# Non-asymptotic Design of Finite State Universal Predictors for Individual Sequences

Amir Ingber and Meir Feder
Department of EE-Systems, Tel Aviv University
Tel Aviv 69978, ISRAEL
{ingber ,meir}@eng.tau.ac.il

**Abstract**

In this work we consider the problem of universal prediction of individual sequences where the universal predictor is a deterministic finite state machine, with a fixed, relatively small, number of states. We examine the case of self-information loss, where the predictions are probability assignments which is equivalent to universal data compression. While previous results in that area are asymptotic only, we examine a class of machine structures and find an optimal method for allocating the probabilities to the machine states which achieves minimal redundancy w.r.t. the constant predictors class. We show analytic bounds for the redundancy of machines from that class, and construct machines with redundancy that is arbitrarily close to these bounds. Finally, we compare our machines to previously proposed machines and show that our machine with 300 states achieves smaller redundancy than the best machine known so far with 6000 states.

## I. Introduction

In this work we consider the problem of universal prediction of individual sequences using deterministic finite state machines. We concentrate on the self-information loss function (a.k.a. log-loss), where the predictions made are probability assignments. Combined with an arithmetic encoder/decoder, such a predictor can be used for universal data compression (see, e.g., the method of [3]). A general objective is to find finite state machines constrained to have a fixed number of states $K$, that achieve the minimal worst-case redundancy w.r.t. the best non-universal predictor chosen from a reference class. Example for reference class can be the constant predictors, $L$-order Markov predictors and so on.

In our search for such finite-state universal predictors, we analyze predictors from the classes *counter* and *semi-counter* (which will be formally defined later) where the reference class is the constant predictors class. The results presented here provide methods for specific, non-asymptotic design of finite state machine (FSM) predictors, with specific and concrete number of states, say, hundreds of states. The resulting predictors presented here achieve a smaller redundancy than all known FSM predictors with this number of states.

## II. Definitions and previous work

### A. Basic definitions

Upon observing a binary word $x_1^i$, a universal predictor $u$ comes up with a prediction about the next coming bit $x_{i+1}$. In the log-loss (or self-information) case, the prediction made is the probability $p(\cdot)$ of the next outcome to be $1$ or $0$. The 'penalty' or the codelength

Proceedings of the Data Compression Conference (DCC'06)

associated with this prediction is[1] $\log \frac{1}{p(x_{i+1})}$. The average codelength of the entire sequence, i.e., the average accumulated loss, is given by:

$$\mathcal{L}_u(x_1^n) = \frac{1}{n} \sum_{i=0}^{n-1} \log \frac{1}{p(x_{i+1})}. \tag{1}$$

We are interested in prediction of individual sequences, where the predictor should be good for all of the sequences, without any prior knowledge of the sequences. Proposed universal machines are compared by their excess codelength over the codelength of the best non-universal machine from the reference class. This measure is called the *worst case redundancy*. Specifically, the redundancy of a predictor $u$ with respect to a reference group of predictors $B$ over sequences of length $n$ is:

$$R_n(B,u) \triangleq \max_{x_1^n} \left[ \mathcal{L}_u(x_1^n) - \min_{b \in B} \mathcal{L}_b(x_1^n) \right], \tag{2}$$

and the asymptotic redundancy (or, just the redundancy of $u$ w.r.t. $B$) is

$$R(B,u) \triangleq \lim_{n\to\infty} R_n(B,u). \tag{3}$$

Now, consider the case where the reference class is the set of all constant (non-universal) predictors, i.e. predictors that give a constant prediction for the probability of the next bit to be $1$. We denote that set by $B_{const}$. In this case it is well known that the best constant probability is the *empirical probability* of the sequence. The resulting reference codelength is given by the empirical entropy:

$$\min_{b \in B_{const}} \mathcal{L}_b(x_1^n) = h_{emp}(x_1^n) = h\left(\frac{n_1(x_1^n)}{n}\right), \tag{4}$$

where $h(\cdot)$ denotes the binary entropy function and $n_1(x_1^n)$ is the number of ones in $x_1^n$. Therefore, the redundancy w.r.t. the constant predictors reference class is given by:

$$R_n(u) \triangleq \max_{x_1^n} \left[ \mathcal{L}_u(x_1^n) - h_{emp}(x_1^n) \right]. \tag{5}$$

### *B. Known results - universal predictors with unlimited resources*

The optimal universal predictor, or universal probability assignment scheme, w.r.t a reference class of source coders, was proposed by Shtarkov [6]. The assigned probability for the entire sequence is the normalized maximum-likelihood solution, over the reference class. This solution, while universal, is not completely sequential since it requires to know the sequence length $n$ in advance.

