Universal Coding and Prediction
on Martin-Löf Random Points
The Case of Stationary Ergodic Measures

Łukasz Dębowski ¹  Tomasz Steifer ²

¹Institute of Computer Science, Polish Academy of Sciences
²Institute of Fundamental Technological Research, Polish Academy of Sciences

virtual seminar in Moscow, 15th June 2020;
virtual seminar in Warsaw, 4th February 2021
The aim of our research

An algorithmic philosophical perspective on prediction:
Prediction must be computable but predicted phenomena needn’t.

Universal estimators, codes, or predictors:
A procedure is called universal if it is optimal for typical random results generated by stochastic sources belonging to some class.

- The theory of almost sure universal coding and prediction is (quite) well established for stationary and ergodic measures, which are typically uncomputable.
- We will lift these results to Martin-Löf random sequences using the effective Birkhoff ergodic theorem and randomness for uncomputable measures.
1 Introduction

2 Classical theory
   • Preliminaries
     • Source coding and prediction
     • Induced prediction
     • PPM measure

3 Effectivization program
   • Preliminaries
     • Effectivized theory

4 Conclusion
Notation

- Measurable space $(X^\mathbb{Z}, \mathcal{X}^\mathbb{Z})$ of two-sided infinite sequences over a finite alphabet $X = \{a_1, \ldots, a_D\}$, where $D \geq 2$.
- Points are infinite sequences $x = (x_i)_{i \in \mathbb{Z}} \in X^\mathbb{Z}$.
- Strings are finite sequences $x_{j}^{k} = (x_i)_{j \leq i \leq k}$, where $x_{j}^{j-1} = \lambda$.
- $X^* = \bigcup_{n \geq 0} X^n$ is the set of strings, $X^0 = \{\lambda\}$.
- Random variables $X_k((x_i)_{i \in \mathbb{Z}}) := x_k$.
- $P$ and $R$ denote probability measures on $(X^\mathbb{Z}, \mathcal{X}^\mathbb{Z})$.
- $P(x_1^n) := P(X_1^n = x_1^n)$.
- $P(x_j^n|x_1^{j-1}) := P(X_j^n = x_j^n | X_1^{j-1} = x_1^{j-1})$.
Stationary and ergodic measures

Shift operation $T((x_i)_{i \in \mathbb{Z}}) := (x_{i+1})_{i \in \mathbb{Z}}$ for $(x_i)_{i \in \mathbb{Z}} \in X^\mathbb{Z}$.

**Definition (stationary and ergodic measures)**

A probability measure $P$ on $(X^\mathbb{Z}, \mathcal{X}^\mathbb{Z})$ is called:

- **stationary** if $P(T^{-1}(A)) = P(A)$ for all events $A \in \mathcal{X}^\mathbb{Z}$;
- **ergodic** if $P(A) \in \{0, 1\}$ for all events $A \in \mathcal{X}^\mathbb{Z}$ such that $T^{-1}(A) = A$. 
Borel-Cantelli and Barron lemma

Theorem (Borel-Cantelli lemma)

Let $P$ be a probability measure. If a sequence of events $U_0, U_1, \ldots \in \mathcal{X}^\mathbb{Z}$ satisfies $\sum_{i=1}^{\infty} P(U_n) < \infty$ then $\sum_{i=1}^{\infty} 1\{x \in U_n\} < \infty$ on $P$-almost every point $x$.

From Barron inequality (Barron, 1985) and Borel-Cantelli lemma:

Theorem (Barron lemma)

For any probability measure $P$ and any semi-measure $R$, $P$-almost surely we have

$$\lim_{n \to \infty} \left[ -\log R(X^n_1) + \log P(X^n_1) + 2 \log n \right] = \infty. \quad (1)$$
Ergodic theorems

Theorem (Birkhoff ergodic theorem)

For a stationary ergodic measure $P$ and a random variable $G$ such that $E|G| < \infty$, $P$-almost surely

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} G \circ T^i = E \left( \lim_{n \to \infty} G_n \right).$$  \hfill (2)

Theorem (Breiman ergodic theorem)

