On interim rationality, belief formation and learning in decision problems with bounded memory

Pathikrit Basu† Kalyan Chatterjee‡
July 11, 2015

Abstract

We consider decision problems where the agent has bounded memory and makes decisions using a finite state automaton. Under a natural notion of belief consistency and in the case of continuous signals we show in an investment problem, there exist optimal automata which fail to be modified multiself consistent. For any automaton, we show there exist performance equivalent automata with significantly lesser amount of randomization. We introduce a new criterion for automaton performance called weak admissibility and demonstrate if memory states are not wasted, the nature of transitions are uniquely defined by the criterion. We show for any optimal automaton, there exists a performance equivalent automaton that is weakly admissible and involves no wastage of memory states. We describe more general environments (POMDP) and establish a revelation principle for optimality. Under a changing worlds environment, we discuss conditions under which unconstrained optimum can be attained using finite state automata. In these environments, we demonstrate characteristics of the constrained optimal automaton.

1 Introduction

In standard state space models of uncertainty, rationality assumes that decision makers apply Bayes’ rule to update beliefs and undertake actions that are optimal under the

†Preliminary and Incomplete
‡Pennsylvania State University. E-mail - pub143@psu.edu
§Pennsylvania State University. E-mail - kchatterjee@psu.edu
beliefs held. An assumption implicit in such models is that agents are able to process information with unbounded capacity. However, in reality, available information may be too sophisticated to process completely and a simplification or an endogenous coarsening of it is performed to keep the act of decision making simple. Agents have bounds to understanding and assimilating new information and given these bounds, contemplate optimal actions. A natural way to incorporate this feature in standard models would be to impose the bound on the agent first and then study constrained rational behaviour. The current paper is concerned precisely with this line of study. We consider standard dynamic environments where the agent’s objective is to learn an unobservable payoff relevant parameter based on information (signals) that arrives at every point in time. To impose the constraint on processing ability, we assume that they use a finite state automaton to perform inference and take actions. We study optimal behaviour under this framework and moreover, the characteristics of the optimal decision making process namely the nature of decision making and belief formation

In dynamic environments, one important issue concerning optimal plans is that of time consistency. In the present model, the agent using an automaton finds himself in a decision problem with imperfect recall. Piccione and Rubinstein (1997)[6] point out that in such problems, optimal plans may not be time consistent but satisfy a weaker notion of time consistency that they call modified multiself consistency. It requires that agents not have profitable one shot deviations from the chosen strategy at any information set. If we interpret the problem as a game with information sets treated as separate players with a common interest, modified multiself consistency is equivalent to an equilibrium of the game. In a framework with agents using a finite state automaton as in the present context, Wilson (2014)[10] establishes a natural analog of it in a model of simple hypothesis testing and Kocer (2010)[11] establishes in more general environments concerning partially observable markov decision processes. We analyse a single person investment problem and find this result to be sensitive to the underlying notion of belief consistency and that in addition, under continuous signals, the result no longer holds. Moreover, as we shall see, belief consistency is hard to define. The difficulty of defining an objective notion of belief consistency here stems from the problem of defining probabilities in an information set with infinitely many histories that may not be mutually exclusive. The model in Piccione and Rubinstein (1997) consider a finite horizon and is hence free from the problem. Wilson (2014) introduces a termination chance and Kocer (2010) interprets the discount factor as a continuation probability to avoid it. In the model studied here, we argue that there may be natural ways to define belief consistency which preclude modified multiself consistency
of the optimal plan.

Within the environment of the investment problem, we introduce a new criterion for performance of an automaton that we call "weak admissibility". The criterion requires that at any memory state reached with positive probability, committing to transition probabilities in one payoff relevant state of the world (hypothesis), the agent under the constraint created by the commitment, necessarily chooses transitions optimally for the other state of the world. If additionally, states are not wasted (in the sense that for any two memory states reached with positive probability, the continuation values differ and neither pair of continuation values dominates the other in terms of weak dominance), then the weak admissibility condition implies that the transitions are uniquely defined and exhibit a simple "stair-case" structure. This is achieved by interpreting the weak admissibility condition as a linear program, which can be uniquely solved using a proposed algorithm which assigns transition probabilities to memory states by ranking the product of continuation values of memory states and likelihood ratios of signals. Additionally, with regard to optimality, it is shown that for any optimal automaton, there exists an equivalent automaton with the same ex-ante value that is weakly admissible and does not waste memory states.

Reinterpreting the discount factor as a stopping probability as in Kocer(2010), we establish a revelation principle for memory states reached with positive probability which states that at any memory state reached with positive probability, it is indeed optimal to stay there before any further information is received via signals under the beliefs held about the state of the world. In a generalised (compared to Wilson (2014)) investment problem with changing worlds (the state of the world evolves non-interactively according to a markov chain), this result proves to be very useful in studying characteristics of the optimal automaton and in particular the nature of transitions taking place in it. In particular, it garantess categorisation of beliefs in the sense that posterior beliefs lie in one of finitely many categories indexed by memory states of the automaton. Each category is exactly the set of beliefs for each its corresponding memory state is optimal. The optimal transitions can then be shown to follow simple rules depending on the signal. With each signal we associate a belief that is the fixed point of an appropriate map and transitions take place depending on whether the current belief is placed left or right to the fixed point. We also demonstrate conditions under which the unconstrained optimal decision rule can be implemented by a deterministic finite state automaton. As in Monte and Said (2014), this demonstrates the lack of need for any sophistication in the decision rule in special environments. However, unlike their model, the model studied here uses discounting,
allows for asymmetry, a more general signalling structure and admits an unconstrained optimal memory of more than 2 states. As an example, we also demonstrate that a simple variant (investment problem) of the framework in Shiryaev’s change point detection problem (See Shiryaev (1978)[8]) allows for optimal unconstrained finite memory under simple conditions. The belief process without bounds on information processing follows an ergodic markov process and admits a unique stationary distribution. Lastly, on a separate but relevant note, we show that any automaton can be modified to one which gives the same performance but significantly reduces the extent of randomization involved. We derive an upper bound on randomisation effectively required and show that if the signal space is large enough, the automaton is “close” to a deterministic one. In the context of the investment problem with unchanging worlds, it is shown that requiring weak admissibility tightens the upper bound and further reduces the extent of randomisation involved.

**Related Literature:**

The strand of literature most related to our work pertain to the issue of learning with a finite state automaton. It’s starting point is marked by a seminal paper due to Hellman and Cover (1970)[4] which solves the simple hypothesis problem with finite memory. In the field of economics, as stated above, Wilson(2014), Kocer (2010), Monte and Said (2014) study related problems. Additionally, Kocer(2014) studies a bandit problem with finite memory and Monte (2007) studies a repeated sender receiver game with the receiver using a finite state automaton. Piccione and Rubinstein(1997), Aumann Hart and Perry (1997) and Halpern(1997) discuss the notion of belief consistency in decision problems with imperfect recall. Shiryaev (1978) solves the quick point change detection problem but without memory constraints. Kalai and Solan (2003) discuss the need for randomisation in simple plans (automata) and establish that in non-interactive environments, randomisation in the decisions is not needed but may be required in the transitions of the optimal automaton.

Dow (1996) studies a single person problem of a consumer searching for a low price but has limited memory to remember prices observed. Recent literature on coarse information and categorisation include Al-Najjar and Pai (2013), Peski (2011) and Jackson and Fryer (2005). However, these models are all based in static environments and differ in terms contexts and approaches consider within them.

2 Investment Problem

Every period, i.i.d payoffs are realised from an unknown distribution $\theta \in \{0, 1\}$. The sample space is $X \subset \mathbb{R}$. Under $\theta = 0$, payoffs are realised according to distribution $F_0$. Under $\theta = 1$ they are realised according to $F_1$. The distributions are assumed to have density $f_0$, $f_1$ (viewed as Radon-Nikodym derivatives according to a measure $\varphi$, this covers the cases of discrete and continuous signals both$^1$). We shall assume for convenience that the pair of distributions satisfies the monotone likelihood ratio property i.e $\frac{f_0(x)}{f_1(x)}$ is strictly increasing in $x^2$. It is assumed that the expectations for the two distributions satisfy $E_0(x) > 0$ and $E_1(x) < 0$. We shall write them as $E_0$ and $E_1$. We can interpret $f_0$ as a good distribution yielding a positive payoff on average. Similarly, we interpret $f_1$ as a bad distribution. Furthermore, we shall refer to $\theta = 0$ as the good state of the world and $\theta = 1$ as the bad state of the world. Lastly, let $\pi \in (0, 1)$ be the prior probability that the true distribution is $\theta = 0$.

There is a single decision maker to whom $\theta$ is unknown. Every period, before the payoffs are realised, he chooses (based on available information) one of two actions in the set $\{I, NI\}$. If $I$ is chosen and $x$ is realised, the decision maker gets a payoff $x$ in that period. If $NI$ is chosen, then he gets 0. Action $I$ is interpreted as decision to invest and $NI$ as a decision to not invest. If the actual realisation is $(x_t)_t \in X^\infty$ and the decision path along $(x_t)_t$ is $(d_t)_t \in \{I, NI\}^\infty$, then the decision maker gets payoff:

$$\sum_{t=0}^\infty (1 - \delta) \delta^t x_t \mathbb{I}_{[d_t = I]}$$

Where $\mathbb{I}$ is the indicator function which takes value 1 when $d_t = I$ and 0 when $d_t = NI$. The decision maker wishes to choose a decision plan to maximise discounted expected utility. Note that in doing so, it will be relevant for the decision maker to make inferences about

$^1$Consider absolute continuity with respect to the counting measure for the discrete case and the uniform measure for the continuous case

$^2$This assumption is introduced for simplicity of exposition and is not crucial for obtaining results
θ using the payoff realisations \( x \) as signals. In this note, we shall consider decision makers who use finite state automata to makes decisions and inferences. We shall be interested in the nature of the optimal automaton and also whether it is modified multiself consistent.

### 2.1 Finite state automaton

An automaton is defined as a tuple \( \tau = < M, X, \sigma, d, i_0 > \) where \( M = \{1, ..., m\} \) is a finite set of memory states, \( \sigma : M \times X \rightarrow \Delta(M) \) is a transition function, \( d : M \rightarrow \{I, NI\} \) is a decision function and \( i_0 \) is the initial state. The interpretation of the automaton is as follows: At a given period, the agent at state \( i \) decides whether to invest or not invest determined by \( d(i) \). If he observes a payoff realisation \( x \), he updates the state randomly accordingly to \( \sigma(i, x) \) and depending on the new state decides whether or not to invest according to \( d \). Denote the ex-ante expected value from the automaton \( \tau \) to be \( V(\tau) \). This can be derived as follows:

- The transition function \( \sigma : M \times X \rightarrow \Delta(M) \) defines two transition matrices on the state space \( M \), \( (\alpha_{ij}), (\beta_{ij}) \) as

\[
\alpha_{ij} = \int_X \Pr(f(i, x) = j)f_0(x)d\varphi(x)
\]

and

\[
\beta_{ij} = \int_X \Pr(f(i, x) = j)f_1(x)d\varphi(x)
\]

Let their stationary distributions be denoted as \( \mu_0, \mu_1 \).

- Derive the values \( V_{\theta 0}, V_{\theta 1} \). These are the long run expected payoffs from state \( i \) when \( \theta = 0 \) and \( \theta = 1 \). Then for \( i \in M \), we have:

\[
V_{\theta 0} = E_0.I_{[d(i)=1]} + \delta \sum_{j \in M} \alpha_{ij}V_{\theta j}
\]

\[
V_{\theta 1} = E_1.I_{[d(i)=1]} + \delta \sum_{j \in M} \beta_{ij}V_{\theta j}
\]

- \( V(\tau) = \pi V_{i0} + (1 - \pi)V_{i1} \)

An optimal automaton is one which solves the maximisation problem:

\[
\max_{\tau} V(\tau)
\]
Notice that we focus on automata with a deterministic initial state and decision function. Even if one were to allow for randomisation, it can be shown that there exists an optimal automaton with a deterministic initial state and decision function. Hence, we can focus on the class of automata. An automaton with randomisation would be defined as \( \tau = \langle M, X, \sigma, d, g_0 \rangle \) where \( M = \{1, \ldots, m\} \) is a finite set of states, \( \sigma : M \times X \rightarrow \Delta(M) \) is a transition function, \( d : M \rightarrow \Delta[I, NI] \) is a decision function and \( g_0 \in \Delta(M) \) is the initial randomisation over states.

