Expected Return of Coin Tossing with Unknown Bias

Minkyu Son *

August 29, 2022

Introduction

Coin tossing is one of the easiest way to introduce randomness, and is one of the simplest real-life action to model probabilistically. Properties of coin tossing has thus been studied extensively and used to introduce concepts of probability to students at all levels. Several games was invented using coin tossing, both in practical and theoretical setting, to serve as an interesting game in a bar or as a problem or a model of mathematical interest.

In this paper, we introduce a novel game involving a coin with unknown bias. Then, we will explore several strategies for the game, ultimately culminating in a tight bound for the maximum return of the game under optimal play. We will also explore, but not prove, a simple rule for approximating the optimal play that requires very little computation.

Author would like to acknowledge Gregory Pylypovych[1] for his contribution to this paper, in conceiving this problem in the first place and discussing several intuitions for the game. In particular, the analysis of fixed length heuristic and tolerance distribution was largely kick-started through a discussion with Gregory.

Problem Statement

Alice has a special coin that turns up heads with probability p and tails with probability $1-p$. Alice, without knowing the value of p, offers Bob to a game with the following rule.

1. Bob tosses the coin once at a time. He may stop at any point, but can only toss the coin at most n times

2. Bob's score is $H-T$ where H is the number of coin tosses that turned up head and T is that of tail.

What strategy can Bob employ to maximize his expected value? And what is this maximum?

Updating Probability Distribution

Initially, We have no particular bias for this belief on p. Therefore, it is reasonable to model p as a random variable with uniform distribution $\mathrm{U}[0,1]$.[2] Furthermore, each subsequent tosses will update our knowledge about the distribution of p. In particular, with $aHbT$ denoting the event of observing a Heads and b Tails in the first $a+b$ tosses,

$$\mathbb{P}(aHbT|p=p_0) = \binom{a+b}{a} p_0^a (1-p_0)^b$$

Thus, by Bayes's Theorem,

$$\mathbb{P}(p \le p_0 | aHbT) = \frac{\int_0^{p_0} p^a(1-p)^b \mathrm{d}p}{\int_0^1 p^a(1-p)^b \mathrm{d}p}$$

and $(p|aHbT)$ is distributed as $\mathrm{Beta}(a+1, b+1)$

*Undergraduate, University of Oxford
email address: minkyu.son@st-annes.ox.ac.uk

[1]Undergraduate, MIT

[2]In real life, we might not expect the coin to be biased too heavily to one side. In this paper however, we will assume uniformity.

Now, since

$$\mathbb{E}(p|aHbT) = \frac{a+1}{a+b+2}$$

for beta distributions, We may use this as the local probability of heads for each turn. For shorthand, denote $\alpha_{a,b} = \frac{a+1}{a+b+2}$ and $\beta_{a,b} = \frac{b+1}{a+b+2}$

Maximizing EV

Knowing the probability of heads in each turn, we may set up a following recursion on expected values:

Define $\mathbb{E}_{\max}(m, a, b)$ as the maximum expected value of a coin toss if $aHbT$ was observed with m tosses remaining. Then, $\mathbb{E}_{\max}(0, a, b) = a - b$. Also, if we decide to flip another coin with m flips remaining, $\mathbb{E}_{\max}(m, a, b)$ is

$$\alpha_{a,b}\mathbb{E}_{\max}(m-1, a+1, b) + \beta_{a,b}\mathbb{E}_{\max}(m-1, a, b+1)$$

Otherwise, if we decide to stop, our EV is simply $a - b$. Thus, for maximum EV strategy, we choose the maximum of the two options and compute the EV accordingly.

Lastly, note the following facts:

1. Having more heads than tails mean $\alpha_{a,b} > 0.5$. Thus, we expect to gain value by continuing. Therefore, if $a > b$, we should continue with the game
2. If $a = b$, our expectations are unbiased. However, our ability to choose the stopping point gives us an edge in the future. So, we should still continue.
3. If recursion tells us to continue at a_0HbT, we should continue at $aHbT$ for any $a > a_0$, since our prospect is strictly better.

Dynamic programming allows this recursion to be carried out in $O(n^2)$ time to compute the EV and the decision boundary (i.e. when to stop the game). Furthermore, as total number of tosses, $n = m+a+b$ stays fixed, we store exactly $\binom{n+2}{2}$ EVs during recursion for total of $O(n^2)$ space complexity. However, by incrementally deleting unnecessary EVs and noting (3) above for decision boundary, we can reduce our space complexity to $O(n)$.

