On the optimal stopping time of learning

Anna Fedyszak-Koszela

November 11, 2008
Abstract

The goal of this thesis is to study the economics of computational learning. Attention is also paid to applications of computational learning models, especially Valiant’s so-called ‘probably approximately correctly’ (PAC) learning model, in econometric situations.

Specifically, an economically reasonable stopping time model of learning is the subject of two attached papers. In the first paper, Paper A, the economics of PAC learning are considered. It is shown how a general form of the optimal stopping time bounds can be achieved using the PAC convergence rates for a ‘pessimistic-rational’ learner in the most standard binary case of passive supervised PAC model of finite Vapnik-Chervonenkis (VC) dimension.

The second paper, Paper B, states precisely and improves the ideas introduced in Paper A and tests them in a specific and mathematically simple case. Using the maxmin procedure of Gilboa and Schmeidler the bounds for the stopping time are expressed in terms of the largest expected error of recall, and thus, effectively, in terms of the least expected reward. The problem of locating a real number $\theta$ by testing whether $x_i \leq \theta$, with $x_i$ drawn from an unknown probability measure, is studied. Examples of stopping time bounds calculated for a range of term rates, sample costs and rewards/penalties from a recall are included. The standard econometric situations, such as product promotion, market research, credit risk assessment, and bargaining and tenders, where such bounds could be of interest, are pointed.

These two papers are the essence of this thesis, and form it together with an introduction to the subject of learning.

Acknowledgements

I would like to thank my colleagues at the School of Education, Culture and Communication, Division of Applied Mathematics (formerly the Department of Mathematics and Physics) at Mälardalen University, for their support, and especially the School Management for giving me the necessary time to write this thesis. I also thank my supervisors, Richard Bonner and Dmitrii Silvestrov, for their supervision, encouragement and inspiration.
# Contents

1 Introduction
   1.1 Background ........................................... 2
   1.2 Learning in economics ............................... 5
   1.3 Computational Learning Theory ....................... 8
   1.4 Stopping time, rationality, learning ............... 11

2 Summary of papers ............................... 14
   2.1 When to stop learning? Bounding the stopping time in the
       PAC model ........................................ 14
   2.2 Learning a real number, rationally ................ 15

3 Problems for further study ....................... 17

Bibliography .................................. 22
Chapter 1

Introduction

This thesis aims to construct an economically reasonable stopping time model of learning. Chapter 1 introduces economic and computational learning emphasizing stopping time problems. In Chapter 2, two papers, that together form the main part of this thesis, are discussed - Paper A: ‘When to stop learning? Bounding the stopping time in the PAC model’ [4], and Paper B: ‘Learning a real number, rationally’ [5]. They can be found in their entirety after the Bibliography. Conclusively, Chapter 3 presents a selection of questions for the future.

1.1 Background

What is learning? How is learning defined in modern science? What are the differences and similarities between the learning models used in the social, the natural and the computing sciences? How is the problem of stopping time modeled? Questions concerning the process of learning, its aim, fundamental properties and consequences have accompanied mankind for a long time. The answers have been sought by philosophers, artists, scholars, politicians, and all others who saw the importance of learning, not only for individual human beings, but also for society as a whole.

The primary element of learning, common to all these approaches, is change. A non-changing system does not learn. Not every change, however, leads to learning; as forgetting is also a form of change. In learning, the change must result in an improvement, e.g. a rise of system efficiency in performing a
function, according to a preset criterion. When people learn, their tasks are
determined by, in addition to biological factors, society and social institu-
tions. The ecological environment or, for example, an animal trainer, assign
roles to animals in the process of learning. As for a machine or computer
program, their functions are defined by the designer, programmer or simply
the user. Sometimes changes improve the system in respect to some func-
tions, and make the system worse regarding others. Not even every positive
change of the system can be considered learning - any improvement must be
autonomous; the learning system must undergo improvements by itself. This
condition is, however, inadequate; any autonomous improvements that occur
accidentally and spontaneously cannot be treated as learning. For any such
improvement to be considered in this way, it must be induced by the envi-
ronment in which the system performs. This can take place either through
observations performed by the system, or information obtained from the en-
v
vironment, related to its functions, and results in changes to its functionality.

A simplified, yet fundamental, definition of learning can be made as follows:
*learning is any autonomous change in the system as a result of experiences,
observations and/or new information gathered from the environment, that
leads to improvements in system functionality.*

Note that when introducing the above improvement condition, the assump-
tion is made that change can be qualitatively measured. While system
changes can be proven in a more or less objective manner, the necessary cri-
teria for establishing whether such changes are improvements become rather
arbitrary and strongly depend on the observer’s point of view.