**Proposition 1** There exists an optimal automaton with a deterministic initial state and decision function.

**Proof**: It can be established that there always exists an optimal automaton with randomisation. Let \( \tau \) be optimal. Denote as \( \Pr^{\theta,t}_\theta(i) \), the probability of the automaton \( \tau \), entering \( i \) at time \( t \) under hypothesis \( \theta \). The value generated by the automaton if the state of the world is \( \theta \in \{0, 1\} \) is:

\[
\sum_{t=0}^{\infty} (1 - \delta) \delta^t E_0 d(i) \Pr^{0,i}_\theta(i) = \sum_{t=0}^{\infty} (1 - \delta) \delta^t E_\theta d(i) \Pr^{0,i}_\theta(i) = \sum_{i \in M} d(i) \sum_{t=0}^{\infty} (1 - \delta) \delta^t E_\theta \Pr^{0,i}_\theta(i)
\]

Hence the ex-ante value is:

\[
\pi \sum_{i \in M} d(i) \sum_{t=0}^{\infty} (1 - \delta) \delta^t E_0 \Pr^{0,i}_\theta(i) + (1 - \pi) \sum_{i \in M} d(i) \sum_{t=0}^{\infty} (1 - \delta) \delta^t E_1 \Pr^{1,i}_\theta(i)
\]

\[
= \sum_{i \in M} d(i) \left[ \pi \sum_{t=0}^{\infty} (1 - \delta) \delta^t E_0 \Pr^{0,i}_\theta(i) + (1 - \pi) \sum_{t=0}^{\infty} (1 - \delta) \delta^t E_1 \Pr^{1,i}_\theta(i) \right]
\]

Defining \( e(i) := \pi \sum_{t=0}^{\infty} (1 - \delta) \delta^t E_0 \Pr^{0,i}_\theta(i) + (1 - \pi) \sum_{t=0}^{\infty} (1 - \delta) \delta^t E_1 \Pr^{1,i}_\theta(i) \), the above expression becomes:

\[
\sum_{i \in M} d(i) e(i)
\]
Now since the above expression is linear in \( d(i) \) and \( \tau \) is optimal, it must be the case that 
\[ d(i) = 1 \text{ if } e(i) > 0 \text{ and } d(i) = 0 \text{ if } e(i) < 0. \]
Hence, the following decision rule is optimal as well:

\[ d'(i) = d(i) I_{e(i) \neq 0} \]

Hence, the automata \( \tau' = \tau \backslash \{ d \} \cup \{ d' \} \) is optimal. Now under \( \tau' \), let the value from state \( i \) under hypothesis \( \theta \) be denote as \( V_{i\theta} \). Then the ex-ante value of the automaton is:

\[
\sum_{i \in M} g_0(i)[\pi V_{i0} + (1 - \pi)V_{i1}]
\]

Let \( i_o \in \arg\max_{i \in M} \pi V_{i0} + (1 - \pi)V_{i1} \). Since \( \tau' \) is optimal, the initial randomisation \( g_0 \) must be optimal and hence from the above expression it is clear that
\[ V(\tau') = \pi V_{i_o0} + (1 - \pi)V_{i_11}. \]

Hence, then automaton \( \tau'' = \tau \backslash \{ g_0 \} \cup \{ \delta_{i_0} \} \) is optimal (here \( \delta_{i_0} \in \Delta(M) \) is the delta measure on \( i_o \)) and is an automaton with deterministic initial state and decision function.

\[ \blacksquare \]

### 2.2 Modified Multiself Consistency

We discuss here the notion of modified multiself consistency. First we discuss two ways of defining consistent beliefs about the state of the world \( \theta \) at a given memory state \( i \in M \) reached with positive probability.

1. Consider an automaton \( \tau \). Let \((\alpha_{ij}, \beta_{ij})\) be the transition matrices induced by \( \tau \). Let \( \mu^0, \mu^1 \) be the corresponding stationary distributions. When the agent is at state \( i \), we use the stationary distributions to define posteriors (\( \pi(i) \)) about the state of the world \( \theta \) by:

\[
\pi(i) = \frac{\pi \mu^0(i)}{\pi \mu^0(i) + (1 - \pi)\mu^1(i)} \quad (1)
\]

Additionally, when the agent is at state \( i \) and receives signal \( x \), his posterior, generated by the aggregate signal \((i, x)\) is defined as:

\[
\pi(i, x) = \frac{\pi \mu^0(i) f_0(x)}{\pi \mu^0(i) f_0(x) + (1 - \pi)\mu^1(i) f_1(x)} \quad (2)
\]

2. We interpret the decision problem as in Kocer (2014 [11]) where the discount factor is interpreted as a continuation probability. This is as follows: At the beginning of every period, a coin of bias \( \delta \) is tossed, if tail comes up (with probability \( 1 - \delta \)), then the decision problem terminates and the cumulative payoff \( \sum_{\tau=0}^T x_\tau I_{d(t) = 1} \) is obtained.
Under this interpretation, the probability of entering state $i$ of the automaton under hypothesis $\theta$ is the following:

$$\hat{\mu}^\theta(i) = \sum_{t=0}^{\infty} (1 - \delta)\delta^t \Pr^\theta(i)$$

Where $\Pr^\theta(0)(i)$ is given by the initial distribution over states of the automaton (deterministic and independent of $\theta$) and $\Pr^\theta(i)(i)$ is the probability of being in state $i$ at time $t$ under the markov chain induced by the automaton in hypothesis $\theta$ (these are $(\alpha_{ij}), (\beta_{ij})$ as defined before). It has been shown by Wilson(2014,[10]) that $\hat{\mu}^0$ and $\hat{\mu}^1$ are the respective stationary distributions of the following perturbed markov chains:

$$w^0_{ij} = (1 - \delta) \Pr^0(j) + \delta\alpha_{ij}$$
$$w^1_{ij} = (1 - \delta) \Pr^1(j) + \delta\beta_{ij}$$

When the agent is at state $i$, we use the stationary distributions to define posteriors ($\hat{\pi}(i)$) about the state of the world $\theta$ by:

$$\hat{\pi}(i) = \frac{\pi \hat{\mu}^0(i)}{\pi \hat{\mu}^0(i) + (1 - \pi)\hat{\mu}^1(i)}$$ (3)

When the agent is state $i$ and receives signal $x$, his posterior, generated by the aggregate signal $(i,x)$ is defined as:

$$\hat{\pi}(i, x) = \frac{\pi \hat{\mu}^0(i)f_0(x)}{\pi \hat{\mu}^0(i)f_0(x) + (1 - \pi)\hat{\mu}^1(i)f_1(x)}$$ (4)

We can now define the notion of modified multiself consistency of an automaton. We do so for both notions of belief consistency defined above:

**Definition** An automaton is said to be *modified multiself consistent* if for all $i \in M$ and $x \in X$:

- If $\Pr(f(i, x) = j) > 0$ then

$$\hat{\pi}(i, x)V_{j0} + (1 - \hat{\pi}(i, x))V_{j1} \geq \hat{\pi}(i, x)V_{k0} + (1 - \hat{\pi}(i, x))V_{k1}$$

for all $k \in M$. 

9
• \( d(i) = 1 \) if
\[
\hat{\pi}(i)E_0 + (1 - \hat{\pi}(i))E_1 > 0
\]
and \( d(i) = 0 \) if
\[
\hat{\pi}(i)E_0 + (1 - \hat{\pi}(i))E_1 < 0
\]

The above definition of modified multiself consistency uses the notion for consistent beliefs as defined in (3) and (4) (Kocer (2014)) above. One may also define belief consistency as in (1) and (2) and define modified multiself consistency. We provide a definition below:

**Definition**: An automaton \( \tau \) is said to be \(*\)-modified multiself consistent if the conditions in the above definition are satisfied with beliefs defined as in (1) and (2).

It is important to note that the right notion of modified multiself consistency ought to depend on the interpretation of the discount factor in the model. If the discount factor \( \delta \) is assumed as factor that is used to measure next period payoffs from the point of the view of the current period, it seems implausible to consider it has a continuation probability that enters the extensive form of the environment. Hence, the appropriate beliefs to consider would be the ones defined as in (1) and (2). However, in the definition of the environment, if the discount factor is indeed a continuation probability, beliefs should correspond to ones defined by (3) and (4). From a purely technical standpoint, both notions are well defined and may be used in defining interim rationality. The use of the right notion depends on the specification of the environment.

It has been shown in Wilson(2014) that optimal automata are modified multiself consistent when considering beliefs defined by (3) and (4). In the next section we argue that there exist optimal automata which fail to be \(*\)-modified multiself consistent i.e beliefs defined by the stationary distribution of the markov chains induced by the transitions of the automaton. We demonstrate the existence of the counter-example both in the case of finite and continuous signals.

### 2.3 Two state automaton

#### 2.3.1 Finite signals

In this section we shall assume that the payoff space \( X \) is finite and focus on automata with two states \( M = \{H, L\} \). We first provide a characterisation of the optimal automaton
and argue that optimality need not imply *-modified multiself consistency. We begin by performing a preliminary analysis of the two state case.

We first argue that WLOG we can focus on automata with decision functions which satisfy \( d(H) = I \) and \( d(L) = NI \). Consider an automaton where you invest (choose action \( I \)) in both states \( H, L \). Such an automaton always invests whenever called upon to act. This behaviour can be replicated by an automaton where one transitions to the state \( H \) with probability one irrespective of the current state and signal and \( d(H) = I \). A similar argument works for automata where one chooses to not invest every time.

Notice that since there are only two states the transition function of the automaton can be thought of as a pair of functions \(< \sigma_H, \sigma_L >\) where \( \sigma_i : X \to [0, 1] \) and \( \sigma_L : X \to [0, 1] \). Here, \( \sigma_i(x) \) is the probability of transitioning from state \( i \in \{H, L\} \) to state \( H \). We use the pair of functions to define the markov chains over the states for each hypothesis \( \theta \in \{0, 1\} \). Define the values

\[
\alpha = \sum_{x \in X} \sigma_H(x)f_0(x)dx, \quad \gamma = \sum_{x \in X} (1 - \sigma_L(x))f_0(x)dx, \quad \beta = \sum_{x \in X} \sigma_H(x)f_1(x)dx \quad \text{and} \quad \lambda = \sum_{x \in X} (1 - \sigma_H(x))f_1(x)dx.
\]

Hence, transition matrices corresponding to \( \theta = 0 \) and \( \theta = 1 \) respectively are as follows:

\[
\begin{array}{c|cc}
 & H & L \\
\hline
H & \alpha & 1 - \alpha \\
L & 1 - \gamma & \gamma \\
\end{array}
\quad \begin{array}{c|cc}
 & H & L \\
\hline
H & \beta & 1 - \beta \\
L & 1 - \lambda & \lambda \\
\end{array}
\]

We use \( \alpha, \beta, \gamma, \lambda \) to define the values \( V_{H0}, V_{L0}, V_{H1}, V_{L1} \). It can be shown that:

- \( V_{H0}, V_{L0} \) are both strictly increasing in \( \alpha \) and strictly decreasing in \( \gamma \)
- \( V_{H1}, V_{L1} \) are both strictly decreasing in \( \beta \) and strictly increasing in \( \lambda \)
- \( V_{H1} < V_{L1} < 0 < V_{L0} < V_{H0} \)

Depending on the initial state \( i_0 \in \{H, L\} \) of the automaton, the value generated by it is either

\[
\pi V_{H0} + (1 - \pi)V_{H1}
\]

or

\[
\pi V_{L0} + (1 - \pi)V_{L1}
\]
both of which are strictly increasing in $\alpha$ and $\lambda$ and strictly decreasing in $\beta$ and $\gamma$. We now establish a characterisation of the optimal automaton in the proposition below:

