Discrete MDL Predicts in Total Variation

Marcus Hutter

RSISE @ ANU and SML @ NICTA

Canberra, ACT, 0200, Australia

marcus@hutter1.net www.hutter1.net

September 2009

Abstract

The Minimum Description Length (MDL) principle selects the model that has the shortest code for data plus model. We show that for a countable class of models, MDL predictions are close to the true distribution in a strong sense. The result is completely general. No independence, ergodicity, stationarity, identifiability, or other assumption on the model class need to be made. More formally, we show that for any countable class of models, the distributions selected by MDL (or MAP) asymptotically predict (merge with) the true measure in the class in total variation distance. Implications for non-i.i.d. domains like time-series forecasting, discriminative learning, and reinforcement learning are discussed.

Contents

Keywords

minimum description length; countable model class; total variation distance; sequence prediction; discriminative learning; reinforcement learning.

1 Introduction

The *minimum description length* (MDL) principle recommends to use, among competing models, the one that allows to compress the data+model most [Grü07]. The better the compression, the more regularity has been detected, hence the better will predictions be. The MDL principle can be regarded as a formalization of Ockham's razor, which says to select the simplest model consistent with the data.

Multistep lookahead sequential prediction. We consider sequential prediction problems, i.e. having *observed sequence* $x \equiv (x_1,x_2,...,x_\ell) \equiv x_{1:\ell}$, *predict* $z \equiv (x_{\ell+1},...,x_{\ell+h}) \equiv x_{\ell+1:\ell+h}$, then observe $x_{\ell+1} \in \mathcal{X}$ for $\ell \equiv \ell(x) = 0,1,2,...$. Classical prediction is concerned with $h=1$, multi-step lookahead with $1<h<\infty$, and total prediction with $h=\infty$. In this paper we consider the last, hardest case. An infamous problem in this category is the Black raven paradox [Mah04, Hut07]: Having observed ℓ black ravens, what is the likelihood that *all* ravens are black. A more computer science problem is (infinite horizon) reinforcement learning, where predicting the infinite future is necessary for evaluating a policy. See Section 6 for these and other applications.

Discrete MDL and Bayes. Let $\mathcal{M} = \{Q_1,Q_2,...\}$ be a *countable class of models*=theories=hypotheses=probabilities over sequences $\mathcal{X}^\infty$, sorted w.r.t. to their *complexity=codelength* $K(Q_i) = 2\log_2 i$ (say), containing the *unknown true sampling distribution* P. Our main result will be for arbitrary measurable spaces $\mathcal{X}$, but to keep things simple in the introduction, let us illustrate MDL for finite $\mathcal{X}$.

In this case, we define $Q_i(x)$ as the Q_i-probability of data sequence $x \in \mathcal{X}^\ell$. It is possible to code x in $\log P(x)^{-1}$ bits, e.g. by using Huffman coding. Since x is sampled from P, this code is optimal (shortest among all prefix codes). Since we do not know P, we could select the $Q \in \mathcal{M}$ that leads to the shortest code on the observed data x. In order to be able to reconstruct x from the code we need to know which Q has been chosen, so we also need to code Q, which takes $K(Q)$ bits. Hence x can be coded in $\min_{Q\in\mathcal{M}}\{-\log Q(x)+K(Q)\}$ bits. MDL selects as model the minimizer

$$\mathrm{MDL}^x \;:=\; \arg\min_{Q\in\mathcal{M}}\{-\log Q(x) + K(Q)\}$$

Given x, the true predictive probability of z is $P(z|x) = P(xz)/P(x)$. Since P is unknown we use $\mathrm{MDL}^x(z|x) := \mathrm{MDL}^x(xz)/\mathrm{MDL}^x(x)$ as a substitute. Our main concern is how close is the latter to the former. We can measure the distance between two predictive distributions by

$$d_h(P,Q|x) \;=\; \sum_{z\in\mathcal{X}^h} \big|P(z|x) - Q(z|x)\big| \tag{1}$$

for $h<\infty$ and $d_\infty = \lim_{h\to\infty} d_h = \sup\{d_1,d_2,...\}$. It is easy to see that d_h is monotone increasing and that d_∞ is twice the total variation distance (tvd) defined in (3).

MDL is closely related to Bayesian prediction, so a comparison to existing results for Bayes is interesting. Bayesians use $\mathrm{Bayes}(z|x)$ for prediction, where $\mathrm{Bayes}(x) :=$

$\sum_{Q\in\mathcal{M}} Q(x)w_Q$ is the Bayesian mixture with prior weights $w_Q > 0\,\forall Q \in \mathcal{M}$ and $\sum_{Q\in\mathcal{M}} w_Q = 1$. A natural choice is $w_Q \propto 2^{-K(Q)}$.

Results. The following results can be shown

$$\begin{array}{ll} \sum_{\ell=0}^{\infty} \mathbf{E}[d_h(P, \mathrm{MDL}^x|x_{1:\ell})] \leq 21\, h \cdot 2^{K(P)}, & d_\infty(P, \mathrm{MDL}^x|x) \to 0 \\ \sum_{\ell=0}^{\infty} \mathbf{E}[d_h(P, \mathrm{Bayes}|x_{1:\ell})] \leq h \cdot \ln w_P^{-1}, & d_\infty(P, \mathrm{Bayes}|x) \to 0 \end{array} \quad \left\{ \begin{array}{l} \text{almost surely} \\ \text{for } \ell(x) \to \infty \end{array} \right. \tag{2}$$

where the expectation $\mathbf{E}$ is w.r.t. $P[\cdot|x]$. The left statements for $h < \infty$ imply $d_h \to 0$ almost surely, including some form of convergence rate. For Bayes it has been proven in [Hut03]; for MDL the proof in [PH05] can be adapted. As far as asymptotics is concerned, the right results $d_\infty \to 0$ are much stronger, and require more sophisticated proof techniques. For Bayes, the result follows from [BD62]. The proof for MDL is the primary novel contribution of this paper; more precisely for arbitrary measurable $\mathcal{X}$ in total variation distance. Another general consistency result is presented in [Grü07, Thm.5.1]. Consistency is shown (only) in probability and the predictive implications of the result are unclear. A stronger almost sure result is alluded to, but the given reference to [BC91] contains only results for i.i.d. sequences which do not generalize to arbitrary classes. So existing results for discrete MDL are far less satisfactory than the elegant Bayesian prediction in tvd.