A review of various philosophical theories through the eyes of contempo-
rary learning theory can be found in Churchland [9]. In the search for more
cr
crete answers to how humans learn, knowledge about learning in nature
seems helpful. Thus we have psychologists studying learning in animals and
humans [8, 13], psychologists and neurobiologists studying the relationship
between physiology and learning in animals [42, 37] or geneticists studying
evolution on the genetical level [35, 10, 34, 47].

In psychology the notion of learning is usually defined as a *relatively per-
manen t change in behavioral potentiality that occurs as a result of reinforced
practice* (Kimble [33]). Thus learning is defined by its consequences - changes
in behaviour. Many psychologists agree that learning cannot be observed directly. Different learning theories can only be evaluated based on their effects on behaviour.

When trying to describe economic behaviour, economists often use models from psychology to help them understand and predict economic behaviour. During a long time period the rationality of behaviour was more important than the process of learning.

In the last decades this situation has changed. For example, in Sargent [43] one can find arguments for ‘bounded rationality’ in economic behaviour illustrated by changes in Eastern Europe in the early 1990s, while Evans and Honkapohja (in [22]) present a new statistical learning approach helpful when constructing macroeconomic models with expectational indeterminacy. For a wide review of learning models in relation to various concrete economics situations, ranging from learning of preferences through reinforcement learning in games to the evolution of cooperation, see Brenner [7].

In computer science the purpose of learning theory is to construct computer programs that automatically improve with experience. In other words, an attempt is made to determine whether a self-learning machine is possible to design. To this end, various concepts and results have been adopted from statistics, artificial intelligence, information theory, computational complexity as well as from philosophy, biology, cognitive science and control theory. Here the modeling of learning seems to be much easier in comparison to learning models in economics, since the learner is, in this case, a machine. The designer of a learning machine has a crucial influence both on the machine’s software and hardware designs, which, in turn, determine the learning process rules, forming its preferences and level of rationality. The designer can also predict the performance of the environment - usually a much simpler one - more extensively. It is for these reasons (as well as the close relationship between computer science and mathematics) that computational learning models tend to, in comparison to economic models, be more precise and formally defined.

One can find the theory of computational learning in monographs [3, 12, 32, 51] and survey papers [1, 2, 20, 30, 31].
1.2 Learning in economics

Even though theoretical economics lacks a general and well formalized learning theory, questions related to learning are not unfamiliar to economists. In seeking answers to various questions, they often consider learning methods mainly found in the statistical or machine learning theories (see survey papers [46, 16] and papers with particular applications [25, 23, 19, 21, 36]). Common questions, inherently linked to learning theory, that economists try to answer, might be whether imitation can be the result of rational choice, whether people can learn enough to make useful decisions by observing the choices and experiences of others, or how institutions can be designed to maximize learning efficiency, and if such institutions can be designed to perform well even in the presence of incomplete learning.

The learning setting can be classified on the basis of three characteristics: the strategic environment; the way in which agents collect information; and the degree of rationality of the agent. Any discussion of learning must begin by assuming that agents need to learn something. Consequently, agents in learning models must fail to be fully informed or fully rational. Agents can collect information passively or actively; the collection process can be free or costly. Passive information collection takes place as an outcome of an adaptive process. Agents are unable to influence the quantity or quality of the information that they obtain. Learning is active when agents’ choices determine the flow of information.

Models of learning based on limited information neglect other possible reasons for learning. If agents have limited ability to solve constrained optimization problems, then the learning process could describe how they look for optima. If agents have limited ability to collect information, then the learning process could describe how they look for information. A general rule is that the more complicated the strategic environment, the simpler the information collection technology and the simpler the behavior followed by the agents. The models usually feature simplicity over accuracy. Likewise, information collection by agents is simulated in the simplest possible way.

In general the learning models considered in economics can be divided into three categories (see even Sobel [46]): models in which there is a single agent, models of observational and social learning, and (game-theoretical) models
in which the actions of one agent have a direct influence on the payoffs of the other agents.