**Proposition 2**: The optimal two state automaton is of the following nature:

$$\sigma^*_H(x) = \begin{cases} 
1 & \text{if } x > \bar{x} \\
q & \text{if } x = \bar{x} \\
0 & \text{if } x < \bar{x} 
\end{cases}$$

and

$$\sigma^*_L(x) = \begin{cases} 
1 & \text{if } x > \bar{x} \\
q' & \text{if } x = \bar{x} \\
0 & \text{if } x < \bar{x} 
\end{cases}$$

**Proof**: Let $\tau^*$ be an optimal automaton. Suppose $< \sigma^*_H, \sigma^*_L >$ are the transitions prescribed by it. We know that irrespective of the initial state, the value generated by the automaton is strictly increasing in $\alpha, \lambda$ and strictly decreasing in $\beta, \gamma$. Let $\alpha^*, \beta^*, \gamma^*, \lambda^*$ be the transition probabilities defined by $< \sigma^*_H, \sigma^*_L >$. Now for $\alpha^*$, consider the set:

$$K(\alpha^*) = \{ \beta \in [0, 1] : \exists \sigma_H \text{ s.t } \alpha^* = \sum_{x \in X} \sigma_H(x)f_0(x)dx, \beta = \sum_{x \in X} \sigma(x)f_1(x)dx. \}$$

Since the value generated by the automaton is strictly decreasing in $\beta$ it must be the case that:

$$\beta^* = \min K(\alpha^*)$$

using $\sigma^*_H$, we already know that $\beta^* \in K(\alpha^*)$. Hence, $\beta^* \geq \min K(\alpha^*)$. Suppose it is true that $\beta^* > \min K(\alpha^*)$. Then there exists a transition $\sigma'_H$ which yields $\alpha^*$ but a strictly lower $\beta$ proving $\sigma^*$ to be sub-optimal. Hence $\beta^* = \min K(\alpha^*)$.

This means the $\sigma^*_H$ is a solution to the following linear program.

$$\min_{\sigma_H} \sum_{x \in X} \sigma_H(x)f_1(x)$$

subject to

$$\sum_{x \in X} \sigma_H(x)f_0(x) = \alpha^*$$

$$0 \leq \sigma_H(x) \leq 1$$
This is equivalent to the following linear program:

\[
\min_y \sum_{x \in X} y(x) \frac{f_1(x)}{f_0(x)} \sum_{x \in X} y(x) = \alpha^* \\
0 \leq y(x) \leq f_0(x)
\]

It can be checked that if \( y^* \) is a solution to the second program, \( \sigma_H = y^* f_0 \) is a solution to the first. And, if \( \sigma_H^* \) is a solution to the first program, \( y = \frac{\sigma_H^*}{f_0} \) is a solution to the second.

We solve for the optimal solution to the second. Note that \( \frac{f_1(x)}{f_0(x)} \) is decreasing in \( x \). Now let \( X = \{x_1, ..., x_n\} \) such that \( x_i < x_{i+1} \). We use the following algorithm:

1. Start increasing \( y(x_n) \) till we hit \( \alpha^* \). If we can, then define \( y^* = (0, ..., q') \) where \( q' = \alpha \). If not then go to next step.

2. Start increasing \( y(x_{n-1}) \) till we can hit \( \alpha \). If we can we can, then define \( y^* = (0, ..., q', f_0(x_n)) \) where \( q' + f_0(x_n) = \alpha \). If we can’t, go to \( n-2 \) and continue until we hit \( \alpha \) at some \( k \in \{1, ..., n\} \).

The final solution will be of the form:

\[
y^*(x) = \begin{cases} 
  f_0(x) & \text{if } x > \bar{x} \\
  q' & \text{if } x = \bar{x} \\
  0 & \text{if } x < \bar{x}
\end{cases}
\]

And will be optimal. Hence the following function is optimal for the first problem:

\[
\sigma^*_H(x) = \begin{cases} 
  1 & \text{if } x > \bar{x} \\
  q & \text{if } x = \bar{x} \\
  0 & \text{if } x < \bar{x}
\end{cases}
\]

But the above linear program has a unique solution of the form described in the statement of the proposition (the monotone likelihood ratio property is used here), thus establishing the first half of the claim. A similar argument holds for \( \sigma^*_L \). ■

The above proposition states that in the optimal automaton, for each state, randomization in the transitions will occur for at most one signal. In the proof we construct the set:
\[ K(\alpha) = \{ \beta \in [0, 1] : \exists \sigma_H \text{ s.t. } \alpha = \sum_{x \in X} \sigma_H(x)f_0(x)dx, \beta = \sum_{x \in X} \sigma(x)f_1(x)dx. \} \]

Hence if the optimal automaton yields \( \alpha \) as the probability of transitioning from \( H \) to \( H \) in \( \theta = 0 \), the optimal value of \( \beta \) is equal to min \( K(\alpha) \). Hence, if we know the optimal value of \( \alpha \), we immediately can compute the optimal value of \( \beta \). Therefore, it shall be useful to treat \( \beta \) as a function \( \alpha \) by defining the function:

\[ \beta(\alpha) = \min K(\alpha) \]

By the same token, one may treat \( 1 - \lambda \) as a function of \( 1 - \gamma \). The construction of these functions will be useful in the analysis of the next of section where we provide an example of a decision problem where \( X = \{h, l\} \) i.e there are only two payoff values (signals) and prove the existence of an optimal automaton which fails to be \( * \)-modified multiself consistent.

### 2.3.2 Two signals

\( X = \{h, l\}, l < 0 < h \) and the signal structure is

\[
\begin{array}{c|cc}
 & l & h \\
\hline
f_0 & 1-p & p \\
f_1 & p & 1-p \\
\end{array}
\]

The values \( p, h, l \) are such that \( E_0 > 0 \) and \( E_1 < 0 \). For simplicity, one may consider the values \( h = -l = 1 \) and \( p > \frac{1}{2} \). The functions \( \alpha \) and \( \beta \):

\[
\beta(\alpha) = \begin{cases} 
\frac{1-p}{p} \alpha & \text{if } \alpha \leq p \\
\frac{p}{1-p}(\alpha - p) + (1-p) & \text{if } p \leq \alpha.
\end{cases}
\]

\[
1 - \beta(\alpha) = \begin{cases} 
\frac{2p-1}{p} + \frac{1-p}{p}(1-\alpha) & \text{if } \alpha \leq p \\
\frac{p}{1-p}(1-\alpha) & \text{if } p \leq \alpha.
\end{cases}
\]

The functions \( \gamma \) and \( \lambda \):

\[
1 - \lambda = \begin{cases} 
\frac{1-p}{p}(1-\gamma) & \text{if } 1-\gamma \leq p \\
\frac{p}{1-p}(1-\gamma - p) + (1-p) & \text{if } p \leq 1-\gamma.
\end{cases}
\]

**Proposition 3**: There exists a \( \pi \) such that there exist optimal automata that is not \( * \)-modified multiself consistent.
Proof: For π = 1, the optimal automaton yields (α, γ) = (1, 0) and for π = 0 we have (β, λ) = (0, 1) which enforces (α, γ) = (0, 1). By continuity, we can find π ∈ (0, 1) such that the optimal automaton yields p < α < 1 and the initial state is H. (the values for β and λ can be derived from the functions above). Denote the values of the states i ∈ {H, L} at θ ∈ {0, 1} as V_{iθ}. We know that:

\[ V_{H1} < V_{L1} < 0 < V_{L0} < V_{H0} \]

This implies that only at a unique π* do we get π'V_{H0} + (1 − π')V_{H1} = π'V_{L0} + (1 − π')V_{L1}. For π' > π* we have π'V_{H0} + (1 − π')V_{H1} > π'V_{L0} + (1 − π')V_{L1} and for π' < π* we have π'V_{H0} + (1 − π')V_{H1} < π'V_{L0} + (1 − π')V_{L1}. Now since p < α < 1, the optimal automaton is of the following nature:

\[
\begin{array}{c|c|c}
 H & x & 1 \\
 L & . & . \\
\end{array}
\]

where 0 < x < 1. This implies that there exists strict randomization at (H, l). Hence, if the automaton is indeed *-modified multiself consistent, the posteriors at this combined signal must satisfy π(H, l) = π*. However, we know from Kocer (2010) and Wilson(2014) that the optimal automaton is modified multiself consistent with respect to discounted beliefs ̂π(., .) as defined by the (3) and (4). Hence it must be the case that π(H, l) = ̂π(H, l).

\[
\frac{p_1}{p_0} \frac{p}{1-p} = \frac{1-p_1}{1-p_0} \frac{p}{1-p}
\]

Where \((p_0, 1-p_0) = \left( \frac{1-γ}{(1−γ)+(1−α)}, \frac{1-α}{(1−γ)+(1−α)} \right)\) and \((p_1, 1-p_1) = \left( \frac{1-λ}{(1−λ)+(1−β)}, \frac{1-β}{(1−λ)+(1−β)} \right)\) are the stationary distributions for the markov chains generated by the two hypotheses. Substituting in the above equation we get

\[
\begin{align*}
1-γ &= \frac{1-λ}{1-β} \\
1-λ &= \frac{p}{1-p}(1-γ)
\end{align*}
\]

But (7) contradicts the fact that \(1-λ = \frac{p}{1-p}(1-γ) + \frac{1-2p}{1}\) which follows from the functions defined above.

□
2.4 Continuous signals

This section deals with the case of a continuum of signals and shows an example of an optimal automaton that is not modified mutually consistent.

2.4.1 Two state automaton

Now assume \( X = [a, b] \) where \( a < 0 < b \) and \( f_0(x), f_1(x) > 0 \) for all \( x \in X \). We also assume that \( \frac{f_0(x)}{f_1(x)} \) is increasing in \( x \). We show that a threshold rule is optimal in this case as well. It can also be established that the counterpart of claim 1 is true in this case.

**Proposition 3** There exists a optimal automaton of the following form:

\[
\sigma_H(x) = \begin{cases} 
1 & \text{if } x \geq x_H \\
0 & \text{if } x < x_H 
\end{cases}
\]

\[
\sigma_L(x) = \begin{cases} 
1 & \text{if } x \geq x_L \\
0 & \text{if } x < x_L 
\end{cases}
\]

where \( \bar{x} \in X \)

**Proof**: Notice that \( \beta(\alpha) \) is the optimal value of the optimization problem\(^3\):

\[
\min_s \int_X s(x)f_1(x)dx - \int_X s(x)f_0(x)dx = \alpha
\]

\( s : X \to [0, 1] \) is measurable.

We claim that for \( \alpha \), the optimal \( s \)

\[
s^*(x) = \begin{cases} 
1 & \text{if } x \geq \bar{x} \\
0 & \text{if } x < \bar{x} 
\end{cases}
\]

\(^3\)Notice that the optimization problem is the same as in the Neyman-Pearson lemma in hypothesis. One can interpret “\( \alpha \)” as the size of the test and the objective that type-1 error.
where $\bar{x}$ such that $1 - F_0(\bar{x}) = \alpha$. Notice that $s^*$ is measurable and $\int_X s'(x)f_0(x)dx = \alpha$. Suppose that $s^*$ is not optimal. Then there exists an $s : X \rightarrow [0, 1]$ measurable such that $\int_X s(x)f_0(x)dx = \alpha$ and $\int_X s(x)f_1(x)dx < \int_X s'(x)f_1(x)dx$. Now this implies that $\int_a^b s(x)f_0(x)dx = \int_a^b s'(x)f_0(x)dx$. Given the definition of $\bar{x}$, this implies :

$$\int_a^\bar{x} (s(x) - 0)f_0(x)dx = \int_x^b (1 - s(x))f_0(x)dx$$

(6)

We also have :

$$\int_a^\bar{x} (s(x) - 0)f_1(x)dx < \int_x^b (1 - s(x))f_1(x)dx$$

(7)

The above two equalities and the fact that $\frac{f_1(x)}{f_0(x)}$ is decreasing in $x$ implies :

$$\frac{f_1(\bar{x})}{f_0(\bar{x})} \int_a^\bar{x} (s(x) - 0)f_0(x)dx \leq \int_a^\bar{x} (s(x) - 0)f_0(x)\left(\frac{f_1(x)}{f_0(x)}\right)dx$$

$$< \int_x^b (1 - s(x))f_0(x)\left(\frac{f_1(x)}{f_0(x)}\right)dx$$

$$\leq \frac{f_1(\bar{x})}{f_0(\bar{x})} \int_a^\bar{x} f(x)dx$$

This is a contradiction. Hence $s^*$ is optimal. □

The above claim shows that a threshold rule is optimal even in the case of a continuous payoff space.

