more specifically $\frac{n}{2} \leq k \leq n - \sqrt{n}$, for our initial estimates. For the tails we only need consider k in the range $0 \leq k < \sqrt{n}$, again, because of symmetry.

We first derive estimates for the main term. If we let $k = \frac{n}{2} + m$, we get $0 \le m \le \frac{n}{2} - \sqrt{n}$. Then clearly

$$\sqrt{n/2} \le \sqrt{n/2+m} \le \sqrt{n-\sqrt{n}} \le \sqrt{n},$$

so that $(n/2+m)^{-1/2} = O(n^{-1/2})$. Similarly we have

$$\sqrt[4]{n} \le \sqrt{n/2-m} \le \sqrt{n/2},$$

so that $(n/2-m)^{-1/2} = O(n^{-1/4})$. Also note that $(1+O(n^{-1/2}))(1+O(n^{-1/4})) = (1+O(n^{-1/4}))$.

Using Stirling's Formula three times, we get that

$$\begin{aligned}\binom{n}{\frac{n}{2}+m} &= \frac{n!}{(n/2-m)!(n/2+m)!} \\ &= \frac{n^n\sqrt{n}}{\sqrt{2\pi}(n/2+m)^{n/2+m+1/2}(n/2-m)^{n/2-m+1/2}}\left(1+O(n^{-1/4})\right).\end{aligned}$$

Now pick m such that $\binom{n}{n/2+m} \ge (2^n/(n+1))^\lambda$ but $\binom{n}{n/2+m+1} < (2^n/(n+1))^\lambda$. Since logarithmic functions are one to one, we have that

$$\#\left\{0 \le k \le n : \binom{n}{k} \ge \left(\frac{2^n}{n+1}\right)^\lambda\right\} = \#\left\{0 \le k \le n : \ln\binom{n}{k} \ge \ln\left(\frac{2^n}{n+1}\right)^\lambda\right\},$$

so we can consider the inequality $\ln\binom{n}{n/2+m} \ge \ln(2^n/(n+1))^\lambda$ and simplify. So we have that

$$\begin{aligned}\lambda n\ln 2 - \lambda\ln(n+1) &\le n\ln n + \frac{1}{2}\ln n - \frac{1}{2}\ln(2\pi) - \left(\frac{n}{2}+m\right)\ln\left(\frac{n}{2}+m\right) - \left(\frac{n}{2}-m\right)\ln\left(\frac{n}{2}-m\right) \\ &\quad - \frac{1}{2}\ln\left(\frac{n}{2}+m\right) - \frac{1}{2}\ln\left(\frac{n}{2}-m\right) + \ln(1+O(n^{-1/4})) \\ &\le n\ln n + \left(\frac{n}{2}+m\right)\ln\left(\frac{n}{2}+m\right) - \left(\frac{n}{2}-m\right)\ln\left(\frac{n}{2}-m\right) \\ &\quad + O(\ln n) + O(n^{-1/4}).\end{aligned}$$

This implies that

$$\lambda n\ln 2 + O(\ln n) \le n\ln n + \left(\frac{n}{2}+m\right)\ln\left(\frac{n}{2}+m\right) - \left(\frac{n}{2}-m\right)\ln\left(\frac{n}{2}-m\right) + O(\ln n) + O(n^{-1/4}).$$

After dividing by n, and rearranging, we get that

$$\lambda\ln 2 \le \ln n - \left(\frac{1}{2}+\frac{m}{n}\right)\ln\left(\frac{n}{2}+m\right) - \left(\frac{1}{2}-\frac{m}{n}\right)\ln\left(\frac{n}{2}-m\right) + O\left(\frac{\ln n}{n}\right) + O(n^{-5/4}).$$