For a stationary ergodic measure $P$ and random variables $(G_i)_{i \geq 0}$ such that $E \sup_n |G_n| < \infty$ and $\lim_{n \to \infty} G_n$ exists $P$-almost surely, $P$-almost surely

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} G_i \circ T^i = E \left( \lim_{n \to \infty} G_n \right).$$  \hfill (3)
Levy law and SMB theorem

**Theorem (Lévy law)**

For a stationary probability measure $P$, $P$-almost surely there exist limits

$$P(x_0|X_{-\infty}^{-1}) := \lim_{n \to \infty} P(x_0|X_{-n}^{-1}).$$  \hspace{1cm} (4)

**Theorem (SMB theorem)**

For a stationary ergodic probability measure $P$, $P$-almost surely we have

$$\lim_{n \to \infty} \frac{1}{n} \left[ - \log P(X_1^n) \right] = \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left[ - \log P(X_1^n) \right].$$  \hspace{1cm} (5)
From Azuma inequality (Azuma, 1967) and Borel-Cantelli lemma:

**Theorem (Azuma theorem)**

*For a probability measure $P$ and real random variables $(Z_n)_{n \geq 1}$ such that $|Z_n| \leq \epsilon_n \sqrt{n / \ln n}$ with $\lim_{n \to \infty} \epsilon_n = 0$, $P$-almost surely we have*

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \left[ Z_i - \mathbb{E}(Z_i \mid X_1^{i-1}) \right] = 0.
$$

(6)
Introduction

Classical theory
- Preliminaries
- Source coding and prediction
  - Induced prediction
  - PPM measure

Effectivization program
- Preliminaries
- Effectivized theory

Conclusion
Denote the entropy rate

\[ h_P := \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left[ -\log P(X_1^n) \right] = \lim_{k \to \infty} \mathbb{E} \left[ -\log P(X_{k+1} | X_1^k) \right]. \]  

(7)

From the SMB theorem and Barron lemma:

**Theorem (source coding)**

*For any stationary ergodic measure \( P \) and any probability measure \( R \), \( P \)-almost surely we have*

\[ \liminf_{n \to \infty} \frac{1}{n} \left[ -\log R(X_1^n) \right] \geq h_P. \]  

(8)
Universal coding

Definition (universal measure)

A probability measure $R$ is called almost surely universal if for any stationary ergodic probability measure $P$, $P$-almost surely we have

$$\lim_{n \to \infty} \frac{1}{n} \left[ - \log R(X_1^n) \right] = h_P.$$ (9)

- Computable almost surely universal measures exist if the alphabet $X$ is finite. (Example: PPM discussed later.)
Source prediction

A predictor is an $f : \mathbb{X}^* \rightarrow \mathbb{X}$. Predictor induced by measure $P$ is

$$f_P(x_1^n) := \underset{x_{n+1} \in \mathbb{X}}{\arg \max} P(x_{n+1}|x_1^n),$$  \hspace{1cm} (10)

where $\underset{x \in \mathbb{X}}{\arg \max} g(x) := \min \{ a \in \mathbb{X} : g(a) \geq g(x) \text{ for } x \in \mathbb{X} \}$. From the Azuma theorem, Levy law, and Breiman ergodic theorem:

**Theorem (source prediction)**

For any stationary ergodic measure $P$ and any predictor $f$, $P$-almost surely we have

$$\lim \inf_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} 1\{X_{i+1} \neq f(X_1^i)\} \geq u_P := \lim_{n \to \infty} \mathbb{E} \left[ 1 - \max_{x_0 \in \mathbb{X}} P(x_0|X_{-n}^0) \right].$$ \hspace{1cm} (11)

Moreover, (11) holds with the equality for $f = f_P$. 
Universal prediction

Definition (universal predictor)

A predictor $f$ is called almost surely universal if for any stationary ergodic probability measure $P$, $P$-almost surely we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} 1\{X_{i+1} \neq f(X_i)\} = u_P. \quad (12)$$

- Computable almost surely universal predictors exist for finite alphabet $\mathcal{X}$. (Example: $f_{PPM}$ discussed later.)
Introduction

Classical theory
- Preliminaries
- Source coding and prediction
  - Induced prediction
- PPM measure

Effectivization program
- Preliminaries
- Effectivized theory

Conclusion
Following the work of Ryabko (2008), cf. Ryabko, Astola, and Malyutov (2016), we can ask a very natural question whether predictors induced by universal measures are also universal.