Let $M = \{H, L\}$ and $X$ be continuous i.e $X = [a, b]$, $a < 0 < b$. An automaton is now a pair of functions (one for each state), $s(x), \tau(x) : X \rightarrow [0, 1]$ where $s(x), \tau(x)$ are the probabilities of transitioning from $H$ to $H$ and $L$ to $H$. Define $\alpha = \int_X s(x)f_0(x)dx$, $\gamma = \int_X (1 - \tau(x))f_0(x)dx, \beta = \int_X s(x)f_1(x)dx$ and $\lambda = \int_X (1 - \tau(x))f_1(x)dx$. The decision rule is fixed $d(H) = I$ and $d(L) = NI$. The two transition matrices are :

<table>
<thead>
<tr>
<th></th>
<th>$H$</th>
<th>$L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H$</td>
<td>$\alpha$</td>
<td>$1 - \alpha$</td>
</tr>
<tr>
<td>$L$</td>
<td>$1 - \gamma$</td>
<td>$\gamma$</td>
</tr>
</tbody>
</table>
We use $\alpha, \beta, \gamma, \lambda$ to define the values $V_{H0}, V_{L0}, V_{H1}, V_{L1}$. It can be shown that:

- $V_{H0}, V_{L0}$ are both strictly increasing in $\alpha$ and strictly decreasing in $\gamma$
- $V_{H1}, V_{L1}$ are both strictly decreasing in $\beta$ and strictly increasing in $\lambda$
- $V_{H1} < V_{L1} < 0 < V_{L0} < V_{H0}$

Since $\pi \in (0, 1)$, at the optimum, $\alpha, \beta, \lambda, \gamma \in (0, 1)$ and $\beta < \alpha$ and $\gamma < \lambda$ and the optimal $< \alpha_H \ast (x), \sigma_L^\ast (x) >$ are necessarily solutions to the following optimization problem:

\[
\min_{\pi : \int_X \pi f_0(x) dx = \alpha} \int_X s(x) f_1(x) dx
\]

\[
\min_{\tau : \int_X 1 - \tau(x) f_0(x) dx = \lambda} \int_X 1 - \tau(x) f_1(x) dx
\]

By arguments from the previous section it can be shown that the optimal $< \alpha_H^\ast, \sigma_L^\ast >$ is unique and there exist $x_H, x_L \in [a, b]$ such that:

\[
\sigma_H^\ast (x) = I_{\{x \geq x_H\}} \quad \sigma_L^\ast (x) = I_{\{x < x_L\}}
\]

We shall now demonstrate a class of optimal automata which are not sequentially rational. It shall be important establish the following claim first:

**Proposition 5**: Let $< \alpha_H^\ast, \sigma_L^\ast >$ be an optimal automaton. If it is additionally *-modified multiself consistent, then

- $x_L \geq x_H$
- It cannot be the case $\{x : x \geq x_H\} \subseteq \{x : \frac{f_0(x)}{f_1(x)} \geq 1\}$

**Proof**: The stationary distributions corresponding the two transition matrices above are \(\left(\frac{1-\gamma}{(1-\gamma)+(1-\alpha)}, \frac{1-\gamma}{(1-\gamma)+(1-\alpha)}\right)\) and \(\left(\frac{1-\alpha}{(1-\alpha)+\beta}, \frac{1-\alpha}{(1-\alpha)+\beta}\right)\). Now define $p_0 = \frac{1-\gamma}{(1-\gamma)+(1-\alpha)} = Pr(i = H|\theta = 0)$ and $p_1 = \frac{1-\alpha}{(1-\alpha)+\beta} = Pr(i = H|\theta = 1)$. Since $\beta < \alpha$ and $\gamma < \lambda$, we have that $p_0 > p_1$. Now let $\pi^\ast = \min\{\pi' : \pi' V_{L0} + (1 - \pi')V_{L1} \leq \pi' V_{H0} + (1 - \pi')V_{H1}\}$. *-modified multiself consistency implies $\{x : x \geq x_H\} = \{x : Pr(\theta = 0|L, x) \geq \pi^\ast\}$. Similarly, $\{x : x \geq x_H\} = \{x : Pr(\theta = 0|H, x) \geq \pi^\ast\}$
\[
\Pr(\theta = 0|H, x) = \frac{1}{1 + \frac{1-\pi}{\pi}(\frac{p_1}{p_0})(\frac{f_1(x)}{f_0(x)})}
\]
\[
\Pr(\theta = 0|L, x) = \frac{1}{1 + \frac{1-\pi}{\pi}(\frac{1-p_1}{1-p_0})(\frac{f_1(x)}{f_0(x)})}
\]

Since \(p_0 > p_1\), and \(\frac{f_1}{f_0}\) and decreasing in \(x\), it means that both \(\Pr(\theta = 0|L, x)\) and \(\Pr(\theta = 0|H, x)\) are decreasing in \(x\) and \(\Pr(\theta = 0|L, x) \leq \Pr(\theta = 0|H, x)\) for all \(x \in \mathcal{X}\). Hence, \(x_L \geq x_H\)

We now prove the second part of the claim. Let \(x\) be such that \(\frac{f_0(x)}{f_1(x)} = 1\). Then we get \(\{x : x \geq x\} = \{x : \frac{f_0(x)}{f_1(x)} \geq 1\}\). We consider two cases

- **CASE 1**: \(\pi V_{L,0} + (1-\pi)V_{L,1} \leq \pi V_{H,0} + (1-\pi)V_{H,1}\). Now there exists \(x < x\) such that \(\frac{p_1}{p_0} \frac{f_1(x)}{f_0(x)} < 1\) so that \(\Pr(\theta = 0|H, x) > \pi\), which implies that the optimal decision at \((H, x)\) should be to transition to \(H\) as opposed to the optimal strategy prescription \(L\).

- **CASE 2**: \(\pi V_{L,0} + (1-\pi)V_{L,1} > \pi V_{H,0} + (1-\pi)V_{H,1}\). By a similar argument, there exists \(x \geq x\) such that \(\frac{1-p_1}{1-p_0} \frac{f_1(x)}{f_0(x)} \geq 1\). Hence \(\Pr(\theta = 0|L, x) \leq \pi\), which implies at the \((L, x)\), the transition to \(H\) is strictly better than a transition to \(H\) which is what the strategy prescribes since \(x_L \geq x_H\).

\(\square\)

**Proposition**: There exists an optimal automaton which is not sequentially rational

**Proof**: It can be shown that the optimal threshold \(x_H(\pi)\) is continuous is \(p\) and satisfies :

- \(\lim_{\pi \to 0} x_H(\pi) = b\)
- \(\lim_{\pi \to 1} x_H(\pi) = a\)

Now consider any \(\pi\) such that \(x_H(\pi) \geq x\). This implies \(\{x : x \geq x_H(\pi)\} \subseteq \{x : \frac{f_0(x)}{f_1(x)} \geq 1\}\). But this means sequential rationality cannot be satisfied.

\(\square\)

### 3 Limits to randomization

We show in this section that any finite state automata can be modified to give the same performance yet significantly reduce the extent of randomization required for implementation. As the number of states and number of signals become large virtually no
randomisation is required and the automaton is “close” to a deterministic one. We require the following definition first:

**Definition** Two finite state automata \( \tau \) and \( \tau' \) are said to be equivalent if they generate the same transition probabilities for all \( \theta \in \{0, 1\} \), have the same decision function and initial state.

Before proceed towards the result, we would need to define a measure of the amount of randomization taking place in the automaton \( \tau \). For each state \( i \in M \), consider the number \( n_i^\tau = |\{(x, j) : \sigma(i, x)(j) > 0\}| \). The measure of randomization is:

\[
 n(\tau) = \frac{\sum_i n_i^\tau}{|M||M||X|}
\]

The result is presented as follows

**Proposition 6** : Let \( |M| \geq 3, |X| \geq 3 \), then for any automaton \( \tau \), there exists an equivalent automaton \( \tau' \) such that:

\[
\frac{1}{|M|} \leq n(\tau') \leq \frac{1}{|M|} + \frac{2}{|X|}
\]

**Proof** : Let \( \tau \) be an automaton yielding transition probabilities \( \alpha_{ij}, \beta_{ij} \). For any equivalent automaton and \( i \in M \), the transition function \( \dot{\sigma}_i(x, j) = \sigma(i, x)(j) \) must satisfy the following conditions:

\[
\begin{align*}
\sum_x \dot{\sigma}_i(x, j)f_1(x) &= \alpha_j \ \forall j \\
\sum_j \dot{\sigma}_i(x, j)f_1(x) &= \beta_x \ \forall j \\
\sum_j \dot{\sigma}_i(x, j) &= 1 \ \forall x \\
\dot{\sigma}_i(x, j) &\geq 0 \ \forall x, j.
\end{align*}
\]

Viewing \( \dot{\sigma} \) as an element in \( \mathbb{R}^{X \times M} \), we observe that \( \{ \dot{\sigma} : \dot{\sigma} \text{ satisfies (7)} \} \) is a polytope of the form \( \{ x \in \mathbb{R}^{X \times M} : Ax \leq b \} \). It is true (see Schrijver(2013) [7]) that if \( z \) is an extreme point of the submatrix \( A_z \) defined by the set of inequalities which bind has \( \text{Rank}(A_z) = |X| \times |M| \). In the above problem, there are \( 2|M| + |X| + |M| \times |X| \) inequalities. Consider an extreme point \( \sigma' \) of the polytope. For this point \( \text{Rank}(A_{\sigma'}) = |X| \times |M| \). Since the first \( 2|M| + |X| \) constraints are all equalities, to satisfy the rank condition at least \( |M| \times |X| - 2|M| - |X| \geq 0 \) of the last constraints should hold at equality. Let \( \tau' \) be the automaton that replaces the transition function in \( \tau \) with \( \sigma' \). Then we have \( n_i^{\tau'} \leq 2|M| + |X| \), hence:
\[ n(\tau) = \frac{\sum_i n_{i}^{T}}{|M||M||X|} \geq \frac{|M|(2|M| + |X|)}{|M||M||X|} = \frac{1}{|M|} + \frac{2}{|X|} \]

The lower bound \(\frac{1}{|M|}\) comes from the fact that the measure of randomization for deterministic automata is the lowest and equals \(\frac{1}{|M|}\).

\[ \square \]

4 Weak admissibility and structure of optimal M-state automata

In this section, we define a new criterion for an automaton to be satisfied in the context of the investment problem called \textit{weak admissibility}. The criterion has a flavor of an "admissibility" requirement as in statistical decision theory and classical hypothesis testing. Suppose the set of states of the automaton is \(M = \{1, ..., m\}\). Automaton \(\tau\) generates the following value function equations:

\[ V_{0}^{i} = d(i)E_{0} + \delta \sum_{j \in M} \alpha_{ij}V_{0}^{j} \]

\[ V_{1}^{i} = d(i)E_{1} + \delta \sum_{j \in M} \beta_{ij}V_{1}^{j} \]