Simplifying the right hand side, we see that

$$\begin{aligned}
&\ln n - \left(\frac{1}{2}+\frac{m}{n}\right)\ln\left(\frac{n}{2}+m\right) - \left(\frac{1}{2}-\frac{m}{n}\right)\ln\left(\frac{n}{2}-m\right) + O(n^{-5/4})\\
&= \ln n - \left(\frac{1}{2}+\frac{m}{n}\right)\ln n - \left(\frac{1}{2}+\frac{m}{n}\right)\ln\left(\frac{1}{2}+\frac{m}{n}\right) - \left(\frac{1}{2}-\frac{m}{n}\right)\ln n\\
&\quad - \left(\frac{1}{2}-\frac{m}{n}\right)\ln\left(\frac{1}{2}-\frac{m}{n}\right) + O(n^{-5/4})\\
&= -\left(\frac{1}{2}+\frac{m}{n}\right)\ln\left(\frac{1}{2}+\frac{m}{n}\right) - \left(\frac{1}{2}-\frac{m}{n}\right)\ln\left(\frac{1}{2}-mn\right) + O(n^{-5/4}).
\end{aligned}$$

We then have the inequality

$$\left(\frac{1}{2}+\frac{m}{n}\right)\ln\left(\frac{1}{2}+\frac{m}{n}\right) + \left(\frac{1}{2}-\frac{m}{n}\right)\ln\left(\frac{1}{2}-mn\right) + O(n^{-5/4}) \geq \lambda \ln 2. \tag{2}$$

Because of our choice of m, (2) implies that

$$\left(\frac{1}{2}+\frac{m+1}{n}\right)\ln\left(\frac{1}{2}+\frac{m+1}{n}\right) + \left(\frac{1}{2}-\frac{m+1}{n}\right)\ln\left(\frac{1}{2}-\frac{m+1}{n}\right) + O(n^{-5/4}) \leq -\lambda \ln 2. \tag{3}$$

Simplifying the left side of the inequality (3), we get

$$\left(\frac{1}{2}+\frac{m}{n}\right)\ln\left(\frac{1}{2}+\frac{m}{n}\right) + \left(\frac{1}{2}-\frac{m}{n}\right)\ln\left(\frac{1}{2}-\frac{m}{n}\right) + O\left(\frac{1}{n}\right) \leq -\lambda \ln 2,$$

and this implies that

$$\left(\frac{1}{2}+\frac{m}{n}\right)\ln\left(\frac{1}{2}+\frac{m}{n}\right) + \left(\frac{1}{2}-\frac{m}{n}\right)\ln\left(\frac{1}{2}-\frac{m}{n}\right) = -\lambda \ln 2 + O\left(\frac{1}{n}\right). \tag{4}$$

Now, let $f(t) := (\frac{1}{2}+t)\ln(\frac{1}{2}+t) + (\frac{1}{2}-t)\ln(\frac{1}{2}-t)$. If $t = \frac{m}{n}$, then $0 \leq t \leq \frac{1}{2} - n^{-1/2}$ and $f : [0, \frac{1}{2} - n^{-1/2}] \to \mathbb{R}$. Otherwise, f is well defined on the interval $[0, \frac{1}{2}]$, with $f(0) = -\ln 2$ and $f(\frac{1}{2}) = 0$. Looking at $f'(t) = \ln(\frac{1}{2}+t) - \ln(\frac{1}{2}-t)$, we see that $f'(t) > 0$ for $t > 0$. This means that f is a strictly increasing function on the interval $[0, \frac{1}{2}]$, and that there exists a unique $t_\lambda \in (0, \frac{1}{2} - n^{-1/2})$ such that $f(t_\lambda) = -\lambda \ln 2$. So if $f(m/n) = -\lambda \ln 2 + O(n^{-1/2}) = f(t_\lambda) + O(n^{-1/2})$, then this implies that $m/n = t_\lambda + O(n^{-1/2})$, so that $m = nt_\lambda + O_\lambda(\sqrt{n})$.

Now, we have that

$$\begin{aligned}
G(\lambda, n) &= \frac{2\#\left\{0 \leq m \leq \frac{n}{2} - \sqrt{n} : \binom{n}{n/2+m} \geq \left(\frac{2^n}{n+1}\right)^\lambda\right\}}{n+1} + \frac{2\#\left\{0 \leq k \leq \sqrt{n} : \binom{n}{k} \geq \left(\frac{2^n}{n+1}\right)^\lambda\right\}}{n+1}\\
&= \frac{2t_\lambda n + O(n^{1/2})}{n+1} + \frac{O(\sqrt{n})}{n+1}\\
&= \frac{2t_\lambda n}{n+1} + O(n^{-1/2}).
\end{aligned}$$

Then taking the limit, we have have that

$$\lim_{n\to\infty} G(\lambda, n) = \lim_{n\to\infty} \frac{2t_\lambda n}{n+1} + O(n^{-1/2}) = 2t_\lambda,$$

so that the limit $G(\lambda)$ exists.

