

# Behaviour of Sequential Predictors of Binary Sequences

T. M. Cover

Publishing House of the Czechoslovak Academy of Sciences, Prague 1967

## 1 Introduction

This paper explores the effectiveness of sequential predictors on finite binary sequences. Prior work by Robbins, Hannan, and Blackwell has demonstrated algorithms capable of asymptotically achieving high scores by leveraging the empirical distribution of the sequences. Fogel has notably extended this by adapting the predictive model based on the sequence observed, achieving near-perfect accuracy on sequences displaying non-random patterns, such as prime numbers. This study investigates the capacity of sequential predictors to consistently generate high scores across a broad spectrum of sequences, delving into the inherent limits and potential of such algorithms in systematic and non-random environments.

## 2 Deterministic Predictors

Consider the set of  $2^n$  sequences  $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_n) \in \{0, 1\}^n$ .

At stage  $k$ , after the observation  $\Theta_1, \Theta_2, \dots, \Theta_{k-1}$ , the prediction 1 or 0 will be made with probability  $p_k$  and  $1 - p_k$  respectively.

A **sequential predictor** on  $\{0, 1\}^n$  will be completely specified by the set of functions

$$p_1, p_2(\Theta_1), p_3(\Theta_1, \Theta_2), \dots, p_n(\Theta_1, \Theta_2, \dots, \Theta_{n-1})$$

taking values in  $[0, 1]$ .

- If the  $p_i$ s are restricted to  $\{0, 1\}$ , the predictor is called a **deterministic predictor**.
- If the  $p_i$ s are independent of the  $\Theta$ s, the predictor is called a **memoryless predictor**.
- If the  $p_i$ s are also independent of  $i$ , the predictor is called a **constant/time invariant predictor**.

Let  $\delta = (\delta_1, \delta_2, \dots, \delta_n) \in \{0, 1\}^n$  be the sequence of R.V.s resulting from the predictor  $p = (p_1, p_2, \dots, p_n)$  and the sequence  $\Theta \in \{0, 1\}^n$ .

Then the empirical average score (the fraction of correct predictions) is given by

$$s = \frac{1}{n} \sum_{i=1}^n (\delta_i \Theta_i + (1 - \delta_i)(1 - \Theta_i)) \quad (2.1)$$

and the expected empirical average score is given by

$$\bar{s} = \mathbb{E}_p(s) = \frac{1}{n} \sum_{i=1}^n (p_i \Theta_i + (1 - p_i)(1 - \Theta_i)) \quad (2.2)$$

**(I) Any sequential deterministic predictor attains a score of  $\frac{k}{n}$  on precisely  $\binom{n}{k}$  sequences in  $\{0, 1\}^n$  where  $k \in [n]$ . For any deterministic predictor, there exists a sequence upon which a score of 0 is attained.**

### 3 Sequential Betting Systems

#### 3.1 Achievable Winnings in Sequential Betting

A series of  $n$  bets  $b = (b_1, b_2, \dots, b_n)$  is made by a gambler on the outcomes of a sequence  $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_n) \in \{0, 1\}^n$ . The gambler's net gain at bet  $k$  is  $b_k$  if  $\Theta_k = 1$  and  $-b_k$  if  $\Theta_k = 0$ . Hence, his net winnings  $w(\Theta)$  using strategy  $b$  against sequence  $\Theta$  is

$$w(\Theta) = \sum_{k=1}^n (b_k \Theta_k - b_k(1 - \Theta_k)) = \sum_{k=1}^n b_k(2\Theta_k - 1), \quad (3.1)$$

where, in general,  $b_k$  will be a real valued function of  $\Theta$ .

Notice that a gambler may win any preassigned amount  $w(\Theta)$  if  $\Theta$  is known a priori. For example, any  $w$  could be achieved with the betting system

$$\begin{aligned} b_1 &= w(\Theta)\Theta_1 - w(\Theta)(1 - \Theta_1), \\ b_2 &= b_3 = \dots = b_n = 0. \end{aligned} \quad (3.2)$$

