The truncated coin flips martingale maximizes escape probability. Preliminary Rough Draft

Thomas P. Hayes

February 24, 2003

Abstract

Let $\mathcal{H}$ be the class of discrete real-valued martingales $X_0 = 0, X_1, \ldots, X_n$ such that for each i, $|X_i - X_{i-1}| \leq 1$. Define, for all non-negative integers n, and $a \in \mathbb{R}_{\geq 0}$, the "truncated coin flips" martingale, which, conditioned on any value $X_{i-1} = x < a$, maximizes the probability that $a - X_i$ is either zero or a positive integer with the same parity as the number of steps remaining, $n - i$. We prove that this maximizes $\Pr(\ X_n \geq a\ )$ over all martingales $(X_i) \in \mathcal{H}$. We also show that this is essentially the unique martingale with this extremal property. The proof uses "retrograde analysis," to prove that, under certain circumstances, a gambler's best strategy results in the truncated coin flips martingale.

We give a recurrence formula for $f_n(a) := \max\limits_{(X_i)\in\mathcal{H}} \Pr(\ X_n \geq a\ )$, and prove that, for every n, f_n is piecewise rational and piecewise convex, nowhere increasing, and continuous except at n. When a is an integer, $f_n(a)$ is the sum of the $n + 1 - a$ smallest normalized binomial coefficients, $\binom{n}{i}2^{-n}$, i.e.,

$$f_n(a) = 2^{-n}\left(\sum_{i=0}^{\lfloor\frac{n-a}{2}\rfloor}\binom{n}{i} + \sum_{i=\lceil\frac{n+a+1}{2}\rceil}^{n}\binom{n}{i}\right).$$

Our work gives a simple alternative proof of the well-known Hoeffding-Azuma inequality.

1 The optimal martingale

We begin by defining the truncated coin-flips martingale, which maximizes $\Pr(\, X_n \geq a \,)$ over martingales $(X_n) \in \mathcal{H}$.

Definition 1.1. Let n be a positive integer, and $a \in [0, n]$. Set $X_0 = 0$. Suppose X_{i-1} has been chosen already. When $X_{i-1} = a$, set $X_i = X_{i-1}$. When $X_{i-1} < a$, let T be the greatest integer having the same parity as $n - i$ and less than $a - X_{i-1}$. Set $A = \max\{X_i - 1, a - T - 2\}$, $B = \min\{X_i + 1, a - T, a\}$. Then select X_{i+1} so that

$$X_{i+1} = \begin{cases} A & \text{with probability } \frac{B-x}{B-A} \\ B & \text{with probability } \frac{x-A}{B-A} \end{cases}$$

We call (X_n) the *truncated coin flips martingale*, and observe that it is a martingale in $\mathcal{H}$.

Remark 1.2. Note that if a is an integer, the truncated coin flips martingale is just the usual sum of n coin flips martingale, except that it stops whenever it reaches a. If a is not an integer, the martingale is designed to maximize the probability of moving either to a or to an integer distance from a having the same parity as the number of steps remaining after the move.

Remark 1.3. The truncated coin flips martingale is also uniquely defined by the following properties. $X_0 = 0$. Conditioned on X_i, the distribution of X_{i+1} maximizes $\Pr(\, X_{i+1} \in a + n + i + 2\mathbb{Z} \,)$ subject to the constraints

1. $\Pr(\, X_{i+1} > a \,) = 0$ (since there's no benefit to going beyond a) and
2. $\mathrm{E}(X_{i+1} \mid X_i) = X_i$ and $|X_{i+1} - X_i| \leq 1$ (so $(X_n) \in \mathcal{H}$).
3. If $X_i \in a + n + i + 2\mathbb{Z}$, then make $X_{i+1} = X_i \pm 1$ with probability $1/2$ each. This is the only case when the preceding rules are ambiguous, and can only arise when a is an integer of the opposite parity from n. These are also the unique circumstances in which the TCF martingale is not the unique optimal martingale; we could alternatively set $X_{i+1} = X_i$ at any step, even a randomly determined step, and still have the same probability of success.