Individual learning assumes that agents follow active learning rules. The agents have an accurate model of their environment and make optimal decisions within their model. Two-armed bandit problems are an example of individual learning models. An agent only chooses between two options that differ in quality, and he does not know which choice is better. When making a decision the agent observes whether the job was done well or badly. His payoff is the discounted sum of payments associated with the quality of choices already made. The agent only observes a performance of the current choice. This costly observation - as the only way to determine the results of a choice is to make it - enables the agent to pick the better choice. The optimal policy thus leads to frequent switching between options. Another group of individual learning models are statistical learning models, where information arrives each time independently of choices made by an agent. The agent will eventually obtain all available information and, given this information, all his decisions will be optimal. The models provide the necessary requirements regarding experiment conditions, payoff functions and the environment state, for the agent to either learn completely, or at least enough to make the optimal decision. However, these models do not describe the actual behaviour of an agent in real-life economics. To gather new information does not require any effort by the agent, though he usually pays a price in terms of foregone utility for taking advantage of this information. Thus, the agent will not necessarily learn the optimal decision in cases when the cost of acquiring new information becomes comparable with the benefit. Moreover, the agent’s recall is bounded in a stationary but natural environment, which obviously differs significantly from the artificial environment used in the models.

Social learning concerns many agents. There is no strategic interaction between them - the actions of one agent do not affect the payoff of another. The information about actions and payoffs of agents is available to them and provides data about the state of the world. Agents do not choose how much information they receive and their actions do not change the environment. In the observational learning models, that are a simple subset of the social learning, each agent only makes one decision, and all decisions are made sequentially. Each agent has access to information about past decisions. The
agent may also have private information. Each agent’s payoff depends only on the underlying state of the world and on his decision, but does not depend directly on decisions made by other agents. The main feature of these models is that learning is incomplete. They, however, often show informational cascade situations when agents, based on observing others, make the same choices independently of private information. In these cases a collective decision sometimes leads to completely incorrect results. More complex models of social learning take into account the possibility that agents may act simultaneously, and some agents may repeat their actions. The word-of-mouth transmission of information models provide conditions - on the amount and weight of information the agents receive, and on the way they make decisions - under which the behaviour of agents leads to efficient learning instead of just to information cascades. The observational learning models give reasonable and useful predictions. They were successfully used to explain a wide range of phenomena in economics, from bank runs, when large numbers of bank customers withdraw their deposits in a state of panic, to popular dining choices of social groups.

When a group of various agents attempts to learn from the experiences of other, preceding, groups, specific social interactions take place. It is these interactions that Manski [36], among others, provides an accurate analysis of. It is shown that the social learning process features a sequential reduction in ambiguity over time. To illustrate, Manski uses the problem of learning about innovations.

In the game-theoretical models agents must learn in fully strategic environments, and then they determine their actions using a simple model of their opponents’ behavior. The beginnings of evolutionary game theory can be found in applications of the mathematical theory of games to biological contexts. A key idea is the introduction of a strategic aspect to biological evolution. The field has recently become of increased interest to economists and social scientists. In evolutionary game models applied to economics, agents seldom change strategies, and when they do, a strategy based on its recent success is most often adopted. Agents do not actively gather information. Their strategic decisions determine the payoffs of other strategies, and by doing this they influence future decisions. The models have been used to solve the selection problem in general games with multiple, but not exact, equilibria. Due to adaptive learning, games in which there is a unique efficient
payoff tend to drift away from any inefficient outcome. When the population reaches an efficient outcome it tends to stay there. The evolutionary game models, in which learning is described by the Bayes rule, provide surprisingly powerful results in economics. The probability of an event A conditional on another event B is in general different from the probability of B conditional on A. Bayes’ theorem defines the relationship between the two. In these models agents begin playing with beliefs about their opponents’ strategies that are not necessarily correct but reasonably realistic. These beliefs will, when updated - using observations about how the game proceeds, and Bayes’ rule - usually generate nearly correct predictions about the future of the game. The accuracy of these predictions depends on Bayesian updating, but not on the mentioned assumptions about how other players select their strategies.

1.3 Computational Learning Theory

Traditionally, in computational sciences, the process of learning has been studied mainly by two schools: Statistical Learning Theory (Vapnik [50], Harman and Kulkarni [29]) and Computational Learning Theory (Valiant [48], Anthony and Barlett [3], Kearns and Vazirani [32]). For a mathematical account of the ground ideas, see Smale and others [44, 11, 12, 38], and see [17, 18, 15, 41, 45, 40] for examples of the algorithms involved.

In rational learning - according to Bayesian reasoning, as mentioned in Section 1.2 - one assumes that the quantities of interest are governed by probability distributions and that optimal decisions can be made by reasoning about these probabilities together with observed data. In Bayesian methods, knowledge about the prior probabilities of alternate hypotheses, as well as about the probability of observing various data given these hypothesis, is required. These methods allow assigning a posterior probability to each candidate hypothesis, based on this knowledge and observed data.