All that is left is to show $G(\lambda)$ satisfies (1). We have that

$$\begin{aligned}
1 - \frac{1}{2\ln 2}&(1+G(\lambda))\ln(1+G(\lambda)) + (1-G(\lambda))\ln(1-G(\lambda)) \\
&= 1 - \frac{(1+2t_\lambda)\ln(1+2t_\lambda) + (1-2t_\lambda)\ln(1-2t_\lambda)}{2\ln 2} \\
&= 1 - \frac{2(\frac{1}{2}+t_\lambda)(\ln 2 + \ln(\frac{1}{2}+t_\lambda)) + 2(\frac{1}{2}-t_\lambda)(\ln 2 + \ln(\frac{1}{2}-t_\lambda))}{2\ln 2} \\
&= 1 - \frac{2f(t_\lambda) + 2\ln 2}{2\ln 2} \\
&= 1 - \frac{-2\lambda\ln 2 + 2\ln 2}{2\ln 2} \\
&= 1 - (1-\lambda) \\
&= \lambda.
\end{aligned}$$

□

Remark 3. There are asymptotic estimates for the binomial coefficients that involve the binary entropy function (see [3]). This helps to explain why the function f in the proof of Theorem 2, as well as the relation that $G(\lambda)$ satisfies, is reminiscent of the binary entropy function.

5 The distribution of values for the Stern sequence

Stern [14] investigated the properties of a sequence he constructed in a way similar to Pascal's triangle: taking two values a and b which form the first row, and then the next row is formed by rewriting the previous row and inserting the sum $a+b$ between its summands. This triangular construction is called the diatomic array. The Stern sequence, denoted $s(n)$, also satisfies the recurrences

$$s(2n) = s(n), \quad s(2n+1) = s(n+1) + s(n) \quad \text{with } s(0) = 0 \text{ and } s(1) = 1.$$

Stern also showed that $s(3n)$ is always even, and that these are the only even terms in the sequence. Lucas [8] observed that when written in the diatomic array, the largest value of a row is a Fibonacci number, and he proved that the maximum value occurs for the integers n closest to $4/3 \cdot 2^r$ and $5/3 \cdot 2^r$. Stern also showed that the sum of the values in a row of the diatomic array is a power of 3, so that

$$\sum_{n=2^r}^{2^{r+1}-1} s(n) = 3^r. \tag{5}$$

This implies that the average value of the r-th row is roughly $(3/2)^r$. This also implies that, for fixed N, we have that

$$\frac{1}{N}\sum_{n=0}^{N} s(n) \asymp N^{\beta-1}, \tag{6}$$

where $\beta = \log_2 3$. For more information on the Stern sequence, D.H. Lehmer [7] summarizes Stern's results. Reznick [9] also gives a summary of results on the Stern sequence, or see entry A002487 [11]. For some new results on the Stern sequence, see [6].

For the distribution of values of the Stern sequence, we then count the number of terms in a row of the diatomic array that are larger than the average value. We define the counting function

$$H(\lambda, N) := \frac{\#\left\{2^N \le n < 2^{N+1} : s(n) \ge \lambda\left(\frac{3^N}{2^N}\right)\right\}}{2^N}.$$

The data in Figure 3 suggests that $H(\lambda, N)$ converges to a smooth function, but it is not clear if it actually does. Overall, the general shape of the graphs looks like e^{-ax^2-bx}, but the

