ON THE PROBABILITY OF REACHING A GIVEN HEAD TO TAIL RATIO

TAMÁS LENGYEL*
Department of Mathematics
Occidental College
Los Angeles, CA 90041, USA

The head to tail ratio converges to 1 with probability one when a fair coin is flipped. We show that the limit probability of reaching the ratio \( \frac{q}{q+m} \) is \( \frac{2}{2+m} \), as \( q \to \infty \) and \( q \) and \( m \) are co-primes.

1. INTRODUCTION

We flip a balanced coin. Let \( X \) and \( Y \) denote the number of heads and tails, respectively. It is well known from the theory of random walks that the probability of ever visiting the line \( Y = X - m \) is 1 for any integer \( m \). For instance, if the line is reached when \( Y = n \) then \( X = n + m \) and the probability of this happening is \( p_n = P(Y = X - m) = \left( \frac{2n+m}{n} \right) / 2^{2n+m} \). It follows that \( 1 - 1/(1 + \sum_{n=1}^{\infty} p_n) \) is the probability that the line \( Y = X - m \) is ever reached [3]. By binomial identities (cf. identities (5.72) and (5.78) in [4], p. 203), we obtain for \( |x| < 1/2 \) that

\[
\sum_{n=0}^{\infty} \binom{2n+m}{n} x^{2n+m} = \left( \frac{1 - \sqrt{1 - 4x^2}}{2x} \right)^m / \sqrt{1 - 4x^2}.
\]

If \( x = 1/2 \), then the sum is divergent, therefore the line will be reached with probability 1. We might as well be interested in calculating the probability of reaching a given ratio instead of a difference. By the theory of recurrent events [3], the probability of reaching the ratio one (or equivalently, a difference of \( m = 0 \)) is 1, though the expected number of flips needed is infinite. In this paper we discuss the extreme value of the probability of reaching a given head to tail ratio which is different from 1.

We note that the case of an unbalanced coin has been discussed in the literature ([3], Exercise 4, p. 339). In general, let \( h \) and \( t \) denote the probability of getting a head and a tail, respectively, where \( h + t = 1 \). The event that the accumulated number of heads equals \( \lambda \) times the accumulated number of tails is \textit{persistent}, i.e., it has probability one, if and only if the head/tail probability ratio, \( h/t \), is equal to \( \lambda \). Other ratios are usually not discussed.

* Present address: T. Lengyel, Dept. Math., Occidental College, 1600 Campus Road, Los Angeles, CA 90041, USA
In this paper we consider the head to tail ratio $X/Y$ for a balanced coin. We like to know how large the probability of ever reaching a given head to tail ratio, $q/p$, is where $p$ and $q$ are co-primes, i.e., the ratio $q/p$ is given in lowest terms. We assume that $q < p$, since for a balanced coin, the ratios $q/p$ and $p/q$ can be reached with the same probability. We set $r = p + q$.

Numerical evidence suggests that the second largest probability is around $2/3$ and it does not exceed $2/3$. Hence there is a gap between $1$ and the second largest probability of reaching a given ratio $q/p$. We prove that for every positive $\epsilon$ and integer $m$, this probability is less than $\frac{2}{2+m} + \epsilon \leq \frac{2}{3} + \epsilon$ for ratios of form $\frac{q}{p} = \frac{q}{q+m}$ with large values of $q$, where $q$ and $m$ are co-primes. Actually, the limit probability is $\frac{2}{2+m}$. Let $u(p, q) = \sum_{n=1}^{\infty} \left(\frac{r_n}{q_n}\right)^{2-rn}$ be. The probability of ever reaching the ratio $q/p$ is $w(p, q) = 1 - \frac{1}{1+u(p, q)}$. The infinite series $\sum_{n=1}^{\infty} \left(\frac{r_n}{q_n}\right)^{2-rn}$ diverges if $p = q$, and it converges otherwise.

Note, that instead of the head to tail ratio we might consider the head to total ratio. The head to tail ratio $1$ corresponds to the head to total ratio $1/2$.