We are now ready to state our main result.

Theorem 1.4. *Let n be a positive integer, and $a \in [0, n]$. Then*

$$\max_{\mathbf{X} \in \mathcal{H}} \Pr (X_n \geq a)$$

is achieved by the truncated coin flips martingale with parameters n, a.

When a is an integer, there is a simple description of $f_n(a)$ as a normalized sum of binomial coefficients. The following notation will be convenient.

Notation 1.5. Arrange the $n+1$ binomial coefficients $\binom{n}{i}$ in increasing order:

$$\binom{n}{0}, \binom{n}{n}, \binom{n}{1}, \binom{n}{n-1}, \ldots, \binom{n}{\lceil \frac{n}{2} \rceil}.$$

For $0 \leq k \leq n+1$, let $S_{n,k}$ denote the sum of the first k terms in this sequence.

Corollary 1.6. *Let $a \leq n$ be non-negative integers. Then*

$$\max_{\mathbf{X} \in \mathcal{H}} \Pr (X_n \geq a) = S_{n,n+1-a} 2^{-n}.$$

Proof: We induct on n. When $n = 0$ or $a = 0$, the result is trivial. Let $n > 0$ and assume the result holds for $n-1$. Let $a \in [n]$. By Theorem 1.4, we know that the first step of the optimal martingale in $\mathcal{H}$ is a ± 1 coin flip. Hence, by induction,

$$f_n(a) = (f_{n-1}(a-1) + f_{n-1}(a+1))/2 = 2^{-n}(S_{n-1,n+1-a} + S_{n-1,n-1-a}),$$

where we define $S_{n-1,-1} = 0$. Now, by Pascal's identity, $\binom{n}{i} = \binom{n-1}{i-1} + \binom{n-1}{i}$, it follows easily that $S_{n-1,n+1-a} + S_{n-1,n-1-a} = S_{n-1,n+1-a}$, which completes the proof. ∎

In Section 3, we present a short and elementary proof of Hoeffding's famous inequality, as a consequence of Corollary 1.6.

2 Proof of Theorem 1.4

In this section, we prove our main result.

Our approach is to prove by reverse induction on i, that the truncated coin-flips martingale maximizes $\Pr(X_n \geq a \mid X_i = x)$ for every i and x.

This may be thought of as an example of retrograde analysis, in which the optimal strategy for a game such as chess is found by reverse induction on the nodes of the game tree, starting with the leaves (final positions). In this case the game is a randomized solitaire game in which the player attempts to move a token from 0 to a along the real line. Initially, the token is at position $X = 0$. For each of n "moves" the player specifies a probability distribution on $[X - 1, X + 1]$ with expectation X, where X is the current position of the token. The next position of the token is then drawn at random according to this distribution. After n moves, the player wins iff $X \geq a$.

Alternatively, we can view the problem as a variant of the classical Gambler's Ruin problem, in which a rich gambler tries to amass an even larger target fortune within n bets. Legal limits restrict the maximum win or loss to a fixed amount, and the gambler's initial fortune is large enough to have no chance of running out. Subject to these constraints, and otherwise allowed to choose any fair proposition to bet on, the gambler must adaptively select the sequence of bets maximizing his chance to reach the target fortune.

We will use the following notations

Definition 2.1. Let n be a non-negative integer. The function $f_n : \mathbb{R} \to [0, 1]$ is defined by $f_n(a) := \max_{(X_i) \in \mathcal{H}} \Pr(\ X_n \geq a\ )$.

By definition, $f_0(a) = \begin{cases} 1 & \text{if } a \leq 0 \\ 0 & \text{if } a > 0 \end{cases}$. It is clear that, for every n, f_n is a non-increasing function, and for every a, $f_n(a)$ is a non-decreasing sequence. Other obvious facts are that $f_n(a) = 1$ whenever $a \leq 0$, and $f_n(a) = 0$ whenever $a > n$.