Motivation. The results above hold for completely arbitrary countable model classes $\mathcal{M}$. No independence, ergodicity, stationarity, identifiability, or other assumption need to be made.

The bulk of previous results for MDL are for continuous model classes [Grü07]. Much has been shown for classes of independent identically distributed (i.i.d.) random variables [BC91, Grü07]. Many results naturally generalize to stationary-ergodic sequences like (kth-order) Markov. For instance, asymptotic consistency has been shown in [Bar85]. There are many applications violating these assumptions, some of them are presented below and in Section 6.

One can often hear the exaggerated claim that (e.g. unlike Bayes) MDL can be used even if the true distribution P is not in $\mathcal{M}$. Indeed, it can be used, but the question is wether this is any good. There are some results supporting this claim, e.g. if P is in the closure of $\mathcal{M}$, but similar results exist for Bayes. Essentially P needs to be at least close to some $Q \in \mathcal{M}$ for MDL to work, and there are interesting environments that are not even close to being stationary-ergodic or i.i.d.

Non-i.i.d. data is pervasive [AHRU09]; it includes all time-series prediction problems like weather forecasting and stock market prediction [CBL06]. Indeed, these are also perfect examples of non-ergodic processes. Too much green house gases, a massive volcanic eruption, an asteroid impact, or another world war could change the climate/economy irreversibly. Life is also not ergodic; one inattentive second in a car can have irreversible consequences. Also stationarity is easily violated in multi-agent scenarios: An environment which itself contains a learning agent is non-stationary (during the relevant learning phase). Extensive games and multi-agent reinforcement learning are classical examples [WR04].

Often it is assumed that the true distribution can be uniquely identified asymptotically. For non-ergodic environments, asymptotic distinguishability can depend on the realized observations, which prevent a prior reduction or partitioning of $\mathcal{M}$. Even if principally possible, it can be practically burdensome to do so, e.g. in the presence of approximate symmetries. Indeed this problem is the primary reason for considering *predictive* MDL. MDL might never identify the true distribution, but our main result shows that the sequentially selected models become predictively indistinguishable.

The countability of $\mathcal{M}$ is the severest restriction of our result. Nevertheless the countable case is useful. A semi-parametric problem class $\bigcup_{d=1}^{\infty}\mathcal{M}_d$ with $\mathcal{M}_d=\{Q_{\theta,d}:\theta\in\mathbb{R}^d\}$ (say) can be reduced to a countable class $\mathcal{M}=\{P_d\}$ for which our result holds, where P_d is a Bayes or NML or other estimate of $\mathcal{M}_d$ [Grü07]. Alternatively, $\bigcup_d\mathcal{M}_d$ could be reduced to a countable class by considering only computable parameters θ. Essentially all interesting model classes contain such a countable topologically dense subset. Under certain circumstances MDL still works for the non-computable parameters [Grü07]. Alternatively one may simply reject non-computable parameters on philosophical grounds [Hut05]. Finally, the techniques for the countable case might aid proving general results for continuous $\mathcal{M}$, possibly along the lines of [Rya09].

Contents. The paper is organized as follows: In Section 2 we provide some insights how MDL and Bayes work in restricted settings, what breaks down for general countable $\mathcal{M}$, and how to circumvent the problems. The formal development starts with Section 3, which introduces notation and our main result. The proof for finite $\mathcal{M}$ is presented in Section 4 and for denumerable $\mathcal{M}$ in Section 5. In Section 6 we show how the result can be applied to sequence prediction, classification and regression, discriminative learning, and reinforcement learning. Section 7 discusses some MDL variations.

2 Facts, Insights, Problems

Before starting with the formal development, we describe how MDL and Bayes work in some restricted settings, what breaks down for general countable $\mathcal{M}$, and how to circumvent the problems. For deterministic environments, MDL reduces to learning by elimination, and the four results in (2) can easily be understood. Consistency of MDL for i.i.d. (and stationary-ergodic) sources is also intelligible. For general $\mathcal{M}$, MDL may no longer converge to the true model. We have to give up the idea of model identification, and concentrate on predictive performance.

Deterministic MDL = elimination learning. For a countable class $\mathcal{M}=\{Q_1,Q_2,...\}$ of *deterministic* theories=models=hypotheses=sequences, sorted w.r.t. to their complexity=codelength $K(Q_i)=2\log_2 i$ (say) it is easy to see why MDL works: Each Q is a model for one infinite sequence $x^Q_{1:\infty}$, i.e. $Q(x^Q)=1$. Given the true observations $x\equiv x^P_{1:\ell}$ so far, MDL selects the simplest Q consistent with $x^P_{1:\ell}$ and

for $h=1$ predicts $x^Q_{\ell+1}$. This (and potentially other) Q becomes (forever) inconsistent if and only if the prediction was wrong. Assume the true model is $P=Q_m$. Since elimination occurs in order of increasing index i, and Q_m never makes any error, MDL makes at most $m-1$ prediction errors. Indeed, what we have described is just classical Gold style learning by elimination. For $1<h<\infty$, the prediction $x^Q_{\ell+1:\ell+h}$ may be wrong only on $x^Q_{\ell+h}$, which causes h wrong predictions before the error is revealed. (Note that at time ℓ only x^P_ℓ is revealed.) Hence the total number of errors is bounded by $h\cdot(m-1)$. The bound is for instance attained on the class consisting of $Q_i=1^{ih}0^\infty$, and the true sequence switches from 1 to 0 after having observed $m\cdot h$ ones. For $h=\infty$, a wrong prediction gets *eventually* revealed. Hence each wrong Q_i ($i<m$) gets eventually eliminated, i.e. P gets eventually selected. So for $h=\infty$ we can (still/only) show that the number of errors is finite. No bound on the *number* of errors in terms of m only is possible. For instance, for $\mathcal{M}=\{Q_1=1^\infty,Q_2=P=1^n0^\infty\}$, it takes n time steps to reveal that prediction 1^∞ is wrong, and n can be chosen arbitrarily large.