However, if he knows only  $\Theta_1, \Theta_2, \dots, \Theta_{k-1}$  when he must place his bet  $b_k$ , his set of achievable winnings  $w$  on  $\{0, 1\}^n$  is limited. For, if  $\{b_1, b_2, \dots, b_n\}$  achieves  $w$ , then manipulation of the above sum, noting the functional independence of  $b_k$  and  $\Theta_k$ , yields

$$w(\Theta_1, \dots, \Theta_{n-1}, 1) + w(\Theta_1, \dots, \Theta_{n-1}, 0) = 2 \sum_{k=1}^{n-1} b_k(2\Theta_k - 1), \quad (3.3)$$

and

$$w(\Theta_1, \dots, \Theta_{n-1}, 1) - w(\Theta_1, \dots, \Theta_{n-1}, 0) = 2b_n. \quad (3.4)$$

So,  $b_n$  is determined and 3.1 is replaced by 3.3 for the determination of  $b_{n-1}$ . Proceeding, we find

$$\sum_{\Theta} w(\Theta) = 0 \quad (3.5)$$

**Proof:**

$$\begin{aligned} \sum_{\Theta} w(\Theta) &= \sum_{\Theta} \sum_{k=1}^n b_k(2\Theta_k - 1) \\ &= \sum_{k=1}^n b_k \sum_{\Theta} (2\Theta_k - 1) \\ &= \sum_{k=1}^n b_k \cdot 0 = 0. \end{aligned}$$

and

$$b_k = \left(\frac{1}{2}\right)^{n-k+1} \sum_{(\Theta_k, \Theta_{k+1}, \dots, \Theta_n) \in \{0,1\}^{n-k+1}} w(\Theta)(2\Theta_k - 1) \quad (k = 1, 2, \dots, n). \quad (3.6)$$

**Proof:**

For  $k = n$ , we need to show that

$$b_n = \left(\frac{1}{2}\right) \sum_{\Theta_n \in \{0,1\}} w(\Theta_1, \dots, \Theta_{n-1}, \Theta_n)(2\Theta_n - 1)$$

we have from 3.4 that

$$\begin{aligned} w(\Theta_1, \dots, \Theta_{n-1}, 1) - w(\Theta_1, \dots, \Theta_{n-1}, 0) &= 2b_n, \\ b_n &= \frac{1}{2}(w(\Theta_1, \dots, \Theta_{n-1}, 1) - w(\Theta_1, \dots, \Theta_{n-1}, 0)). \end{aligned}$$

For general  $b_k$  it holds that

$$\begin{aligned} &\sum_{(\Theta_k, \dots, \Theta_n)} w(\Theta_1, \dots, \Theta_{k-1}, \Theta_k, \dots, \Theta_n)(2\Theta_k - 1) = \\ &\sum_{(\Theta_{k+1}, \dots, \Theta_n)} w(\Theta_1, \dots, \Theta_{k-1}, 1, \Theta_{k+1}, \dots, \Theta_n) - w(\Theta_1, \dots, \Theta_{k-1}, 0, \Theta_{k+1}, \dots, \Theta_n) = \\ &2 \cdot b_k \cdot 2^{n-(k+1)+1} = 2 \cdot b_k \cdot 2^{n-k} = b_k \cdot 2^{n-k+1} \\ \implies &b_k = \left(\frac{1}{2}\right)^{n-k+1} \sum_{(\Theta_k, \dots, \Theta_n) \in \{0,1\}^{n-k+1}} w(\Theta)(2\Theta_k - 1). \end{aligned}$$

Hence, for  $w(\Theta)$  to be achievable by a sequential betting scheme, it is necessary and sufficient (3.5) be satisfied. The betting scheme achieving  $w$  is unique and is given by (3.6).