2. The result

Let $\gcd(q, m)$ denote the greatest common divisor of the positive integers $q$ and $m$. We prove

**Theorem.** $\lim_{q \to \infty} w(q+1, q) = 2/3$, and in general, for every fixed $m \geq 1$,

$$\lim_{q \to \infty, \gcd(q, m) = 1} w(q+m, q) = \frac{2}{2+m}.$$  

Theorem 1 shows the somewhat surprising fact that $u(p, q)$ is not a continuous function of the ratio $q/p$. To illustrate this, we compare two ratios that are close. Say, the first pair is $(q+1, q)$, i.e., $m = 1$, while the other is $(q+2, q)$, with $m = 2$. By selecting a sufficiently large odd $q$, the two ratios can be arbitrarily close, though the probabilities of reaching them stay apart since $w(q+1, q) \approx 2/3$, while $w(q+2, q) \approx 1/2$.

In this paper we use the following notations and assumptions.

Let $m$ be a fixed positive integer. Assume that $p = q + m$, i.e., $r = 2q + m$, such that $\frac{m^2}{2p} < 1$.

From now on, $c_1(p, m, n), c_2(p, m, n)$, and $c_3(p, m, n)$ denote bounded functions of the variables $p, m, n$. Similarly, $c_4(p, m, N), c_5(p, m, N), c_6(p, m, N), c_7(p, m, N), c_8(p, m, N), c_9(p, m, N)$, and $c_{10}(p, m)$ are bounded functions of the variables indicated in parentheses.
Lemma 1 utilizes the Stirling formula in order to asymptotically evaluate \( g(p, q, n) = \binom{r_n}{qn} 2^{-n} \).

It will be applied to the sum \( u(p, q) = \sum_{n=1}^{\infty} g(p, q, n) \).

**Lemma 1.** In addition to the previous conditions on \( p, q, \) and \( m \), let \( q > m \) be. Then

\[
g(p, q, n) = \left( \frac{1}{2} \frac{p - q}{p - q} \right)^n \sqrt{2 \pi \frac{p}{2(p - q)}} \sqrt{\frac{1}{n!} \left( 1 + c_1(p, q, n) \frac{1}{pn} \right)}.
\]

We omit the proof of Lemma 1 but note that it can be proved similarly to the asymptotical formula

\[
\binom{(a + b)n}{an} \sim \frac{(a + b)^{n(a + b) + 1/2}}{a^{n+1/2} b^{n+1/2}} \frac{1}{\sqrt{2\pi n}}
\]

for positive integers \( a \) and \( b \) (cf. [1], Exercise 2, p. 292).

By introducing the notation \( \frac{1}{2} \frac{p}{p} = \frac{1}{2} - \epsilon \), we get \( \epsilon = \frac{m}{2p} \) and \( 2p\epsilon^2 = \frac{m^2}{2p} < 1 \). Lemma 1 yields

\[
g(p, q, n) = \sqrt{\frac{2}{\pi} \frac{1}{\sqrt{p}}} \left( 1 - 2\epsilon^2 + c_2(p, m, n)\epsilon^4 \right)^{pn} \sqrt{n!} \left( 1 + c_1(p, m, n) \frac{1}{pn} \right) \left( 1 + 2\epsilon^2 + c_3(p, m, n)\epsilon^4 \right).
\]

We set \( S_N(p, q) = \sum_{n=1}^{N} \binom{r_n}{qn} \frac{1}{2^n} \). The Theorem will be proven in three steps. We shall need Lemmas 2 and 3 to approximate the sum \( u(p, q) \). We select a large \( N \) in identity (2) to get a close approximation to \( u(p, q) = \sum_{n \geq 1} g(p, q, n) \) by the finite sum \( S_N(p, q) \). Next, we need a sufficiently large \( p \) in equation (3) to approximate \( S_N(p, q) \) by another sum which is easier to calculate. Formula (4) suggests that we choose large \( p \) and \( N \) in order to have a meaningful approximation when using Euler’s formula. The proof follows as we combine identities (2) and (5).