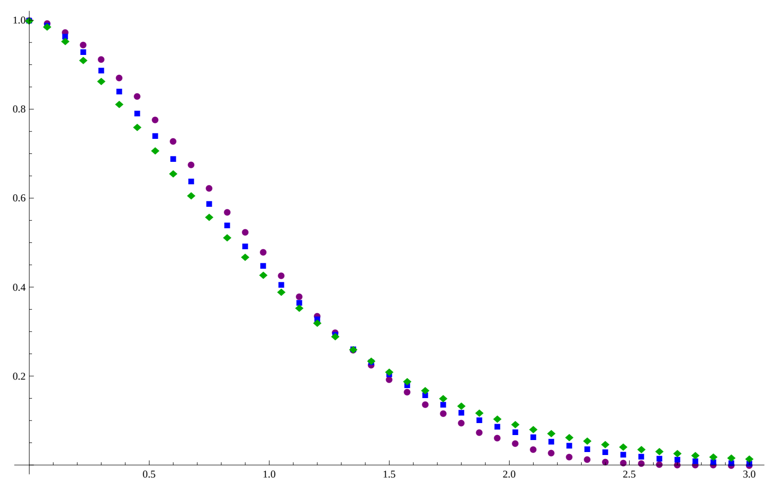


Figure 3: $H(\lambda, 2^{12})$ (circles), $H(\lambda, 2^{17})$ (squares), and $H(\lambda, 2^{22})$ (diamonds)

data does not stay close to the curve. We leave understanding the nature of this limiting distribution as an open problem.

As it turns out, the binomial coefficients are not similar enough to gain any information for the Stern sequence. The next step after this problem would be to understand the distribution of gaps of the Stern sequence.

6 Acknowledgments

The author thanks Bruce Reznick and Alexandru Zaharescu for their helpful insight. The author would also like to thank the referee for his helpful feedback.

References

[1] E. Alkan, M. Xiong, and A. Zaharescu, Pair correlation of sums of rationals with bounded height. *J. Reine Angew. Math.* **641** (2010), 21–67.

[2] P. Gallagher, On the distribution of primes in short intervals, *Mathematatika* **23** (1976), 4–9.

[3] R. Graham, M. Grötschel, and L. Lovász, *Handbook of Combinatorics*, Vol. 2, MIT Press, 1995.

[4] C. Hooley, An asymptotic formula in the theory of numbers, *Proc. London Math. Soc.* **7** (1957), 396–413.

[5] C. Hooley, On the intervals between consecutive terms of sequences, *Proc. Symp. Pure Math.* **24** (1973), 129–140.

[6] J. Lansing, On the Stern sequence and a related sequence, Ph. D. dissertation, in preparation.

[7] D. H. Lehmer, On Stern's diatomic series, *Amer. Math. Monthly* **36** (1929), 59–67.

[8] E. Lucas, Sur les suites de Farey. *Bull. Soc. Math. France* **6** (1878), 118–119.

[9] B. Reznick, Regularity properties of the Stern enumeration of the rationals, *J. Integer. Seq.* **11** (2008), Article 08.4.1.

[10] Z. Rudnick and A. Zaharescu, The distribution of spacings between fractional parts of lacunary sequences, *Forum Math.* **14** (2002), 691–712.

[11] N. J. A. Sloane, The Online Encyclopedia of Integer Sequences. Published electronically at http://oeis.org.

[12] V. Sós, On the distribution mod 1 of the sequence $n\alpha$, *Ann. Univ. Sci. Budapest, Eötvös Sect. Math.* **1** (1958), 127–134.

[13] P. Stănică, Good lower and upper bounds on binomial coefficients, *J. Inequal. Pure Appl. Math.* **2** (2001), Article 30.

[14] M. A. Stern, Über eine zahlentheoretische Funktion, *J. Reine Angew. Math.* **55** (1858), 193–220.

[15] S. Swierczkowski, On successive settings of an arc on the circumference of a circle, *Fundam. Math.* **46** (1958), 187–189.

[16] H. Weyl, Über die Gleichverteilung von Zalhen mod. Eins, *Math. Annal.* **77** (1916), 313–352.

2010 *Mathematics Subject Classification*: Primary 11B65; Secondary 11B05.
Keywords: binomial coefficients, Stern sequence, distribution.

(Concerned with sequences A002487 and A007318.)

Received September 18 2012; revised version received January 25 2013; February 22 2013. Published in *Journal of Integer Sequences*, March 2 2013.

Return to Journal of Integer Sequences home page.