Ryabko was close to demonstrate this implication, showing that:

**Theorem**

Let $\mathbf{R}$ be an almost surely universal measure and $\mathbf{P}$ be a stationary ergodic measure. We have $\mathbf{P}$-almost surely

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E} \left| \mathbf{P}(X_{i+1} | X_0^i) - \mathbf{R}(X_{i+1} | X_0^i) \right| = 0.
$$

(13)

**Problem:**

$$
\lim_{n \to \infty} \mathbb{E} |Y_n| = 0 \text{ does not imply } \lim_{n \to \infty} Y_n = 0 \text{ almost surely.}
$$
**Theorem (Pinsker inequality)**

Let \( p \) and \( q \) be two probability distributions over a countable alphabet \( X \). We have

\[
\left( \sum_{x \in X} |p(x) - q(x)| \right)^2 \leq (2 \ln 2) \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} .
\]  

(14)

**Theorem (prediction inequality)**

Let \( p \) and \( q \) be two probability distributions over a countable alphabet \( X \). For \( x_p = \arg \max_{x \in X} p(x) \) and \( x_q = \arg \max_{x \in X} q(x) \), we have inequality

\[
0 \leq p(x_p) - p(x_q) \leq \sum_{x \in X} |p(x) - q(x)| .
\]

(15)
From the Levy law and Breiman ergodic theorem:

**Theorem (conditional SMB theorem)**

Let the alphabet be finite and let $P$ be a stationary ergodic probability measure. We have $P$-almost surely

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \left[ - \sum_{x_{i+1} \in X} P(x_{i+1} | X_i^i) \log P(x_{i+1} | X_1^i) \right] = h_P. \quad (16)
$$
From the Azuma theorem:

**Theorem (conditional universality)**

Let the alphabet be finite and let $P$ be a stationary ergodic probability measure. If measure $R$ is almost surely universal and satisfies

$$- \log R(x_{n+1}|x_1^n) \leq \epsilon_n \sqrt{n/\ln n}, \quad \lim_{n \to \infty} \epsilon_n = 0$$

(17)

then $P$-almost surely we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \left[ - \sum_{x_{i+1} \in X} P(x_{i+1}|X_1^i) \log R(x_{i+1}|X_1^i) \right] = h_P.$$ 

(18)
Induced universal prediction

From the conditional universality, conditional SMB theorem, Pinsker inequality, prediction inequality, and source prediction:

**Theorem (induced universal prediction)**

*If measure $R$ is almost surely universal and satisfies*

$$- \log R(x_{n+1}|x_1^n) \leq \epsilon_n \sqrt{n/\ln n}, \quad \lim_{n \to \infty} \epsilon_n = 0 \quad (19)$$

*then the induced predictor $f_R$ is almost surely universal.*
Proof

By the conditional universality and SMB theorem,

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \left[ \sum_{x_{i+1}} P(x_{i+1}|X_1^i) \log \frac{P(x_{i+1}|X_1^i)}{R(x_{i+1}|X_1^i)} \right] = 0. \quad (20)
\]

Hence by Pinsker inequality,

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \left[ \sum_{x_{i+1}} \left| P(x_{i+1}|X_1^i) - R(x_{i+1}|X_1^i) \right| \right]^2 = 0. \quad (21)
\]

Thus by \( \mathbb{E} Y^2 \geq (\mathbb{E} Y)^2 \),

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \sum_{x_{i+1}} \left| P(x_{i+1}|X_1^i) - R(x_{i+1}|X_1^i) \right| = 0. \quad (22)
\]

Finally we apply the prediction inequality and Azuma theorem.
# Introduction

## Classical theory
- Preliminaries
- Source coding and prediction
- Induced prediction
- PPM measure

## Effectivization program
- Preliminaries
- Effectivized theory

# Conclusion
An example of a universal measure

Definition (PPM measure)

Let the alphabet be $\mathbb{X} = \{a_1, \ldots, a_D\}$, where $D \geq 2$.