For illustrative purposes, here is the (piecewise) formula for f_5.

$$f_5(a) = \begin{cases} (15a^2 + 21a + 8)/8(a + 1)^3 & \text{for } a \in [0, 1] \\ (3a^2 - 17a + 25)/8(3 - a) & \text{for } a \in [1, 2] \\ (4a - 5)/8(a - 1)^2 & \text{for } a \in [2, 3] \\ (8a^4 - 136a^3 + 871a^2 - 2492a + 2689)/16(a - 5)^3 & \text{for } a \in [3, 4] \\ 1/16(a - 3) & \text{for } a \in [4, 5] \\ 0 & \text{for } a > 5. \end{cases}$$

Figure 2 shows plots of $f_1, \ldots, f_7$.

Definition 2.2. The operator T is defined pointwise by

$$Tf(a) = \max \mathrm{E}(\ f(a + Y)\ ),$$

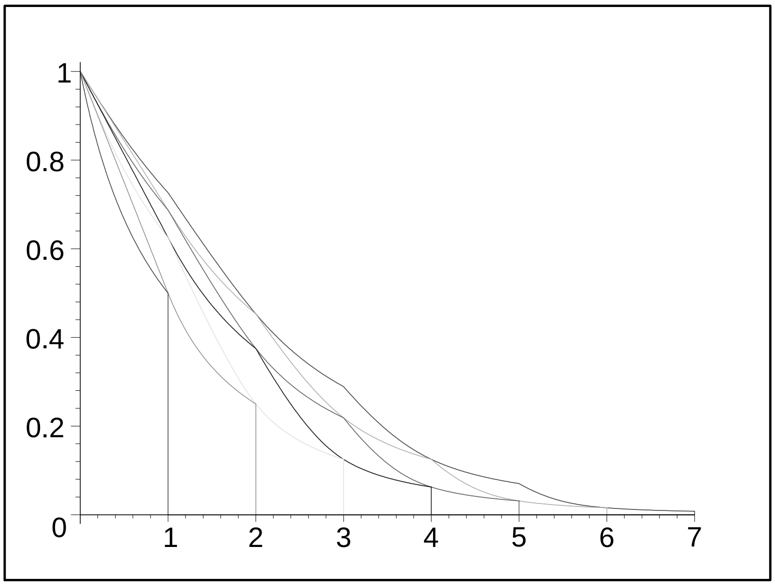

Figure 1: Plots of $f_1, \ldots, f_7$. f_n has a discontinuity at n; the plot indicates this with a vertical line segment. Each f_n is left-continuous everywhere, right-continuous except at i, convex and C^1 except at 0 and integers of the same parity as n, and C^∞ except at integers.

where the maximum is taken over all random variables Y such that $|Y| \leq 1$ and $\mathrm{E}(\ Y\ ) = 0$.

Observation 2.3. In Definition 2.2, we may assume without loss of generality that Y takes on at most two values: for any 3 values y_1, y_2, y_3, where y_2 and y_3 have the opposite sign from y_1, one of the two chords from $(y_1, f(y_1))$ to $(y_i, f(y_i))$, where $i = 2, 3$, must lie above the other (not necessarily strictly), and the other y_i can be dropped from the range of Y.

Almost by definition, $f_n = T^n(f_0)$: if Y maximizes $\mathrm{E}(\ f_{n-1}(a+Y)\ )$ subject to the constraints, then $-Y$ is the best first move in the game. Most of the properties we require of f_n are immediate consequences of more general properties of T, which we now prove.