On the other hand, learning - as understood in computational sciences - does not assume ‘a priori’ probability distribution knowledge, and involves optimal decisions made by non-Bayesian reasoning. It is therefore often called ‘learning with bounded rationality’ or simply ‘empirical’ learning. In the 1980s Valiant [48] and others, developed a theory known as CoLT (Computational Learning Theory), based on statistical theories from Vapnik [50, 49] and uni-
form limit theorems [14, 39]. It is in CoLT that the ‘probably approximately
correctly’ learning model, PAC - the basis of the study of computational
learning economics undertaken in this thesis - is defined.

In PAC learning, one considers algorithms that learn target functions, from
some function class, using training examples drawn at random according to
an unknown, but fixed, probability distribution. A formulation of PAC learn-
ing can be found in Paper B, while a more intuitive summary of the model
follows.

Consider the following: a learner is tasked with creating a hypothesis that will
simulate a specific function \( f \) in a specific function class \( F \). He has been given
samples \((x_i, y_i)\), which are information about \( f \) in points \( x_i \), i.e. \( y_i = f(x_i) \)
for \( i = 1, 2, ..., n \). The learner’s task could be defined as constructing a hypo-
thetical function \( h_n(z_1, ..., z_n) \) (also belonging to class \( F \)), which would yield
the same results as the sought after function \( f \) using the samples given to
him, i.e. such a function \( h_n(z_1, ..., z_n)(x_i) = f(x_i) = y_i \) for \( 1 \leq i \leq n \). He also has very limited knowledge about the following:

(1) a measurable space \( X \) with a class \( M \) of probability measures,

(2) a measurable metric space \((Y, d)\) and a class \( F \) of measurable functions
on \( X \) with values in \( Y \), and,

(3) a class \( H \) of consistent learning algorithms for \( F \).

The sets \( X \) and \( Y \) come with \( \sigma \)-algebras of subsets, on which all measures
live, and the metric \( d \) is measurable on the product \( Y \times Y \). A learning
algorithm for \( F \) is a sequence \( h \) of functions \( h_n \) on \((X \times Y)^n\), \( n = 1, 2, ..., \), with values in \( F \); one writes the value of \( h_n(z_1, ..., z_n) \) at \( x \in X \) as
\( h(z_1, ..., z_n; x) \), and \( h_n(z_1, ..., z_n) \) as \( h_n \). An algorithm \( h \) is consistent for \( F \) if
\( h((x_1, f(x_1)), ..., (x_n, f(x_n)); x_i) = f(x_i) \) for \( i \leq n \) and all \( n \).

The environment now picks a distribution \( \mu \in M \), a sequence \( x_1, x_2, \ldots \) of
independent \( \mu \)-distributed points in \( X \), and a function \( f \in F \). The learner
first picks a learning algorithm \( h \in H \) and two (small) positive numbers \( \epsilon \)
and \( \delta \), and then sequentially queries the environment for examples \((x, f(x))\),
that is, he observes the sequence \((x_1, f(x_1)), (x_2, f(x_2)), \ldots \) for as long as he
likes, and, using his algorithm $h$, he chooses ‘hypotheses’ $h_1$, $h_2$, ... in $F$, each consistent with $f$ on all past observations.

He may compute, for every positive integer $n$, every function $f$ in $F$, and every probability measure $\mu$ in $M$, the probability

$$p_n(h) = p_n(\mu, f; h_1, x_2, \ldots, x_n; \epsilon)$$

that for any $m > n$ the expectation of the error $d(f(x_m), h_n(x_m))$ of his prediction of $f(x_m)$ as $h_n(x_m)$ is less than $\epsilon$.

He may stop observing the random sequence the first time $p_n(h)$ is larger than $1 - \delta$ uniformly in $f \in F$ and $\mu \in M$, claiming to have learned the unknown function $f$ in time $n = n(F, h; M; \epsilon, \delta)$ with accuracy not worse than $\epsilon$ and confidence not less than $1 - \delta$.

In function of its two last arguments, the number $n(F, h; M; \epsilon, \delta)$ tells the learner how fast, in the worst case, his algorithm $h$ learns a function $f \in F$ in a state $\mu \in M$ of the world. If, in estimating the expected error, the algorithm $h$ is ‘maximized away’ over the class $H$ of all consistent algorithms, and should $M$ contain all probability measures on $X$, then the time bound $n = n(F; \epsilon, \delta)$ for learning elements of $F$ in a stable but otherwise unknown environment, is a ‘measure of complexity’ of $F$.

A class $F$ is then ‘learnable’ if $n(F; \epsilon, \delta)$ is finite for all $\epsilon, \delta > 0$; the class is learnable in polynomial time if $n(F; \epsilon, \delta)$ is moreover bounded by a polynomial in $1/\epsilon$ and $1/\delta$; etc.