An alternative, fully sequential machine, that achieves the same asymptotic redundancy, is based on the Krichevsky-Trofimov [1] probability assignment (which is an extension of the well known Laplace assignment):

$$p_{i+1} = \frac{n_1(x_1^i) + 1/2}{i+1}. \tag{6}$$

Clearly, both solutions require infinite memory for the probability assignment.

[1]Unless specified otherwise, the base of the logarithm is 2.

### C. Known results - deterministic finite-state universal predictors

This problem was first addressed by Rajwan [4] and Rajwan and Feder [5]. An important observation made therein is the relation between the worst-case redundancy and the minimal circles:

*Definition 1:* A *circle* of a FSM is a closed set of states $\{S_i\}_{i=0}^{L-1}$ and input bits $\{b_i\}_{i=1}^{L}$, that rotate the machine between the states. A *minimal circle* is a circle that does not contain the same state more than once.

*Theorem 1:* The worst data sequence for a given FSM, is the one that takes the machine to its minimal circle with highest redundancy, and rotates in it endlessly.

The proof is given in detail in [4]. This theorem specifies the worst sequence for a FSM predictor, and calculates the worst-case redundancy of that machine. As will be shown later in the paper, this theorem will help in designing machines with low redundancy.

In [4] there is also a description of a finite state machine that achieves an asymptotic (in the number of states $K$) redundancy w.r.t. the constant predictors class of $\Theta\left(\frac{\log(K)}{\sqrt{K}}\right)$. A better machine, that achieves a redundancy of $O\left(\frac{1}{K^{2/3}}\right)$ was proposed in [2]. Another important result in [2] is a lower bound of $\Omega\left(\frac{1}{K^{4/5}}\right)$ on the redundancy of any FSM predictor.

Essentially, then, all existing previous work provided asymptotic results. There is no specific answer to the question 'what is the best achievable redundancy for a universal coder with $K$ states'. While this general question remains open, in this work we substantially improve the best known achievable redundancy for values of $K$ up to about 6000 states.

## III. Designing FSM predictors with minimal redundancy

First we note that the FSM design task can be divided into two separate tasks: design the machine's structure (the transitions between states), and find the probabilities associated with the states. We start by searching for the optimal probability set for a given structure, and by that, we find the optimal redundancy that can be achieved by a machine with that structure.

### A. The saturated counter

The first FSM we study is the saturated counter. This machine is defined by a linear array of $K$ states, with each state changing to the next state on its right on input $1$, and left on input $0$ (See Fig. 1).

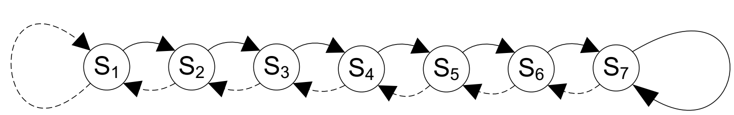


Fig. 1. Saturated counter FSM predictor. Solid and dashed lines correspond to '1' and '0' transitions, respectively.

The question to be asked, is what probabilities are best to place on the states in order to achieve the minimal redundancy.

Recalling Theorem 1, instead of searching over all sequences of infinite length for the one with the maximal redundancy, it is sufficient to check the redundancies of all the minimal circles in the FSM, and take the maximal redundancy over these circles. Therefore, we need to assign probabilities that result in a minimal maximal circle redundancy.