Deterministic Bayes = majority learning. Bayesian learning is at the same time, closely related to and very different from MDL. Bayes predicts with a w_Q-weighted average of the models (rather than with a single one). For a deterministic class, Bayes is similar to prediction by majority: Consider the models consistent with the true observation $x^P_{1:\ell}$, having total weight W, and take the weighted majority prediction (this is the Bayes-optimal decision under 0-1 loss, Bayesian prediction would randomize). For $h=1$, making a wrong prediction means that Q's contributing to at least half of the total weight W get eliminated. Since $P=Q_m$ never gets eliminated, we have $w_P\leq W\leq 2^{-\#\text{Errors}}$, hence the number of errors is bounded by $\log_2 w_P^{-1}$. For probabilistic Bayesian prediction proper, it is also easy to see that the expected number of errors is bounded by $\ln w_P^{-1}$. One can show that these bounds are essentially sharp. (e.g. for Q_i defined as the digits after the comma of the binary expansion of $(i-1)/2^n$ for $i=1..m$ and $m=2^n-1$.) With the same reasoning as in the MDL case, for $h>1$ we have to multiply the bound by h; and for $h=\infty$ we get correct prediction eventually, but no explicit bound anymore.

Comparison of deterministic↔probabilistic and MDL↔Bayes. The flavor of results carries over to some extent to the probabilistic case. On a very abstract level even the line of reasoning carries over, although this is deeply buried in the sophisticated mathematical analysis of the latter. So the special deterministic case illustrates the more complex probabilistic case. For instance for $h=1$ and $w_i\propto 1/i^2$, we see that "Bayes" makes only $2\log_2 m$ errors, while MDL can make up to the m errors. This carries over to the probabilistic case. Also the multiplier h for $1<h<\infty$ and the lack of an explicit bound for $h=\infty$ carries over. Cf. the bounds in (2). The reader is invited to reveal other relations not explicitly mentioned here. The differences are as follows: In the probabilistic case, the true P can in general not be identified anymore. Further, while the Bayesian bound trivially follows from the 1/2-century old classical merging of opinions result [BD62], the corresponding

MDL bound we prove in this paper is more difficult to obtain.

Consistency of MDL for stationary-ergodic sources. For an i.i.d. class $\mathcal{M}$, the law of large numbers applied to the random variables $Z_t := \log[P(x_t)/Q(x_t)]$ implies $\frac{1}{\ell}\sum_{t=1}^{\ell} Z_t \to \mathrm{KL}(P||Q) := \sum_{x_1} P(x_1)\log[P(x_1)/Q(x_1)]$ with P-probability 1. Either the Kullback-Leibler (KL) divergence is zero, which is the case if and only if $P=Q$, or $\log P(x_{1:\ell}) - \log Q(x_{1:\ell}) \equiv \sum_{t=1}^{\ell} Z_\ell \sim \mathrm{KL}(P||Q)\ell \to \infty$, i.e. asymptotically MDL does not select Q. For countable $\mathcal{M}$, a refinement of this argument shows that MDL eventually selects P [BC91]. This reasoning can be extended to stationary-ergodic $\mathcal{M}$, but essentially not beyond. To see where the limitation comes from, we present some troubling examples.

Trouble makers. For instance, let P be a Bernoulli(θ_0) process, but let the Q-probability that $x_t=1$ be θ_t, i.e. time-dependent (still assuming independence). For a suitably converging but "oscillating" (i.e. infinitely often larger and smaller than its limit) sequence $\theta_t \to \theta_0$ one can show that $\log[P(x_{1:t})/Q(x_{1:t})]$ converges to but oscillates around $K(Q)-K(P)$ w.p.1, i.e. there are non-stationary distributions for which MDL does not converge (not even to a wrong distribution).

One idea to solve this problem is to partition $\mathcal{M}$, where two distributions are in the same partition if and only if they are asymptotically indistinguishable (like P and Q above), and then ask MDL to only identify a partition. This approach cannot succeed generally, whatever particular criterion is used, for the following reason: Let $P(x_1)>0\ \forall x_1$. For $x_1=1$, let P and Q be asymptotically indistinguishable, e.g. $P=Q$ on the remainder of the sequence. For $x_1=0$, let P and Q be asymptotically distinguishable distributions, e.g. different Bernoullis. This shows that for non-ergodic sources like this one, asymptotic distinguishability depends on the drawn sequence. The first observation can lead to totally different futures.

Predictive MDL avoids trouble. The Bayesian posterior does not need to converge to a single (true or other) distribution, in order for prediction to work. We can do something similar for MDL. At each time we still select a single distribution, but give up the idea of identifying a single distribution asymptotically. We just measure predictive success, and accept infinite oscillations. That's the approach taken in this paper.

3 Notation and Main Result

The formal development starts with this section. We need probability measures and filters for infinite sequences, conditional probabilities and densities, the total variation distance, and the concept of merging (of opinions), in order to formally state our main result.

Measures on sequences. Let $(\Omega := \mathcal{X}^\infty, \mathcal{F}, P)$ be the space of infinite sequences with natural filtration and product σ-field $\mathcal{F}$ and probability measure P. Let $\omega \in \Omega$ be an infinite sequence sampled from the true measure P. Except when mentioned

otherwise, all probability statements and expectations refer to P, e.g. almost surely (a.s.) and with probability 1 (w.p.1) are short for with P-probability 1 (w.P.p.1). Let $x=x_{1:\ell}=\omega_{1:\ell}$ be the first ℓ symbols of ω.

For countable $\mathcal{X}$, the probability that an infinite sequence starts with x is $P(x):=P[\{x\}\times\mathcal{X}^\infty]$. The conditional distribution of an event A given x is $P[A|x]:=P[A\cap(\{x\}\times\mathcal{X}^\infty)]/P(x)$, which exists w.p.1. For other probability measures Q on Ω, we define $Q(x)$ and $Q[A|x]$ analogously. General $\mathcal{X}$ are considered at the end of this section.

Convergence in total variation. P is said to be *absolutely continuous* relative to Q, written

$$P \ll Q \quad :\Leftrightarrow \quad [Q[A]=0 \text{ implies } P[A]=0 \text{ for all } A\in\mathcal{F}]$$

P and Q are said to be *mutually singular*, written $P\perp Q$, iff there exists an $A\in\mathcal{F}$ for which $P[A]=1$ and $Q[A]=0$. The *total variation distance* (tvd) between Q and P given x is defined as

$$d(P,Q|x) \;:=\; \sup_{A\in\mathcal{F}}\big|Q[A|x]-P[A|x]\big| \tag{3}$$

Q is said to *predict P in tvd* (or merge with P) if $d(P,Q|x)\to 0$ for $\ell(x)\to\infty$ with P-probability 1. Note that this in particular implies, but is stronger than one-step predictive on- and off-sequence convergence $Q(x_{\ell+1}=a_{\ell+1}|x_{1:\ell})-P(x_{\ell+1}=a_{\ell+1}|x_{1:\ell})\to 0$ for any a, not necessarily equal ω [KL94]. The famous Blackwell and Dubins convergence result [BD62] states that if P is absolutely continuous relative to Q, then (and only then [KL94]) Q merges with P:

$$\text{If} \quad P\ll Q \quad \text{then} \quad d(P,Q|x)\to 0 \quad \text{w.p.1} \quad \text{for} \quad \ell(x)\to\infty$$