Consider the transition from a state \(i \in M\) that is reached with positive probability. Suppose under the automaton transitions dictate probabilities \(\beta_{i}\) in the state of the world \(\theta = 1\). Fixing \(\beta_{i}\), it would be natural to require that the transitions in the good state of the world are optimal i.e:

\[ \max_{\alpha} \sum_{j} \alpha_{ij}V_{0}^{j} \]

s.t. \(\alpha \in K(\beta_{i})\)

Where for a probability distribution \(\beta \in \Delta(M)\), \(K(\beta)\) is defined as:

\[ K(\beta) = \{ \alpha \in \Delta(M) : \exists \sigma : X \rightarrow \Delta(M) \text{ s.t} \sum_{x} f_{0}(x)\sigma(x) = \alpha \text{ and } \sum_{x} f_{1}(x)\sigma(x) = \beta \} \]

The above condition has the flavor of optimal tests in classical hypothesis testing. The transition probability vector \(\beta_{i}\) may be interpreted as the "size of the test" and we choose optimally from the constrained choice of \(\alpha_{i}\). Weak admissibility is hence defined as follows:
Definition : An automaton $\tau$ is said to be *weakly admissible* if at every state $i$ reached with positive probabilitiy, the transition probabilities $(\alpha_i, \beta_i)$ satisfy:

$$\alpha_i \in \arg \min \alpha \sum_j \alpha_j V_j^0$$

s.t. $\alpha \in K(\beta_i)$

We refine the class of automata of interest by requiring that no wastage of states occurs in the performance of the automaton. By this, it is meant that no two states reached with positive probabilitiy generate the same continuation values in both states of the world $\theta \in \{0, 1\}$ and also that one does not “weakly dominate” in terms of continuation values. We present a formal definition below:

Definition : An automaton $\tau$ satisfies *no wastage* if for any two states $i$ and $j$ reached with positive probability:

1. $(V_i^0, V_i^1) \neq (V_j^0, V_j^1)$
2. $(V_i^0, V_i^1) \not\preceq (V_j^0, V_j^1)$ and $(V_j^0, V_j^1) \not\preceq (V_i^0, V_i^1)$

4.1 Structure of weakly admissible automata with no wastage

The optimization problem posed by the weak admissibility constraint is a linear program and for every $\beta \in \Delta(M)$, $K(\beta)$ is a polytope in $\mathbb{R}^M$. It can be written more explicitly as follows:

$$\max_\sigma \sum_j \sum_x \sigma(x, j) f_0(x) V_j^0$$

s.t. $\sum_x \sigma(x, j) f_1(x) = \beta_j \forall j$

$$\sum_j \sigma(x, j) = 1 \forall x$$

$$\sigma(x, j) \geq 0 \forall x, j.$$  \(10\)

Here, the $\sigma(x, j)$ denotes the probability of transition to state $j$ upon receiving signal $x$. By thinking of $\sigma$ as a matrix of dimension $|X| \times |M|$ with non-negative entries, the first constraint says that the columns up to a vector of ones and the rows sum to the vector $\beta$. 

22
It can be shown the above problem is equivalent to the following problem:

$$\max_u \sum_j \sum_x u(x,j) \frac{f_0(x)}{f_1(x)} V^0_j$$

s.t. 
$$\sum_x u(x,j) = \beta_j \forall j$$
$$\sum_j u(x,j) = f_1(x) \forall x$$
$$u(x,j) \geq 0 \forall x,j.$$  \hspace{1cm} (11)

The equivalence is guaranteed by noting the relation \( u(x,j) = \sigma(x,j) f_1(x) \). Note that the feasible set is the set of all non-negative bi-stochastic matrices whose columns add to \( f_1 \) and rows add to \( \beta \). Note also that since the above problem is considered from a state \( i \in M \) reached with positive probability, then clearly any \( j \in M \) with \( \beta_j > 0 \) implies that \( j \) is reached with positive probability as well. Hence, from no wastage, all the vectors in \( \{(V^0_j, V^0'_j) : j \text{ such that } \beta_j > 0\} \). A solution exists due to compactness of \( K(\beta) \) and continuity of the objective and the following algorithm guarantees a unique optimal solution \( u^* \).

**Algorithm:**

1. Define \( u_1 = 0 \)

2. Suppose \( u_n \) is achieved at the \( n \)th step of algorithm. Search for \((x,j)\) with the highest value of \( \frac{f_0(x)}{f_1(x)} V^0_j \) such that \( \min\{f_1(x), \beta_j\} - u_n(x,j) > 0 \). If non exists, return \( u_n \) otherwise return the matrix \( u_{n+1} \) defined by:

\[
\begin{align*}
    u_{n+1}(x', j') = & \begin{cases} 
        \min\{f_1(x), \beta_j\} & \text{if } (x', j') = (x, j) \\
        u_n(x', j') & \text{o.w}
    \end{cases}
\end{align*}
\]

The above algorithm terminates and returns the unique optimal solution with the following "staircase" form where the order of row components i.e \( x \)'s is done via likelihood ratios \( \frac{f_0(x)}{f_1(x)} \) and the ordering over column components i.e \( j \)'s is done via the continuation.
values $V_j^0$ (from no wastage this ordering is unique).

\[
\sigma = \begin{pmatrix}
  * & * & 0 & 0 & \ldots & 0 \\
  0 & * & 0 & 0 & \ldots & 0 \\
  0 & * & * & \ldots & 0 \\
  0 & 0 & 0 & * & \ldots & 0 \\
  \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \\
  0 & 0 & 0 & 0 & \ldots & * \\
  0 & 0 & 0 & 0 & \ldots & * 
\end{pmatrix}.
\]

(12)

This leads us to the following proposition :

**Proposition 7** : Every automaton that is weakly admissible and satisfies no wastage necessarily admits a transition function of a staircase form as in (13). Moreover, is the unique solution to the LP problem imposed by the weak admissibility constraints.

**Proof** : The algorithm above terminates uniquely to say $\sigma^*$ since from no wastage the values $\frac{f_0(x)}{f_j(x)} V_j^0$ are distinct over all $(x, j)$'s. Consider any optimal solution $\hat{\sigma}$. It is sufficient to show that either the first row or the first column has the intersecting term $((x, j)$ with highest value of $\frac{f_0(x)}{f_j(x)} V_j^0)$ as the only non zero elements. If not, then there exists a $(y, k)$ such that $\frac{f_0(y)}{f_j(y)} < \frac{f_0(x)}{f_j(x)}$ and $V_k^0 < V_j^0$ such that $\hat{\sigma}(y, j) > 0$ and $\hat{\sigma}(x, k) > 0$. Now clearly, $\epsilon > 0$ amount of weight can be taken from $\hat{\sigma}(y, j) > 0$ and shifted to $\hat{\sigma}(x, j) > 0$ and $\hat{\sigma}(y, k) > 0$ to strictly improve the solution. Hence, $\hat{\sigma}$ agrees with the first step of the algorithm. By induction, it can be shown that it agrees with all steps. Hence $\hat{\sigma} = \sigma^* \square$

Using the above result, we can now show that weakly admissible automata with no wastage improve the upper bound on the extent of randomization involved when using the measure of randomisation has introduced in section 5. We present it as the following result.

**Proposition 8** : Let $\tau$ be a weakly admissible automaton with no-wastage, then :

\[
\frac{1}{|M|} \leq n(\tau') \leq \frac{1}{|M|} + \frac{1}{|X|}
\]

**Proof** : The algorithm terminates to an extreme point of the feasible set in the linear program. The argument is similar to the proof of proposition 6 but with tighter constraints. $\square$

24
Now, for optimal automata, the following can be said:

**Proposition 9**: For every optimal automaton \( \tau \), there exists an equivalent automaton with the same ex-ante value that is weakly admissible and satisfies no wastage.

**Proof**: Shift transition probabilities between equivalent states of the automaton and use modified multiself consistency. □

## 5 Changing worlds and optimal finite memory

In this section we study a more general decision problem where the underlying state of the word \( \theta \) can potentially change over time evolving according to a Markov process \( P \).

Consider \( \Theta = \{0, 1\} \) and suppose that the underlying state evolves according to:

\[
P(\theta = 0|\theta = 0) = 1 - \epsilon \\
P(\theta = 1|\theta = 1) = 1 - \Delta
\]

where \( 0 < \Delta, \epsilon < 1 \) and the stationary distribution is \((\frac{\Delta}{\Delta + \epsilon}, \frac{\epsilon}{\Delta + \epsilon})\).

### 5.1 Unconstrained optimal automata

The decision problem every period is to invest or not every period depending on the current state and maximise long run discounted expected expected utility. The utilities for actions are: \( u(I, \theta) = E_0 \) and \( u(NI, \theta) = 0 \) and \( E_0 > 0 > E_1 \). However, the states are not observable and the agent can only observe an informative signal from a finite set \( X \) about the current state via the signalling structures \(< f_0(x), f_1(x) >\), where \( f_0(x), f_1(x) > 0 \).

We assume that payoffs are unobservable every period hence omitting the possibility of inference of the true state from the period payoff. When \( \Delta, \epsilon \) are small, the decision problem can be imagined as one close to the investment problem in section 1, where the probability of staying at a state is close to one.

Such models are also called hidden Markov models and fall under the category of non-interactive partially observable Markov decision processes. The memory unconstrained
ex-ante optimal decision rule is simple: At time $t$ if beliefs about the current state is $\pi_t$, then invest and only if $\pi_t \geq \pi^*$, where $\pi^*$ is such that $\pi^*E_0 + (1 - \pi^*)E_1 = 0$.

Clearly the decision path is completely described by the belief process. Hence, to understand the optimal decision rule, it is important to understand the dynamics of the belief process. At a given time if the belief about $\theta = 0$ is $\pi$, then the belief about the next state after receiving the signal $x$ will be:

$$\hat{\pi}(\pi, x) = \frac{\pi(1 - \epsilon)f_0(x) + (1 - \pi)\Delta f_1(x)}{\pi(1 - \epsilon)f_0(x) + (1 - \pi)\Delta f_0(x) + \pi\epsilon f_1(x) + (1 - \pi)(1 - \Delta)f_1(x)}$$ (13)

Before proceeding to the result the following definition would be important:

**Definition**: A signalling structure $< f_0(x), f_1(x) >$, has full support if for all $\theta \in \{0, 1\}$, for all $x \in X$, $f_\theta(x) > 0$.

**Definition**: A signalling structure $< f_0(x), f_1(x) >$ with full support is said to be informative if $f_0 \neq f_1$.

We now present the result:

**Proposition 10**: If $\frac{\Delta}{\Delta + \epsilon} \neq \pi^*$, and $1 - \Delta > \epsilon$, then there exists an informative signalling structure $< f_0(x), f_1(x) >$, with full support such that the optimal decision rule can be implemented by a deterministic finite state automaton.