Upper Bound on Maximum EV

We now turn to estimating the maximum EV. For a game allowing a total of n tosses, if our tosses were such that we toss all n coins, we will end up at $kH(n-k)T$ for some k. Take any legal ordering [3] of k heads and $(n-k)$ tails. The probability of occurrence of this specific ordering of tosses is computed by multiplying all the local probabilities. At each toss, we multiply by $\alpha_{a,b}$ if we get heads, incrementing a by 1 or by $\beta_{a,b}$ if we get tails, incrementing b by 1. Thus, in the total of n tosses we perform, our denominator will be the product of $2, 3, \ldots, n+1$ while our numerator will be of $1, 2, \ldots, k$ and $1, 2, \ldots, n-k$ each coming from $\alpha_{a,b}$ and $\beta_{a,b}$ respectively. Thus, the probability of a specific ordering of toss is $\frac{(k+1)!(n-k+1)!}{(n+1)!}$ Notably, this probability is independent of the order of heads and tails, only depending on the final head/tail count. Thus, the probability of ending the game with $kH(n-k)T$ is $\frac{k!(n-k)!}{(n+1)!}$ multiplied by the number of legal ordering of $kH(n-k)T$. Clearly, the number of legal ordering is bounded above by $\binom{n}{k}$, and thus probability of reaching $kH(n-k)T$ is bounded above by $\frac{1}{n+1}$.

Now, observe that our maximum EV is the linear sum of all the scores of events where we stop the game, with the coefficients determined by the probability of the event. Also, these termination points are either where we use up all n tosses, or where we deem it beneficial to stop midway. Thus,

$$\mathbb{E}_{\max}(n, 0, 0) = \sum_{\text{midway}} q_{a,b}(a-b) + \sum_{n\text{tosses}} q_{k,n-k}(2k-n)$$

where $q_{a,b}$ denote the probability of ending up at $aHbT$. Now, by point (1) and (2) in previous section, we only stop midway when $a < b$. Thus,

$$\begin{aligned}\mathbb{E}_{\max}(n, 0, 0) &\leq \sum_{n\text{tosses}} q_{k,n-k}(2k-n) \\ &\leq \sum_{k=\lceil \frac{n}{2} \rceil}^{n} q_{k,n-k}(2k-n)\end{aligned}$$

[3] legal meaning recursion would not have stopped midway

where the last inequality follows since we are removing only the negative valued events. Finally, since $q_{k,n-k} \leq \frac{1}{n+1}$ by previous argument, we conclude that

$$\mathbb{E}_{\max}(n,0,0) \leq \sum_{k=\lceil \frac{n}{2} \rceil}^{n} \frac{2k-n}{n+1} \leq \frac{n+1}{4}$$

Hence, our maximum EV is bounded above by $\frac{n+1}{4}$.

Fixed Length Heuristic

We finish by finding a suitable lower bound on the maximum EV. Since maximum EV by definition cannot be over come using a different strategy, any strategy relying on a heuristic will generate a lower bound on the maximum EV. In particular, we will consider the Fixed Length Heuristic where we choose a fixed number $m < n$ and proceed with the game in the following way:

1. We toss the first m coins without stopping
2. If half or more coins turns up heads, we continue with rest of the $n-m$ tosses. Otherwise, we stop

The intuition behind the heuristic is that the first m tosses provides us with the estimate of p and that we continue when $p \geq 0.5$.

Denote $\mathbb{E}_{\mathrm{H}}(n,m)$ as the expected value of our fixed length heuristic if we toss total of n coins, stopping at m^{th} toss to review our estimate. Now, since we initially have no bias about the value of p, our belief on probability of Head is $\frac{1}{2}$. Also, since our strategy is independent of the values of the first m tosses, our belief stays the same for the first m tosses. Thus, the EV of the first m tosses is just 0.

Furthermore, noting that,

$$\mathbb{P}(kH(m-k)T | p = p_0) = \binom{m}{k} p_0^k (1-p_0)^{m-k}$$

We compute the probability $\mathbb{P}(kH(m-k)T)$ with uniform p as the integral with uniform weight:

$$\mathbb{P}(kH(m-k)T) = \binom{m}{k} \int_0^1 p^k (1-p)^{m-k} \mathrm{d}p$$

In particular, it is an well known fact that the integral simplifies to $\frac{k!(m-k)!}{(m+1)!}$. Thus, $\mathbb{P}(kH(m-k)T) = \frac{1}{m+1}$.

