Entropy of Continued Fractions (Gauss-Kuzmin Entropy)

Linas Vepstas <linasvepstas@gmail.com>

June 2008
(trivial updates April 2012)

Abstract

This short note provides a numerical exploration of the entropy of the Gauss-Kuzmin distribution, confirming that it seems to have a value of 3.432527514776... bits. Some information-theoretic questions regarding the distribution of rationals are explored. In particular, one may define a "de facto" entropy for fractions with a small denominator; it is not clear that this de-facto entropy approaches the above in the limit of large denominators.

1 Introduction

Let x be a real number, $0 \leq x \leq 1$. Let

$$x = [a_1, a_2, a_3, \cdots] = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \cdots}}}$$

be the continued fraction expansion of x. Given the uniform distribution of the reals on the unit interval, the Gauss-Kuzmin distribution gives the probability $\Pr(a_n = k)$ of an integer k appearing in any given place a_n of the expansion. This probability distribution has been famously studied by Kuz'min, Levy, Khinchin and Wirsing; it is given by[4]

$$\Pr(a_n = k) = p_k = -\log_2 \left[1 - \frac{1}{(k+1)^2} \right]$$

The continued fraction expansion can be viewed as a discrete random variable, which may be sampled; the n'th sampling giving the value a_n. Given a discrete random variable with N possible discrete states, the (information-theoretic) entropy[1] is defined as

$$H = -\sum_{k=1}^{N} p_k \log_2 p_k$$

where p_k is the probability of the k'th state occurring in a measurement of the random variable. The entropy is measured in bits; and so $\log_2 p = \log p / \log 2$ is the base-2 logarithm. For continued fractions, one has an infinite number of possible states, and so $N = \infty$ and so

$$H = -\sum_{k=1}^{\infty} p_k \log_2 p_k$$

Here, the probability p_k is given by the Gauss-Kuzmin distribution. This entropy shall be termed the "*Gauss-Kuzmin entropy*", as it is uniquely fixed by the Gauss-Kuzmin distribution. It appears to have been first defined and calculated by N. M. Blachman in 1984[2]. The numerical value of H may be obtained by computation; it is

$$H = 3.432527514776 \cdots \text{ bits}$$

or alternately, in terms of the natural logarithm,

$$H \log_e 2 = 2.379246769061 \cdots \text{ nats}$$

These numbers are not obviously related to any previously known constants, according to Plouffe's Inverter[5].

The above values were computed with the GNU MP multiple precision library[3], and should be accurate to approximately the last two digits. They were obtained by means of brute-force summation, together with quadratic extrapolation, up to values of $k = 4.09 \times 10^{10}$. The quadratic extrapolation may be performed as follows: let

$$t = 1 - \sum_{k=1}^{N} p_k$$

and

$$H(t) = -\sum_{k=1}^{N} p_k \log p_k$$

It can be readily seen that $\lim_{N\to\infty} t = 0$. An explicit form for t as a function of N is given in the next section. It is also straightforward to observe that $H(t)$ is very nearly a linear function of t. Thus, one can readily estimate the value of $H = \lim_{t\to 0} H(t)$ by means of a quadratic extrapolation in t to the limit $t = 0$. Such extrapolation offers several additional decimal digits of precision over the raw value of the sum, terminated at a finite N.

There does not appear to be any simple or straightforward way to rewrite the sums to allow high-precision (more than 10 decimal places) calculation.

