On decrease of a priori randomness deficiency

Alexey Chernov
Jürgen Schmidhuber

Technical Report No. IDSIA-12-05-2005
March 2005

IDSIA / USI-SUPSI
Dalle Molle Institute for Artificial Intelligence
Galleria 2, 6928 Manno, Switzerland

This work was sponsored by SNF grant 200020-100259/1.

On decrease of a priori randomness deficiency

Alexey Chernov
Jürgen Schmidhuber*

March 2005

Let M be a priori enumerable semimeasure.

The a priori randomness deficiency of $x \in \{0,1\}^*$ with respect to a computable measure μ is

$$d_\mu(x) = -\log_2 \frac{\mu(x)}{M(x)}\,.$$

For sequence prediction (see [1, Problem 3.13] for details) it is important to bound the value

$$\log_2 \frac{\mu(y|x)}{M(y|x)} = d_\mu(x) - d_\mu(y)$$

in terms of complexity of μ conditional to x. Here we present two results in this direction.

The prefix complexity is denoted by KP.

Lemma 1. *There is a constant C such that for any computable measure μ, for any $x, y \in \{0,1\}^*$ such that x is a prefix of y, and for any $l \le \ell(x)$, it holds*

$$\log_2 \frac{\mu(y|x)}{M(y|x)} \le KP(\mu \mid x_{1:l}) + KP(l, \lceil d_\mu(x) \rceil) + C\,.$$

Proof. Informally speaking, the idea is as follows. Given l, d (numbers), and p (the conditional description of μ given $x_{1:l}$), we define a new enumerable semimeasure ν_{dlp}: enumerating z of length $\ell(z) \ge l$, we are looking for z with the deficiency $d_{\mu(p,z)}(z) > d$, where $\mu(p,z)$ is the measure constructed from the description p and the condition $z_{1:l}$; on such z and its prolongations, ν_{dlp} is 2^d times as much as μ'; on the other sequences, ν_{dlp} is defined by the semimeasure property. Since the measure of all sequences with large deficiency ("non-random" sequences) is small, the condition $\nu_{dlp}(\Lambda) \le 1$ will not be violated. This ν_{dlp} is less than the a priori semimeasure M up to a multiplicative constant (which is in essence is the complexity of d, l, and p), and the required bound follows.

Now let us accurately check technical details.

Let U be a universal Turing machine (from the definition of the conditional complexity). Let p_μ be a conditional description of μ, that is $U(p_\mu, x_{1:l})$ is a program that enumerates μ. For any $p \in \{0,1\}^*$ and for any $v \in \{0,1\}^*$, let $\mu_{p,v}$ be a semimeasure that is enumerated (from below) by the program given by $U(p,v)$ (with corrections if the semimeasure conditions are violated; the same construction is used to enumerate all semimeasures). Thus, $\mu_{p_\mu, x_{1:l}} = \mu$, and $\mu_{p,v}$ depends effectively on p, v.

*IDSIA & TU Munich, Boltzmannstr. 3, 85748 Garching, München, Germany

For any integer d, any natural l, and any $p \in \{0,1\}^*$, put

$$S_{dlp} = \{z \in \{0,1\}^{\geq l} \mid \sum_{\substack{v\in\{0,1\}^{\ell(z)},\\ v\neq z}} \mu_{p,z_{1:l}}(v) + 2^{-d}M(z) > 1\}\,.$$

The set S_{dlp} is enumerable given d, l, and p. If $z \in S_{dlp}$, then $d_{\mu_{p,z_{1:l}}}(z) > d$.

Let $\bar{S}_{dlp}$ be the set of all prolongations of elements of S_{dlp}. Put

$$\nu_{dlp}(z) = \begin{cases} 2^d \mu_{p,z_{1:l}}(z), & \text{if } z \in \bar{S}_{dlp}, \\ \nu_{dlp}(z0) + \nu_{dlp}(z1), & \text{otherwise.} \end{cases}$$

(Formally speaking, this definition may lead to an infinite recursion, but actual enumerating of ν_{dlp} works as follows: at the first step, ν_{dlp} is equal to zero everywhere, and when a new element of S_{dlp} is enumerated, ν_{dlp} is changed on all the prefixes and prolongations according the formula above.)