Bayesian prediction. This result can immediately be utilized for Bayesian prediction. Let $\mathcal{M}:=\{Q_1,Q_2,Q_3,...\}$ be a countable (finite or infinite) class of probability measures, and $\text{Bayes}[A]:=\sum_{Q\in\mathcal{M}}Q[A]w_Q$ with $w_Q>0\;\forall Q$ and $\sum_{Q\in\mathcal{M}}w_Q=1$. If the model assumption $P\in\mathcal{M}$ holds, then obviously $P\ll\text{Bayes}$, hence Bayes merges with P, i.e. $d(P,\text{Bayes}|x)\to 0$ w.p.1 for all $P\in\mathcal{M}$. Unlike many other Bayesian convergence and consistency theorems, no (independence, ergodicity, stationarity, identifiability, or other) assumption on the model class $\mathcal{M}$ need to be made. Good convergence rates for the weaker $d_{h<\infty}$ distances have also been shown [Hut03]. The analogous result for MDL is as follows:

Theorem 1 (MDL predictions) *Let $\mathcal{M}$ be a countable class of probability measures on $\mathcal{X}^\infty$ containing the unknown true sampling distribution P. No (independence, ergodicity, stationarity, identifiability, or other) assumptions need to be made on $\mathcal{M}$. Let*

$$MDL^x \;:=\; \arg\min_{Q\in\mathcal{M}}\{-\log Q(x)+K(Q)\} \quad \textit{with} \quad \sum_{Q\in\mathcal{M}}2^{-K(Q)}<\infty$$

be the measure selected by MDL at time ℓ given $x\in\mathcal{X}^\ell$. Then the predictive distributions $MDL^x[\cdot|x]$ converge to $P[\cdot|x]$ in the sense that

$$d(P,MDL^x|x) \;\equiv\; \sup_{A\in\mathcal{F}}\big|MDL^x[A|x]-P[A|x]\big| \;\to\; 0 \quad \textit{for} \quad \ell(x)\to\infty \quad \textit{w.p.1}$$

$K(Q)$ is usually interpreted and defined as the length of some prefix code for Q, in which case $\sum_Q 2^{-K(Q)}\leq 1$. If $K(Q):=\log_2 w_Q^{-1}$ is chosen as complexity, by Bayes rule $\Pr(Q|x)=Q(x)w_Q/\text{Bayes}(x)$, the maximum a posteriori estimate $\text{MAP}^x:=\text{argmax}_{Q\in\mathcal{M}}\{\Pr(Q|x)\}\equiv\text{MDL}^x$. Hence the theorem also applies to MAP. The proof of the theorem is surprisingly subtle and complex compared to the analogous Bayesian case. One reason is that $\text{MDL}^x(x)$ is not a measure on $\mathcal{X}^\infty$.

Arbitrary $\mathcal{X}$. For arbitrary $\mathcal{X}$, definitions are more subtle. The casual reader satisfied with countable $\mathcal{X}$ can skip this paragraph. We can consider even more generally $x_t\in\mathcal{X}_t$ [BD62]. Let $\mathcal{B}_t$ be a σ-field of subsets of $\mathcal{X}_t$ for $t=1,2,3,....$ Let $\mathcal{F}_\ell$ be the σ-field for $\mathcal{X}^\ell:=\mathcal{X}_1\times...\times\mathcal{X}_\ell$ generated by (i.e. the smallest σ-field containing) $\mathcal{B}_1\times...\times\mathcal{B}_\ell$ for $\ell\leq\infty$. Let $(\Omega:=\mathcal{X}^\infty,\mathcal{F}=\mathcal{F}_\infty,P)$ be a probability space. Let P_ℓ be the marginal distribution on $(\mathcal{X}^\ell,\mathcal{F}_\ell)$, i.e. $P_\ell[A]:=P[A\times\mathcal{X}_{\ell+1}\times\mathcal{X}_{\ell+2}\times...]$ for $A\in\mathcal{F}_\ell$. The predictive distribution $P^\ell[A|x_{1:\ell}]$ is (a version of) the conditional distribution of the future "$x_{\ell+1:\infty}$" given past $x_{1:\ell}$, implicitly defined by $\int P^\ell[A|x_{1:\ell}]dP_\ell(x_{1:\ell}):=P[A]$ $\forall A\in\mathcal{F}$. Similarly define Q_ℓ and Q^ℓ for the other $Q\in\mathcal{M}$. See [Doo53] for details.

Let M be a measure on Ω such that Q is absolutely continuous (see below) relative to M for all $Q\in\mathcal{M}$. For instance $M[\cdot]=\text{Bayes}[\cdot]$ has this property. Now define the density (Radon-Nikodym derivative) $Q_\ell(x_{1:\ell})$ (round brackets) of measure $Q_\ell[\cdot]$ (square brackets) relative to $M_\ell[\cdot]$. It is important to note that all essential quantities, in particular MDL^x, are independent of the particular choice of M. We therefore plainly speak of the Q-density or even Q-probability of x.

For countable $\mathcal{X}$ and counting measure M, $Q^\ell[A|x]$ and $Q_\ell(x)$ coincide with $Q[A|x]$ and $Q(x)$ above. In the following, we drop the sup&superscripts ℓ, since they will always be clear from the argument. Note that by Carathodory's extension theorem, $\{Q(x):x\in\mathcal{X}^*\}$ uniquely defines $Q[A]$ $\forall A\in\mathcal{F}$.

4 Proof for Finite Model Class

We first prove Theorem 1 for finite model classes $\mathcal{M}$. For this we need the following Definition and Lemma:

Definition 2 (Relations between Q and P) *For any probability measures Q and P, let*

- *$Q^r+Q^s=Q$ be the Lebesgue decomposition of Q relative to P into an absolutely continuous non-negative measure $Q^r\ll P$ and a singular non-negative measure $Q^s\perp P$.*

- $g(\omega) := dQ^r/dP = \lim_{\ell\to\infty}[Q(x_{1:\ell})/P(x_{1:\ell})]$ *be (a version of) the Radon-Nikodym derivative, i.e.* $Q^r[A]=\int_A g\,dP$.
- $\Omega^\circ := \{\omega : Q(x_{1:\ell})/P(x_{1:\ell})\to 0\} \equiv \{\omega : g(\omega)=0\}$.
- $\vec{\Omega} := \{\omega : d(P,Q|x)\to 0$ *for* $\ell(x)\to\infty\}$.