2 Analytic results

It is strightforward to sum the cumulative distribution function. The cumulative distribution is the partial sum

$$C(N) = \sum_{k=1}^{N} p_k$$

Note that

$$1 - \frac{1}{(k+1)^2} = \frac{k(k+2)}{(k+1)^2}$$

and so

$$\begin{aligned} C(N) =& -\sum_{k=1}^{N} \log_2 \left[\frac{k(k+2)}{(k+1)^2} \right] \\ =& -\log_2 \left[\prod_{k=1}^{N} \frac{k(k+2)}{(k+1)^2} \right] \\ =& 1 - \log_2 \left[\frac{N+2}{N+1} \right] \end{aligned}$$

The extrapolation parameter t is then

$$t = 1 - C(N) = \log_2 \left[\frac{N+2}{N+1} \right] = \log_2 \left[1 + \frac{1}{N+1} \right]$$

and so, to first order in N,

$$t = \frac{1}{(N+1)\log_e 2} + \mathcal{O}\left(\frac{1}{N^2} \right)$$

3 Typical Sequences

Given any particular value of $x = [a_1, a_2, a_3, \cdots]$, one may ask just how representative the sequence is of a "typical" sequence, where a "typical" sequences is one which has a distribution of a_n close to that of the Gauss-Kuzmin distribution. It is of some interest to see whether the rational numbers are "typical" continued fractions, or not. The question is important, as many numerical explorations of continued fractions must, by necessity, work with either finite-length, or periodic continued fraction expansions. Also of some curiosity is whether well-known transcendental constants, such as π of the Euler-Mascheroni constant γ are "typical" or not.

The standard techniques of discussing typical sequences[1] are not directly applicable, as the Gauss-Kuzmin distribution has infinite mean and mean-square variance. Let p/q be a rational, with a continued-fraction representation of length M. Let m_k be the number of times that the integer k occurs in the continued fraction expansion of p/q. Normalizing, one has a frequency of occurrence:

$$f_k = \frac{m_k}{M}$$

and clearly, $\sum_k f_k = 1$. One may define a relative entropy as

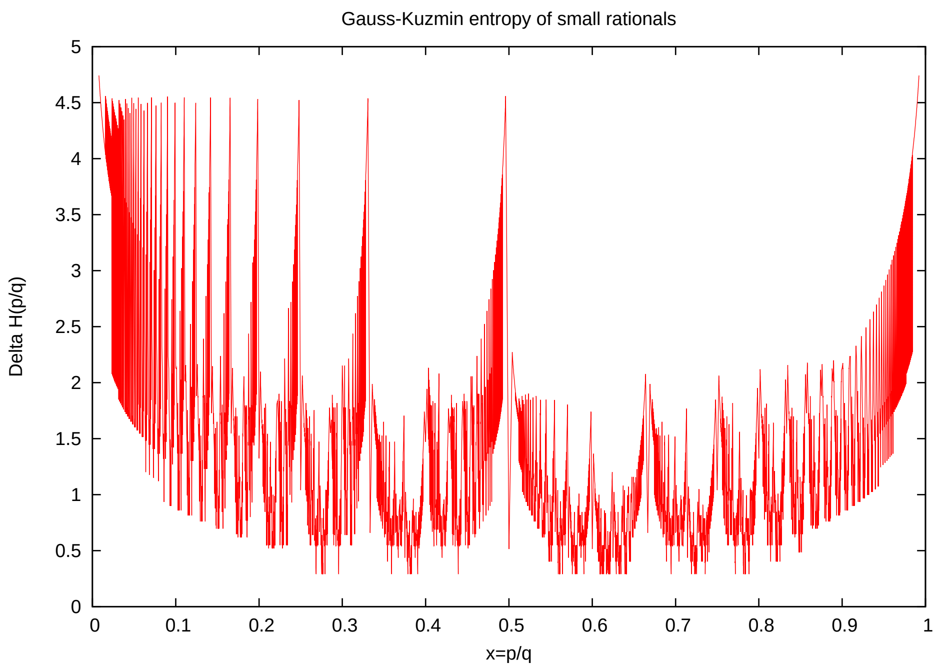
$$\Delta H\left(\frac{p}{q}\right) = -\sum_{k=1}^{\infty} (f_k - p_k) \log_2 p_k \tag{1}$$

This relative entropy is shown in figure 3.