The minimal circles for the saturated counter are numbered from left $0$ to $K$, and are given below with their respective redundancies:

$$\begin{array}{lll} Sequence & States & Redundancy \\ 0; & S_1; & R_0 = \log \frac{1}{1-p_1}, \\ 10; & S_i, S_{i+1}; & R_i = \frac{1}{2}\left(\log \frac{1}{p_i} + \log \frac{1}{1-p_{i+1}}\right) - 1, \\ 1; & S_K; & R_K = \log \frac{1}{p_K}. \end{array} \tag{7}$$

The probability set that minimizes the maximal redundancy, is the one that brings the redundancies to equality (this fact is proved later on, for a more general set of FSM predictors).

*Theorem 2:* For all the predictors of the saturated counter form, the infimum for the redundancy is given by:

$$\mathcal{R}_{SaturatedCounter} = \log_2 \frac{5}{4} \approx 0.3219 bit. \tag{8}$$

*Proof:* We start by showing that no FSM predictor of a saturated counter structure can achieve a redundancy smaller than $\log \frac{5}{4}$. Then we show how to construct machines that achieve redundancy that is arbitrarily close to that lower bound.

To show the lower bound, consider the optimal FSM predictor of a saturated counter structure with a set of probabilities $\{p_1, p_2, ...p_K\}$. Now assume, for contradiction, that its redundancy is less than $\log_2 \frac{5}{4}$. As mentioned before, the minimal redundancy is attained when the redundancies are all equal, i.e. $R_0 = R_1 = ... = R_K = R$. From the equation $R_0 = R_1$, it can be shown that $p_2 > p_1$ if and only if $p_1 > \frac{1}{5}$. By assumption, $R < \log \frac{5}{4}$, therefore

$$R_0 = \log \frac{1}{1-p_1} < \log \frac{5}{4}, \tag{9}$$

which leads to $p_1 < \frac{1}{5}$, and eventually to $p_2 - p_1 < 0$, and $p_2 < \frac{1}{5}$. Since all the probability pairs $p_i, p_{i+1}$ for $i < K$ has a similar relation such as $p_1, p_2$, then if $p_1 < \frac{1}{5}$, all the probabilities are smaller than $\frac{1}{5}$, including $p_{K-1}$. Since $R_K < \log \frac{5}{4}$, we get that $p_K > \frac{4}{5}$ - a contradiction.

In order to complete the proof of the theorem, it is necessary to show a FSM predictor whose redundancy is arbitrarily close to $\log \frac{5}{4}$. Let $R^*$ be the desired redundancy. Since $R^* > \log \frac{5}{4}$, we denote $\Delta R \triangleq R^* - \log \frac{5}{4}$. For convenience, we select $K$, the number of states, to be odd. We denote by $M$ the state that is exactly in the middle, i.e. $M = \frac{K+1}{2}$. The number of states itself will be determined later on. By symmetry, we select $p_M = \frac{1}{2}$. We define the probabilities $p_i$ for $i < M$, and the probabilities on the axis half closer to 1 are defined symmetrically by $p_i = 1 - p_{K+1-i}$.

We start constructing the machine from $p_M$, and define the probabilities $p_{M-1}, p_{M-2}$ etc, each time setting the lowest possible probability induced by the desired redundancy. We recall from Theorem 1, that in order to have redundancy $R^*$, we only need to verify that the redundancy over the minimal circles is less than or equal to $R^*$. We set the probabilities, s.t. the redundancy is equal to $R^*$:

$$p_i = \frac{2^{-2(R^*+1)}}{1-p_{i+1}} \triangleq f(p_{i+1}). \tag{10}$$

This function is applied consecutively, each time forming a new state with reduced probability. We can stop generating new states when the probability is low enough to be a terminal probability $p_1$, i.e., to satisfy the redundancy condition $p_1 < p^* = 1 - \frac{4}{5}2^{-\Delta R}$ (Note that since $\Delta R > 0$, $p^* > \frac{1}{5}$).

The iteration process indeed converges to a low enough probability. Although the iteration process can go on forever, the probabilities themselves have a lower bound which is the fixed point of the function $f(\cdot)$. It can be shown that $f(\cdot)$ has a fixed point $p_{FP}$, and $p_{FP} < \frac{1}{5}$.