For the binary case $Y = \{0, 1\}$ it is well known that the learnable and the polynomially learnable classes $F$ are the same, namely those of finite Vapnik–Chervonenkis (VC) dimension (see Wenocur and Dudley [52]), where VC dimension means that there is a bound on the size of the finite subsets of $X$ from which extrapolation in $F$ is always possible: the VC dimension of $F \subset 2^X$ is the largest cardinality of a finite set $S \subset X$ such that the map $F \rightarrow 2^S$ of restriction to $S$ is surjective.
1.4 Stopping time, rationality, learning

The question of optimal stopping time is the problem of choosing when to take a given action based on sequentially observed random variables in order to maximize an expected payoff or to minimize an expected cost. The action taken may be to test a hypothesis or to estimate a parameter (in statistics), and to replace a machine, hire a secretary, or reorder stock (in operations research). The Bayesian perspective on the optimal stopping problem, where the joint distribution of observed random variables is assumed to be known, is described in an electronic text by Ferguson [24]. A different approach to the optimal stopping time problem, as presented in Papers A and B, is the simple PAC model of learning. This model is also discussed in [6] in relation to rational choice and learning.

Generally, computational learning models have their roots in computer science. In contrast to economics, in this field investigating the lower bounds of the learning time, in function of the accuracy of recall, is an important objective of learning. In economics however, the upper bounds - especially related to the utility of what has been learned, with the learning costs taken into account - are also important.

Suppose now (see also Paper B) that whenever a learner (named Alice in Paper B) encounters a point \( x \in X \), he may choose either to view \( f(x) \) (to ‘learn’), or to predict \( f(x) \), that is, to guess ‘\( f(x) = y \)’ based on what he has learnt about \( f \) so far (to ‘recall’). Suppose further that doing the former costs him \( c(x) \) dollars, while doing the latter earns him \( \rho(x, d(y, f(x))) \) dollars, with the function \( \rho(x, \xi) \) known to him. Suppose finally that the learning occurs at times \( t = 0, 1, 2, \ldots, s - 1 \), and that it is from then on indefinitely followed by recall, whereby for all \( t \geq s \) of the learner’s guesses ‘\( f(x_t) = h_{s-1}(x_t) \)’.

Now the learner, like mentioned in Section 1.3, simulates a specific function \( f \). This time around it is a project where, using data \( x \) supplied by a client, he constructs a hypothesis \( h \) that should correctly compute \( f(x) \).

The clients, each identified by their data \( x \in X \), arrive independently, following an unknown but fixed probability distribution. The learner has knowledge of a certain function class \( F, f \in F \), based on his client cases. Every new case
allows him to narrow $F$ even further; another option would be to purchase additional data from competing learners. Now, is it feasible for the learner to engage in this project with, or without purchasing samples? If samples are bought, how many should they be?

The learner must start by making these decisions (obviously at $t = 0$, because no learning has yet taken place). Any decisions are final: he is not to decide ‘dynamically’ during learning, where subsequent decisions could depend on the consequences of the previous ones.

He must thus in advance propose a numerical value $v(s)$ for the project, in function of its sample size $s = 0, 1, 2, \ldots$, which is normalized so that the value of doing nothing is zero. He commits to the project if and only if $v(s) \geq 0$ for some $s$, which he chooses so as to maximize $v(s)$. He discounts his future indefinitely at a fixed rate.

The learner starts by computing the expectation $\phi$ with respect to $\mu$ of his payoff from sampling $f$ at times $t = 0, 1, 2, \ldots, s-1$, and then guessing $f(x_t)$ indefinitely as $h_{s-1}(x_t)$. He may do this symbolically for every possible value of his ‘ambiguity variable’ $\alpha = (f, h, \mu)$, which he has no way of knowing the true value of, and for every value of his ‘decision variable’ $s$. The expected payoff $\phi$, which represents the value of any prospective learning and recall project, is thus a function of $\alpha$ and $s$. Then $\phi = \phi(\alpha, s)$ meaning ‘a learning and recall process with expected payoff $\phi$’.

The learner now minimises away his ambiguity, in effect employing the so-called maxmin expected utility criterion (of Gilboa and Shmeidler [27]). To this end, he sets $v(s)$ to his maxmin subjective utility

$$\phi^* := \max_s \phi(s) = \max_s \min_\alpha \phi(\alpha, s).$$

Let $\phi(s^\sharp) := \max_s \min_\alpha \phi(\alpha, s)$, assuming that conditions are met for $s^\sharp$ to exist. Call $s^\sharp$ the optimal sample size, or optimal stopping time, and call $\phi^*$, the value of the project.