Lemma 2.4. *1. If f is non-increasing on $\mathbb{R}$, then so is Tf.*

2. *Suppose f is non-increasing on $\mathbb{R}$. If f is right-continuous (resp. left-continuous) at a, then Tf is right-continuous (resp. left-continous) at $a+1$.*

3. *Suppose $Tf(a) = \frac{B-a}{B-A} f(A) + \frac{a-A}{B-A} f(B)$, where $A \in [a-1, a), B \in (a, a+1]$. So $\mathrm{E}(\ f(a+Y)\ )$ is maximized over $\mathrm{E}(\ Y\ ) = 0$, $|Y| \leq 1$ when $a+Y$ takes values A and B. Suppose further that f is convex on (A', B'), where $A' \leq A$, $A' \neq a-1$, $B' \geq B$ and $B' \neq a+1$. Then Tf is convex at a.*

Proof:

1. T clearly preserves non-negativity. T also clearly preserves non-increasingness, because given any valid random variable Y, and $a < a'$, we have $f(Y+a) \geq f(Y+a')$, and taking expectations preserves this ordering.

2. This one is quite tedious considering how easy it is to see. The third part, on the other hand, is much more interesting. Suppose f is non-increasing on $\mathbb{R}$ and left-continuous at a. Let $a_1 < a_2 < \ldots$ be an increasing sequence with limit a, and let $Y_1, Y_2, \ldots$ be a corresponding sequence of random variables, each taking at most two values, such that $\mathrm{E}(\ Y_i\ ) = 0$, and $Tf(a_i) = \mathrm{E}(\ f(a_i + Y_i)\ )$. Since there are only three free variables involved for each i (a_i and the two values of Y_i) we can think of this as a sequence in the unit cube, which is compact. Hence, there must be a convergent subsequence, with a limiting random

variable Y, satisfying $\mathrm{E}(\ Y\ ) = 0$. Now, if f is left-continuous at both values of $a - Y$, then $Tf(a) \geq \lim Tf(a_i)$, with Y as witness. Since Tf is non-increasing, it follows that $Tf(a) = \lim Tf(a_i)$. Otherwise, shifting both values of Y to the left by a tiny amount (and changing their probabilities appropriately) shows that $Tf(a) \geq \lim Tf(a_i) - \epsilon$ for every $\epsilon > 0$, and hence that $Tf(a) \geq \lim Tf(a_i)$. The only thing which can go wrong in the above is that f could be left-discontinuous at $a-1$, and the left value of Y could be -1, in which case it cannot be shifted further left. Thus, if f is non-increasing on $\mathbb{R}$, and left-continuous at $a - 1$, then Tf is left-continuous at a. [1]

The argument that, if f is non-increasing on $\mathbb{R}$ and right-continuous at $a - 1$, then Tf is right-continuous at a is almost the same, so I am omitting it.

3. We use the following characterization of convex functions:

Lemma 2.5. *Let $f : I \to \mathbb{R}$, where I is an open interval.*

(a) Let $x \in I$. f is convex at x iff there exists $\epsilon > 0$ such that if y is a two-value random variable satisfying $|y| < \epsilon$ and $\mathrm{E}(\ y\ ) = 0$, then $\mathrm{E}(\ f(x+y)\ ) \geq f(x)$.

(b) f is convex on I iff for every pair of random variables X, Y such that $\mathrm{E}(Y \mid X) = 0$ and $X+Y \in I$, it follows that $\mathrm{E}(\ f(X+Y)\ ) \geq \mathrm{E}(\ f(X)\ )$.

(c) Assume f is continuous. Then f is convex on I iff for every $x \in I$, f is convex at x.

Now we prove that Tf is convex at a. We apply Lemma 2.5. Suppose y is a suitably small two-value random variable satisfying $\mathrm{E}(\ Y\ ) = 0$. We shall prove that $\mathrm{E}(\ Tf(a+Y)\ ) \geq Tf(a)$. We are given that $Tf(a) = \mathrm{E}(\ Tf(X)\ )$, where X is a random variable satisfying $\mathrm{E}(\ X\ ) = a$, $|X - a| \leq 1$, and the values of X are A and B. Without loss of generality, assume that X and Y are independent.