As $f(\cdot)$ is applied over and over again, $p_i$ approaches $p_{FP}$. Since $p_{FP} < \frac{1}{5} < p^*$, there exists a number $M^*$, s.t. in the iteration process, staring from $p_{M^*}$, $p_1$ will be low enough, i.e. $p_1 \le p^*$ and $R_1 \le R^*$. It is obvious that as the desired redundancy approaches its lower bound, the required number of states is higher. Moreover, it can be shown that the excess redundancy over the lower bound vanishes exponentially with $M$.

Since all the circles (except the ones in the edges) in the FSM have, by construction, redundancy equal to $R^*$, and the circle in the leftmost edge (and, symmetrically, in the rightmost edge) has redundancy less than $R^*$, the described FSM predictor has a maximal redundancy of $R^*$. Since $R^*$ was arbitrarily close to $\log\frac{5}{4}$, this completes the proof. ■

### B. The semi-counter

A more general structure of FSM predictors is the Semi-Counter:

*Definition 2:* A FSM predictor is calld a *Semi-Counter with parameter q* if it has the following form:

- A symmetric array of $K$ states, $\{S_1, ..., S_K\}$.
- For the lower half of the states, i.e. $S_i, i < M = \frac{K+1}{2}$, the transitions are as follows:

$$\begin{array}{l} \gamma(S_i, 0) = S_{i-1}, 1 < i \le M; \\ \gamma(S_1, 0) = S_1; \\ \gamma(S_i, 1) = S_{i+\delta_i}, i < M. \end{array} \tag{11}$$

  where $\gamma(\cdot,\cdot)$ is the transition function and $\delta_i$ is any non-increasing sequence of transition lengths, with $\delta_{M-1} = 1$. The parameter $q$ is defined by the longest transition in the FSM, i.e. $q = \delta_1$.
- For the upper half of the states, the transitions are defined symmetrically.

We denote the set of all Semi-Counter FSM predictors with parameter $q$ by $B_q$. An example for a semi-counter is shown in Figure 2.

For a given structure of a semi-counter FSM predictor, we seek the best probability set that brings the worst-case redundancy to a minimum. As in the saturated counter case, the minimal redundancy is achieved when the redundancies over all the minimal circles are equal, as stated by the following theorem:

*Theorem 3:* Given a semi-counter machine structure, the set of probabilities that brings the worst-case redundancy to a minimum, is the one that brings the redundancies of all the minimal circles to **equality**.

*Proof:* First, we recall that the maximum redundancy over all the sequences is equal to the maximum redundancy over the *minimal circles* of the FSM (Theorem 1). For a FSM

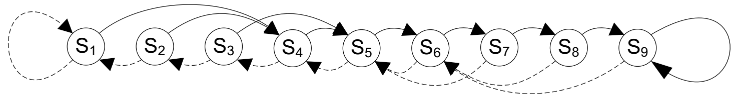


Fig. 2. A semi-counter FSM predictor, with $q = 3$. Solid and dashed lines correspond to '1' and '0' transitions, respectively.

predictor of a semi-counter type, for each state $S_i$, $1 < i < M$, there is only one minimal circle s.t. that its minimal state (in the sense that it has the minimal index) is $i$. This follows from the fact that all the minimal circles contain only one '1' transition. For $i = 1$, there exist two circles: the circle that includes only the state $S_1$, (corresponding to the string $'0'$), and the circle corresponding to the string $1 \cdot 0^{\delta_i}$. We denote the redundancy of the leftmost circle by $R_0$, and the redundancy of the rest of the circles by $R_i$, where $i$ is the minimal state in each circle. The redundancy of the minimal circles is given by:

$$R_0 = \log \frac{1}{1-p_1}, \tag{12}$$

$$R_i = \frac{1}{1+\delta_i}\left[\log\frac{1}{p_i} + \sum_{j=1}^{\delta_i} \log\frac{1}{1-p_{i+j}}\right] - h\left(\frac{1}{1+\delta_i}\right),\ i > 0. \tag{13}$$