The learner now decides as follows. He engages in a project $\phi$ if and only if its value $\phi^*$ is non-negative; he calls such a project feasible. Obviously, for real $q$, $\phi^* \geq q$ if and only if there is such an $s$ that $\phi(\alpha, s) \geq q$ for all $\alpha$.  

12
Hence, a project $\phi$ is feasible if and only if there is a stopping time $s$, for which the expected utility $\phi(\alpha, s)$ is nonnegative irrespective of $\alpha$.

Moreover, he commits to the project without learning if and only if $s^\sharp$ is zero.

If committed to learning, he will of course stop learning at an optimal stopping time $s^\sharp$. Should there be several such times, he chooses the smallest. The learner may bound $s^\sharp$, quite effectively at times, by a simple ‘sandwich argument’:

If $\gamma_1(s) \leq \phi(\alpha, s) \leq \gamma_2(s)$, and as follows $\gamma_1(s) \leq \min_{\alpha} \phi(\alpha, s) \leq \gamma_2(s)$, then $s^\sharp$ is bound to the set where $\max \gamma_1 \leq \gamma_2(s)$.

Of course, this simple trick works only if the last inequality is easy to solve, and if its solution set is not too large.
Chapter 2

Summary of papers

2.1 When to stop learning? Bounding the stopping time in the PAC model

The main purpose of Paper A is to consider the economics of a specific computational model of learning, namely Valiant’s PAC model. In the most standard binary case of passive supervised PAC model of finite Vapnik-Chervonenkis (VC) dimension, a general optimal stopping time bounds form is achieved for a ‘pessimistic-rational’ learner (Paper A, Theorem 4.1):

For any $\epsilon > 0$, there exist positive numbers $B_1(F)$ and $B_2(\epsilon) = B_2(F, \epsilon)$ (depending on the learned concept class $F$ only) such that the optimal stopping time for the least favorable case of the restricted binary PAC learning model of finite Vapnik-Chervonenkis dimension, is contained in the smallest interval with integer endpoints $[\alpha, \beta]$, where $\alpha \leq \beta$ are the two solutions of the equation

\[
(1 - \frac{\delta B_1}{t}) \lambda^t = \max \left\{ (1 - \frac{\delta B_2(\epsilon)}{t^{1-\epsilon}}) \lambda^t, t \geq 0 \right\},
\]

with $\delta = \frac{a+b}{a+1}$ and a correct guess at time $s$ is rewarded while a wrong one is penalized with $a$ and $b$ monetary units, respectively. The discounting factor $0 < \lambda < 1$ is constant. The constants $B_1$ and $B_2(\epsilon)$ bound the expectation $e(s)$ of a wrong guess at time $s$ for the ‘least lucky’ learner learning the least
favourable function in the least favourable state of the world

\[
\frac{B_1}{s} \leq e(s) \leq \frac{B_2(\epsilon)}{s^{1-\epsilon}}.
\]  

(2.2)

The above result was reached for the simplest PAC learning model using only the simplest of tools in the estimations. It is not, however, explicitly solvable, although numerical methods are expected to be effective if fixed parameter values are used.

2.2 Learning a real number, rationally

The goal of Paper B is to state the ideas presented in Paper A more precisely and to test them in a specific and mathematically simple case. Using the maxmin procedure of Gilboa and Schmeidler [27] the bounds for the stopping time are expressed in terms of the largest expected error of recall, and thus, effectively, in terms of the least expected reward. The problem of locating a real number \( \theta \) by testing whether \( x_i \leq \theta \), with \( x_i \) drawn from an unknown probability measure, is studied. Examples of stopping time bounds calculated for a range of term rates, sample costs and rewards/penalties from a recall are included. The standard econometric situations, such as product promotion, market research, credit risk assessment, and bargaining and tenders, where such bounds could be of interest, are pointed.

The first part, Background, introduces the basics of the standard model of ‘learning and recall’ as seen from three fields: computational learning theory, mathematical analysis and economic decision.

The next section, Decision, contains the decision model previously formulated in Paper A. This time, however, it contains a more specified definition leaving no room for obscurity. The main result of Paper A has also been redefined - not only the gains and losses, but also the costs of information \( c \), have been taken into account.
The $\delta$ factor used in Theorem 4.1 (Paper A) takes the form of:

$$\delta = \frac{a+b}{a+c} = \frac{1+\beta}{1+\gamma},$$

in which $\beta = \frac{b}{a}$ (loss-to-gain coefficient) and $\gamma = \frac{c}{a}$ (cost-to-gain coefficient).

