On impossibility of sequential algorithmic forecasting*

V.V. V'yugin

Institute for Information Transmission Problems, Russian Academy of Sciences,
Bol'shoi Karetnyi per. 19, Moscow GSP-4, 101447, Russia
e:mail vyugin@inbox.ru

Abstract. The problem of prediction future event given an individual sequence of past events is considered. Predictions are given in form of real numbers p_n which are computed by some algorithm φ using initial fragments $\omega_1, \dots, \omega_{n-1}$ of an individual binary sequence $\omega = \omega_1, \omega_2, \dots$ and can be interpreted as probabilities of the event $\omega_n = 1$ given this fragment. According to Dawid's *prequential framework* we consider partial forecasting algorithms φ which are defined on all initial fragments of ω and can be undefined outside the given sequence of outcomes. We show that even for this large class of forecasting algorithms combining outcomes of coin-tossing and transducer algorithm it is possible to efficiently generate with probability close to one sequences for which any partial forecasting algorithm is failed by the method of verifying called *calibration*.

1 Introduction

Let a sequence $\omega_1, \omega_2, \dots, \omega_{n-1}$ of outcomes is observed by a forecaster whose task is to give a probability p_n of the future event $\omega_n = 1$. A typical example is that p_n is interpreted as a probability that it will rain. Forecaster is said to be well calibrated if it rains as often as he leads us to expect. It should rain about 80% of the days for which $p_n = 0.8$, and so on. So, for simplicity we consider binary sequences, i.e. $\omega_n \in \{0, 1\}$ for all n. We give a rigorous definition of calibration later.

In the measure-theoretic framework we expect that forecaster has a method for assigning probabilities p_n of a future event $\omega_n = 1$ for all possible finite sequences $\omega_1, \omega_2, \dots, \omega_{n-1}$. In other words, all conditional probabilities

$$p_n = P(\omega_n = 1 | \omega_1, \omega_2, \dots, \omega_{n-1})$$

must be specified and the overall probability distribution on the space Ω will be defined. But in reality, we should recognize that we have only individual sequence $\omega_1, \omega_2, \dots, \omega_{n-1}$ of events and that the corresponding forecasts p_n whose

* The research described in this publication was made possible in part by grants RFBR 03-01-00475, CNRS 02-02-22001; A part of this work was done while the author was in Poncelet Laboratoire LIF CNRS, Marseille, France.

Dagstuhl Seminar Proceedings 06051
Kolmogorov Complexity and Applications
http://drops.dagstuhl.de/opus/volltexte/2006/630

testing is considered may fall short of defining a full probability distribution on the whole space Ω. This is the point of the *prequential principle* proposed by Dawid [1]. This principle says that the evaluation of a probability forecaster should depend only on his actual probability forecasts and the corresponding outcomes. The additional information contained in a probability measure that has these probability forecasts as conditional probabilities should not enter in the evaluation. According to Dawid's prequential framework we do not consider numbers p_n as conditional probabilities generating by some overall probability distribution defined on the all possible events. We start with a forecasting system which is a *partial* function f on the set of all finite sequences. We have to suppose that the valid forecasting system f is defined on all finite initial fragments $\omega_1, \dots, \omega_{n-1}, \dots$ of an analyzed individual sequence of outcomes. A computable forecasting system f is defined on some finite sequence of outcomes if and only if the corresponding algorithm when fed to this sequence finish its work and print out the result, otherwise it is undefined.

First examples of individual sequences for which well-calibrated forecasting is impossible (noncalibrable sequences) were presented in statistical papers [7], [11]. Unfortunately, the methods used in these papers and in [1], [2] do not comply with prequential principle; they depend on some mild assumptions about the measure from which probability forecasts are derived as conditional probabilities. The method of generation the noncalibrable sequences with probability arbitrary close to one presented in [13] also based on the same assumptions. In this paper we modify construction from [13] for the case of partial forecasting systems (forecasting algorithms) do not corresponding to any overall probability distribution. Our main result shows that even for this large class of forecasting algorithms combining outcomes of coin-tossing and transducer algorithm it is possible to generate with probability close to one sequences for which any partial forecasting system is failed by the method of calibration.