Since $\bar{S}_{dlp}$ is enumerable, ν_{dlp} is enumerable too.

The inequality $\nu_{dlp}(z) \geq \nu_{dlp}(z0) + \nu_{dlp}(z1)$ holds by definition for $z \notin \bar{S}_{dlp}$, and is provided by the property of $\mu_{p,z_{1:l}}$ otherwise.

Prove that $\nu_{dlp}(\Lambda) \leq 1$. Let $r(S_{dlp})$ be the prefix-free set such that the set of all prolongations of its elements is $\bar{S}_{dlp}$.[1] If $z \in r(S_{dlp})$, then $z \in S_{dlp}$, then $2^{-d}M(z) > \mu_{p,z_{1:l}}(z)$, since $\sum_{v\in\{0,1\}^{\ell(z)}} \mu_{p,z_{1:l}}(v) \leq 1$; therefore $\nu_{dlp}(z) < M(z)$. Now let $z \in \{0,1\}^*$, $\ell(z) \leq l$. Then, since $r(S_{dlp})$ is prefix-free, $\nu_{dlp}(z) \leq \sum_{zv\in r(S_{dlp})} \nu_{dlp}(zv) < \sum_{zv \in r(S_{dlp})} M(zv) \leq M(z)$. In particular, $\nu_{dlp}(\Lambda) < M(\Lambda) \leq 1$. Thus ν_{dlp} is a semimeasure.

Put

$$\nu_{dl}(z) = \begin{cases} \sum_p 2^{-KP(p|z_{1:l})}\nu_{dlp}(z), & \text{if } \ell(z) \geq l, \\ \nu_{dlp}(z0) + \nu_{dlp}(z1), & \text{otherwise.} \end{cases}$$

To show that ν_{dl} is an enumerable semimeasure it is sufficient to prove that $\nu_{dl}(\Lambda) \leq 1$. Actually, $\sum_{\ell(z)=l} \nu_{dl}(z) = \sum_{\ell(z)=l}\sum_p 2^{-KP(p|z_{1:l})}\nu_{dlp}(z) < \sum_{\ell(z)=l}\sum_p 2^{-KP(p|z)}M(z) \leq \sum_{\ell(z)=l} M(z) \leq 1$.

Finally, put

$$\nu(z) = \sum_{d,l} 2^{-KP(d,l)}\nu_{dl}(z)\,.$$

Obviously, ν is an enumerable semimeasure, and thus there is a constant C such that for all z it holds $\nu(z) \leq C \cdot M(z)$ (here we actually need the universal property of M: the construction gives $\nu_{dlp}(z) < M(z)$ for $z \in S_{dlp}$ but not longer z).

Let μ be an arbitrary computable measure, l be an arbitrary natural, $x, y \in \{0,1\}^*$, $\ell(x) \geq l$ and x is a prefix of y. Let p_μ be a description of μ under the condition $x_{1:l}$.

Put $d = \lceil d_\mu(x) \rceil - 1$, i. e., $d_\mu(x) - 1 \leq d < d_\mu(x)$. Hence $\mu(x) < 2^{-d}M(x)$. Since $\mu = \mu_{p_\mu,x_{1:l}}$ is a measure, $\sum_{v\in\{0,1\}^{\ell(x)}} \mu_{p_\mu,x_{1:l}}(v) = 1$, and therefore $x \in S_{dlp_\mu}$, and $y \in \bar{S}_{dlp_\mu}$. Thus $\nu_{dlp_\mu}(y) = 2^d\mu(y)$, and

$$2^{-KP(d,l)} \cdot 2^{-KP(p_\mu|y_{1:l})} \cdot 2^d\mu(y) \leq \nu(y) \leq C \cdot M(y)\,.$$

After trivial transformations we get

$$\log_2 \frac{\mu(y|x)}{M(y|x)} \leq KP(p_\mu|y_{1:l}) + KP(d,l) + \log_2 C + 1\,.$$

Taking into account that $KP(p_\mu|y_{1:l}) = KP(\mu|x_{1:l}) + O(1)$, we get the required statement. □

[1] A prefix-free set with this property is uniquely defined. Actually, assume there are two such sets, R_1 and R_2. Both are contained in $\bar{S}_{dlp}$. If $z \in R_1$, then there is its prefix $w \in R_2$, and then there is $v \in R_1$ being a prefix of w. Then v is a prefix of z, therefore $v = z = w$, and $R_1 \subseteq R_2$. Thus $R_1 = R_2$.