In the (Learning a real number) section the case of threshold functions on the real line is studied. These functions are then used to illustrate a simple example of the economics of learning: to learn a threshold function $f_\theta$ is to learn the number $\theta$, with $f_\theta(x)$ equal to 0 or 1 according to whether $x \leq \theta$ or not. A sequential estimation of the expectation $e(s)$ of a wrong guess at time $s$ over all real $\theta$, and ‘all world states’ being probability measures, gives the following bounds (section 2.1, formula 2.2):

$$\frac{1}{2s+1} \leq e(s) \leq \frac{1}{s+1}$$

and, consequentially, the following result (Paper B, Theorem 1):

Any optimal stopping time for learning a real number in the sense here considered, is contained in the smallest interval $[t_1, t_2]$ with integer endpoints, containing the two real solutions $t_1 \leq t_2$ of the equation

$$\left(1 - \frac{\delta/2}{1 + t}\right) \lambda^t = \max_{t \geq 0} \left(1 - \frac{\delta}{1 + t}\right) \lambda^t$$

(2.4)

Here bounds for the stopping time are immediate. These bounding intervals are displayed in function of term rate, sample cost, and reward from/penalty of a recall, in the Appendix.

In the Examples section a few standard econometric situations, such as product promotion, market research, credit risk assessment, and bargaining and tenders, in which such bounds could be of interest, are discussed. Common to all these examples is the importance of both the lower and upper stopping time bounds. The lower bound indicates whether starting learning is profitable at all (a positive lower bound would mean learning is profitable), while the upper bound indicates when learning definitively stops being beneficial. All of this is illustrated in the Appendix.
Chapter 3

Problems for further study

In Paper B (Section 6, Closing), various theoretical and technical problems suitable for future study were suggested. Some of these questions are discussed below, as well as some additional ones:

1. The ‘modified dynamic’ learning model.

In this thesis a simple learning model, where the decision is only made once at the start of the learning process, is defined. The next step would be to construct a more dynamic model, so that the decision whether to continue learning is taken after every sample.

The schema (a)-(e) below describes such a model. A similar schema, but with assumed knowledge about sample distribution, can be found in the models described by Ferguson [24].

(a) define the start hypothesis \( h_0 \);

(b) decide whether to start learning or accept the hypothesis \( h_0 \);

(c) if learning, buy the first sample \((x_1, f(x_1))\) and, using it, test \( h_0 \) comparing \( h_0(x_1) \) and \( f(x_1) \);

(d) for \( n \geq 1 \), if \( h_{n-1}(x_n) \neq f(x_n) \) then either change \( h_{n-1} \) to \( h_n \) so that \( h_n(x_n) = f(x_n) \) and continue testing the hypothesis \( h_n \) on additional samples \((x_{n+1}, f(x_{n+1}))\) or accept the hypothesis \( h_{n-1} \);
(e) if \( h_{n-1}(x_n) = f(x_n) \) then either continue testing the hypothesis on additional samples (i.e. continue learning), or accept the current hypothesis. The result of this last choice is then the end of the learning process.

In the decision made in (b), (d) and (e), economical factors such as the cost of each sample, costs related to the hypothesis being incorrect and the gain related to a correct hypothesis, must be taken into account.

In this ‘modified dynamic learning model’ the learner makes a decision about whether to stop learning or continue at each time point \( t \), when he has just bought a sample \( x_s, t = s \), and has used it for testing his current hypothesis \( h_{s-1} \). In order to make a decision the learner estimates, similarly to the procedure shown in Paper B, the expected reward \( r \) from recall at time \( t > s \). However, in contrast to Paper B, the learner knows both the incurred cost \( L(s) \) (equal to the cost of samples \( x_1, . . . , x_s \) already bought) and a value of a requested function \( f \in F \) on these samples. Based on this information he can reduce a class \( F \). Testing the hypothesis \( h \) the learner has also improved his algorithm. As a result the learner is less likely to err both when deciding whether to buy another sample, as well as when estimating their future number. Thus, the lower and the upper bounds of the optimal stopping time \( s_{opt} \) become smaller than in the static model described in Paper B. Note that knowledge of the sample distribution is not assumed here.