By Lemma 1 we obtain

**Lemma 2.** Let \( p = q + m \) and \( r = 2q + m \) be where \( m > 0 \) is a fixed integer such that \( \frac{m^2}{2p} < 1 \). Then

\[
u(p, q) = \sum_{n=1}^{\infty} \binom{r_n}{qn} \frac{1}{2^n} = S_N(p, q) + c_4(p, m, N) \left( \frac{p}{N} \right)^{1/2} \left( 1 - \frac{m^2}{2p} \right)^N,
\]

and

\[
S_N(p, q) = \sqrt{\frac{2}{\pi} \frac{1}{\sqrt{p}}} \sum_{n=1}^{N} \frac{1}{\sqrt{n}} \frac{1 - \left( \frac{m^2}{2p} \right)^n}{\sqrt{n}} + c_5(p, m, N) \frac{\ln N}{p}.
\]
Proof of Lemma 2.

We get an upper bound on \( \sum_{n=N+1}^{\infty} g(p,q,n) \) by using the identity \( \sum_{i=N}^{\infty} z^i = \frac{z^N}{1-z} \) with any \( z \) exceeding \( \left( 1 - \frac{m^2}{2p} \right) \). It follows from identity (1) that \( u(p,q) - S_N(p,q) = \sum_{n=N+1}^{\infty} g(p,q,n) = c_5(p,m,N) \frac{1}{(pN)^{1/2}} \left( 1 - \frac{m^2}{2p} \right)^{2p/m} \). Similarly, identity (1) gives an upper bound on the error term’s contribution to \( \sum_{n=1}^{N} g(p,q,n) \). The error is of magnitude \( \ln N/p \).

We shall need

**Lemma 3.** Under the conditions of Lemma 2,

\[
\sum_{n=1}^{N} \frac{1}{\sqrt{p}} \frac{(1 - \frac{m^2}{2p})^n}{\sqrt{n}} = \sqrt{2\pi} + c_z(p,m,N) \left( \frac{1}{\sqrt{p}} + \sqrt{\frac{p}{N}} \right).
\]

Therefore,

\[
S_N(p,q) = \sum_{n=1}^{N} \left( \frac{r_n}{q_n} \right) \frac{1}{2^{rn}} = \frac{2}{m} + c_5(p,m,N) \frac{\ln N}{p} + c_z(p,m,N) \left( \frac{1}{\sqrt{p}} + \sqrt{\frac{p}{N}} \right) \sqrt{\frac{2}{\pi}}.
\]

Remark. Lemma 3 shows that \( S_N(p,q) \) can get arbitrarily close to \( \frac{2}{m} \), for large \( p \) and \( N \). In fact, we select a sequence \( N = N(p) \) so that \( p/N(p) \to 0 \) and \( \ln N(p)/p \to 0 \), as \( p \to \infty \). By Lemma 2, it follows that \( \sum_{n=1}^{\infty} \left( \frac{r_n}{q_n} \right) \frac{1}{2^{rn}} \) converges to \( \frac{2}{m} \), as \( q \to \infty \) and \( \gcd(q,m) = 1 \).

Proof of Lemma 3.

We shall need an application of Euler’s summation formula ([5], p. 108 or [2]) to derive identity (4). Let \( f(k) = \frac{1}{\sqrt{p}} \left( 1 - \frac{m^2}{2p} \right)^k \) be. Euler’s method yields formula (6) for the difference between \( \int_1^n f(x)dx \) and \( \sum_{1 \leq k < n} f(k) \) if \( f(x) \) is differentiable, i.e.,

\[
\sum_{1 \leq k < n} f(k) = \int_1^n f(y)dy - \frac{1}{2} \left( f(n) - f(1) \right) + \int_1^n B_1(\{y\}) f'(y)dy,
\]

where \( B_1(y) = y - 1/2 \) and \( \{y\} = y - [y] \).