The PPM measure of order $k \geq 0$ is defined as

$$\text{PPM}_k(x^n_1) := D^{-k} \prod_{i=k+1}^{n} \frac{N(x_{i-k}^i|x_1^{i-1}) + 1}{N(x_{i-k}^{i-1}|x_1^{i-2}) + D}, \tag{23}$$

where the frequency of a substring $w_1^k$ in a string $x_1^n$ is

$$N(w_1^k|x_1^n) := \sum_{i=1}^{n-k+1} 1\{x_i^{i+k-1} = w_1^k\}. \tag{24}$$

Subsequently, we define the total PPM measure

$$\text{PPM}(x^n_1) := \sum_{k=0}^{\infty} \left[ \frac{1}{k + 1} - \frac{1}{k + 2} \right] \text{PPM}_k(x^n_1). \tag{25}$$
Universality of PPM and PPM-induced predictor

From a bound by empirical entropy and Birkhoff ergodic theorem:

**Theorem (PPM universality)**

Measure PPM is almost surely universal.

From the definition of PPM:

**Theorem (PPM bounds)**

We have

\[- \log \text{PPM}(x_1^n) \leq \log \frac{\pi^2}{6} + 2 \log n + n \log D, \quad (26)\]

\[- \log \text{PPM}(x_{n+1}|x_1^n) \leq \log \frac{\pi^2}{6} + 3 \log(n + D). \quad (27)\]

Hence, by bound (19), PPM induces an almost surely universal predictor.
1 Introduction

2 Classical theory
   - Preliminaries
   - Source coding and prediction
   - Induced prediction
   - PPM measure

3 Effectivization program
   - Preliminaries
   - Effectivized theory

4 Conclusion
The effectivization program

**Theorem**

We have $\varphi(x)$ for $P$-almost all $x$.

$\downarrow \downarrow \downarrow$

**Theorem**

We have $\varphi(x)$ for all $1-P$-random $x$.

**Known effectivizations:**

- Borel-Cantelli and Barron lemma.
- Birkhoff ergodic theorem.
- Levy law and SMB theorem.
Computability

- **Computably enumerable** is abbreviated as c.e.
- For an \( r \in \mathbb{R} \), the left cut of \( r \) is set \( \{ q \in \mathbb{Q} : q < r \} \).
- A real function \( f \) with arguments in a countable set is called **computable** or **left-c.e.** respectively if the left cuts of \( f(\sigma) \) are uniformly computable or c.e. given an enumeration of \( \sigma \).
- For a sequence \( s \in X^\mathbb{Z} \), we say that real functions \( f \) are **s-computable** or **s-left-c.e.** if they are computable or left-c.e. with oracle \( s \).
Representations of uncomputable measures

Typical stationary ergodic measures are not computable.
—Think of Bernoulli($p$) process, where $p$ is not computable.

A construction by Reimann and Slaman:

- Let $\mathcal{P}(X^\mathbb{Z})$ be the space of probability measures on $(X^\mathbb{Z}, X^\mathbb{Z})$.
- A measure $P \in \mathcal{P}(X^\mathbb{Z})$ is called $s$-computable if real function $(\sigma, \tau) \mapsto P(X_{-|\sigma|+1}^{|\tau|} = \sigma \tau)$ is $s$-computable.
- A representation function is a $\rho : X^\mathbb{Z} \rightarrow \mathcal{P}(X^\mathbb{Z})$ such that real function $(\sigma, \tau, s) \mapsto \rho(s)(X_{-|\sigma|+1}^{|\tau|} = \sigma \tau)$ is computable.
- We say that an infinite sequence $s \in X^\mathbb{Z}$ is a representation of measure $P$ if there exists a representation function $\rho$ such that $\rho(s) = P$.
- Any measure $P$ is $s$-computable for any representation $s$ of $P$. 
1-randomness (Martin-Löf) randomness

**Definition**

A collection of events $U_1, U_2, \ldots \in \mathcal{X}^\mathbb{Z}$ is called uniformly $s$-c.e. if and only if there is a collection of sets $V_1, V_2, \ldots \subset \mathbb{X}^* \times \mathbb{X}^*$ such that $U_i = \left\{ x \in \mathbb{X}^\mathbb{Z} : \exists (\sigma, \tau) \in V_i : x^{|\tau|} - |\sigma| + 1 = \sigma \tau \right\}$ and sets $V_1, V_2, \ldots$ are uniformly $s$-c.e.