Case 1: $A = a - 1$, $B = a + 1$. We know that $\mathrm{E}(\ Tf(a+Y)\ ) \geq \mathrm{E}(\ f(X+Y)\ )$ since $X - a$ has maximum length at most one and expectation zero, and since X and Y are independent. Requiring Y to be

[1] This argument ought to be shortened somehow.

small enough that $X+Y \in (A', B')$, we can apply the convexity of f and Lemma 2.5 to conclude $Tf(a+Y) \geq \mathrm{E}\,(\, f(X+Y)\,) \geq \mathrm{E}\,(\, f(X)\,) = Tf(a)$.

Case 2: $A > a-1$, $B < a+1$. Require Y to be smaller than $\min\{A+1-a, a+1-B\}$. Define X' so that $X'+Y$ has value set $\{A, B\}$ and $\mathrm{E}\,(X' \mid Y) = a$; these conditions uniquely specify the joint distribution of (X', Y). By our requirement on the size of Y, it follows that $X'-a$ has maximum length ≤ 1. Thus $\mathrm{E}\,(\,Tf(a+Y)\,) \geq \mathrm{E}\,(\, f(X'+Y)\,)$. But there is only one distribution on A, B having expectation a: namely, that of X. Thus $\mathrm{E}\,(\,Tf(a+Y)\,) \geq \mathrm{E}\,(\, f(X)\,) = Tf(a)$.

Case 3: Exactly one of A, B, is at distance 1 from a. Without loss of generality, assume $A = a-1$ and $B < a+1$. Require Y to be sufficiently small that $a+Y-A \leq 1$, $B-(a+Y) \leq 1$, and $a+Y-1 > A'$.

For each value y of Y, define X' so that X' is either $a-1$ or $B-y$, with expectation a, conditioned on the event $\{Y = y\}$. By our restriction on $|Y|$, it follows that $X'-a$ has maximum length at most one. Thus $\mathrm{E}\,(\,Tf(a+Y)\,) \geq \mathrm{E}\,(\, f(X'+Y)\,)$. Suppose that $\Pr\,(\, X' = B-Y\,) \geq \Pr\,(\, X = B\,)$.

Assume without loss of generality that $X'+Y = B$ whenever $X = B$. Then, conditioned on the event $X \neq B$, the expectation of $X'+Y$ must be $a-1$, since we know $\mathrm{E}\,(\, X'+Y\,) = \mathrm{E}\,(\, X\,) = a$, and $X = a-1$ when $X \neq B$. By convexity of f, it follows that $\mathrm{E}\,(f(X'+Y) \mid X \neq B) \geq f(a-1) = \mathrm{E}\,(f(X) \mid X \neq B)$. But when $X = B$, $f(X) = f(X'+Y)$. Thus $\mathrm{E}\,(\, f(X'+Y)\,) \geq \mathrm{E}\,(\, f(X)\,) = Tf(a)$, completing the proof. All that remains is to prove that $\Pr\,(\, X' = B-Y\,) \geq \Pr\,(\, X = B\,)$.

Assume without loss of generality that Y takes on exactly two values, $-\delta$ and $+\epsilon$. Then

$$\begin{aligned} \Pr\,(\, X'+Y = B\,) &= \frac{\epsilon}{(\delta+\epsilon)(1+B-a+\delta)} + \frac{\delta}{(\delta+\epsilon)(1+B-a-\epsilon)}, \\ \text{while } \Pr\,(\, X = B\,) &= \frac{1}{1+B-a}. \end{aligned}$$

Subtracting and simplifying gives

$$\Pr\,(\, X'+Y = B\,) - \Pr\,(\, X = B\,) = \frac{\delta\epsilon}{(1+B-a+\delta)(1+B-a-\epsilon)(1+B-a)} > 0.$$

■

Lemma 2.4 has the following immediate important consequences for f_n.