2 Well calibrated forecasting

Let Ω be the set of all infinite binary sequences, and Ξ be the set of all finite binary sequences. Let λ be the empty sequence. For any finite or infinite $\omega = \omega_1 \dots \omega_n \dots$ we denote $\omega^n = \omega_1 \dots \omega_n$, $l(\omega^n) = n$ denotes the length of the sequence ω^n. If x is a finite sequence and ω is a finite or infinite sequence then $x\omega$ denotes the concatenation of these sequences, $x \subseteq \omega$ means that $x = \omega^n$ for some n.

The evaluation of probability forecasts is based on a method called *calibration.* Let us give the correct definitions. A selection rule is a partial function on the set of all finite binary sequences taking values 0 and 1. A selection rule δ is said to select the subsequence $s = n_1 n_2 \dots$ under an infinite binary sequence $\omega = \omega_1 \omega_2 \dots \omega_n \dots$ if

1) $\delta(\omega_1 \omega_2 \dots \omega_{n-1})$ is defined for all n, and
2) $n \in s$ just when $\delta(\omega_1 \omega_2 \dots \omega_{n-1}) = 1$.

We say that a partial forecasting system f is well calibrated for $\omega_1\omega_2\dots\omega_n\dots$ with respect to selection rule δ

1) $f(\omega_1\omega_2\dots\omega_n)$ is defined for all n, and

2) ether the subsequence $n_1n_2\dots$ selected by δ under $\omega_1\omega_2\dots\omega_n\dots$ is finite or for $r\to\infty$

$$\frac{1}{r}\sum_{i=1}^{r}\omega_{n_i}-\frac{1}{r}\sum_{i=1}^{r}f(\omega_1\omega_2\dots\omega_{n_i-1})\longrightarrow 0. \tag{1}$$

A forecasting system f is *well calibrated* for ω if it is well calibrated for it with respect to any partial recursive selection rule δ. We call also such selection rules - computable.

We consider cylinders in the set Ω of all infinite binary sequences of type $\Gamma_\alpha=\{\omega\in\Omega : x\subseteq\omega\}$, where $\alpha\in\Xi$. Any measure P on Ω is unequally defined by its values $P(\alpha)=P(\Gamma_\alpha)$ for all $\alpha\in\Xi$.

Any everywhere defined forecasting system f generates the overall probability distribution on Ω

$$P(\omega_1\dots\omega_n)=f^*(\omega_1)f^*(\omega_1\omega_2)\cdot\ldots\cdot f^*(\omega_1\dots\omega_n),$$

where

$$f^*(\omega_1\dots\omega_n)=\begin{cases} f(\omega_1\omega_2\dots\omega_{n-1}) \text{ if } \omega_n=1\\ (1-f(\omega_1\omega_2\dots\omega_{n-1})) \text{ otherwise}\end{cases}$$

On the other hand, any probability distribution P on Ω such that $P(\omega_1\dots\omega_n)>0$ for all n generates the forecasting system

$$f(\omega_1\dots\omega_{n-1})=P(\omega_n=1|\omega_1\dots\omega_{n-1}),$$

where

$$P(\omega_n=1|\omega_1\dots\omega_{n-1})=P(\omega_1\dots\omega_{n-1}1)/P(\omega_1\dots\omega_{n-1})$$

is the conditional probability of the event $\omega_n=1$ given $\omega_1\dots\omega_{n-1}$.

A variant of the law of large numbers holds for everywhere defined forecasting systems (Dawid's general calibration theorem [1]).