**Definition (Martin-Löf test)**

A uniformly $s$-c.e. collection of events $U_1, U_2, \ldots \in \mathcal{X}^\mathbb{Z}$ is called a Martin-Löf $(s, P)$-test if $P(U_n) \leq 2^{-n}$ for every $n \in \mathbb{N}$.

**Definition (Martin-Löf or 1-randomness)**

A point $x \in \mathbb{X}^\mathbb{Z}$ is called $1$-$(s, P)$-random if for each Martin-Löf $(s, P)$-test $U_1, U_2, \ldots$ we have $x \notin \bigcap_{i \geq 1} U_i$. A point is called $1$-$P$-random if it is $1$-$(s, P)$-random for a representation $s$ of $P$. 
1 Introduction

2 Classical theory
   - Preliminaries
   - Source coding and prediction
   - Induced prediction
   - PPM measure

3 Effectivization program
   - Preliminaries
   - Effectivized theory

4 Conclusion
From Solovay tests (Solovay, 1975):

**Theorem (effective Borel-Cantelli lemma)**

Let $P$ be a probability measure. If a uniformly $s$-c.e. sequence of events $U_0, U_1, \ldots \in \mathcal{X}^\mathbb{Z}$ satisfies $\sum_{i=1}^{\infty} P(U_n) < \infty$ then $\sum_{i=1}^{\infty} 1\{x \in U_n\} < \infty$ on each 1-$(s, P)$-random point $x$.

From Barron inequality (Barron, 1985) and Borel-Cantelli lemma:

**Theorem (effective Barron lemma)**

For any probability measure $P$ and any $s$-computable semi-measure $R$, on 1-$(s, P)$-random points we have

$$
\lim_{n \to \infty} \left[ -\log R(X^n_1) + \log P(X^n_1) + 2 \log n \right] = \infty.
$$

(28)
From Bienvenu et al. (2012) and Franklin et al. (2012)

**Theorem (effective Birkhoff ergodic theorem)**

For a stationary ergodic measure $P$ and an $s$-left-c.e. random variable $G \geq 0$ such that $\mathbb{E} G < \infty$, on $1$-$(s, P)$-random points

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} G \circ T^i = \mathbb{E} G.
$$

(29)

**Theorem (effective Breiman ergodic theorem — our result)**

For a stationary ergodic measure $P$ and uniformly $s$-computable random variables $(G_i)_{i \geq 0}$ such that $G_n \geq 0$, $\mathbb{E} \sup_n G_n < \infty$, and $\lim_{n \to \infty} G_n$ exists $P$-almost surely, on $1$-$(s, P)$-random points

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} G_i \circ T^i = \mathbb{E} \lim_{n \to \infty} G_n.
$$

(30)
Takahashi (2008):

**Theorem (effective Lévy law)**

For a stationary probability measure $P$, on $1-P$-random points there exist limits

$$P(x_0|X^{-1}_{-\infty}) := \lim_{n \to \infty} P(x_0|X^{-1}_{-n}). \quad \text{(31)}$$

Hoyrup (2011):

**Theorem (effective SMB theorem)**

For a stationary ergodic probability measure $P$, on $1-P$-random points we have

$$\lim_{n \to \infty} \frac{1}{n} \left[ - \log P(X^n_1) \right] = \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left[ - \log P(X^n_1) \right]. \quad \text{(32)}$$
From Azuma inequality (Azuma, 1967) and Borel-Cantelli lemma:

**Theorem (effective Azuma theorem)**

For a probability measure $P$ and uniformly $s$-computable real random variables $(Z_n)_{n \geq 1}$ such that $|Z_n| \leq \epsilon_n \sqrt{n / \ln n}$ with $\lim_{n \to \infty} \epsilon_n = 0$, on $1-(s, P)$-random points we have