It is well-known that the Lebesgue decomposition exists and is unique. The representation of the Radon-Nikodym derivative as a limit of local densities can e.g. be found in [Doo53, VII§8]: $Z_\ell^{r/s}(\omega) := Q^{r/s}(x_{1:\ell})/P(x_{1:\ell})$ for $\ell=1,2,3,...$ constitute two martingale sequences, which converge w.p.1. $Q^r \ll P$ implies that the limit Z_∞^r is the Radon-Nikodym derivative dQ^r/dP. (Indeed, Doob's martingale convergence theorem can be used to prove the Radon-Nikodym theorem.) $Q^s \perp P$ implies $Z_\infty^r = 0$ w.p.1. So g is uniquely defined and finite w.p.1.

Lemma 3 (Generalized merging of opinions) *For any Q and P, the following holds:*

(i) $P \ll Q$ *if and only if* $P[\Omega^\circ]=0$
(ii) $P[\Omega^\circ]=0$ *implies* $P[\vec{\Omega}]=1$ *[(i)+[BD62]]*
(iii) $P[\Omega^\circ\cup\vec{\Omega}]=1$ *[generalizes (ii)]*

(i) says that $Q(x)/P(x)$ converges almost surely to a strictly positive value if and only if P is absolutely continuous relative to Q, (ii) says that an almost sure positive limit of $Q(x)/P(x)$ implies that Q merges with P. (iii) says that even if $P \not\ll Q$, we still have $d(P,Q|x)\to 0$ on almost every sequence that has a positive limit of $Q(x)/P(x)$.

Proof. Recall Definition 2.

$(i\Leftarrow)$ Assume $P[\Omega^\circ]=0$: $P[A]>0$ implies $Q[A]\geq Q^r[A]=\int_A g\,dP>0$, since $g>0$ a.s. by assumption $P[\Omega^\circ]=0$. Therefore $P\ll Q$.

$(i\Rightarrow)$ Assume $P\ll Q$: Choose a B for which $P[B]=1$ and $Q^s[B]=0$. Now $Q^r[\Omega^\circ]=\int_{\Omega^\circ} g\,dP=0$ implies $0\leq Q[B\cap\Omega^\circ]\leq Q^s[B]+Q^r[\Omega^\circ]=0+0$. By $P\ll Q$ this implies $P[B\cap\Omega^\circ]=0$, hence $P[\Omega^\circ]=0$.

(ii) That $P\ll Q$ implies $P[\vec{\Omega}]=1$ is Blackwell-Dubins' celebrated result. The result now follows from (i).

(iii) generalizes [BD62]. For $P[\Omega^\circ]=0$ it reduces to (ii). The case $P[\Omega^\circ]=1$ is trivial. Therefore we can assume $0<P[\Omega^\circ]<1$. Consider measure $P'[A] := P[A|B]$ conditioned on $B := \Omega\setminus\Omega^\circ$.

Assume $Q[A]=0$. Using $\int_{\Omega^\circ} g\,dP=0$, we get $0=Q^r[A]=\int_A g\,dP=\int_{A\setminus\Omega^\circ} g\,dP$. Since $g>0$ outside Ω°, this implies $P[A\setminus\Omega^\circ]=0$. So $P'[A]=P[A\cap B]/P[B]=P[A\setminus\Omega^\circ]/P[B]=0$. Hence $P'\ll Q$. Now (ii) implies $d(P',Q|x)\to 0$ with P' probability 1. Since $P'\ll P$ we also get $d(P',P|x)\to 0$ w.P'.p.1.

Together this implies $0\leq d(P,Q|x)\leq d(P',P|x)+d(P',Q|x)\to 0$ w.P'.p.1, i.e. $P'[\vec{\Omega}]=1$. The claim now follows from

$$\begin{aligned} P[\Omega^\circ\cup\vec{\Omega}] &= P'[\Omega^\circ\cup\vec{\Omega}]P[\Omega\setminus\Omega^\circ] + P[\Omega^\circ\cup\vec{\Omega}|\Omega^\circ]P[\Omega^\circ] \\ &= 1\cdot P[\Omega\setminus\Omega^\circ] + 1\cdot P[\Omega^\circ] \;=\; P[\Omega] \;=\; 1 \end{aligned}$$

■

The intuition behind the proof of Theorem 1 is as follows. MDL will asymptotically not select Q for which $Q(x)/P(x)\to 0$. Hence for those Q potentially selected by MDL, we have $\omega\notin\Omega^\circ$, hence $\omega\in\vec{\Omega}$, for which $d(P,Q|x)\to 0$ (a.s.). The technical difficulties are for finite $\mathcal{M}$ that the eligible Q depend on the sequence ω, and for infinite $\mathcal{M}$ to deal with non-uniformly converging d, i.e. to infer $d(P,\mathrm{MDL}^x|x)\to 0$.

Proof of Theorem 1 for finite $\mathcal{M}$. Recall Definition 2, and let g_Q,Ω°_Q,$\vec{\Omega}_Q$ refer to some $Q\in\mathcal{M}\equiv\{Q_1,...,Q_m\}$. The set of sequences ω for which some g_Q for some $Q\in\mathcal{M}$ is undefined has P-measure zero, and hence can be ignored. Fix some sequence $\omega\in\Omega$ for which $g_Q(\omega)$ is defined for all $Q\in\mathcal{M}$, and let $\mathcal{M}_\omega:=\{Q\in\mathcal{M}:g_Q(\omega)=0\}$.

$$\mathrm{MDL}^x := \arg\min_{Q\in\mathcal{M}} L_Q(x), \quad \text{where} \quad L_Q(x) := -\log Q(x) + K(Q).$$

Consider the difference

$$L_Q(x) - L_P(x) \;=\; -\log\frac{Q(x)}{P(x)} + K(Q) - K(P) \stackrel{\ell\to\infty}{\longrightarrow} -\log g_Q(\omega) + K(Q) - K(P)$$

For $Q\in\mathcal{M}_\omega$, the r.h.s. is $+\infty$, hence

$$\forall Q\in\mathcal{M}_\omega\;\exists\ell_Q\forall\ell>\ell_Q : L_Q(x) > L_P(x)$$