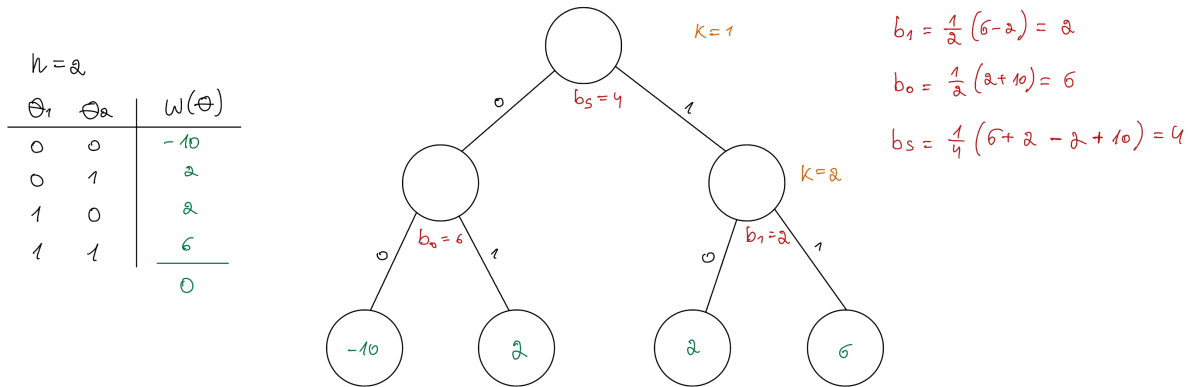


Figure 1: Sequential Betting Scheme

**Summary:**

Consider a betting strategy for a game against sequences:  $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_n) \in \{0,1\}^n$ , which allows the bet  $b_k$  at stage  $k$  to be some element in a subset  $B_k$  of the collection  $B$  of all functions from  $\{0,1\}^n$  to  $\mathbb{R}$ . Let  $w : \{0,1\}^n \rightarrow \mathbb{R}$  be a desired set of net winnings defined for

each sequence  $\Theta$  in  $\{0,1\}^n$ . As before, 3.1 expresses the net winnings  $w(\Theta)$  as a function of  $\{b_1, b_2, \dots, b_n\}$ . Then:

**(II) Trivially, if  $B_k = B$ , any  $w$  is achievable.**

**(III) If  $B_k$  is the set of all functions in  $B$  depending only on  $\Theta_1, \Theta_2, \dots, \Theta_{k-1}$ , then  $w$  is achievable if and only if (3.5) is satisfied.**

**(IV) If, for  $k = 1, 2, \dots, n$ ,  $B_k \subseteq B$  is the set of functions bounded in absolute value by  $b$ , depending only on  $\Theta_1, \Theta_2, \dots, \Theta_{k-1}$ , then  $w$  is achievable if and only if**

$$\sum_{\Theta} w(\Theta) = 0 \quad (3.7)$$

**and if, for  $k = 1, 2, \dots, n$ ,**

$$\left| \left( \frac{1}{2} \right)^{n-k+1} \sum_{(\Theta_k, \Theta_{k+1}, \dots, \Theta_n) \in \{0,1\}^{n-k+1}} w(\Theta)(2\Theta_k - 1) \right| < b \quad (3.8)$$

**for every  $(\Theta_1, \Theta_2, \dots, \Theta_{k-1}) \in \{0,1\}^{k-1}$ . This is the sequential betting scheme with bounded bet size.**

### 3.2 Winnings which are functions of $\sum_{i=1}^n \Theta_i$

We may be interested in winnings  $w$  which are functions only of  $\sum_{i=1}^n \Theta_i$ , the number of 1's in  $\Theta$ . In this case, define, for every  $\Theta \in \{0,1\}^n$ ,

$$\hat{w} \left( \sum_{i=1}^n \Theta_i \right) = w(\Theta). \quad (3.9)$$

Then the conditions of 3.7 and 3.8 become respectively

$$\sum_{k=0}^n \binom{n}{k} \hat{w}(k) = 0 \quad (3.10)$$

and

$$|b_k(i)| = \left| \left( \frac{1}{2} \right)^{n-k+1} \sum_{j=0}^{n-k} (\hat{w}(i+j+1) - \hat{w}(i+j)) \binom{n-k}{j} \right| < b \quad (3.11)$$

For  $i = 0, 1, \dots, k-1$  and  $k = 1, 2, \dots, n$  where  $b_k(i)$  is the bet at stage  $k$  when the sum of the first  $k-1$  outcomes is  $i$ .

**Proof:**

We start by expanding the expression from Equation (3.8) under the assumption that  $\hat{w}(\sum_{i=1}^n \Theta_i) = w(\Theta)$  from Equation (3.9):

$$\begin{aligned}
\text{LHS of (3.8)} &= \left| \left( \frac{1}{2} \right)^{n-k+1} \sum_{(\Theta_k, \Theta_{k+1}, \dots, \Theta_n) \in \{0,1\}^{n-k+1}} w(\Theta)(2\Theta_k - 1) \right| \\
&= \left| \left( \frac{1}{2} \right)^{n-k+1} \sum_{j=0}^{n-k} \left( \sum_{(\Theta_k, \dots, \Theta_n) \in \{0,1\}^{n-k+1}: \sum \Theta_i = j} w(\Theta)(2\Theta_k - 1) \right) \right| \\
&= \left| \left( \frac{1}{2} \right)^{n-k+1} \sum_{j=0}^{n-k} (\hat{w}(i+j+1) - \hat{w}(i+j)) \binom{n-k}{j} \right|,
\end{aligned}$$

where the binomial coefficient  $\binom{n-k}{j}$  counts the number of ways in which  $\sum \Theta_i = j$  can occur. The term inside the summation represents the change in winnings  $\hat{w}$  when the count of 1's increases by one, hence reflecting the derivative (or difference)  $\hat{w}(i+j+1) - \hat{w}(i+j)$ . The multiplication by  $\binom{n-k}{j}$  adjusts for the number of sequences that result in a particular sum  $j$  of 1's among the last  $n-k$  flips.