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \left[ Z_i - \mathbb{E} \left( Z_i \mid X_1^{i-1} \right) \right] = 0. \quad (33)
$$
Denote the entropy rate

\[ h_P := \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left[ - \log P(X_1^n) \right] = \lim_{k \to \infty} \mathbb{E} \left[ - \log P(X_{k+1} | X_1^k) \right]. \]  

(34)

From the SMB theorem and Barron lemma:

**Theorem (effective source coding)**

*For any stationary ergodic measure \( P \) and any \( s \)-computable probability measure \( R \), on \( 1-(s, P) \)-random points we have*

\[ \liminf_{n \to \infty} \frac{1}{n} \left[ - \log R(X_1^n) \right] \geq h_P. \]  

(35)
Universal coding

Definition (universal measure)

A computable (not necessarily stationary) probability measure $R$ is called $1$-universal if for any stationary ergodic probability measure $P$, on $1$-$P$-random points we have

$$\lim_{n \to \infty} \frac{1}{n} \left[ -\log R(X_1^n) \right] = h_P. \quad (36)$$

- $1$-universal measures exist if the alphabet $X$ is finite.

(Example: PPM discussed later.)
Source prediction

A predictor is an $f : X^* \to X$. Predictor induced by measure $P$ is

$$f_P(x^n_1) := \arg \max_{x_{n+1} \in X} P(x_{n+1} \mid x^n_1), \quad (37)$$

where $\arg \max g(x) := \min \{ a \in X : g(a) \geq g(x) \text{ for } x \in X \}$. From the Azuma theorem, Levy law, and Breiman ergodic theorem:

**Theorem (effective source prediction)**

For any stationary ergodic measure $P$ and any $s$-computable predictor $f$, on $1-(s, P)$-random points we have

$$\liminf_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} 1\{X_{i+1} \neq f(X^n_i)\} \geq u_P := \lim_{n \to \infty} E \left[ 1 - \max_{x_0 \in X} P(x_0 \mid X_{-n}^{-1}) \right]. \quad (38)$$

Moreover, if the induced predictor $f_P$ is $s$-computable then (38) holds with the equality for $f = f_P$. 
Universal prediction

Definition (universal predictor)

A computable predictor $f$ is called 1-universal if for any stationary ergodic probability measure $P$, on 1-$P$-random points we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} 1\{X_{i+1} \neq f(X_1)\} = u_P.$$  \hspace{1cm} (39)

- 1-universal predictors exist for finite alphabet $\mathbb{X}$.
  (Example: $f_{\text{PPM}}$ discussed later.)
Effective induced universal prediction

From the conditional universality, conditional SMB theorem, Pinsker inequality, prediction inequality, and source prediction:

Theorem (effective induced universal prediction)

If measure $R$ is 1-universal and satisfies

$$- \log R(x_{n+1}|x_1^n) \leq \epsilon_n \sqrt{n/\ln n}, \quad \lim_{n \to \infty} \epsilon_n = 0 \quad (40)$$

then the induced predictor $f_R$ is 1-universal if $f_R$ is computable.

From a bound by empirical entropy and Birkhoff ergodic theorem:

Theorem (effective PPM universality)

Measure $PPM$ is 1-universal and rational.

Hence, by bound (40), PPM induces a 1-universal predictor.
1 Introduction

2 Classical theory
   • Preliminaries
   • Source coding and prediction
   • Induced prediction
   • PPM measure

3 Effectivization program
   • Preliminaries
   • Effectivized theory

4 Conclusion
Universality of predictor $f_{PPM}$ is expected and intuitive.

The PPM measure satisfies the sufficient condition

$$- \log \text{PPM}(x_{n+1}|x_1^n) \leq \epsilon_n \sqrt{n/\ln n}, \quad \lim_{n \to \infty} \epsilon_n = 0 \quad (41)$$

with a large reserve.

It is an open question whether there are universal measures such that conditional probabilities $R(x_{n+1}|x_1^n)$ converge to zero much faster than for the PPM measure but they still induce universal predictors.

It would be interesting to find such universal measures.

Maybe they have some other desirable properties.