**Proof**: We shall prove this for the case when $\frac{\Delta}{\Delta + \epsilon} < \pi^*$. The proof for $\frac{\Delta}{\Delta + \epsilon} > \pi^*$ will be analogous. It shall be convenient to define notice that beliefs can be represented in the following way:
\[ \hat{\pi}(\pi, x) = \frac{\pi(1 - \epsilon)f_0(x) + (1 - \pi)\Delta f_1(x)}{\pi(1 - \epsilon)f_0(x) + (1 - \pi)\Delta f_0(x) + \pi\epsilon f_1(x) + (1 - \pi)(1 - \Delta)f_1(x)} \]  

(14)

\[ = \frac{1}{1 + \frac{\pi\epsilon f_1(x) + (1 - \pi)(1 - \Delta)f_1(x)}{\pi(1 - \epsilon)f_0(x) + (1 - \pi)\Delta f_0(x)}} \]  

(15)

\[ = \frac{1}{1 + \frac{f_1(x)\pi\epsilon + (1 - \pi)(1 - \Delta)}{f_0(x)\pi(1 - \epsilon) + (1 - \pi)\Delta}} \]  

(16)

\[ = \frac{1 + \frac{f_1(x)\pi\epsilon + (1 - \pi)(1 - \Delta)}{f_0(x)\pi(1 - \epsilon) + (1 - \pi)\Delta} \left( \frac{\pi}{1 - \pi} \right) \left( \frac{1 - \pi}{\pi} \right)}{1 + \frac{f_1(x)\pi\epsilon + (1 - \pi)(1 - \Delta)}{f_0(x)\pi(1 - \epsilon) + (1 - \pi)\Delta}} \]  

(17)

The condition \(1 - \Delta > \epsilon\) guarantees that the function \(\hat{\pi}(\pi, x)\) is strictly increasing in \(\pi\). For \(l > 0\) define the function \(g_l(\pi)\) defined as:

\[ g_l(\pi) = \frac{1}{1 + l \pi\epsilon + (1 - \pi)(1 - \Delta)\frac{\pi}{1 - \pi}} \]  

Notice that \(g\) is continuous, strictly increasing and maps \([0, 1]\) to \([0, 1]\) and has a unique fixed point. The point \(\bar{\pi}\) is a fixed point if and only if:

\[ \frac{1}{1 + l \pi\epsilon + (1 - \pi)(1 - \Delta)\frac{\pi}{1 - \pi}} - \frac{\pi}{1 - \pi} = 0 \]

In addition, the fixed points are strictly increasing and continuous in \(l\). Let \(\bar{\pi}_l\) be the unique fixed point associated with \(f_0(x), f_1(x)\). By continuity of \(\pi_l\) in \(l\), there exists \(l^* > 1\) such that

\[ \frac{\Delta}{\Delta^*} < \bar{\pi}_{l^*} < \pi^* \]

Now, there exists an informative signalling structure \(< f_0(x), f_1(x) >\) with full support such that for max\{\(f_0(x) : x \in X\)\} \(< l^*\) : We show that this signalling structure has the desired property. Order the x’s according to likelihood ratios as \(\{x_1, ..., x_n\}\). We can assume for convenience that likelihood ratios are distinct hence strictly increasing in \(i\). For each \(i \in \{1, ..., n\}\), define \(\bar{\pi}_i\) as the unique fixed point associated with \(\frac{f_0(x_i)}{f_1(x_i)}\). Hence, we have:

\[ 0 < \bar{\pi}_1 < \bar{\pi}_2 < \cdots < \frac{\Delta}{\Delta^*} \leq \bar{\pi}_{n-1} < \bar{\pi}_n < \pi^* < 1 \]

Now, consider the partition \([0, \bar{\pi}_1) \cup [\bar{\pi}_1, \bar{\pi}_n] \cup (\bar{\pi}_n, 1]\) of \([0, 1]\). We can show that if the belief enters the interval \([\bar{\pi}_1, \bar{\pi}_n]\), there it will forever stay there with probability one. If the prior
π belongs to the interval \((\bar{\pi}_n, 1]\), then either if the signal with the highest likelihood ratio is received repeatedly, the belief drops below \(\pi^*\) for sure. Starting in \([0, \bar{\pi}_1]\), we forever stay below \(\pi^*\) irrespective of signals. Notice in all three cases, there is a time \(\bar{t}\) after which the decision maker should not invest for sure. Finite automata can replicate all such decision rules.

\[\square\]

**Corollary** : For any informative signalling structure \(< f_0(x), f_1(x) >\) with full support, let \(\bar{t} = \max_x \frac{f_0(x)}{f_1(x)}\). If \(\bar{\pi}_l < \pi^*\), and \(1 - \epsilon > \Delta\), then the optimal decision rule can be implemented by a deterministic finite state automaton.

We now show the following:

**Proposition 11** : If signals have full support and are informative and \(1 - \epsilon > \delta\), then the belief process is markov and has a unique invariant distribution.

**Proof** : The state space for the process is \([0, 1]\) and transitions are defined by \(\pi(\pi, x)\) making the belief process markov. By definition, it is clear that monotonicity and the feller property are satisfied. The third condition to check is the mixing condition (See SLP [9]). Since signals are informative, the fixed points associated with the likelihood ratios can be ordered according to the likelihood ratios as in the proof of the above proposition. Let \(Q\) be the transition function associated with the markov process then for \(c = \frac{\Delta}{\Delta + \epsilon}\), \(Q^n(0, [c, 1]) > 0\) is positive by repeatedly receiving the highest signal and \(Q^n(1, [0, c]) > 0\) by repeatedly receiving the lowest signal. Hence, the mixing condition is satisfied for \(c\) as defined and \(\delta = \min\{Q^n(0, [c, 1]), Q^n(1, [0, c])\}\). Hence a there exists a unique invariant distribution.

\[\square\]

**Example** : Variant of Shiryaevs Disruption Problem In the shiryaev decision problem the transitions take place according to

\[
\begin{array}{c|cc}
0 & 0 & 1 \\
\hline
0 & 1 - \epsilon & \epsilon \\
1 & 0 & 1 \\
\end{array}
\]

The results presented above depends on the threshold belief \(\pi^*\). In the Shiyraev problem, one can obtain a stronger result which does not depend on the threshold. The statement
is as follows:

**Proposition 12**: Let $\bar{l} = \max \frac{f_0(x)}{f_1(x)}$. If $\bar{l}(1 - p) < 1$, then belief process converges to 0 almost surely. Hence, deterministic finite state automata can execute the optimal decision rule.

**Proof**: For the transition matrix above:

$$g'(\pi) = \frac{\pi l(1 - e)}{\pi(l - 1)(1 - e) + 1}$$

Notice that $g'$ is strictly increasing, continuous and differentiable over $[0, 1]$ with $g'(0) = 0$ for all $l > 0$. The derivative is:

$$g''(\pi) = \frac{l(1 - e)}{(\pi(l - 1)(1 - e) + 1)^2}$$

Since $\pi(l - 1)(1 - e) + 1 = l\pi(1 - e) + (1 - \pi(1 - e)) > 0$ we have that $g''$ is strictly decreasing in $\pi$ for $l > 1$ and strictly increasing in $\pi$ for $l < 1$. Hence, $g'$ is concave for $l > 1$ and convex for $l < 1$.

$$g''(0) = l(1 - p)$$

Now since, $\bar{l}(1 - p) < 1$, for each $x \in X$, $\frac{f_0(x)}{f_1(x)}(\pi) < \pi$ for all $\pi > 0$. Hence, beliefs always strictly decrease even under the effect of the highest signal and as a result converge to 0. □

### 5.2 Constrained optimal automata

In the section above, we focussed on conditions under which the optimal value under full memory could be achieved by a finite state automaton i.e actions taken by decision maker bounded by memory may be replicated by one who is not. Here, we consider the constrained problem, where the agent chooses the optimal automata with a fixed number of states. The definition of an automaton $\tau = < i_0, M, \sigma, d >$ and optimality are as in section 1. Note here that the payoff sequence is completely determined by the evolution of the pair $(\theta, i)$ which gives payoff $E_d(i)$. These transitions follows a markov chain $Q$ on $\Theta \times M$ and take place in the following manner:

$$Q(\theta', i' | \theta, i) = P(\theta' | \theta) \sum_{x \in X} f_{\theta'}(x)\sigma(i, x)(i')$$ (18)
defining the transition probabilities \( \alpha_{ij}^\theta := \sum_{x \in X} f_\theta(x) \sigma(i, x)(j) \), the above simplifies to:

\[
Q(\theta', i' | \theta, i) = P(\theta' | \theta) \alpha_{ii'}^\theta
\]

(19)

Let \( V(\theta, i) \) denote the value generated by \( Q \) starting from \((\theta, i)\). Interpreting the discount factor as a stopping probability, from Kocer(2010) [11] it is known that optimal automata in POMDP environments are modified multiself consistent. In the current environment, it is satisfied for beliefs generated by the unique invariant distribution \( \hat{\mu} \) on \( \Theta \times M \) of the perturbed markov chain:

\[
\hat{Q}(\theta', i' | \theta, i) = (1 - \delta)P(\theta' | \theta)\pi(\theta')\delta_{ii'} + \delta P(\theta' | \theta)\alpha_{ii'}^\theta
\]

(20)

Modified multiself consistency means the following:

**Definition**: An automaton \( \tau \) is said to be *modified multiself consistent* if for the invariant distribution \( \hat{\mu} \), the following is satisfied:

1. **Decisions**: For all \( i \in M \),

\[
d(i) \in \arg \max_{a \in \{I, NI\}} [\hat{\mu}(0|i)E_0 + \hat{\mu}(1|i)E_1]I_{a = I}
\]

(21)

2. **Transitions**: For \((i, x) \in M \times X\), if \( \sigma(i, x)(j) > 0 \), then:

\[
j \in \arg \max_{k \in M} [\hat{\pi}(\hat{\mu}(0|i), x)V(0, k) + (1 - \hat{\pi}(\hat{\mu}(0|i), x))V(0, k)]
\]

(22)

We may modify the optimal automaton in a way such that the values generated by the states with positive probability in the invariant distribution are all different. If we interpret each \( j \) as a decision taken under uncertainty yielding payoff \( V(\theta, j) \) in state \( \theta \in \Theta \), MMC tells us that states \( j \) with \( \hat{\mu}(\theta, j) > 0 \) are admissible in the set:

\[
V = \{(V(0, k), V(1, k)) : k \in M\}
\]

Moreover it can be shown (see appendix) that a *revelation principle* (As in Wilson(2014)) holds here as well i.e at a belief \( \hat{\mu}(0|i) > 0 \), the decision \((V(0, i), V(1, i))\) is optimal.

**Proposition 13 (Revelation Principle)**: Let \( i \in M \) such that \( \hat{\mu}(0|i) > 0 \), then:

\[
i \in \arg \max_{k \in M} \hat{\mu}(0|i)V(0, k) + \hat{\mu}(1|i)V(1, k)
\]

(23)
Proof: Let \( i \in M \) such that \( \hat{\mu}(0|i) > 0 \). Suppose there exists \( j \in M \setminus \{i\} \) such that \( \hat{\mu}(0|i)V(0, j) + \hat{\mu}(1|i)V(1, j) > \hat{\mu}(0|i)V(0, i) + \hat{\mu}(1|i)V(1, i) \). W.l.o.g assume that \( V(0, j) < V(1, j) \). The proof is analogous for the case \( V(0, j) > V(1, j) \). Letting \( \pi \in (0, 1) \) such that

\[
\pi V(0, j) + (1 - \pi)V(1, j) = \pi V(0, i) + (1 - \pi)V(1, i)
\]

This implies that \( \hat{\mu}(0|i) < \pi \).

Define the \( Z = (\Theta \times M \times \{s\} \times \Theta \times M) \cup (\Theta \times M \times \{n\} \times \Theta \times X \times M) \). Notice that \( Z \) can be interpreted as a tree which branches out into distinct paths at history \((\theta, k)\) either through \( s \) or \( n \). We define a probability measure \( \nu \) on \( Z \):

\[
\nu(\theta, l, s, \theta', j) := \hat{\mu}(\theta, l)(1 - \delta)\pi(\theta')\delta_{i_0}(j) \text{ for all } (\theta, k, s, \theta', j) \in \Theta \times M \times \{s\} \times \Theta \times M
\]

\[
\nu(\theta, l, n, \theta', y, j) := \hat{\mu}(\theta, l)\delta P(\theta'|\theta)f_\theta(y)\sigma(l, y)(j) \text{ for all } (\theta, l, s, \theta', y, j) \in \Theta \times M \times \{n\} \times X \times M
\]

Let \( z \in Z \), denote as \( z(\theta), z(l), z(c) \) (where \( c \in \{s, n\} \)) \( z(\theta'), z(y), z(j) \) as the value of the appropriate nodes on the path \( z \). We prove the result in the following 3 steps:

1. Define the set \( Z_i = \{ z \in Z : z(j) = i \} \), \( Z_{\theta''} = \{ z : z(\theta') = \theta'' \} \) then:

\[
\nu(Z_0|Z_i) = \frac{\nu(Z_0 \cup Z_i)}{\nu(Z_i)} = \frac{\sum_{\theta, l} \hat{\mu}(\theta, l)(1 - \delta)P(0|\theta)\pi(0)\delta_{i_0}(i) + \delta P(0|\theta)\pi(0)\sum_{x \in X} \sigma(l, x)(i)}{\sum_{\theta, l} \hat{\mu}(\theta, l)(1 - \delta)P(\theta'|\theta)\pi(\theta')\delta_{i_0}(i) + \delta P(\theta'|\theta)\pi(\theta')\sum_{x \in X} \sigma(l, x)(i)}
\]

\[
= \frac{\sum_{\theta, l} \hat{\mu}(\theta, l)Q(0, i|\theta, l)}{\sum_{\theta, l} \hat{\mu}(\theta, l)Q(0, i|\theta, l) + \sum_{\theta, l} \hat{\mu}(\theta, l)Q(1, i|\theta, l)}
\]

\[
= \frac{\hat{\mu}(0, i)}{\hat{\mu}(0, i) + \hat{\mu}(0, i)} = \hat{\mu}(0|i)
\]

Hence, \( \nu(Z_0|Z_i) < \pi \).