Corollary 2.6. *f_n is nowhere increasing, left-continuous everywhere, and right-continuous except at n. Suppose that f_n is convex except at 0 and at $n, n-2, n-4, \ldots, n-2\lfloor n/2 \rfloor$. Suppose further that the truncated coin flips martingale has been established as optimal for $n+1$ moves. Then f_{n+1} is convex except at 0 and $n+1, n-1, n-3, \ldots$.*

Proof: The first claim is immediate from Lemma 2.4. For the second claim, we note that $f_{n+1} = Tf_n$. Since we are assuming the truncated coin flips martingale is optimal, we know the values of A and B for any given a, and that f_n is convex on the intervals (A, B). Except for integer a, it is clear how to pick (A', B') meeting the criteria of Lemma 2.4, part (3). Thus f_{n+1} is convex except on $0, 1, \ldots, n+1$. For a an integer with the same parity as n, the explicit formula for the success probability of truncated coin flips, Corollary 1.6, establishes that $f_{n+1}(a) = f_n(a) = (f_n(a+1) + f_n(a-1))/2$. Since f_n is convex on $(a-2, a)$ and $(a, a+2)$, we can now define a function $g \geq f_n$ which is convex on $(a-2, a+2)$ by

$$g(x) = \begin{cases} f(x) & \text{if } x \in (a-2, a-1) \cup (a+1, a+2) \\ f_n(a) + (x-a)(f_n(a+1) - f_n(a)) & \text{if } x \in [a-1, a+1] \end{cases}$$

Now, for any random variable Y with $\mathrm{E}(\,Y\,) = 0$ and $|Y| \leq 1$, define a random variable X' with expectation a such that $X + Y$ always takes values in $\{a-1, a, a+1\}$.

■

Our goal will be to prove by induction on n that $f_n(a) = \Pr(\,X_n \geq a\,)$, where (X_n) is the truncated coin flips martingale.

Proof of Theorem 1.4: (Sketch)

Recall that $f_0(a) = \begin{cases} 1 & \text{if } a < 0 \\ 0 & \text{otherwise} \end{cases}$, and $f_n = T^n f_0$. By Lemma 2.4 and induction, it follows that

1. f_n is non-increasing on $\mathbb{R}$.
2. f_n is left-continuous on $\mathbb{R}$ and right-continuous except at n.
3. f_n is convex except at $0, 1, \ldots, n$.

Now, doubtless we could prove more, such as that f_n is C^∞ except at $0, 1, \ldots, n$, and C^1 except at $n, n-2, \ldots, n-2\lfloor n/2 \rfloor$. However, we don't need these properties for our result.

Finally, we establish optimality of the TCF martingale by another induction. The base cases, $n \leq 2$, are all straightforward. Suppose that the TCF martingale is known to be optimal for $\leq n$ steps, and every a. Let $a > 0$ be given. We need to show that $Tf_n(a)$ is witnessed by a random variable Y corresponding to the definition of the TCF martingale. Namely, Y should have a range of size 2, and should maximize $\Pr(\ a + Y \in n + 2\mathbb{Z}\ )$ subject to the constraints $a + Y \geq 0$, $|Y| \leq 1$, and $\mathrm{E}(\ Y\ ) = 0$.

As observed earlier, Y has a range of size 2 without loss of generality. The constraints $|Y| \leq 1$ and $\mathrm{E}(\ Y\ ) = 0$ are mandated by the definition of $\mathcal{H}$. The condition $a + Y \geq 0$ certainly cannot hurt–if Y were ever less than $-a$, we could increase that value to $-a$, decrease some positive value of Y, and since we know f_n is non-increasing, this would only increase $\mathrm{E}(\ f_n(a+Y)\ )$.

Now, since f_n is convex except at $\mathbb{Z}$, the only possible values for Y are ± 1, and values which make $a + Y$ an integer; any value y of Y which did not satisfy both conditions could be replaced by 2 values $y \pm \epsilon$, each chosen with $1/2$ the original probability of y, and by convexity, the expectation of $f_n(a + Y)$ would increase.