All the above redundancies are functions of the probability vector $\mathbf{p} = [p_1, ..., p_K]$. We denote the maximum of these redundancies by $R_{max} = \max_i R_i$. Note that $R_{max}$ is a continuous convex function of the vector $\mathbf{p}$. At the edges of its domain $(0,1)^K$, its value diverges to infinity, therefore the minimum is achieved in the interior domain. The following lemma characterizes the minimum points:

*Lemma 1:* If, for a given probability vector $\mathbf{p} = [p_1, ..., p_K]$, the redundancies $R_0, R_1, ..., R_{M-1}$ are not equal, then $\mathbf{p}$ is not the optimal probability vector, i.e. there exist a vector $\mathbf{p}'$, that achieves a lower maximal redundancy.

*Proof:* Let us assume, for contradiction, that there exists a semi-counter FSM predictor that attains the optimal redundancy where the redundancies over the minimal circles are not equal. We will show, that this machine can be improved, and thus not optimal.

The improvement is done in two steps:

1. Starting from $p_{M-1}$, change the probabilities to the smallest possible value, considering the minimal circle starting at this state. This process ensures that the redundancies $R_1, ..., R_{M-1}$ are equal. The redundancy $R_0$ now must be lower than $R_{max}$.

2. Start in $p_0$. Increase it by a nonzero amount, s.t. $R_0 < R_{max}$ still holds. Now, since $R_1 < R_{max}$, increase the probabilities that are members of that circle, i.e. $p_2, ..., p_{1+\delta_1}$, to decrease $R_2, ..., R_{1+\delta_1}$. Now, repeat the above process with $p_{1+\delta_1}$. Since is was decreased, $R_{1+\delta_1} < R_{max}$, and it is possible to increase $p_j$, where $1 + \delta_1 < j \leq 1 + \delta_1 + \delta_{1+\delta_1}$, in order to decrease the corresponding redundancies. Repeating this process in a similar fashion will reduce *all* the redundancies, and by that achieving a new FSM predictor with modified probabilities, that has a maximal redundancy (over the minimal circles) which is strictly smaller than $R_{max}$.

As shown in the above process, any semi-counter machine with unequal redundancies can be improved to get a machine with better maximal redundancy, thus proving the lemma. ■

We now complete the proof of the Theorem. From above, the remaining vectors in $(0,1)^K$ that may achieve the global minimum, are those that result in equality between the redundancies over the minimal circles. It can be shown, that only a single point satisfies that, i.e., there is only a single probability vector **p** that equates the redundancies, and therefore it is the only vector (probability set) that minimizes $R_{max}$. ■

After characterizing the structure of the optimal semi-counter FSM predictors, we now explore their performance, i.e. the redundancy that is achievable by such machines. As was shown above (Theorem 2), the best redundancy of a saturated counter, which is, in fact, a semi-counter with $q = 1$, cannot be lower than $\log\left(\frac{5}{4}\right)$. This result can be viewed as a corollary of the following more general theorem, that relates the worst-case redundancy to the empirical probability of the minimal circles:

*Theorem 4:* If $c_1$ and $c_2$ are two minimal circles with empirical probabilities $r_1$ and $r_2$ respectively, and there are no minimal circles with empirical probability $r$ s.t. $r_1 < r < r_2$, then the worst case redundancy of the machine is bounded from below:

$$R \geq D(r_1||MP(r_1, r_2)), \tag{14}$$

where $D(\cdot||\cdot)$ is the Kullback-Leibler Distance, and $MP(r_1, r_2)$ is defined by

$$MP(r_1, r_2) = \frac{1}{1 + 2^{\frac{h(r_2)-h(r_1)}{r_2-r_1}}}. \tag{15}$$

*Proof:* We define a *Threshold Sequence*, as in [2]: The sequence is machine dependent, parameterized by $x$ and denoted $TS(x)$. The sequence traverses the FSM as follows: if the probability of the current state is lower than $x$, then the next symbol is '1', otherwise - '0'. This sequence must cause the machine to circle in a cycle of states. We denote the length of this circle with $L$, and the number of ones and zeroes in it by $L_1$ and $L_0$ respectively. It can be shown that the redundancy over $TS(x)$ satisfies

$$R \geq \frac{1}{L}\left[\sum_{i=1}^{L_1} \log\frac{1}{x} + \sum_{i=1}^{L_0} \log\frac{1}{1-x}\right] - h\left(\frac{L_1}{L}\right) = D\left(\frac{L_1}{L}||x\right). \tag{16}$$