We apply this formula to function \( f(k) \). Clearly, \( f(n) \) converges to 0 at a rate faster than \( \frac{1}{\sqrt{n}} \) as \( n \to \infty \), and \( f(1) < \frac{1}{\sqrt{p}} \). We set \( \frac{1}{s} = (1 - \frac{m^2}{2p}) \). Here \( s > 1 \), since \( p \) is large enough to make \( \frac{m^2}{2p} < 1 \). We note that \( f'(y) = \frac{1}{\sqrt{p}} \frac{(1 - \frac{m^2}{2p})^y}{\sqrt{y}} (- \ln s - \frac{1}{2y}) \). Observe that \( \ln s \sim \frac{m^2}{2p} \) as \( p \to \infty \).
First we asymptotically evaluate the first term on the right side in formula (6). A well-known integral equation for the gamma function \([5]\) says that for all \(\alpha > -1\)

\[
\int_0^\infty x e^{-x^\alpha} dv = \frac{1}{x^\alpha} \int_0^\infty e^{-t^\alpha} dt = \frac{1}{x^\alpha}, \ (\alpha + 1).
\]  

(7)

By setting \(x = \ln s\) and \(\alpha = -1/2\), it follows that

\[
\int_0^\infty f(y) dy = \int_0^\infty \frac{1}{\sqrt{p}} e^{-y \ln s} \frac{dy}{\sqrt{y}} = \frac{1}{\sqrt{p}} (\ln s)^{-1/2} \sqrt{\pi}.
\]  

(8)

Therefore, if \(p\) is sufficiently large then \(\ln s \sim \frac{m^2}{2p}\) and the above integral is asymptotically equal to \(\frac{\sqrt{\pi}}{m}\). Hence the term \(\int_1^n f(y) dy\) contributes \(\frac{\sqrt{\pi}}{m} + c_8(p, m, n) \frac{1}{\sqrt{p}} + c_9(p, m, n) \frac{1}{\sqrt{n}}\) to \(\sum_{1 \leq k < n} f(k)\) in formula (6).

For the last term of identity (6) we obtain

\[
\left| \int_1^n B_1(\{y\}) f'(y) dy \right| \leq \int_1^n |f'(y)| dy \leq \int_1^n \frac{1}{\sqrt{p}} \left(1 - \frac{m^2}{2p}\right)^y \frac{dy}{\sqrt{y}} \left(2 \frac{m^2}{2p} + \frac{1}{2y}\right)
\]

\[
\leq 2 \int_0^\infty \frac{1}{\sqrt{p}} \left(1 - \frac{m^2}{2p}\right)^y \frac{m^2}{2p} dy + \int_1^\infty \frac{1}{\sqrt{p}} \left(1 - \frac{m^2}{2p}\right)^y \frac{1}{2y} dy.
\]  

(9)

Similarly to equation (8), identity (7) yields

\[
\int_0^\infty \frac{1}{\sqrt{p}} \left(1 - \frac{m^2}{2p}\right)^y \frac{m^2}{2p} dy = c_{10}(p, m) \frac{m^2}{p}.
\]  

(10)

For the second term, we get

\[
\int_1^\infty \frac{1}{\sqrt{p}} \left(1 - \frac{m^2}{2p}\right)^y \frac{dy}{y^{3/2}} \leq \frac{2}{\sqrt{p}} \int_1^\infty \frac{dy}{y^{3/2}} = \frac{2}{\sqrt{p}}.
\]

These inequalities provide us with an upper bound on \(\int_1^n B_1(\{y\}) f'(y) dy\).

From here it follows that for fixed \(m\), \(\sum_{1 \leq k < n} f(k) = \frac{\sqrt{\pi}}{m} + c_7(p, m, n) \left(\frac{1}{\sqrt{p}} + \frac{1}{\sqrt{n}}\right)\). In fact, we get \(\lim_{q \to \infty} u(q + m, q) = \frac{2}{m}\) and for the probability that the ratio \(q/p\) will ever be reached, we conclude that \(\lim_{q \to \infty} 1 - \frac{1}{1 + u(q + m, q)} = 1 - \frac{1}{1 + 2/m} = \frac{2}{2 + m} \leq \frac{2}{3}\), where the limit is taken over the set of \((q, m)\)-pairs that are co-primes.

**Acknowledgment**

The problem discussed in this paper has been brought to the author’s attention by Professor J. Shavlik of the University of Wisconsin-Madison.
REFERENCES