Next we prove by induction that f_n is convex at integers a of the opposite parity from n. Since we know that f_{n-1} and f_n were realized by TCF, we know that

$$f_n(a) = f_{n-1}(a) = (f_{n-1}(a-1) + f_{n-1}(a+1))/2.$$

Since f_n was realized by TCF, we know that

$$f_n(a \pm \epsilon) = (\epsilon' f_{n-1}(a \pm \epsilon \pm 1) + (1 - \epsilon') f_{n-1}(a)),$$

where $\epsilon' = \epsilon/(1+\epsilon)$. Hence

$$\begin{aligned} f_n(a+\epsilon) + f_n(a-\epsilon) - 2f_n(a) &= \epsilon'(f_{n-1}(a+1+\epsilon) + f_{n-1}(a-1-\epsilon)) + 2(1-\epsilon')f_{n-1}(a) - 2f_{n-} \\ &= \epsilon'(f_{n-1}(a+1+\epsilon) - f_{n-1}(a+1) - f_{n-1}(a-1) + f_{n-1}(a-1-\epsilon \end{aligned}$$

which is positive because by induction, f_{n-1} is convex on the intervals $(a-2, a)$ and $(a, a+2)$, and since we know $f_{n-1}(a) = (f_{n-1}(a-1) + f_{n-1}(a+1))/2$, we can replace f_{n-1} by a straight line on the interval $[a-1, a+1]$ to obtain a convex function on $(a-2, a+2)$ with the same values at the 5 points of interest $(a, a \pm 1, a \pm 1 \pm \epsilon)$.

All that remains is to show that it is correct to prefer to make $a + Y$ an integer of the same parity as n than to move past such an integer. This requires a case analysis.

Case 1: $a \in (0, 1)$ and n odd.

Case 2: $\lfloor a \rfloor$ has the same parity as n.

Case 3: $\lceil a \rceil$ has the same parity as n. Since the first two cases are pretty similar, I will prove only this one for now... probably this proof needs to be broken up into some more lemmas anyway, so I will do that before filling in every detail...

We have two possible first moves to consider: We can set $A = a - 1$, $B = a + 1$, or we can set $A = a - 1, B = \lceil a \rceil$. Then A or B is chosen with probabilities set to make the expected change zero.

Now, for the second move, we know by induction that TCF defines the best strategy. This means that $\lceil a \rceil \pm 1$ will be target values. Writing out the game trees to depth 2 for the two possible first moves we find that with the ± 1 strategy, we end at $a - 2$ with probability $\alpha/2$, $\lceil a \rceil - 1$ with probability $(1 - \alpha)/2$, a with probability $\alpha/2$, and $\lceil a \rceil + 1$ with probability $(1 - \alpha)/2$, where $\alpha = 1 - 1/(\lceil a \rceil - a)$. With the truncating strategy, we end at $a - 2$ with probability α^2, $\lceil a \rceil - 1$ with probability $(1 - \alpha)(\alpha + 1/2)$, and $\lceil a \rceil + 1$ with probability $(1 - \alpha)/2$. If we "take the difference", we find we are comparing $f_{n-1}(a - 2)\alpha/2 + f_{n-1}(a)\alpha/2$ with $f_{n-1}(a - 2)\alpha^2 + f_{n-1}(\lceil a \rceil - 1)\alpha(1 - \alpha)$. But, except for the extra factor of α, this is exactly the comparison which would have been made for the decision at $f_n(a - 1)$, and in that case TCF won. So the second strategy is the better one for $f_{n+1}(a)$ as well. ∎

3 Hoeffding's bound

We now show how Hoeffding's bound, $f_n(a) \leq 2^{n\left(H\left(\frac{a+n}{2n}\right)-1\right)}$, can be derived from Corollary 1.6, at least when a is an integer. Hopefully the non-integer case also follows easily.

The following upper bound is a slight (up to a factor of 2) improvement over a well-known estimate on tails of the binomial distribution (cf. [4, Lemma 10.11.8, p. 310]).