The relative entropy has an obvious self-similarity, with two generators. By examining the graph, one may guess that one of the generators might be:

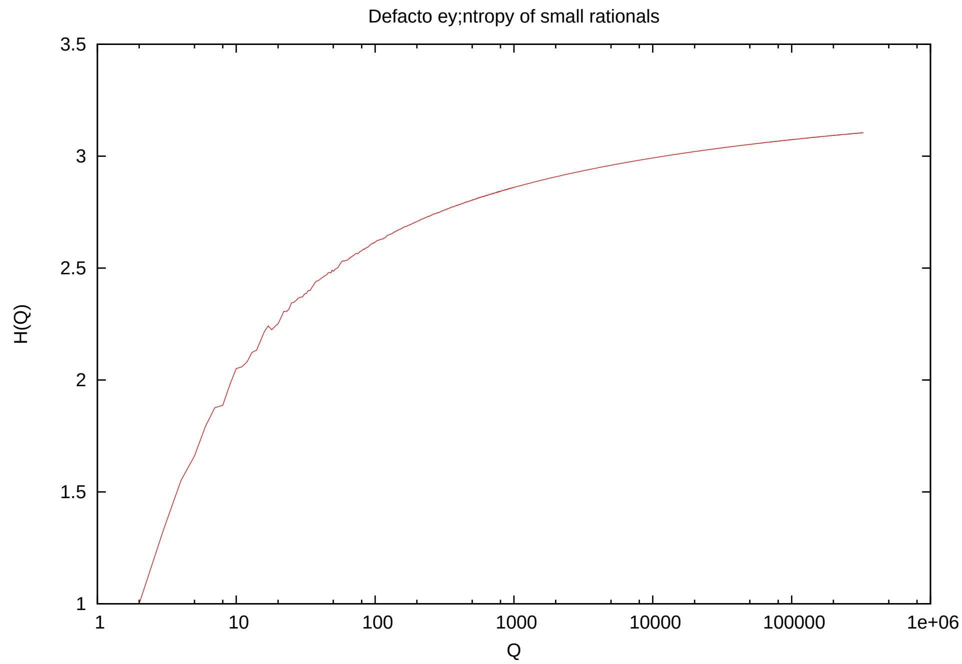
$$\Delta H\left(\frac{p}{q}\right) \approx \Delta H\left(\frac{p}{p+q}\right)$$

Figure 1: Relative Entropy



This graph shows the relative entropy, given by eqn. 1, for all of the rationals p/q for $q \leq 128$. Notice an obvious self-similarity.

Figure 2: De facto entropy for small rationals



This figure shows the de facto entropy, given by eqn. 2. The graph extends to all rationals with denominators $q \leq Q = 4 \times 10^5$. As can be seen, the convergence is very slow. The limiting value might possibly be H=3.43..., but this is hardly clear simply by gazing at the graph.

The form of the other generator is not clear.

Another standard interpretation of entropy is that, given a length ℓ, there are $2^{\ell H}$ "typical" sequences of length ℓ; other sequences are possible, but unlikely. For small rationals, this interpretation can be reversed. Consider, instead, the set of all (irreducible) rationals p/q up to a maximum denominator $q \leq Q$. There are $N(Q)$ such rationals, which, expressed as continued fractions, have an average number of terms $\ell(Q)$. One then defines a *de facto* entropy as

$$H(Q) = \frac{\log_2 N(Q)}{\ell(Q)} \tag{2}$$

For small rationals, the de facto entropy is considerably smaller than the Gauss-Kuzmin entropy. This is illustrated in figure 3.

References

[1] Robert B. Ash. *Information Theory*. Dover Publications, 1965.

[2] N. M. Blachman. The continued fraction as an information source (corresp.). *IEEE Transactions on Information Theory*, 30(4):671– 674, July 1984.

[3] Free Software Foundation. *Gnu Multiple Precision arithmetic library*. http://www.swox.com/gmp/.

[4] A. Ya. Khinchin. *Continued Fractions*. Dover Publications, (reproduction of 1964 english translation of the original 1935 russian edition) edition, 1997.

[5] Simon Plouffe. Plouffe's inverter. http://pi.lacim.uqam.ca/, 199?