Lastly, provided we end up at $kH(m-k)T$ with $k > \frac{m}{2}$, the EV of the remaining $n-m$ tosses are

$$(\alpha_{k,m-k} - \beta_{k,m-k})(n-m) = \frac{(2k-m)(n-m)}{m+2}$$

Thus, $\mathbb{E}_{\mathrm{H}}(n,m)$ is computed as the sum

$$\mathbb{E}_{\mathrm{H}}(n,m) = (n-m) \sum_{k=\lceil m/2 \rceil}^{m} \frac{2k-m}{(m+1)(m+2)}$$

which simplifies to

$$\mathbb{E}_{\mathrm{H}}(n,m) = \begin{cases} \frac{(n-m)m}{4(m+1)} & \text{if } n \text{ even} \\ \frac{(n-m)(m+1)}{4(m+2)} & \text{if } n \text{ odd} \end{cases}$$

Computing the maximum of $\mathbb{E}_{\mathrm{H}}(n,m)$ over m yields $m = \sqrt{n+1} - 1$ and $m = \sqrt{n+2} - 2$ respectively. Then, plugging in integers nearest m will give us the EV of our heuristic and thus the lower bound for a given value of n.

Conclusion

Combining the result of previous two sections gives us a fairly tight estimate on the maximum EV with n tosses. In particular, we see that for m roughly on the order of $\sqrt{n}$,

$$\frac{4\mathbb{E}_{\mathrm{H}}(n,m)}{n} \to 1 \text{ as } n \to \infty$$

Also, we showed that $\mathbb{E}_{\max}(n,0,0)$ is bounded above by $\frac{n+1}{4}$. Thus, it follows by sandwich lemma that

$$\frac{4\mathbb{E}_{\max}(n,0,0)}{n} \to 1 \text{ as } n \to \infty$$

Hence, using the maximum EV strategy, our EV will be on the order of $\frac{n}{4}$. We also note that while maximum EV would require $O(n^2)$ time to fully determine, our heuristic doesn't take anytime to compute while producing an EV with relatively small reduction in return for large values of n.

Expected Returns

n	10	100	1000	10000
$\mathbb{E}_{\max}$	1.674	22.66	244.9	2491.
$\mathbb{E}_{\mathrm{H}}$	1.333	20.48	234.7	2450.

In ending, we list a few natural generalizations of the question

1. Start with prior information about the coin (See Appendix II)
2. Play the game, gaining x points for each Heads and losing 1 point for each Tail
3. Play the game with a dice instead, gaining 5 point for each 6's and losing 1 for the rest
4. Play the game with a dice, gaining/losing some fixed amount of point for each number

Appendix I: Intuitive Argument for Value of $\mathbb{E}(p)$

In the first section, we derived an expression for $\mathbb{E}(p)$ by first showing that p follows a beta distribution. In this section, we show a different approach that requires minimal computation.

Consider the following model for the event of getting an $aHbT$:

1. Choose some p uniformly at random on $[0,1]$
2. define a point x as H if $x < p$ and T if $x > p$

Then, we can model the event $aHbT$ by picking $a+b+1$ points uniformly at random on $[0,1]$ and defining first a points as Heads, next as p and the rest as Tails. Our goal is to estimate the value of p.

First, note that n points chosen uniformly at random will cut the interval $[0,1]$ into $n+1$ sub-intervals. In particular, the lengths of these intervals forms a sequence of $(n+1)$ numbers. Further, all permutation of these lengths should be equally likely, since we are choosing from a uniform distribution. In other words, the expected length of each of these sub-intervals are precisely $\frac{1}{n+1}$. Then, the expected value of each of the n numbers are $\frac{k}{n+1}$ for $k = 1 \dots n$. Therefore, returning to our $a+b+1$ points, the estimate of p, the $(a+1)^{\text{st}}$ point, is $\frac{a+1}{a+b+2}$.

Note that we are essentially evaluating the expected value of the order static of uniform distribution. In particular, $(a+1)^{\text{st}}$ order statistic of $a+b+1$ random sample follows $\mathrm{Beta}(a+1, b+1)$, which is consistent with our previous finding.

Appendix II: Generalization of the Game

Consider the generalization of the game where Bob is allowed to toss the coin a few times before the start of the game. If Bob observes a_0Hb_0T before the game starts, what is the maximum EV of the game then?

First, we observe that our EV is equal to

$$\mathbb{E}_{\max}(n, a_0, b_0) - (a_0 - b_0)$$

by considering the game as having started normally but with scores up until a_0Hb_0T deleted. Thus, our problem reduces to that of finding $\mathbb{E}_{\max}(n, a_0, b_0)$.