Theorem 3.1. *For* $0 \leq k \leq n$*,*

$$S_{n,k+1} \leq \frac{n^n}{k^k(n-k)^{n-k}} = 2^{nH\left(\frac{k}{2n}\right)}.$$

Remark 3.2. Equality holds when $k = 0$ or $k = n$. On the other hand, it follows easily from Stirling's approximation,

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n},$$

that

$$S_{n,k+1} = \Theta\left(n^{-1/2} 2^{nH\left(\frac{k}{2n}\right)}\right),$$

when the ratio $\lambda = k/n \notin \{0, 1\}$ is fixed.

Our proof relies on the following Lemma.

Lemma 3.3. *Let $k \leq n$ be non-negative integers, $n > 0$. Let $r \in [0, 1]$. Then*

$$r^k S_{n,k+1} \leq \left(1 + r^2\right)^n.$$

Equality holds iff $k = r = 0$ (when we assume $0^0 = 1$) or $k = n$.

Proof: The case $r = 0$ is clear, so assume $r > 0$. By the binomial theorem,

$$r^{-k}\left(1 + r^2\right)^n = \sum_{i=0}^{n} \binom{n}{i} r^{2i-k} \geq \sum_{i=0}^{k} \binom{n}{i} r^{2i-k}.$$

"Folding" this sequence, and assuming for notational convenience that k is odd, we obtain

$$= \sum_{i<k/2} \binom{n}{i} r^{2i-k} + \binom{n}{k-i} r^{k-2i}.$$

Now, whenever $i \leq k - i \leq n - i$, we have $\binom{n}{k-i} \geq \binom{n}{i}$. Thus

$$\geq \sum_{i<k/2} \binom{n}{i} \left(r^{2i-k} + r^{k-2i}\right).$$

Finally, since $x + 1/x \geq 2$ for all $x > 0$,

$$\geq 2 \sum_{i<k/2} \binom{n}{i} = S_{n,k+1}.$$

The proof for k even is identical, except that the largest binomial coefficient pairs with itself in the "folding" step. ■

Proof of Theorem 3.1: The result is clear for $k = 0$. Otherwise, apply Lemma 3.3, and set $r = \sqrt{k/(n-k)}$, which minimizes $r^{-k}\left(1+r^2\right)^n$. Simplifying the resulting expression yields the result. ■

Next, we establish that Hoeffding's bound implies the Hoeffding-Azuma bound.

Lemma 3.4. *For all $x \in [0,1]$,*

$$H(x) \le 1 - \frac{(2x-1)^2}{2\ln 2}.$$

Proof: Differentiating, we find $H'(x) = \log_2\left(\frac{1-x}{x}\right)$ and $H''(x) = \frac{-1}{x(1-x)\ln 2} \le \frac{-4}{\ln 2}$. In particular, $H(\frac{1}{2}) = 1$, $H'(\frac{1}{2}) = 0$, and so Taylor's formula gives

$$H(x) \le 1 + \frac{-4}{\ln 2}\frac{(x-1/2)^2}{2} = 1 - \frac{(2x-1)^2}{2\ln 2}.$$

■

The desired inequality is an immediate corollary:

Corollary 3.5. *For all $a \in [-n,n]$,*

$$2^{n\left(H\left(\frac{a+n}{2n}\right)-1\right)} \le e^{-a^2/2n}.$$

References

[1] K. Azuma. *Weighted Sums of Certain Dependent Random Variables.* Tôhoku Math. Journ. **19** (1967) 357–367.

[2] H. Chernoff. *A measure of the asymptotic efficiency for tests of a hypothesis based on the sum of observations.* Annals of Mathematical Statistics **23** (1952) 493–509.

[3] W. Hoeffding. *Probability inequalities for sums of bounded random variables.* J. Am. Stat. Assoc. **58** (1963) 13–30.

[4] F.J. MacWilliams, N.J.A. Sloane. *The Theory of Error-Correcting Codes.* North-Holland, 1977.