Since $\mathcal{M}$ is finite, this implies

$$\forall\ell>\ell_0\,\forall Q\in\mathcal{M}_\omega : L_Q(x) > L_P(x), \quad \text{where} \quad \ell_0 := \max\{\ell_Q : Q\in\mathcal{M}_\omega\} < \infty$$

Therefore, since $P\in\mathcal{M}$, we have $\mathrm{MDL}^x\notin\mathcal{M}_\omega\;\forall\ell>\ell_0$, so we can safely ignore all $Q\in\mathcal{M}_\omega$ and focus on $Q\in\overline{\mathcal{M}}_\omega:=\mathcal{M}\setminus\mathcal{M}_\omega$. Let $\Omega_1:=\bigcap_{Q\in\overline{\mathcal{M}}_\omega}(\Omega^\circ_Q\cup\vec{\Omega}_Q)$. Since $P[\Omega_1]=1$ by Lemma 3(iii), we can also assume $\omega\in\Omega_1$.

$$Q\in\overline{\mathcal{M}}_\omega \quad\Rightarrow\quad g_Q(\omega)>0 \quad\Rightarrow\quad \omega\notin\Omega^\circ_Q \quad\Rightarrow\quad \omega\in\vec{\Omega}_Q \quad\Rightarrow\quad d(P,Q|x)\to 0$$

This implies

$$d(P,\mathrm{MDL}^x|x) \;\leq\; \sup_{Q\in\overline{\mathcal{M}}_\omega} d(P,Q|x) \;\to\; 0$$

where the inequality holds for $\ell>\ell_0$ and the limit holds, since $\mathcal{M}$ is finite. Since the set of ω excluded in our considerations has measure zero, $d(P,\mathrm{MDL}^x|x)\to 0$ w.p.1, which proves the theorem for finite $\mathcal{M}$. ■

5 Proof for Countable Model Class

The proof in the previous Section crucially exploited finiteness of $\mathcal{M}$. We want to prove that the probability that MDL asymptotically selects "complex" Q is small. The following Lemma establishes that the probability that MDL selects a *specific* complex Q infinitely often is small.

Lemma 4 (MDL avoids complex probability measures Q) *For any Q and P we have $P[Q(x)/P(x)\geq c$ infinitly often$]\leq 1/c$.*

Proof.
$$P[\forall\ell_0\exists\ell>\ell_0\!:\!\frac{Q(x)}{P(x)}\geq c] \overset{(a)}{=} P[\overline{\lim_{\ell\to\infty}}\frac{Q(x)}{P(x)}\geq c] \leq$$
$$\overset{(b)}{\leq} \frac{1}{c}\mathbf{E}[\overline{\lim_{\ell}}\frac{Q(x)}{P(x)}] \overset{(c)}{=} \frac{1}{c}\mathbf{E}[\underline{\lim_{\ell}}\frac{Q(x)}{P(x)}] \overset{(d)}{\leq} \frac{1}{c}\underline{\lim_{\ell}}\mathbf{E}[\frac{Q(x)}{P(x)}] \overset{(e)}{=} \frac{1}{c}$$

(a) is true by definition of the limit superior $\overline{\lim}$, (b) is Markov's inequality, (c) exploits the fact that the limit of $Q(x)/P(x)$ exists w.p.1, (d) uses Fatou's lemma, and (e) is obvious. ■

For sufficiently complex Q, Lemma 4 implies that $L_Q(x)>L_P(x)$ for most x. Since convergence is non-uniform in Q, we cannot apply the Lemma to all (infinitely many) complex Q directly, but need to lump them into one $\bar{Q}$.

Proof of Theorem 1 for countable $\mathcal{M}$. Let the $Q\in\mathcal{M}=\{Q_1,Q_2,...\}$ be ordered somehow, e.g. in increasing order of complexity $K(Q)$, and $P=Q_n$. Choose some (large) $m\geq n$ and let $\widetilde{\mathcal{M}}:=\{Q_{m+1},Q_{m+2},...\}$ be the set of "complex" Q. We show that the probability that MDL selects infinitely often complex Q is small:

$$\begin{aligned}
&P[\text{MDL}^x\in\widetilde{\mathcal{M}}\text{ infinitely often}]\\
&\equiv\ P[\forall\ell_0\exists\ell>\ell_0:\text{MDL}^x\in\widetilde{\mathcal{M}}]\\
&\leq\ P[\forall\ell_0\exists\ell>\ell_0\wedge Q\in\widetilde{\mathcal{M}}:L_Q(x)\leq L_P(x)]\\
&=\ P[\forall\ell_0\exists\ell>\ell_0:\sup_{i>m}\tfrac{Q_i(x)}{P(x)}2^{K(P)-K(Q_i)}\geq 1]\\
&\overset{(a)}{\leq}\ P[\forall\ell_0\exists\ell>\ell_0:\tfrac{\bar{Q}(x)}{P(x)}\,\delta\,2^{K(P)}\geq 1]\\
&\overset{(b)}{\leq}\ \delta\,2^{K(P)}\ \overset{(c)}{\leq}\ \varepsilon
\end{aligned}$$

The first three relations follow immediately from the definition of the various quantities. Bound (a) is the crucial "lumping" step. First we bound

$$\sup_{i>m}\frac{Q_i(x)}{P(x)}2^{-K(Q_i)}\ \leq\ \sum_{i=m+1}^{\infty}\frac{Q_i(x)}{P(x)}2^{-K(Q_i)}\ =\ \delta\,\frac{\bar{Q}(x)}{P(x)},$$

$$\delta:=\sum_{i>m}2^{-K(Q_i)}<\infty,\qquad \bar{Q}(x):=\frac{1}{\delta}\sum_{i>m}Q_i(x)2^{-K(Q_i)},$$

While MDL$^{\cdot}[\cdot]$ is not a (single) measure on Ω and hence difficult to deal with, $\bar{Q}$ is a proper probability measure on Ω. In a sense, this step reduces MDL to Bayes. Now we apply Lemma 4 in (b) to the (single) measure $\bar{Q}$. The bound (c) holds for sufficiently large $m=m_\varepsilon(P)$, since $\delta\to 0$ for $m\to\infty$. This shows that for the sequence of MDL estimates

$$\{\text{MDL}^{x_{1:\ell}}:\ell>\ell_0\}\subseteq\{Q_1,...,Q_m\}\quad\text{with probability at least}\quad 1-\varepsilon$$

Hence the already proven Theorem 1 for finite $\mathcal{M}$ implies that $d(P,\text{MDL}^x|x)\to 0$ with probability at least $1-\varepsilon$. Since convergence holds for every $\varepsilon>0$, it holds w.p.1. ■

6 Implications

Due to its generality, Theorem 1 can be applied to many problem classes. We illustrate some immediate implications of Theorem 1 for time-series forecasting, classification, regression, discriminative learning, and reinforcement learning.