Since there are no minimal circles in the FSM that have empirical probability between $r_1$ and $r_2$, the empirical probability $\frac{L_1}{L}$ must satisfy either $\frac{L_1}{L} \leq r_1$ or $\frac{L_1}{L} \geq r_2$. Selecting a value of $x$ s.t. $r_1 < x < r_2$ yields

$$D\left(\frac{L_1}{L}||x\right) > \min\{D(r_0||x), D(r_1||x)\}. \tag{17}$$

The lower bound is maximized by the value that equates the divergences, i.e. the value of $x$ that solves $D(r_0||x) = D(r_1||x)$. The solution is given by $MP(r_1, r_2)$ (equation (15)). Follow equations (16) and (17) to complete the proof. ■

*Corollary 1:* The redundancy of a saturated counter, regardless of its number of states, is bounded from below by

$$D\left(0||MP\left(0, \frac{1}{2}\right)\right) = \log_2\frac{5}{4} \approx 0.3219 bit. \tag{18}$$

*Proof:* Immediate from Theorem 4 and from the fact that the saturated counter has minimal circles with empirical probabilities $0$ and $1/2$ only. This is in fact an alternative proof for the first part of Theorem 2. ■

*Corollary 2:* The redundancy of any semi-counter with parameter $q$, regardless of its number of states, is bounded from below by

$$D\left(0||MP\left(0,\frac{1}{q}\right)\right) = \log_2\left[1+\frac{q^q}{(q+1)^{q+1}}\right] \approx \frac{1}{e\cdot\ln 2}\frac{1}{q}. \qquad (19)$$

*Proof:* Immediate from Theorem 4 and from the fact that the semi-counter with parameter $q$ does not have circles with empirical probabilities between $0$ and $1/q$. ■

It appears that this bound is achieved for small values of $q$. One specific example of a machine that was constructed, is described as follows (only the lower half of the states): The machine has $q$ states with transition length $1$, $q$ states with transition length $2$, $q$ with transition length $3$ and so on, until $q$ states with transition length $q-1$. All the remaining states have transition length of $q$. As the number of states grows, the redundancy approaches the lower bound of corollary 2 (See Figure 3).

The design of semi-counter FSM predictors can be viewed as 'bridging' the gap between the minimal probability, $p_1$, and $p_M = 1/2$. Connecting these two probabilities is done by selecting the appropriate transition length for each state, or, equivalently, at which states the transition length increases. This approach yields the following greedy algorithm for designing semi-counter FSM predictors:

Let $R^*$ be the desired redundancy. Calculate $p^*$, the minimal required probability for the FSM, i.e. the one that solves $R^* = \log\frac{1}{1-p^*}$.

Set the middle probability $p_M = 1/2$, the current transition length $t = 1$, and the current next probability $i = M-1$. Then repeat, until $p_i \le p^*$:

- evaluate the probability $p_i$, in the two following cases:
  1) The transition length is $t$, or
  2) The transition length is $t+1$,

  Select the one that minimizes $p_i$.
- Set $t$ to be the selected transition length, and decrease $i$, for the next iteration.

At the end of the above loop, the current probability, $p_i$ is low enough to be the minimal probability in the machine, i.e. the last formed circle when $p_i$ is the minimal probability has a redundancy that is lower than the desired redundancy $R^*$. This probability is named $p_1$, and thus the value of $M$ is determined. The resulting machine has maximal redundancy $R^*$, and a low number of states.