Proposition 1. *Let f be a forecasting system and P be the corresponding overall probability distribution. Then f is well calibrated for P-almost all infinite sequences $\omega_1\omega_2\dots$.*

Oakes [7] proposed arguments (see Dawid [3] for different proof) that no deterministic forecasting system can be well calibrated for all possible sequences. For any everywhere forecasting system f we can define a sequence $\omega=\omega_1\omega_2\dots$ such that

$$\omega_i=\begin{cases} 1 \text{ if } p_i<0.5\\ 0 \text{ otherwise}\end{cases}$$

where $p_i=f(\omega_1\dots\omega^{i-1})$, $i=1,2,\dots$. The corresponding selections rules are

$$\delta_j(\omega^{i-1})=\begin{cases} j \text{ if } p_i<0.5\\ 1-j \text{ otherwise}\end{cases}$$

for $j = 0, 1$. It is easy to see that for some $j = 0$ or $j = 1$ the selection rule δ_j selects under ω an infinite sequence $n_1, n_2, \dots$ such that condition (1) of calibration fails. The lack of this example is in that the selection rule δ_j can be noncomputable with respect to p_i even when the forecasting system f is computable. It is well known that no algorithm exists deciding whether $r < 0.5$ for an arbitrary real number r. A possible objection is that we can truncate values of forecasting system f and consider some its approximation taking only rational values. We will use this idea to construct counter-example in Section 3.

Much efforts were devoted for develop *computable* methods and algorithms for "universal forecasting". The problem with Oakes's example was overcome by Foster and Vohra [4] and others (see also [9] who constructed a randomized forecasting system that is well calibrated for any sequence in the mean. Kakade and Foster [5] constructed deterministic universal forecasting algorithms that are well calibrated for any sequence, where the notion of calibration is weakened - only "smooth" selection rules are used. Vovk [12] developed ideas of "smooth" selection rules in general setting based on game-theoretic approach to probability theory [10]. Let us explain these ideas in more details and their relation for our Theorem 1 below. Let us consider the corresponding *Game of forecasting*.

Players: Realty, Forecaster, Sceptic
Let $\mathcal{K}_0 = 0$.
Protocol:
FOR $n = 1, 2, \dots$:
Sceptic announces continuous real function $S_n(p)$.
Forecaster announces a real number p_n.
Realty announces $x_n \in \{0, 1\}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + S_n(p_n)(x_n - p_n)$
END FOR.

Sceptic wins if $\mathcal{K}_n \geq 0$ for all n and $\mathcal{K}_n$ is unbounded, otherwise Forecaster wins. It was proved in [10] that in this case the law of large numbers of type (1) and condition of calibration hold. Vovk [12] proved that in case of continuous $S_n(p)$ there exists a winning strategy for the Forecaster, i.e. a sequence of real numbers p_n such that $\mathcal{K}_n \geq 0$ for all n and $\mathcal{K}_n$ is bounded. Moreover, this strategy satisfies $\mathcal{K}_n \leq \mathcal{K}_{n-1}$ for all $n = 1, 2, \dots$. Also, in most cases this forecasting strategy $p_n = f(\omega_1, \dots, \omega_{n-1})$ is computable. The idea of corresponding strategy is very simple: define p_n as the root of $S(p) = 0$, if there are no roots define $p_n = (1 + sign(S_n))/2$.

A typical example of selection rule used in whether forecasting is the following. Let f be some forecasting system and p^* be any real number between zero and one. Define

$$\delta_{p^*}(\omega_1\omega_2\dots\omega_{n-1}) = \begin{cases} 1 \text{ if } f(\omega_1\omega_2\dots\omega_{n-1}) \in I_{p^*} \\ 0 \text{ otherwise} \end{cases}$$

Here $f(\omega_1\omega_2\dots\omega_{n-1}) \in I_{p^*}$ means that the forecast $f(\omega_1\omega_2\dots\omega_{n-1})$ belongs to some "neighborhood" I_{p^*} of the real number p^*. A forecasting system f is well calibrated for $\omega_1\omega_2\dots$ in the small if it is well calibrated for $\omega_1\omega_2\dots$ with respect to any selection rule of type δ_{p^*}.