Thus, we have shown that:

$$\text{LHS of (3.8)} = \text{RHS of (3.11)},$$

Now we will find a sufficient bound on  $|b_k(i)|$  such that the conditions of 3.11 are satisfied.

Letting  $k = n$  in (3.11) we have the condition:

$$|b_n(i)| = \frac{1}{2} |\hat{w}(i+1) - \hat{w}(i)| < b \quad (3.12)$$

for  $i = 0, 1, \dots, n-1$ .

All other conditions of (3.11) are consequences of (3.12), since, assuming (3.12) true,

$$\begin{aligned}
|b_k(i)| &= \left| \left( \frac{1}{2} \right)^{n-k+1} \sum_{j=0}^{n-k} (w(i+j+1) - w(i+j)) \binom{n-k}{j} \right| \leq \\
&\left( \frac{1}{2} \right)^{n-k+1} \sum_{j=0}^{n-k} \binom{n-k}{j} 2b = b.
\end{aligned} \quad (3.13)$$

**(V) A terminal score  $\hat{w}$  depending only on  $\sum_{i=1}^n \Theta_i$  is achievable by a sequential betting scheme with bounded bet size  $b$  if and only if**

$$\sum_{k=0}^n \binom{n}{k} \hat{w}(k) = 0 \quad (3.14)$$

**and**

$$|\hat{w}(k+1) - \hat{w}(k)| < 2b, \quad k = 1, 2, \dots, n-1. \quad (3.15)$$

### 3.3 Examples

#### 3.3.1 Example 1 - foreknowledge of a sequence which will not occur

Consider a gambler betting on a binary sequence of length  $n$ , consisting of 1's and 0's. The gambler can choose his bet amount at each stage, based on the sequence observed so far. Suppose the gambler knows in advance that a specific sequence, say  $\Theta^*$ , will not occur. The question

is whether the gambler can guarantee a profit. The answer is yes; he can potentially win an infinite amount.

For any goal function  $w(\Theta)$ , there exists a betting strategy that guarantees a win of  $w(\Theta)$  for any sequence  $\Theta \neq \Theta^*$ . This is achieved by setting:

$$w(\Theta^*) = - \sum_{\Theta \neq \Theta^*} w(\Theta)$$

and applying the betting strategy outlined in section 3.6. This example illustrates that, under certain conditions, a gambler can manipulate his bets based on a probability distribution over possible outcomes to achieve a desired terminal wealth distribution.

### 3.3.2 Example 2 - independent flips of a fair coin

Let  $\Theta_1, \dots, \Theta_n$  be independent flips of a fair coin. For a desired distribution function  $F$ , there exists a sequential betting scheme achieving a terminal distribution  $F_n$  such that

$$\sup_x |F(x) - F_n(x)| < \frac{1}{2^n}$$

To achieve  $F_n$  in the case of continuous  $F$ , choose  $w_i$  such that  $F(w_i) = \frac{i}{2^n}$  for  $i = 1, 2, \dots, 2^n - 1$  and  $w_0 = - \sum_{i=1}^{2^n-1} w_i$ . Associate the winnings  $w$  with the outcomes  $\Theta$ 's in an arbitrary fashion and use the betting scheme of 3.6.

## 3.4 Summary

Our investigations reveal that while almost any probability distribution for terminal capital can theoretically be achieved through sequential betting strategies, in practice, most terminal distributions are not appealing to gamblers. This disinterest largely stems from the nature of popular gambling systems like Martingale and Progression Systems, which typically offer small gains offset by a significant risk of substantial losses.

Most betting systems fail to optimize gambler's utilities because they do not sufficiently reward the risk of extreme outcomes, leading to inherently suboptimal strategies. The utility functions, if properly utilized to influence betting decisions, should account for the potential gains at the extremes of the distribution. However, the increase at a few terminal points, as suggested by theoretical models, often does not compensate for the overall risk, making these strategies less favorable.