One can show that the additional term proportional to $KP(\ell(x))$ is unavoidable if we use standard notion of prefix complexity $KP(\mu|x)$. Informally speaking, $KP(x|y)$ uses the information about the length of the condition y, but in the case of sequence prediction this length is chosen by chance and is absolutely irrelevant to the task. So we need a new kind of conditional complexity.

Motivation is as follows. Consider a Turing machine with two input tapes. Inputs are provided without delimiters, so the size of input is defined by the machine itself. Let us define new conditional complexity *KPM* (with respect to this machine): consider the work of the machine on all possible inputs; if the machine halted and outputed an object x having read exactly p on the first tape and y on the second tape, we say that $KPM(x|y') \le \ell(p)$ for all y' that are prolongations of y. The following formal definition is not based on machines, but formalizes the same idea.

Definition 1. An enumerable set E of triples of finite binary sequences is called *KPM-correct* if it satisfies the following requirements:

1. if $\langle p, y, x_1\rangle \in E$ and $\langle p, y, x_2\rangle \in E$, then $x_1 = x_2$;
2. if $\langle p, y, x\rangle \in E$, then $\langle p', y', x\rangle \in E$ for all p', y' such that p is a prefix of p' and y is a prefix of y'.

A *KPM*-correct set E is called *optimal* if for any *KPM*-correct set E' there is a constant C such that for all x, y

$$\min\{\ell(p) \mid \langle p, y, x\rangle \in E\} \le \min\{\ell(p) \mid \langle p, y, x\rangle \in E'\} + C\,.$$

Let E be an optimal *KPM*-correct set. The *prefix monotonically conditional complexity* of x under a condition y is

$$KPM(x|y) = \min\{\ell(p) \mid \langle p, y, x\rangle \in E\}$$

An optimal *KPM*-correct set exists by the usual argument. And as usual, the complexity *KPM* is defined up to a bounded additive term.

Lemma 2. *There is a constant C such that for any computable measure μ, for any $x, y \in \{0,1\}^*$ such that x is a prefix of y, it holds*

$$\log_2 \frac{\mu(y|x)}{M(y|x)} \le KPM(\mu \mid x) + KP(KPM(\mu \mid x)) + KP(\lceil d_\mu(x)\rceil) + C\,.$$

Proof. Let E be the optimal *KPM*-correct set. For any $q \in \{0,1\}^*$, let μ_q be an enumerable semimeasure and if q is a program that enumerates a semimeasure, then μ_q is the same semimeasure. Let $E(p,z)$ be the partial computable function such that if $\langle p, z, q\rangle \in E$, then $E(p,z)$ is defined and is equal to q.

For any integer d and any $p \in \{0,1\}^*$, put

$$S_{dp} = \{z \mid E(p,z) \text{ is defined}, \sum_{\substack{v\in\{0,1\}^{\ell(z)},\\ v\neq z}} \mu_{E(p,z)}(v) + 2^{-d}M(z) > 1\}\,.$$

The set S_{dp} is enumerable given d and p. If $z \in S_{dp}$, then $d_{\mu_{E(p,z)}}(z) > d$.

Let $\bar{S}_{dp}$ be the set of all prolongations of elements of S_{dp}. If $z \in \bar{S}_{dp}$, then $E(p,z)$ is defined, since $E(p,z_1)$ is defined for some prefix z_1 of z. Put

$$\nu_{dp}(z) = \begin{cases} 2^d \mu_{E(p,z)}(z), & \text{if } z \in \bar{S}_{dp},\\ \nu_{dlp}(z0) + \nu_{dlp}(z1), & \text{otherwise.}\end{cases}$$

(Formally speaking, this definition may lead to an infinite recursion, but actual enumerating of ν_{dp} works as follows: at the first step, ν_{dp} is equal to zero everywhere, and when a new element of S_{dp} is enumerated, ν_{dp} is changed on all the prefixes and prolongations according the formula above.)