This algorithm achieved very good results (see Figure 3), better than all the examined machines above. However, it has one important flaw - the redundancy it achieved never goes under a certain level, for any number of states $K$. In fact, this flaw is intrinsic for all the semi-counters:

*Corollary 3:* The redundancy of any semi-counter, regardless of its parameter and number of states, is bounded from below by

$$D\left(\frac{1}{3}||MP\left(\frac{1}{3},\frac{1}{2}\right)\right) = \log_2\frac{1753}{1728} \approx 0.0207bit. \qquad (20)$$

*Proof:* Immediate from Theorem 4 and from the fact that any semi-counter does not have minimal circles with empirical probabilities between $1/3$ and $1/2$. ■

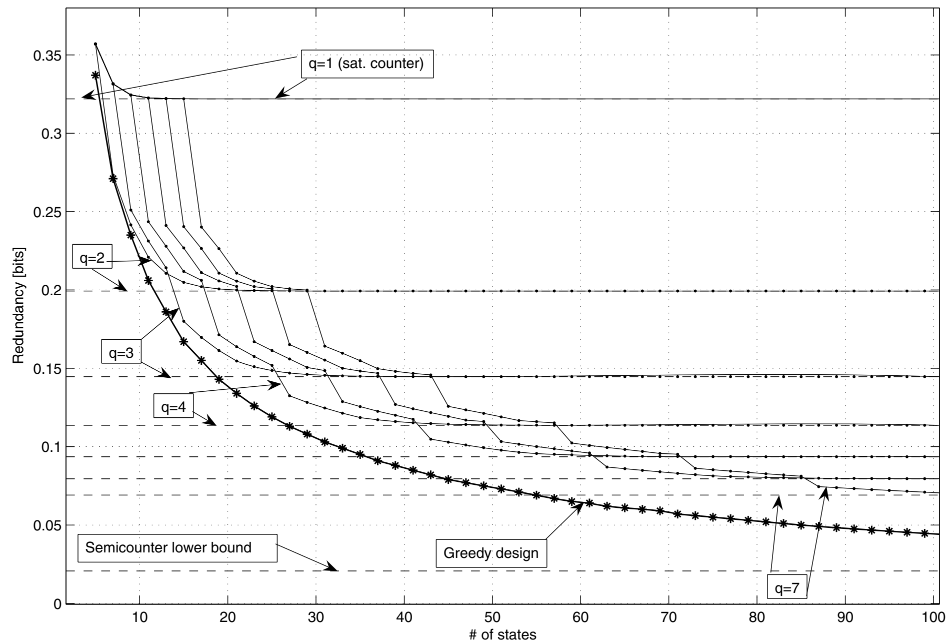


Fig. 3. The redundancy of the 'Back-q' machines versus the number of states. Note the lower bound for each value of $q$ and the convergence of the redundancy towards it. The best semi-counter (designed with the greedy alg.) is also shown, with the general bound for the semi-counter redundancy.

The greedy algorithm cannot achieve redundancy that is lower than the lower bound in equation (20), but practically achieves it within a few hundreds of states. This is the lowest redundancy that has been shown for this number of states, as seen in Figure 4. The best machine known so far (the Exponentially Decaying Memory machine [2]) achieves this redundancy at approx. $6000$ states, compared to only $300$ states in the semi-counter. The finite-state Krichevsky-Trofimov machine [5] achieves this redundancy at over $25,000$ states.

## IV. FURTHER RESEARCH

Every FSM predictor has a set of minimal circles with corresponding (rational) empirical probabilities. A general lower bound on the redundancy for such a machine is the maximum of all the lower bounds (Theorem 4) on all the adjacent pairs of empirical probabilities in the machine. Therefore, in order to achieve a vanishing redundancy, the density of the rational probabilities must increase.