## 4 Random Predictors

### 4.1 Average Scores of Random Predictors as Winnings in Sequential Betting

In 3.1 we have seen that the net winnings  $w(\Theta)$  of a series of  $n$  bets  $b = (b_1, b_2, \dots, b_n)$  on the outcomes of a sequence  $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_n) \in \{0, 1\}^n$  can be expressed as

$$w(\Theta) = \sum_{k=1}^n (b_k \Theta_k - b_k(1 - \Theta_k)) = \sum_{k=1}^n b_k(2\Theta_k - 1)$$

We now consider the case where the bets  $b_k$  are random variables, and the gambler's strategy is to choose the  $b_k$ 's to maximize his expected winnings. The random predictor  $p = (p_1, p_2, \dots, p_n)$  yields an average score

$$\bar{s}(\Theta) = \frac{1}{n} \sum_{k=1}^n (p_k \Theta_k + (1 - p_k)(1 - \Theta_k)) \quad (4.1)$$

The characterization of the set of all achievable scores for all  $p \in P$  is a consequence of the previous section under the correspondence of bets with probabilities given by

$$b_k = \frac{1}{n}(p_k - \frac{1}{2}) \quad \left( \Rightarrow \quad p_k = \frac{1}{2} + nb_k \right) \quad (4.2)$$

The probability constraints  $0 \leq p_k \leq 1$  for all  $k \in [n]$  imply bounded sets

$$|b_k| \leq \frac{1}{2n} \quad (4.3)$$

Thus, if  $b = (b_1, b_2, \dots, b_n)$  yields  $w(\Theta)$ , the corresponding predictor  $p = (p_1, p_2, \dots, p_n)$  yields score

$$\begin{aligned} \bar{s}(\Theta) &= \frac{1}{n} \sum_{k=1}^n (p_k \Theta_k + (1 - p_k)(1 - \Theta_k)) \\ &= \frac{1}{n} \sum_{k=1}^n \left( \left( \frac{1}{2} + nb_k \right) \Theta_k + \left( \frac{1}{2} - nb_k \right) (1 - \Theta_k) \right) \\ &= \frac{1}{2} + \frac{1}{n} \sum_{k=1}^n nb_k (2\Theta_k - 1) = \frac{1}{2} + w(\Theta) \end{aligned} \quad (4.4)$$

**(VI) We deduce from previous sections (3.5) that an average score function  $\bar{s}(\Theta)$  for  $\Theta \in \{0, 1\}^n$  is achievable by a sequential predictor  $p \in P$  if and only if:**

$$\sum_{\Theta} w(\Theta) = 0 \quad \Longleftrightarrow \quad \sum_{\Theta \in \{0,1\}^n} \bar{s}(\Theta) = \frac{1}{2} 2^n \quad \Longleftrightarrow \quad \left( \frac{1}{2} \right)^n \sum_{\Theta} \bar{s}(\Theta) = \frac{1}{2} \quad (4.5)$$

**and**

$$\left| \left( \frac{1}{2} \right)^{n-k+1} \sum_{(\Theta_k, \dots, \Theta_n) \in \{0,1\}^{n-k+1}} \bar{s}(\Theta) (2\Theta_k - 1) \right| \leq b \leq \frac{1}{2n} \quad (4.6)$$

**for all  $\Theta \in \{0, 1\}^n$  and each  $k = 1, 2, \dots, n$ .**

**(VII) A score**

$$\bar{s}(\Theta) = \hat{s} \left( \frac{1}{n} \sum_{k=1}^n \Theta_k \right) \quad (4.7)$$

**depending only on the weight of the sequence  $\Theta$ , is achievable by a sequential predictor  $\hat{p} \in P$  if and only if:**

$$\left( \frac{1}{2} \right)^n \sum_{k=0}^n \binom{n}{k} \hat{s} \left( \frac{k}{n} \right) = \frac{1}{2} \quad (4.8)$$

**and**

$$\left| \hat{s} \left( \frac{k+1}{n} \right) - \hat{s} \left( \frac{k}{n} \right) \right| \leq 2b \leq \frac{1}{n} \quad (4.9)$$

**for  $k = 0, 1, 2, \dots, n-1$ .**

In accordance with the previous notation, let the predictor at stage  $k$  given history  $i$  be denoted by

$$\hat{p}_k(i) = Pr\{\delta_k = 1 \mid \sum_{j=1}^{k-1} \Theta_j = i\} \quad (4.10)$$

for  $i = 1, 2, \dots, k-1$  and  $k = 1, 2, \dots, n$ .