2. Define the sets \( Z_s = \{ z \in Z : z(c) = s \} \), \( Z_n = \{ z \in Z : z(c) = n \} \), \( Z_k = \{ z \in Z : z(l) = k \} \) , \( Z_x = \{ z \in Z : z(y) = x \} \). Clearly \( Z_s \cap Z_n = \emptyset \). Now consider the set \( L = \{(k, x) :
\( \sigma(k, x)(i) > 0 \). Note that \((k, x) \in L \iff \nu(Z_k \cap Z_x \cap Z_i \cap Z_n) > 0 \iff \nu(Z_k \cap Z_x \mid Z_i \cap Z_n) > 0 \).

For all \((k, x) \in L\) we have:

\[
\nu(Z_0 \mid Z_k \cap Z_x \cap Z_i \cap Z_n) = \frac{\nu(Z_0 \cap Z_k \cap Z_x \cap Z_i \cap Z_n)}{\nu(Z_k \cap Z_x \cap Z_i \cap Z_n)} \sum_{\theta} \hat{\mu}(\theta, k) \delta P(0 \mid \theta) f_0(x) \sigma(k, x)(i)
\]

\[
= \sum_{\theta} \hat{\mu}(\theta, k) \delta P(0 \mid \theta) f_0(x) \sigma(k, x)(i) + \sum_{\theta} \hat{\mu}(\theta, k) \delta P(1 \mid \theta) f_1(x) \sigma(k, x)(i)
\]

\[
= \frac{\sum_{\theta} \hat{\mu}(\theta, k) P(0 \mid \theta) f_0(x) + \sum_{\theta} \hat{\mu}(\theta, k) P(1 \mid \theta) f_1(x)}{\sum_{\theta} \hat{\mu}(0 \mid \theta) P(0 \mid \theta) + \hat{\mu}(0 \mid \theta) P(1 \mid \theta)} f_0(x)
\]

\[
= \hat{\pi}(\hat{\mu}(0 \mid k), x)
\]

By modified multiself consistency, we get that transitioning to \(i\) must be optimal at any \((k, x) \in L\) and hence must be better than \(j\). This implies \(\nu(Z_0 \mid Z_k \cap Z_x \cap Z_i \cap Z_n) = \hat{\pi}(\hat{\mu}(0 \mid k), x) \geq \underline{\pi}\) which in turn implies that

\[
\nu(Z_0 \mid Z_i \cap Z_n) = \sum_{(k, x) \in L} \nu(Z_0 \mid Z_k \cap Z_x \cap Z_i \cap Z_n) \nu(Z_k \cap Z_x \cap Z_i \cap Z_n) \geq \underline{\pi}
\]

3. We now prove the result. There are two cases to consider:

(a) **Case 1**: \(i \neq i_0\). In this case, \(\delta_{i_0}(i) = 0\), so \(\nu(Z_0 \mid Z_i) = 0\), hence:

\[
\hat{\mu}(0 \mid i) = \nu(Z_0 \mid Z_i) = \nu(Z_0 \mid Z_i \cap Z_n) \geq \underline{\pi}
\]

which is a contradiction.

(b) **Case 2**: \(i = i_0\). By optimality it must be the case that that \(\pi \geq \underline{\pi}\). Also in this case \(\nu(Z_0 \mid Z_i) > 0\). Now we have:

\[
\hat{\mu}(0 \mid i) = \nu(Z_0 \mid Z_i) = \nu(Z_0 \mid Z_i \cap Z_0) \nu(Z_0 \mid Z_i) + \nu(Z_0 \mid Z_i \cap Z_n) \nu(Z_n \mid Z_i) \geq \underline{\pi}
\]

which is again a contradiction.

\(\Box\)

Hence, if an individual reaches a state that recurs, his optimal choice is indeed to stay.
there. Using arguments from Wilson(2014) the following useful result can be shown:

**Proposition 14 (Categorisation):** There exist thresholds $0 = \pi_1 < \ldots \leq \pi_n \leq \ldots < \pi_{n+1} = 1$ such that $\sigma(i, x)(j) > 0$ only if $\hat{\pi}(\hat{\mu}(0|i), x) \in [\pi_j, \pi_{j+1}]$

**Proof:** Since the values are $(V(0, k), V(1, k))$ are all distinct, we can order the states in $M$ according to the value $V(0, k)$. Notice that the revelation principle implies that:

$$V(0, k) > V(0, l) \implies \hat{\mu}(0|k) \geq \hat{\mu}(0|l)$$

because if $\hat{\mu}(0|k) < \hat{\mu}(0|l)$, then the agent strictly prefers $l$ over $k$ at state $k$. Hence, the ordering is consistent with the beliefs at the states of the automaton. Now, define the thresholds as:

$$\pi_j := \min\{\pi' : i \in \arg\max_k \pi' V(0, k) + (1 - \pi') V(1, k)\}$$

$$\pi_{n+1} = 1$$

We shall show that the values defined above are indeed the desired thresholds. First we make the following observations:

1. For state $1 \in M$, $\pi_1 = 0$. This is true since $V(1, 1) > V(1, k)$ for all $k \in M \setminus \{1\}$.

2. For each $i \in M$, $\pi_{i+1} = \max\{\pi' : i \in \arg\max_k \pi' V(0, k) + (1 - \pi') V(1, k)\} =: c$. Suppose not. By definition of $\pi_{i+1}$, $\pi_{i+1} V(0, i + 1) + (1 - \pi_{i+1}) V(1, i + 1) = \pi_{i+1} V(0, i) + (1 - \pi_{i+1}) V(1, i)$ and $c V(0, i + 1) + (1 - c) V(1, i + 1) = c V(0, i) + (1 - c) V(1, i)$. Hence $c = \pi_{i+1}$

3. Notice that $\pi_i \leq \pi_{i+1}$. This is true since $V(0, i + 1) > V(0, i)$.

Now we show the result. Suppose $\sigma(i, x)(j) > 0$. Then by modified multiself consistency, it must be the case that $\hat{\pi}(\hat{\mu}(0|i), x) \in \{\pi' : j \in \arg\max_k \pi' V(0, k) + (1 - \pi') V(1, k)\}$. From 2 above, it follows that $\hat{\pi}(\hat{\mu}(0|i), x) \in [\pi_j, \pi_{j+1}]$.

$\Box$

The above result shows that posterior beliefs pre and post receipt of a signal always lie in some category of beliefs. This observation is very useful to understand the nature of the transitions taking place in an optimal automaton (as we shall see below). First, we define some notation. As in the previous section the fixed points induced by the signals $X = \{x_1, \ldots, x_n\}$ will be denote $\pi_j$ or alternatively as $\bar{\pi}_x$ (the fixed point associated with the signal $x \in X$). For each $x \in X$ define the set $M^-_x = \{k \in M : \pi_k < \bar{\pi}_x\}$, and the sets $M^+_x = \{k \in M : \pi_k > \bar{\pi}_x\}$. 

33
Proposition 15: (Optimal Automata) For the optimal automata $\tau^*$ the following are true:

1. (Bounded Jumps) For each signal $x \in X$:
   
   (a) If $i \in M^-_x$, then
   
   If $\sigma(i, x)(j) > 0$, then $j \notin \{k \in M : k < i\} \cup M^+_x$
   
   (b) If $i \in M^+_x$, then
   
   If $\sigma(i, x)(j) > 0$, then $j \notin \{k \in M : k > i\} \cup M^-_x$

2. (Bias) For $x \in X$, if $\bar{\pi}_x < \pi$, then for all $i \geq i_0$

   If $\sigma(i, x)(j) > 0$ then $j \leq i$

Hence, it is possible to go to a lower state even when the likelihood ratio of the signal is strictly greater than 1.

Proof:

1. We only prove (a). The proof for (b) is analogous. Let $i \in M^-_x$, then $\hat{\pi}(\hat{\mu}(0|i), x) = g^x(\hat{\mu}(0|i))$. We show the claim using properties of the function $g^x(\cdot)$ discussed in the previous section. There are two cases to consider:

   (a) CASE 1: Suppose that $\hat{\mu}(0|i) \leq \bar{\pi}_x$. If equality is satisfied, $\hat{\mu}(0|i) = \hat{\pi}(\hat{\mu}(0|i), x) = \bar{\pi}_x$ and we are done. Suppose $\hat{\mu}(0|i) < \bar{\pi}_x$ From the properties of the function $g^x(\cdot)$, we know that $\bar{\pi}_x$ is the unique fixed point of $g^x(\cdot)$ and that $\hat{\mu}(0|i) < \hat{\pi}(\hat{\mu}(0|i), x) = g^x(\hat{\mu}(0|i)) < \bar{\pi}_x$. This implies from Proposition 10 that $\pi_j < \hat{\pi}(\hat{\mu}(0|i), x)$. Hence, there cannot be transitions to $\{k \in M : k < i\}$. Since $\hat{\pi}(\hat{\mu}(0|i), x) < \bar{\pi}_x$, we have $\hat{\pi}(\hat{\mu}(0|i), x) < \bar{\pi}_x$ for all $k \in M^+_x$. Hence, transitions to the set $M^+_x$ are not possible.

   (b) CASE 2: Now suppose that $\bar{\pi}_x < \hat{\mu}(0|i)$. Again, from properties of the function $g^x(\cdot)$, it is the case that $\bar{\pi}_x < \hat{\pi}(\hat{\mu}(0|i), x) < \hat{\mu}(0|i) \leq \bar{\pi}_x$ for all $k \in M^+_x$. Hence,

2. Note that it is sufficient to show the claim for the initial state. We consider two cases:

   (a) CASE 1: Suppose $\hat{\mu}(0|i_0) \leq \bar{\pi}_x$. By optimality, we know that $\pi \in [\bar{\pi}_0, \bar{\pi}_{i_0+1}]$. Hence, for all $j > i_0$ it is the case that $\hat{\pi}(\hat{\mu}(0|i_0), x) \bar{\pi}_x < \pi \leq \bar{\pi}_j$. Hence transitions to $j > i_0$ cannot happen.

   (b) CASE 2: Suppose $\hat{\mu}(0|i_0) > \bar{\pi}_x$, then the result follows from arguments in 1.
References


6 Appendix

6.1 A revelation principle for POMDP under bounded memory

This section establishes a revelation principle in the sense of Wilson(2014) for optimal plans under bounded memory in environments where the decision maker faces a partially observable markov decision process (POMDP). POMDP environments are general decision problems with state unobservability and encompass the investment problems with unchanging and changing worlds discussed in the paper. We develop the framework as in
Kocer (2010) and then establish a revelation principle using modified multiself consistency.