Time-series forecasting. Classical online sequence prediction is concerned with predicting $x_{\ell+1}$ from (non-i.i.d.) sequence $x_{1:\ell}$ for $\ell=1,2,3,....$ Forecasting farther into the future is possible by predicting $x_{\ell+1:\ell+h}$ for some $h>0$. One can show that $0\leq d_1\leq d_h\leq d_{h+1}\leq d_\infty=2d\leq 2$, see (1) and (3). Hence Theorem 1 implies good asymptotic (multi-step) predictions. Offline learning is concerned with training a predictor on $x_{1:\ell}$ for fixed ℓ in-house, and then selling and using the predictor on $x_{\ell+1:\infty}$ without further learning. Theorem 1 shows that for enough training data, predictions "post-learning" will be good.

Classification and Regression. In classification (discrete $\mathcal{X}$) and regression (continuous $\mathcal{X}$), a sample is a set of pairs $D=\{(y_1,x_1),...,(y_\ell,x_\ell)\}$, and a functional relationship $\dot{x}=f(\dot{y})+\text{noise}$, i.e. a conditional probability $P(\dot{x}|\dot{y})$ shall be learned. For reasons apparent below, we have swapped the usual role of $\dot{x}$ and $\dot{y}$. The dots indicate $\dot{x}\in\mathcal{X}$ and $\dot{y}\in\mathcal{Y}$), while $x=x_{1:\ell}\in\mathcal{X}^\ell$ and $y=y_{1:\ell}\in\mathcal{Y}^\ell$. If we assume that also $\dot{y}$ follows some distribution, and start with a countable model class $\mathcal{M}$ of joint distributions $Q(\dot{x},\dot{y})$ which contains the true joint distribution $P(\dot{x},\dot{y})$, our main result implies that $\text{MDL}^D[(\dot{x},\dot{y})|D]$ converges to the true distribution $P(\dot{x},\dot{y})$. Indeed since/if samples are assumed i.i.d., we don't need to invoke our general result.

Discriminative learning. Instead of learning a generative [Jeb03] joint distribution $P(\dot{x},\dot{y})$, which requires model assumptions on the input $\dot{y}$, we can discriminatively [LSS07] learn $P(\cdot|\dot{y})$ directly without any assumption on y (not even i.i.d). We can simply treat $y_{1:\infty}$ as an oracle to all Q, define $\mathcal{M}'=\{Q'\}$ with $Q'(x):=Q(x|y_{1:\infty})$, and apply our main result to $\mathcal{M}'$, leading to $\text{MDL}'^x[A|x]\to P'[A|x]$, i.e. $\text{MDL}^{x|y_{1:\infty}}[A|x,y_{1:\infty}]\to P[A|x,y_{1:\infty}]$. This not yet useful since $y_{1:\infty}$ is never known completely. If $x_1,x_2,...$ are conditionally independent, we can write

$$Q(x|y)\;=\;\prod_{t=1}^{\ell}Q(x_t|y_t)\;=\;\sum_{x_{\ell+1:m}}\prod_{t=1}^{m}Q(x_t|y_t)\;=\;\sum_{x_{\ell+1:m}}Q(x_{1:m}|y_{1:m})\;=\;Q(x_{1:\ell}|y_{1:m})$$

Taking the limit $m\to\infty$ we get $Q(x|y)=Q(x|y_{1:\infty})$. This is a generic property satisfied for all *causal* processes, that a future y_t for $t>l$ does not influence past observations $x_{1:\ell}$. Hence for a class of conditionally independent distributions, we

get $\mathrm{MDL}^{x|y}[A|x,y] \to P[A|x,y]$. Since the x given y are *not* identically distributed, classical MDL consistency results for i.i.d. or stationary-ergodic sources do *not* apply. The following corollary formalizes our findings:

Corollary 5 (Discriminative MDL) *Let $\mathcal{M} \ni P$ be a class of discriminative causal distributions $Q[\cdot|y_{1:\infty}]$, i.e. $Q(x|y_{1:\infty}) = Q(x|y)$, where $x = x_{1:\ell}$ and $y = y_{1:\ell}$. Regression and classification are typical examples. Further assume $\mathcal{M}$ is countable. Let $MDL^{x|y} := \mathrm{argmin}_{Q\in\mathcal{M}}\{-\log Q(x|y) + K(Q)\}$ be the discriminative MDL measure (at time ℓ given x,y). Then $\sup_A \big|MDL^{x|y}[A|x,y] - P[A|x,y]\big| \to 0$ for $\ell(x) \to \infty$, $P[\cdot|y_{1:\infty}]$ almost surely, for* every *sequence $y_{1:\infty}$.*

For finite $\mathcal{Y}$ and conditionally independent x, the intuitive reason how this can work is as follows: If $\dot{y}$ appears in $y_{1:\infty}$ only finitely often, it plays asymptotically no role; if it appears infinitely often, then $P(\cdot|\dot{y})$ can be learned. For infinite $\mathcal{Y}$ and deterministic $\mathcal{M}$, the result is also intelligible: Every $\dot{y}$ might appear only once, but probing enough function values $x_t = f(y_t)$ allows to identify the function.

Reinforcement learning (RL). In the agent framework [RN03], an agent interacts with an environment in cycles. At time t, an agent chooses an action y_t based on past experience $x_{<t} \equiv (x_1,...,x_{t-1})$ and past actions $y_{<t}$ with probability $\pi(y_t|x_{<t}y_{<t})$ (say). This leads to a new perception x_t with probability $\mu(x_t|x_{<t}y_{1:t})$ (say). Then cycle $t+1$ starts. Let $P(xy) = \prod_{t=1}^{\ell} \mu(x_t|x_{<t}y_{1:t})\pi(y_t|x_{<t}y_{<t})$ be the joint interaction probability. We make no (Markov, stationarity, ergodicity) assumption on μ and π. They may be POMDPs or beyond.