Since $\bar{S}_{dp}$ is enumerable, ν_{dp} is enumerable too.

The inequality $\nu_{dp}(z) \geq \nu_{dp}(z0) + \nu_{dp}(z1)$ holds by definition for $z \notin \bar{S}_{dp}$, and is provided by the property of $\mu_{E(p,z)}$ otherwise.

Prove that $\nu_{dp}(\Lambda) \leq 1$. Let $r(S_{dp})$ be the prefix-free set such that the set of all prolongations of its elements is $\bar{S}_{dp}$.[2] If $z \in r(S_{dp})$, then $z \in S_{dp}$, then $2^{-d}M(z) > \mu_{E(p,z)}(z)$, since $\sum_{v\in\{0,1\}^{\ell(z)}} \mu_{E(p,z)}(v) \leq 1$; therefore $\nu_{dp}(z) < M(z)$. Hence $\nu_{dp}(\Lambda) \leq \sum_{z\in r(S_{dp})} \nu_{dp}(z) < \sum_{z\in r(S_{dp})} M(z) \leq 1$. Thus ν_{dp} is a semimeasure.

Put

$$\nu(z) = \sum_{p,d} 2^{-KP(d)} 2^{-KP(p)} \nu_{dp}(z)\,.$$

Clearly, ν is an enumerable semimeasure. Thus there is a constant C such that for all z it holds $\nu(z) \leq C \cdot M(z)$ (here we actually need the universal property of M: the construction gives $\nu_{dp}(z) < M(z)$ for $z \in S_{dp}$ but not longer z).

Let μ be an arbitrary computable measure, $x, y \in \{0,1\}^*$ such that x is a prefix of y. Let p_μ be a sequence such that $KPM(\mu|x) = \ell(p_\mu)$, $E(p_\mu, z)$ is defined and $\mu = \mu_{E(p_\mu,z)}$.

Put $d = \lceil d_\mu(x) \rceil - 1$, i. e., $d_\mu(x) - 1 \leq d < d_\mu(x)$. Hence $\mu(x) < 2^{-d}M(x)$. Since $\mu = \mu_{E(p_\mu,z)}$ is a measure, $\sum_{v\in\{0,1\}^{\ell(x)}} \mu_{E(p_\mu,z)}(v) = 1$, and therefore $x \in S_{dp_\mu}$, and $y \in \bar{S}_{dp_\mu}$. Thus $\nu_{dp_\mu}(y) = 2^d\mu(y)$, and

$$2^{-KP(d)} 2^{-KP(p)} \cdot 2^d \mu(y) \leq \nu(y) \leq C \cdot M(y)\,.$$

After trivial transformations we get

$$\log_2 \frac{\mu(y|x)}{M(y|x)} \leq KP(p_\mu) + KP(d) + \log_2 C + 1\,.$$

Taking into account that $KP(p_\mu) \leq \ell(p_\mu) + KP(\ell(p_\mu)) + O(1)$, we get the required statement. □

One can easily obtain the following result.

Lemma 3. *For any $x, y \in \{0,1\}^*$, for any natural l, it holds*

$$KPM(x|y) \leq KP(x|y_{1:l}) + KP(l) + O(1)\,.$$

The authors are grateful to Marcus Hutter for many useful discussions.

References

[1] M. Hutter. *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability.* Springer, Berlin, 2004. (On J. Schmidhuber's SNF grant 20-61847).

[2] A prefix-free set with this property is uniquely defined. Actually, assume there are two such sets, R_1 and R_2. Both are contained in $\bar{S}_{dlp}$. If $z \in R_1$, then there is its prefix $w \in R_2$, and then there is $v \in R_1$ being a prefix of w. Then v is a prefix of z, therefore $v = z = w$, and $R_1 \subseteq R_2$. Thus $R_1 = R_2$.