Furthermore, we may bound the maximum EV of the modified game similarly using the fixed length heuristic. In particular, our first m flips will now produce an EV of

$$\frac{(a_0 - b_0)m}{a_0 + b_0 + 2}$$

and our probability $\mathbb{P}(kH(m-k)T)$ will again be given by an integral, this time with p distributed over $\mathrm{Beta}(a_0+1, b_0+1)$. Thus,

$$\mathbb{P}(kH(m-k)T) = \binom{m}{k} \frac{\int_0^1 p^{k+a_0}(1-p)^{m-k+b_0}\mathrm{d}p}{\int_0^1 p^{a_0}(1-p)^{b_0}\mathrm{d}p}$$

Now, simplifying the integral yields

$$\frac{(a_0+b_0+1)!m!(k+a_0)!(m-k+b_0)!}{a_0!b_0!(m+a_0+b_0+1)!k!(m-k)!}$$

and reducing to asymtotic behavior for large k and m with fixed a_0, b_0, we have [4]

$$\mathbb{P}(kH(m-k)T) \approx \frac{(a_0+b_0+1)!}{a_0!b_0!}\left(\frac{k}{m}\right)^{a_0}\left(1-\frac{k}{m}\right)^{b_0}$$

[4]Technically, for $k = m$, $\left(1 - \frac{k}{m}\right)$ is not the limiting behavior as b_0 is relevant in this case. However, we can safely ignore it by "throwing away" $k = (m - \sqrt{m}) \sim m$. Our integral will still end at 1 and $m - k > \sqrt{m}$ will outsize b_0

Finally, our heuristic EV is

$$\frac{(a_0 - b_0)m}{a_0 + b_0 + 2} + (n - m) \sum_{k=m_0}^{m} (2k - m)\mathbb{P}(kH(m - k)T)$$

where $m_0 = \lceil \frac{m+a_0+b_0}{2} \rceil$ (since we expect net positive EV when more than half of $m + a_0 + b_0$ flips are heads). Here, we see that for large n, the summation represents the Riemann sum of the integral

$$\frac{(a_0 + b_0 + 1)!}{a_0!b_0!} \int_{\frac{1}{2}}^{1} (2x - 1)x^{a_0}(1 - x)^{b_0)}\mathrm{d}x$$

Defining

$$B(a, b) = \int_{\frac{1}{2}}^{1} x^a(1 - x)^b \mathrm{d}x$$

and

$$I(a, b) = \frac{(a + b + 1)!}{a!b!}(2B(a + 1, b) - B(a, b))$$

we see that

$$\frac{\text{EV}}{n} \to I(a_0, b_0) \text{ as } n \to \infty$$

for the fixed length heuristic, provided we choose m such that $m \to \infty$ and $\frac{m}{n} \to 0$ as $n \to \infty$ (ex. $m \approx \sqrt{n}$). Therefore, by preservation of weak inequality,

$$I(a_0, b_0) \leq \lim_{n\to\infty} \frac{\mathbb{E}_{\max}(n, a_0, b_0)}{n} \tag{1}$$

Now, it follows from simple algebra that

$$I(a, b) = \alpha_{a+1,b}I(a + 1, b) + \beta_{a,b+1}I(a, b + 1)$$

Also, for large enough n, $\mathbb{E}_{\max}$ will favor continuing tossing over stopping the game since as the number of tosses grows, we have greater opportunity to make up for our initial losses. Thus, as $n \to \infty$ $\mathbb{E}_{\max}(n, a, b)$ will follow the same recursion above.

We will finish by proving that our inequality in (1) is in fact an equality by inducting on a and b. Indeed, for $a = b = 0$, $I(0, 0) = \frac{1}{4}$ and our result follows from previous section. Now, assuming equality for some c and d, note that

$$\begin{aligned}
\alpha_{c+1,d}I(c + 1, d) &= I(c, d) - \beta_{c,d+1}I(c, d + 1) \\
&\geq \lim_{n\to\infty} \frac{\mathbb{E}_{\max}(n, c, d)}{n} - \beta_{c,d+1}\frac{\mathbb{E}_{\max}(n, c, d + 1)}{n} \\
&= \alpha_{c+1,d} \lim_{n\to\infty} \frac{\mathbb{E}_{\max}(n, c + 1, d)}{n}
\end{aligned}$$

where the inequality follows from the inductive hypothesis and (1). Combining this result with (1) shows our desired equality.
Therefore, if Bob starts the game after having flipped a_0Hb_0T, then Bob's maximum EV will be approximately $I(a_0, b_0)n$ for large values of n.