The semi-counter structure leads to many 'gaps' that cannot be eliminated by using more states. The widest 'gap' in the sense of this lower bound is between $1/3$ and $1/2$. Therefore, in order to achieve low redundancy for a machine based on the semi-counter, adding a probability between $1/3$ and $1/2$ is needed. A natural candidate seems to be $2/5$ as it is the rational number with the lowest denominator in this range. Circles with low denominator are desirable since they require a small number of states. Unfortunately, the machine itself results in minimal circles that have a more complex form, such that

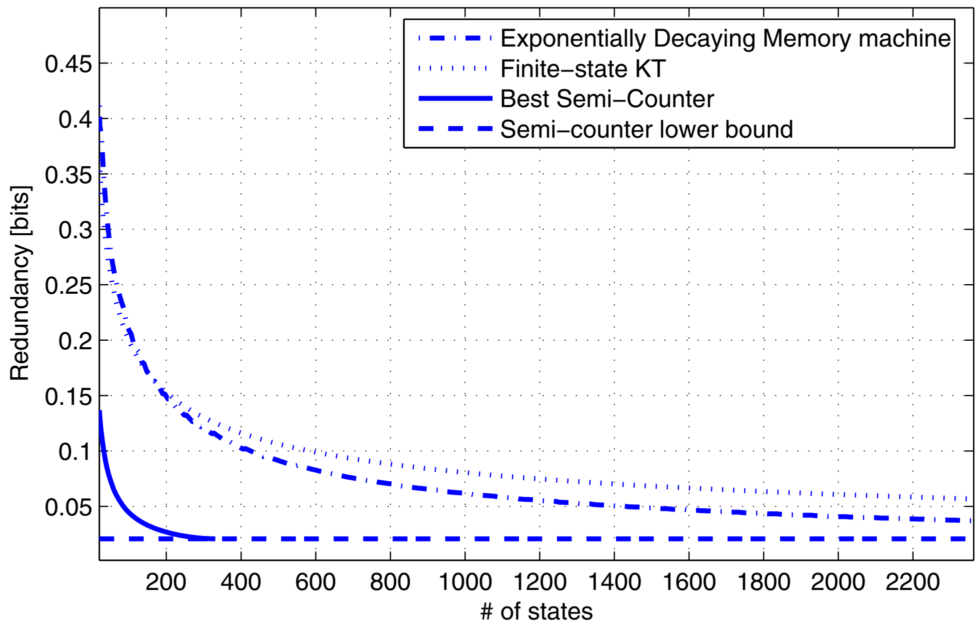


Fig. 4. Redundancy versus the number of states. Although the best semi-counter cannot achieve redundancy below $0.02bit$, its performance is superior to that of the other known predictors for number of states up to about 6000.

the principles of Theorem 3 no longer apply. Therefore, it is needed to find the optimal probability set to assign to each state.

A general method for selecting the next probability to be added to the machine is needed. It can be shown, that the rational with the lowest denominator between $a/b$ and $c/d$ is $(a+c)/(b+d)$. Repeating this process yields a non-uniform 'quantizer' that is arbitrarily dense, and hopefully circles that are not too long.

A method of connecting two regions with different empirical probabilities, e.g. connection circles with probability $5/14$ to circles with probability $4/11$, is also needed. Although for small numbers such a 'sewing' was found, a general method is still elusive.

We conjecture that this construction will lead to efficient machines, hopefully machines with the best possible redundancy for a given number of states. For this we have to specify the explicit construction ("sewing") and prove that indeed these machines have optimal performance. All these topics are now under investigation.

## References


[1] R. E. Kriechevski and V. E. Trofimov. The performance of universal encoding. IT-27:199–207, March 1981.

[2] Eado Meron and Meir Feder. Finite-memory universal prediction of individual sequences. *IEEE Transactions on Information Theory*, 50(7):1506–1523, 2004.

[3] W. B. Pennebaker and J. L. Mitchell. Probability estimation for the q-coder. *IBM J. Res. Dev.*, 32(6):737–752, 1988.

[4] D. Rajwan. Universal finite memory coding of binary sequences. Master's thesis, Tel Aviv University, 2000.

[5] Doron Rajwan and Meir Feder. Universal finite memory machines for coding binary sequences. In *Data Compression Conference*, pages 113–122, 2000.

[6] Y. M. Shtar'kov. Universal Sequential Coding of Single Messages. *Problemy Pered. Inform. (Problems of Inform. Trans.)*, 24:175–186, 1988.



Proceedings of the Data Compression Conference (DCC'06)
0-7695-2545-8 /06 $20.00 © 2006 IEEE