Corollary 6 (Single-agent MDL) *For a fixed policy=agent π, and a class of environments $\{\nu_1,\nu_2,...\} \ni \mu$, let $\mathcal{M} = \{Q_i\}$ with $Q_i(x|y) = \prod_{t=1}^{\ell} \nu_i(x_t|x_{<t}y_{1:t})$. Then $d(P[\cdot|y], MDL^{x|y}) \to 0$ with joint P-probability 1.*

The corollary follows immediately from the previous corollary and the facts that the Q_i are causal and that with $P[\cdot|y_{1:\infty}]$-probability 1 $\forall y_{1:\infty}$ implies w.P.p.1 jointly in x and y.

In reinforcement learning [SB98], the perception $x_t := (o_t,r_t)$ consists of some regular observation o_t and a reward $r_t \in [0,1]$. Goal is to find a policy which maximizes accrued reward in the long run. The previous corollary implies

Corollary 7 (Fixed-policy MDL value function convergence) *Let $V_P[xy] := \mathbf{E}_{P[\cdot|xy]}[r_{\ell+1} + \gamma r_{\ell+2} + \gamma^2 r_{\ell+3} + ...]$ be the future γ-discounted P-expected reward sum (true value of π), and similarly $V_{Q_i}[xy]$ for Q_i. Then the MDL value converges to the true value, i.e. $V_{MDL^{x|y}}[xy] - V_P[xy] \to 0$, w.P.p.1. for any policy π.*

Proof. The corollary follows from the general inequality

$$\big|\mathbf{E}_P[f] - \mathbf{E}_Q[f]\big| \;\leq\; \sup|f| \cdot \sup_A \big|P[A] - Q[A]\big|$$

by inserting $f:=r_{\ell+1}+\gamma r_{\ell+2}+\gamma^2 r_{\ell+3}+...$ and $P=P[\cdot|xy]$ and $Q=\text{MDL}^{x|y}[\cdot|xy]$, and using $0\leq f\leq 1/(1-\gamma)$ and Corollary 6. ■

Since the value function probes the infinite future, we really made use of our convergence result in total variation. Corollary 7 shows that MDL approximates the true value asymptotically arbitrarily well. The result is weaker than it may appear. Following the policy that maximizes the estimated (MDL) value is often not a good idea, since the policy does not explore properly [Hut05]. Nevertheless, it is a reassuring non-trivial result.

7 Variations

MDL is more a general principle for model selection than a uniquely defined procedure. For instance, there are *crude* and *refined* MDL [Grü07], the related *MML* principle [Wal05], a *static*, a *dynamic*, and a *hybrid* way of using MDL for prediction [PH05], and other variations. For our setup, we could have defined multi-step lookahead prediction as a product of single-step predictions:

$$\text{MDLI}(x_{1:\ell}) \;:=\; \prod_{t=1}^{\ell}\text{MDL}^{x_{<t}}(x_t|x_{<t}), \qquad \text{MDLI}(z|x) \;=\; \text{MDLI}(xz)/\text{MDLI}(x)$$

which is a more *incremental* MDL version. Both, MDL^x and MDLI are 'static' in the sense of [PH05], and each allows for a dynamic and a hybrid version. Due to its incremental nature, MDLI likely has better predictive properties than MDL^x, and conveniently defines a *single* measure over $\mathcal{X}^\infty$, but inconveniently is $\notin\mathcal{M}$. One reason for using MDL is that it can be computationally simpler than Bayes. E.g. if $\mathcal{M}$ is a class of MDPs, then MDL^x is still an MDP and hence tractable, but MDLI like Bayes are a nightmare to deal with.

Acknowledgements. My thanks go to Peter Sunehag for useful discussions.

References

[AHRU09] M.-R. Amini, A. Habrard, L. Ralaivola, and N. Usunier, editors. *Learning from non-IID data: Theory, Algorithms and Practice (LNIDD'09)*, Bled, Slovenia, 2009.

[Bar85] A. R. Barron. *Logically Smooth Density Estimation.* PhD thesis, Stanford University, 1985.

[BC91] A. R. Barron and T. M. Cover. Minimum complexity density estimation. *IEEE Transactions on Information Theory*, 37:1034–1054, 1991.

[BD62] D. Blackwell and L. Dubins. Merging of opinions with increasing information. *Annals of Mathematical Statistics*, 33:882–887, 1962.

[CBL06] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[Doo53] J. L. Doob. *Stochastic Processes*. Wiley, New York, 1953.

[Grü07] P. D. Grünwald. *The Minimum Description Length Principle*. The MIT Press, Cambridge, 2007.

[Hut03] M. Hutter. Convergence and loss bounds for Bayesian sequence prediction. *IEEE Transactions on Information Theory*, 49(8):2061–2067, 2003.

[Hut05] M. Hutter. *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*. Springer, Berlin, 2005. 300 pages, http://www.hutter1.net/ai/uaibook.htm.

[Hut07] M. Hutter. On universal prediction and Bayesian confirmation. *Theoretical Computer Science*, 384(1):33–48, 2007.

[Jeb03] T. Jebara. *Machine Learning: Discriminative and Generative*. Springer, 2003.

[KL94] E. Kalai and E. Lehrer. Weak and strong merging of opinions. *Journal of Mathematical Economics*, 23:73–86, 1994.

[LSS07] P. Long, R. Servedio, and H. U. Simon. Discriminative learning can succeed where generative learning fails. *Information Processing Letters*, 103(4):131–135, 2007.

[Mah04] P. Maher. Probability captures the logic of scientific confirmation. In C. Hitchcock, editor, *Contemporary Debates in Philosophy of Science*, chapter 3, pages 69–93. Blackwell Publishing, 2004.

[PH05] J. Poland and M. Hutter. Asymptotics of discrete MDL for online prediction. *IEEE Transactions on Information Theory*, 51(11):3780–3795, 2005.

[RN03] S. J. Russell and P. Norvig. *Artificial Intelligence. A Modern Approach*. Prentice-Hall, Englewood Cliffs, NJ, 2nd edition, 2003.

[Rya09] D. Ryabko. Characterizing predictable classes of processes. In *Proc. 25th Conference on Uncertainty in Artificial Intelligence (UAI'09)*, Montreal, 2009.

[SB98] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998.

[Wal05] C. S. Wallace. *Statistical and Inductive Inference by Minimum Message Length*. Springer, Berlin, 2005.

[WR04] M. Weinberg and J. S. Rosenschein. Best-response multiagent learning in non-stationary environments. In *Proc. 3rd International Joint Conf. on Autonomous Agents & Multi Agent Systems (AAMAS'04)*, pages 506–513, 2004.