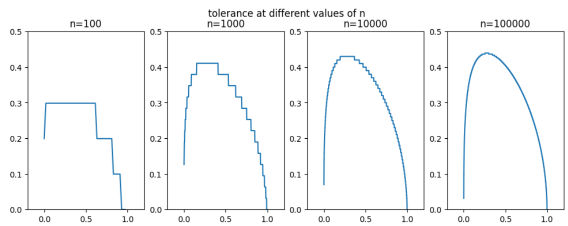
Some values of $I(a, b)$

a \ b	0	1	2	3
0	1/4	1/12	1/32	1/80
1	5/12	3/16	7/80	1/24
2	17/32	23/80	5/32	19/224
3	49/80	3/8	51/224	35/256

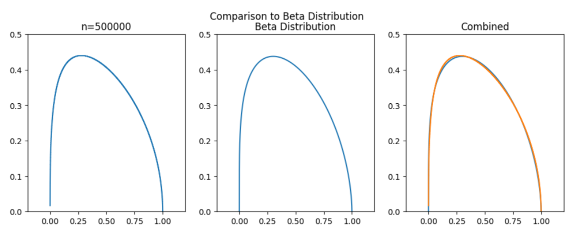
Appendix III: Observation on Decision Boundary

Observe that, assuming same number of Heads, we are in a strictly better position when we have less Tails. Thus, our decision of either continuing or stopping the game comes down to the question of "How many more Tails than Heads will I tolerate?" So, given we have tossed H Heads, we wish to find the tolerance $C(H)$ representing the maximum values of $T - H$ allowed at H Heads. This tolerance distribution represents the decision boundary for the optimal strategy. Thus, it is in our interest to investigate properties of the tolerance distribution. While not providing any proofs, we list some empirical results.

Intuitively, we need to have about as many heads as tails to be confident continuing with the game since our expected return is negative otherwise. Conversely, if we have more than $\frac{n}{2}$ heads in a game with with n tosses, we continue to the end of game. Also, we expect our tolerance to be sub-linear in n since we can't let the tolerance to scale on the order of n as this would mean that we might have much more tails than heads in the long run. In fact, as we increase n, our tolerance distribution seems to converge when scaled by $n/2$ in the number of Heads and by $\sqrt{n}$ in the tolerance.



Furthermore, the limiting distribution looks very close to a beta distribution, and fitting the tolerance graph with $n = 500000$ gives us a beta distribution $\mathrm{Beta}(1.23989379, 1.55704953)$ scaled by 0.34452932.



However, while the two graphs are close, they do not align perfectly. So, tolerance graph might converge to distribution that is slightly different from a beta distribution.

Regardless, utilizing this beta distribution produces a expected return far better than the fixed length heuristic.

Expected Returns

n	10	100	1000	10000
$\mathbb{E}_{\max}$	1.6744	22.660	244.90	2490.6
Beta	1.5950	22.655	244.90	2490.3
$\mathbb{E}_{\mathrm{H}}$	1.3333	20.475	234.68	2450.5

So, by approximating the tolerance with the beta distribution, we can greatly improve our expected return from the fixed length heuristic while greatly reducing the computational complexity. In addition, considering the height of the distribution, maximum EV strategy will bound our losses of playing the game with n tosses at approximately $0.45\sqrt{n}$. Lastly, our distribution can be used for generalized game discussed in Appendix II by simply starting our tolerance at $H = a_0$, which provides an advantage over Fixed Length Heuristic where computing the the optimal value of m ever more complicated for large values of a and b.

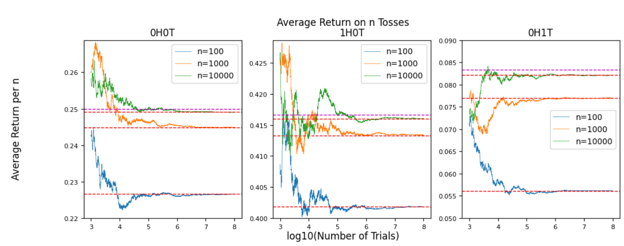
Appendix IV: Average Return of Simulated Games

In order to test our results about the expected return, we have simulated several versions of the game to estimate their return using their average. In particular, we simulated 3 different initial states, $0H0T$, $1H0T$, and $0H1T$, and 3 different number of tosses $n = 100, 1000, 10000$. The table below shows the computed values of maximum expected return scaled down by n.

Expected Returns

n	100	1000	10000	∞
$0H0T$	0.2266	0.2449	0.2491	0.25
$1H0T$	0.4018	0.4133	0.4160	0.4167
$0H1T$	0.0562	0.0770	0.0821	0.0833

Graph below shows the results of our simulation. Each simulation ran a total of 10^8 trials and we plotted the running average of the first 10^3 to 10^8 trials in a logarithmic scale. The red lines represents the max EV and the magenta line represents $I(a, b)$, both scaled down by n.



As shown in the graph, each of the running average converges to their respective expected return, empirically verifying our result.