There is a finite set of states of the world $\Theta$. The agent takes an action every period from a finite set $A$ and receives per period payoffs according to utility function $u : A \times \Theta \to \mathbb{R}$. Hence, if in a given period action $a \in A$ is chosen and the current state of the world is $\theta \in \Theta$, the agent receives payoff $u(a, \theta)$ in that period. The state of the world is however unobservable to the agent, evolves randomly and can be inferred from a signals received from finite set $X$. The state evolution and signal generation process is determined jointly according to a transition function $P : \Theta \times A \to \Delta(\Theta \times X)$. The interpretation is as follows: If the current state is $\theta \in \Theta$ and current action taken is $a \in A$, then the next state $\theta' \in \Theta$ and signal $x \in X$ are both jointly determined according to probability $P(\theta', x|\theta, a) \equiv P(\theta, a)(\theta', x)$. The objective of the agent is to maximise long run discounted expected utility according to discount factor $\delta \in (0, 1)$ i.e maximise:

$$\mathbb{E}\left[ \sum_{t=0}^{\infty} (1 - \delta)^t u(a_t, \theta_t) \right]$$

(24)

The agent uses an automaton $\tau = \langle g_0, M, \sigma, d \rangle$ to maximise (24). Here, $g_0 \in \Delta(M)$, $\sigma : M \times X \to \Delta(M)$ and $d : M \to \Delta(A)$. Note that once an automaton is fixed the long run payoffs are completely determined by the evolution of the pair $(\theta, i)$ over time. This happens according to a markov process:

$$Q(\theta', i'|\theta, i) = \sum_{a \in A} d(i)(a) \sum_{x \in X} P(\theta', x|\theta, a)\sigma(i, x)(i')$$

(25)

Interpreting the discount factor as a continuation probability that explicitly enters the extensive form of the decision making environment, we obtain a long run distribution $\hat{\mu}$ over pairs in $\Theta \times M$ determined by:

$$\hat{\mu}(\theta, i) = \sum_{t=0}^{\infty} (1 - \delta)^t \mathbb{P}^t(\theta, i)$$

(26)

where $\mathbb{P}^t(\theta, i)$ denotes the probability of the pair $(\theta, i)$ occurring at time $t$ under the transitions induced by $Q$ above in (25). It can be shown that $\hat{\mu}$ is the unique invariant distribution on $\Theta \times M$ of the perturbed markov chain:

$$\hat{Q}(\theta', i'|\theta, i) = (1 - \delta)\mathbb{P}^0(\theta')g^0(i') + \delta Q(\theta', i'|\theta, i)$$

(27)
Denote the continuation values defined by each pair \((\theta, i)\) under an automaton be denoted as \(V(\theta, i)\). Kocer(2010) has shown modified multiself consistency of optimal automata under beliefs generated by \(\hat{\mu}\). A formal definition is given below:

**Definition**: An automaton \(\tau\) is said to be *modified multiself consistent* if for the invariant distribution \(\hat{\mu}\), the following conditions are satisfied:

1. **Decisions**: For all \(i \in M\) with \(\hat{\mu}(i) > 0\), \(d(i)(a) > 0\) implies:
   \[
   a \in \arg\max_{a' \in A} \sum_{\theta} \hat{\mu}(\theta|i)[u(a', \theta) + \delta \sum_{(\theta', x) \in \Theta \times X} P(\theta', x|\theta, i)V(\theta', i')]
   \]  
   (28)

2. **Transitions**: For \((i, x) \in M \times X\), if \(d(i)(a) > 0\) and \(\sigma(i, x)(j) > 0\), then:
   \[
   j \in \arg\max_{k \in M} \sum_{\theta' \in \Theta} \hat{\pi}(\hat{\mu}(\cdot|i), a, x)(\theta')V(\theta', k)
   \]  
   (29)

   where \(\hat{\pi}(\hat{\mu}(\cdot|i), a, x) \in \Delta(\Theta)\) is the belief about the next state when action \(a \in A\) is taken and signal \(x \in X\) is received:

   \[
   \hat{\pi}(\hat{\mu}(\cdot|i), a, x)(\theta') = \frac{\sum_{\theta} \hat{\mu}(\theta|i)d(i)(a)P(\theta', x|\theta, a)}{\sum_{\theta'' \in \Theta} \sum_{\theta} \hat{\mu}(\theta|i)d(i)(a)P(\theta'', x|\theta, a)} = \frac{\sum_{\theta} \hat{\mu}(\theta|i)P(\theta', x|\theta, a)}{\sum_{\theta'' \in \Theta} \sum_{\theta} \hat{\mu}(\theta|i)P(\theta'', x|\theta, a)}
   \]  
   (30)

We now state and prove the following revelation principle:

**Proposition 16**: For each \(i \in M\) such that \(\hat{\mu}(i) > 0\):

\[
\begin{align*}
i & \in \arg\max_{k \in M} \sum_{\theta} \hat{\mu}(\theta|i)V(\theta, k) \\
\end{align*}
\]  
(31)

**Proof**: Suppose not. Then there exists a \(j \in M\setminus\{i\}\) such that:

\[
\begin{align*}
\sum_{\theta} \hat{\mu}(\theta|i)V(\theta, j) & > \sum_{\theta} \hat{\mu}(\theta|i)V(\theta, i)
\end{align*}
\]  
(32)
Defining the set \( \Pi(j, i) = \{ \pi \in \Delta(\Theta) : \sum_{\theta \in \Theta} \pi(\theta)V(\theta, j) > \sum_{\theta \in \Theta} \pi(\theta)V(\theta, i) \} \) we observe that \( \hat{\mu}(\cdot|i) \in \Pi(j, i) \). Notice that \( \Pi(j, i), \Pi(j, i)^c \) are both convex.

Now define the set \( Z = (\Theta \times M \times [s] \times \Theta \times M) \cup (\Theta \times M \times [n] \times A \times \Theta \times X \times M) \). Notice that \( Z \) can be interpreted as a tree which branches out into distinct paths at history \((\theta, k)\) either through \( s \) or \( n \). We define a probability measure \( \nu \) on \( Z \):

\[
\nu(\theta, l, s, \theta'', j) := \hat{\mu}(\theta, l)(1 - \delta)\pi(\theta'')(g^0(j)) \quad \text{for all} \quad (\theta, k, s, \theta'', j) \in \Theta \times M \times [s] \times \Theta \times M
\]

\[
\nu(\theta, l, n, \theta'', y, j) := \hat{\mu}(\theta, l)\delta d(l)(a)P(\theta'', x|\theta, a)\sigma(l, y)(j) \quad \text{for all} \quad (\theta, l, n, a, \theta'', y, j) \in \Theta \times M \times [n] \times A \times \Theta \times X \times M
\]

For \( z \in Z \), denote as \( z(\theta), z(l), z(c) \) (where \( c \in [s, n] \)) \( z(\theta''), z(y), z(j) \) as the value of the appropriate nodes on the path \( z \). We now establish the result via the following steps:

(a) Define the sets \( Z_i = \{ z \in Z : z(j) = i \}, Z_{\theta'} = \{ z : z(\theta'') = \theta' \} \) then:

\[
\nu(Z_{\theta'}|Z_i) = \frac{\nu(Z_{\theta'} \cap Z_i)}{\nu(Z_i)} = \frac{\sum_{\theta, l} \hat{\mu}(\theta, l)(1 - \delta)\pi(\theta'')(g^0(i)) + \delta \sum_{a \in A} d(l)(a) \sum_{x \in X} P(\theta'', x|\theta, a)\sigma(i, x)(i'))}{\sum_{\theta, l} \hat{\mu}(\theta, l)(1 - \delta)\pi(\theta'')(g^0(i)) + \delta \sum_{a \in A} d(l)(a) \sum_{x \in X} P(\theta'', x|\theta, a)\sigma(i, x)(i'))}
\]

\[
= \frac{\sum_{\theta, l} \hat{\mu}(\theta, l)\hat{Q}(\theta'', i|\theta, l)}{\sum_{\theta, l} \hat{\mu}(\theta, l)\hat{Q}(\theta'', i|\theta, l)} = \frac{\hat{\mu}(\theta', i)}{\sum_{\theta'} \hat{\mu}(\theta'', i)} = \hat{\mu}(\theta'|i)
\]

Hence we have \( \{ \nu(Z_{\theta'}|Z_i) \}_{\theta' \in \Theta} \in \Pi(j, i) \).

(b) Define the sets \( Z_s = \{ z \in Z : z(c) = s \}, Z_n = \{ z \in Z : z(c) = n \}, Z_a = \{ z \in Z : z(a') = a \}, Z_k = \{ z \in Z : z(l) = k \}, Z_x = \{ z \in Z : z(y) = x \} \). Clearly \( Z_s \cap Z_n = \emptyset \).
Now consider the set \( L = \{(k,a,x) : \nu(Z_i \cap Z_n \cap Z_k \cap Z_a \cap Z_x) > 0\} \). Note that if 
\((k,a,x) \in L \) then \(d(k)(a) > 0\) and \(\sigma(k,x)(i) > 0\). For all \((k,a,x) \in L \) we have:

\[
\nu(Z_{\theta'}|Z_i \cap Z_n \cap Z_k \cap Z_a \cap Z_x) = \frac{\nu(Z_{\theta'} \cap Z_i \cap Z_n \cap Z_k \cap Z_a \cap Z_x)}{Z_i \cap Z_n \cap Z_k \cap Z_a \cap Z_x} \\
\sum_{\theta} \mu(\theta,k)\delta d(k)(a)P(\theta', x|\theta,a)\sigma(k,x)(i) \\
\sum_{\theta} \mu(\theta,k)\delta d(k)(a)P(\theta'', x|\theta,a)\sigma(k,x)(i) \\
\sum_{\theta} \mu(\theta,k)P(\theta', x|\theta,a) \\
\sum_{\theta''} \mu(\theta,k)P(\theta'', x|\theta,a) \\
= \hat{\pi}(\mu(.|k), a,x)(\theta')
\]

By modified multiself consistency, we get that transitioning to \( i \) must be optimal at any \((k,a,x) \in L \) and hence must be better than \( j \). This implies \( \nu(Z_{\theta'}|Z_i \cap Z_n \cap Z_k \cap Z_a \cap Z_x)_{\theta' \in \Theta} = \hat{\pi}(\mu(.|k), a,x) \in \Pi(j,i)^c \). From the convexity of \( \Pi(j,i)^c \) it follows that:

\[
< \nu(Z_{\theta'}|Z_i \cap Z_n) >_{\theta'} = < \nu(Z_{\theta'}|Z_i \cap Z_n \cap Z_k \cap Z_a \cap Z_x)\nu(Z_k \cap Z_a \cap Z_x|Z_i \cap Z_n) >_{\theta' \in \Theta}
\]

belongs to the set \( \Pi(j,i)^c \).

(c) We now prove the result. There are two cases to consider:

i. Case 1: \( g^0(i) = 0 \). In this case, \( \nu(Z_i|Z_i) = 0 \). So, \( \nu(Z_n|Z_i) = 1 \) hence:

\[
< \hat{\mu}(\theta'|i) >_{\theta' \in \Theta} = < \nu(Z_{\theta'}|Z_i) >_{\theta' \in \Theta} = < \nu(Z_{\theta'}|Z_i \cap Z_n) >_{\theta' \in \Theta} \in \Pi(j,i)^c
\]

which contradicts (10).

ii. Case 2: \( g^0(i) > 0 \). By optimality it must be the case that \( \pi^0 \in \Pi(j,i)^c \). Moreover, \( \pi^0 = < \nu(Z_{\theta'}|Z_i \cap Z_n) >_{\theta' \in \Theta} \). Also in this case \( \nu(Z_i|Z_i) > 0 \). Now
from the convexity of $\Pi(j, i)^c$, we have:

$$< \hat{\mu}(\theta'|i) >_{\theta' \in \Theta} =< \nu(Z_{\theta'}|Z_i) >_{\theta' \in \Theta}$$

$$= < \nu(Z_{\theta'}|Z_i \cap Z_n) \nu(Z_{\theta'}|Z_i) + \nu(Z_{\theta'}|Z_i \cap Z_{m}) \nu(Z_{n}|Z_i) >_{\theta' \in \Theta}$$

$$= < \pi(\theta') \nu(Z_{\theta'}|Z_i) + \nu(Z_{\theta'}|Z_i \cap Z_n) \nu(Z_{n}|Z_i) >_{\theta' \in \Theta}$$

$$\in \Pi(j, i)^c$$

which again contradicts (10).

\[ \square \]