Obviously, the above model requires some in-detail development, especially regarding employment of the maxmin expected utility criterion (of Gilboa and Schmeidler [27]). In its current basic state, use of the model in the simple case examined in Paper B (using the different, static learning model) where the agent is learning a threshold function \( f_0 \) on the real line, a sequential estimation of the expectation \( e(s, f_0, \mu) \) of a wrong guess at time \( s \) over all functions \( f_0 \), and over ‘all world states’ \( \mu \), would appear to give the following bounds: \( \frac{1/2}{(s+1)!} \leq e(s, f_0, \mu) \leq \frac{1}{(s+1)!} \). The lower and upper bounds of the optimal stopping time would be smaller for the reasons discussed above.

Decision-making in dynamic situations is widely discussed both theoretically and practically and the developed methods have many implementations (see [21, 23, 24, 26, 28, 36]). Perhaps the proposed way to treat economics of learning will add something to the field.
2. An ‘active learning model’.

Another way to improve the model is to assume that the learner, when buying a sample \( x_s \), can make a choice between \( n \) different types of samples. Then, the environment provides him with a random sample of the chosen type: \( x^k_s, k = 1, \ldots, n \). How does this influence the optimal stopping time estimation in both considered models (static and dynamic)? An example of such a situation is a tv-gameshow where the contestant picks the topic of the next question she or he has to answer.

3. A ‘finite time horizon model’.

The question about the finite time horizon model was asked in Paper B. This is also discussed in Ferguson [24], but from a Bayesian perspective. How will the change of time horizon from infinite to finite and fixed (known to the learner) change the estimate of the optimal stopping time in this model? What about if the time horizon is finite but changing (randomly or not) throughout the learning? Does the ‘sandwich lemma’ still apply to these cases?

These questions are not purely theoretical. For an example, consider the problem of higher education system profitability below.

4. A higher education system profitability case.

Consider a higher education system case, where external parameters are assumed (any prospective salary evolution, student loan repayment rate, etc.) to calculate the optimal study length on selected higher education programmes. This value would then be compared to both the planned programme length as well as the actual time spent by the average programme student. Such a model could help determine whether studying is, in the economical sense, profitable - and if so - which programmes possess the highest economical utility.

This problem is an example of the finite horizon model. Here, the time horizon consists of three parts: learning, recalling and retirement, when the agent’s payoff is based on his wages during the time spent working.
5. The ‘sandwich lemma’.

As discussed above (Chapter 1, section 1.4), the following naive ‘sandwich argument’ was used to estimate the optimal learning time $s^♯$ in Paper B:

If $\gamma_1(s) \leq e(s) \leq \gamma_2(s)$, then $s^♯$ is bound to the set where

$$\max_{s \geq 0} (1 - \delta \gamma_1(s))\lambda^s \leq (1 - \delta \gamma_2(s))\lambda^s,$$

(for $\delta$ see Chapter 2, section 2.2).

This argument can be helpful in learning a real number only if firstly the expectation $e(s)$ of a wrong guess at time $s$ over ‘all world states’, being probability measures, has these bounds, and secondarily if the solution interval of the above inequality is easy to find and not too large.

What happens when some of the ‘lucky assumptions’ above are not valid? Is it then possible to find the effective bounds of optimal stopping time? This could be the subject of future work (of more technical character).

As for learning a number hold for all Borel probability measures (not only for Lebesgue absolutely continuous measures) - are there better estimations than the ‘sandwich argument’?

6. ‘Ambiguity variables’ $f$, $h$ and $\mu$.

The variables $f$, $h$ and $\mu$ are ‘ambiguity variables’ in the presented model. They are treated in the same way, even though there are significant differences between them. The variable $\mu$ describes ‘the state of the world’ in which the learner acts, upon which he may or may not have ‘a priori’ views - but one he has no interest in. After the learning has ended, his knowledge of the world at large remains unchanged.

The variable $f$ is also independent of the learner’s actions. During learning he predicts and calculates its values, allowing him to discover more about its properties and verify his ‘a priori’ views. After the learning his knowledge about $f$ has increased.

The variable $h$ is the learner’s own way of theoretically interpreting received data. He calculates its value, checks its difference from $f$, modifies and cor-
rects it - \( h \) is, in other words, under his control.

How will the maxmin procedure change when these three ‘ambiguity variables’ are treated differently from each other? Will this affect the optimal stopping time bounds? The search for these answers could be the subject of future work (of more theoretical character).

7. The term rate \( \lambda \).

In the model presented in this text, the term rate \( \lambda \) is assumed to be known to the learner and fixed throughout the learning. If the learner does not know the value of \( \lambda \) he can add \( \lambda \) to the ‘ambiguity variables’, which will undoubtedly affect the maxmin procedure. An example could be the problem of ranking projects by their value with respect to all term rates. A more complex model will have the term rate changing throughout the learning (and/or throughout the recall). How will this affect the maxmin procedure?
Bibliography