In the definition of selection rule δ_{p^*} it is convenient to consider "smooth" intervals $I_{p^*}(p) = K(p^*, p)$, where $K(p^*, p)$ is a continuous function - Mercer kernel, such that $K(p^*, p^*) = 1$ and $K(p^*, p) = 0$ outside some finite interval containing p^*. Then the condition $f(\omega_1\omega_2 \dots \omega_{n-1}) \in I_{p^*}$ in the definition of δ must be replaced on $I_{p^*}(f(\omega_1\omega_2 \dots \omega_{n-1})) = 1$

The $K29$ algorithm (with parameter K) defined in [12] uses the following Sceptic's strategy

$$S_n(p) = \sum_{i=1}^{n-1} K(p, p_i)(\omega_i - p_i).$$

It was proved in [12] (Appendix, Corollary 1) that Forecaster has a computable winning strategy $p_i = f(\omega_1 \dots \omega_{i-1})$ (like that defined above) such that calibration in the small

$$\frac{\sum\limits_{i=1}^{n} I_{p^*}(p_i)(\omega_i - p_i)}{\sum\limits_{i=1}^{n} I_{p^*}(p_i)} \to 0$$

holds for $\omega_1\omega_2 \dots$ for any point p^* such that the denominator tends to infinity.

In Section 3 we generalize Oakes's example in other direction. We overcome the problem of non-computability of the selection rules δ_j using some its computable with respect to ω^{i-1} approximation, but for all that we lose the property of its continuity and computability with respect to p_i. Recall, that according to Dawid's prequential principle we consider partial forecasting systems and there is no efficient procedure for extending partial recursive functions to total recursive functions. Also, there is no algorithm deciding whether algorithm computing an arbitrary $f(\omega_1 \dots \omega^{i-1})$ finishes its work. The main result of this paper possesses an advantage of computer effectiveness and universality over Oakes - type examples: we can effectively generate noncalibrable sequences with probability close to one. More correctly, the probabilistic machine constructed in Theorem 1 below outputs with probability $1 - \epsilon$ (where ϵ is arbitrary small) an infinite sequence $\omega_1\omega_2 \dots$ such that any partial recursive forecasting system φ does not calibrated for $\omega_1\omega_2 \dots$ since for some partial recursive selection rule δ defined in the proof of Theorem 1

$$\frac{1}{s_n} \sum_{i=1}^{n} \delta(\omega^{i-1})(\omega_i - \varphi(\omega^{i-1})) > c > 0$$

holds for some c for all n, where $s_n = \sum\limits_{i=1}^{n} \delta(\omega^{i-1}) \to \infty$ as $n \to \infty$. In particular, Theorem 1 shows that the condition of continuity of $I_{p^s}(p)$ can not be omitted in computable forecasting. The plot of $\delta(\omega^{i-1})$ against $p_i = \varphi(\omega^{i-1})$ can not be presented as a continuous curve.

3 Non-calibrable sequences

We consider very broad class of computable forecasting systems - predictions can be computed by algorithms φ using only initial fragments $\omega_1, \dots, \omega_{n-1}$ of the analyzed individual sequence $\omega = \omega_1, \omega_2, \dots$. According to Dawid's *prequential framework* we consider partial forecasting systems φ which are obliged to be defined only on all initial fragments of ω and can be undefined outside this sequence. The system of this type may not correspond to any overall probability distribution in the set of all binary sequences. The following theorem shows that even for this large class of forecasting algorithms combining outcomes of coin-tossing and transducer algorithms it is possible to generate with probability close to one sequences for which any partial forecasting system is failed by the method of calibration with respect to selection rules which are represented by rational approximations of selection rules proposed by Oakes.

We need a concept of *computable operation* on $\Xi \bigcup \Omega$ [14, 15]. Let $\hat{F}$ be a recursively enumerable set of ordered pairs of finite sequences satisfying the following properties:

- $(x, \lambda) \in \hat{F}$ for any x, where λ is the empty sequence;
- if $(x, y) \in \hat{F}$, $(x', y') \in \hat{F}$ and $x \subseteq x'$ then $y \subseteq y'$ or $y' \subseteq y$.