Then, utilizing (3.6), (4.2) and (4.6) (as we did in (3.11))

$$\begin{aligned} \hat{p}_k(i) &= \frac{1}{2} + nb_k(i) \\ &= \frac{1}{2} + n \left(\frac{1}{2}\right)^{n-k+1} \sum_{j=0}^{n-k} \left( \hat{s}\left(\frac{i+j+1}{n}\right) - \hat{s}\left(\frac{i+j}{n}\right) \right) \binom{n-k}{j} \end{aligned}$$

## 4.2 Examples and uniform distribution

### 4.2.1 Example 1:

$\bar{s}(\Theta) = \frac{1}{2}$  for all  $\Theta \in \{0, 1\}^n$  is achievable by the predictor  $p = (\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2})$ .

proof:

$$\begin{aligned} \bar{s}(\Theta) &= \frac{1}{n} \sum_{k=1}^n (p_k \Theta_k + (1 - p_k)(1 - \Theta_k)) \\ &= \frac{1}{n} \sum_{k=1}^n \left( \frac{1}{2} \Theta_k + \frac{1}{2} (1 - \Theta_k) \right) \\ &= \frac{1}{n} \sum_{k=1}^n \frac{1}{2} = \frac{1}{2} \end{aligned}$$

Intuitively, the predictor  $p = (\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2})$  is the most conservative strategy, and for every flip we will predict correctly with probability  $\frac{1}{2}$ .

### 4.2.2 Example 2:

$$\hat{s}\left(\frac{k}{n}\right) = \frac{k}{n}$$

$\eta + (1 - \frac{k}{n})(1 - \eta)$  for  $\eta \in [0, 1]$ , satisfies the achievability constraints and is achieved by the constant predictor  $p = (\eta, \eta, \dots, \eta)$ .

**proof:**

Given the score function

$$\hat{s}\left(\frac{k}{n}\right) = \frac{k}{n},$$

we need to verify the conditions stated in the given constraints for achievability by a sequential predictor.

**Condition 1:**

$$\left(\frac{1}{2}\right)^n \sum_{k=0}^n \binom{n}{k} \hat{s}\left(\frac{k}{n}\right) = \frac{1}{2}$$

This condition simplifies as follows:

$$\left(\frac{1}{2}\right)^n \sum_{k=0}^n \binom{n}{k} \frac{k}{n} = \left(\frac{1}{2}\right)^n \frac{1}{n} \sum_{k=0}^n k \binom{n}{k} = \left(\frac{1}{2}\right)^n \frac{n 2^{n-1}}{n} = \frac{1}{2}$$



which holds true by the binomial theorem where the expected value of  $k$  in a binomial distribution  $B(n, \frac{1}{2})$  is  $\frac{n}{2}$ .

**Condition 2:**

$$\left| \hat{s} \left( \frac{k+1}{n} \right) - \hat{s} \left( \frac{k}{n} \right) \right| = \left| \frac{k+1}{n} - \frac{k}{n} \right| = \frac{1}{n} < 2b \leq \frac{1}{n}$$

This also holds as the absolute difference between successive terms of  $\hat{s}$  is  $\frac{1}{n}$ , which is within the bounds specified by the constraint.

**Construction of Sequential Predictor  $\hat{p}$**  Given  $\hat{s}$ , the predictor  $\hat{p}_k$  for each stage  $k$  based on the history of sums can be designed as:

$$\hat{p}_k(i) = \frac{1}{2} + n \left( \frac{1}{2} \right)^{n-k+1} \sum_{j=0}^{n-k} (\hat{s}(i+j+1) - \hat{s}(i+j)) \binom{n-k}{j}$$

simplifies to  $\hat{p}_k(i) = \frac{1}{2} + n \left( \frac{1}{2} \right)^{n-k+1} \frac{1}{n} \left( \binom{n-k}{0} - \binom{n-k}{n-k} \right) = \frac{1}{2}$  as all terms cancel out except the boundary cases, which simplifies back to  $\frac{1}{2}$ , making it a uniform predictor that always bets  $\frac{1}{2}$  regardless of the history, consistent with the uniform distribution over binary outcomes.

## 5 Conclusions