A computable operation F is defined as follows

$$F(\omega) = \sup\{y \mid x \subseteq \omega \text{ and } (x, y) \in \hat{F} \text{ for some } x\},$$

where $\omega \in \Omega \bigcup \Xi$ and sup is in the sense of the partial order $\subseteq$ on Ξ.

Informally, the computable operation F is defined by some algorithm which when fed with an infinite or a finite sequence ω takes it sequentially bit by bit, processes it and produces an output sequence also sequentially bit by bit.

By *probabilistic algorithm* we mean any pair (P, F), where P is a computable measure in the set of all binary sequences and F is a computable operation. In the following $P = L$, where $L(x) = L(\Gamma_x) = 2^{-l(x)}$ is the uniform measure in Ω.

A sequence $\omega_1\omega_2 \dots \omega_n \dots$ is *calibrable* if some partial computable f is well calibrated for it; otherwise, $\omega_1\omega_2 \dots \omega_n \dots$ is noncalibrable. The following theorem is the main result of this paper.

Theorem 1. *For any $\epsilon > 0$ a probabilistic algorithm (L, F) can be constructed which with probability $\geq 1 - \epsilon$ outputs an infinite binary sequence ω such that the following property holds: for each partial recursive forecasting system φ defined on all initial fragments of the sequence ω there exists a computable selection rule which selects under ω an infinite subsequence $\omega_{n_1}, \omega_{n_2}, \dots$ such that*

$$\frac{1}{r}\sum_{i=1}^{r}\omega_{n_i} - \frac{1}{r}\sum_{i=1}^{r}\varphi(\omega_1\omega_2\dots\omega_{n_i-1}) \not\longrightarrow 0$$

as $r \to \infty$.

In other words, the probabilistic machine (L, F) generates with probability close to one an infinite noncalibrable sequence.

References

1. A.P. Dawid. (1982) The well-calibrated Bayesian [with discussion], *J. Am. Statist. Assoc.* **77**, 605–613.
2. A.P. Dawid. (1985) Calibration-based empirical probability [with discussion] *Ann. Statist.* **13**, 1251–1285.
3. A.P. Dawid. (1985) The impossibility of inductive inference, *J. Am. Statist. Assoc.* **80**, 340–341.
4. D.P. Foster, R. Vohra. (1998) Asymptotic calibration, *Biometrika* **85**, 379–390.
5. S.M. Kakade, D.P. Foster. (2004) Deterministic calibration and Nash equilibrium, In Jon Shawe Taylor and Yoram Singer, editors *Proceedings of the Seventeenth Annual Conference on Learning Theory* Volume **3120** of Lecture Notes in Computer Science, 33-48, Heidelberg, Springer.
6. *Li M., Vitányi P.* An Introduction to Kolmogorov complexity and its applications. New York: Springer–Verlag, 1997.
7. D. Oakes. (1985) Self-calibrating priors do not exists [with discussion], *J. Am. Statist. Assoc.* **80**, 339–342.
8. H. Rogers. (1967) Theory of recursive functions and effective computability, New York: McGraw Hill.
9. A. Sandroni, R. Smorodinsky, R. Vohra. (2003) Calibration with many checking rules, *Mathematics of Operations Research* **28**, 141–153.
10. G. Shafer, V. Vovk. Probability and Finance. It's Only a Game! New York. Wiley. 2001.
11. M.J. Schervish. (1985) Comment [to Oakes, 1985], *J. Am. Statist. Assoc.* **80**, 341–342.
12. Vladimir Vovk, Akimichi Takemura, Glenn Shafer. (2004) Defensive forecasting, http://www.probabilityandfinance.com
13. V.V. V'yugin. (1998) Non-stochastic infinite and finite sequences, *Theor. Comp. Science.* **207**, 363–382.
14. V.A Uspensky, A.L. Semenov, A.Kh. Shen. (1990) Can an individual sequence of zeros and ones be random, *Russian Math. Surveys* **45**, No. 1, 121–189.
15. A.K. Zvonkin and L.A. Levin. (1970) The complexity of finite objects and the algorithmic concepts of information and randomness, *Russ. Math. Surv.* **25**, 83–124.