# On Sequences with Non-Learnable Subsequences

Vladimir V. V'yugin

Institute for Information Transmission Problems
Russian Academy of Sciences

CSR-2008, Moscow, 8-12 June 2008

Let a binary sequence

$$\omega_1, \omega_2, \ldots, \omega_{n-1}$$

of outcomes is observed by a forecaster whose task is to give a probability $p_n$ of a future event $\omega_n = 1$.

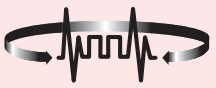
Example: $p_n$ is interpreted as a probability that it will rain.

A minimal requirement for testing of any prediction algorithm is that it should be calibrated (A.P.Dawid).

Informally: a 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.

## Some notations

- $\Omega$ - the set of all infinite binary sequences;
- $\Xi$ - the set of all finite binary sequences and $\lambda$ be the empty sequence;
- For any finite or an infinite sequence $\omega = \omega_1 \dots \omega_n \dots$, we write $\omega^n = \omega_1 \dots \omega_n$;
- $l(\omega^n) = n$ - the length of the sequence $\omega^n$;
- $x \sqsubseteq \omega$ means that $x = \omega^n$ for some $n$

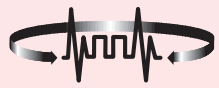


In the measure-theoretic framework we expect that the forecaster has a method for assigning probabilities $p_n$ of a future event $\omega_n = 1$ for all possible finite sequences $\omega_1, \omega_2, \ldots, \omega_{n-1}$.

In other words, all conditional probabilities

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

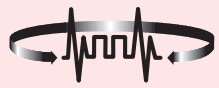
must be specified and the overall probability distribution $P$ in the space $\Omega$ of all infinite binary sequences will be defined.



In reality,

- we have only individual sequence $\omega_1, \omega_2, \ldots, \omega_{n-1}$ of outcomes;
- we can not define a probability distribution in the whole space $\Omega$.

Dawid's prequential principle: the evaluation of a probability forecaster should depend only on his actual probability forecasts $p_1, \ldots, p_n, \ldots$ and the corresponding outcomes $\omega_1, \ldots, \omega_n, \ldots$.

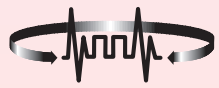


A deterministic forecasting system is a partial function $f : \Xi \to [0,1]$ such that

$$p_n = f(\omega_1, \omega_2, \ldots, \omega_{n-1})$$

We require that the valid forecasting system is defined on all initial fragments $\omega_1, \ldots, \omega_{n-1}$ of a given infinite sequence of outcomes $\omega$; it can be undefined on some $x \not\sqsubset \omega$.

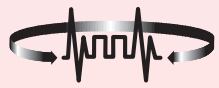
We suppose that this function is computable (partial recursive).



The main goal declared by Dawid was to construct an universal forecasting method which gives coherent forecasts for any sequence of outcomes regardless of the nature of a source generating data.

Dawid's ideas can be considered as some specifications of earlier more general ideas of Solomonoff and Levin who defined an universal probability pretending for the role of the universal forecasting method.

Solomonoff's and Levin's universal probability is non-computable.



The evaluation of probability forecasts is based on a method called calibration.

A sequence of forecasts $p_1, p_2, \ldots$ is well-calibrated for an infinite sequence $\omega_1 \omega_2 \ldots$ if for the characteristic function $I(p)$ of any subinterval of $[0,1]$ the calibration error tends to zero, i.e.,

$$\frac{\sum_{i=1}^{n} I(p_i)(\omega_i - p_i)}{\sum_{i=1}^{n} I(p_i)} \longrightarrow 0$$

as the denominator of this relation tends to infinity; in fact, we must add a requirement that the denominator must grow sufficiently fast (see below).

Here, $I(p_i)$ determines some "selection rule" which defines moments of time where we compute the deviation between forecasts $p_i$ and outcomes $\omega_i$.

Drawback:

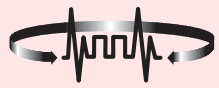
Any total deterministic forecasting system $f$

$$p_n = f(\omega_1, \omega_2, \ldots, \omega_{n-1})$$

is not calibrated for the sequence $\omega = \omega_1 \omega_2 \ldots$, where

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

and $p_i = f(\omega_1 \ldots \omega_{i-1})$, $i = 1, 2, \ldots$. The condition of calibration fails for this $\omega$, where $I = [0, 0.5)$ or $I = [0.5, 1]$.



*A randomized forecasting system* is a partial random variable $f$ taking values in $[0,1]$ such that

$$\tilde{p}_n = f(\alpha : \omega_1, \omega_2, \ldots, \omega_{n-1}),$$

where $\alpha$ belongs to some probability space depending on $\omega^{n-1}$, which is a parameter of this variable.
Usually we omit the variable $\alpha$.

Random rounding is a specific method for defining random forecasts.

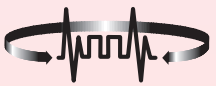
Kakade and Foster show that an observer can only randomly round with a precision of rounding $\Delta$ some deterministic forecast in order to calibrate:

For any infinite sequence $\omega = \omega_1 \omega_2 \ldots$ and for the characteristic function $I(p)$ of any subinterval of $[0,1]$ the overall probability $Pr$ (generated by this randomization) of the event

$$\left| \frac{1}{n} \sum_{i=1}^{n} I(\tilde{p}_i)(\omega_i - \tilde{p}_i) \right| \leq \Delta$$

tends to one as $n \to \infty$, where $\tilde{p}_i$ is the random rounding of some deterministic forecast $p_i$ up to $\delta$.

In fact, more accurate calculations show that $n$ in the denominator can be replaced on $\alpha(n)\sqrt{n}$, where $\alpha(n)$ is any unbounded nondecreasing function.

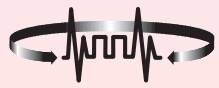


- A function $I : [0,1] \to \{0,1\}$ - *forecast-based* checking rule;
- A function $\delta : \Xi \to \{0,1\}$ - *outcome-based* checking rule.

Lehrer and Sandrony et al. extended the class of checking rules to combination of forecast- and outcome-based checking rules: A checking rule is a function

$$c(\omega^{i-1}, p) = \delta(\omega^{i-1})I(p),$$

where $\delta : \Xi \to \{0,1\}$ is an outcome-based checking rule, and $I(p)$ is the characteristic function of any subinterval of $[0,1]$.



Let a sequence $\{\delta_k\}$ of outcome-based checking rules and a sequence $\{I_k\}$ of characteristic functions of subintervals of $[0,1]$ be given, $k = 1,2,\ldots$ . Let also $\Delta > 0$.

Sandrony et al. defined a randomized universal forecasting system $f$ such that for $\tilde{p}_i = f(\omega^{i-1})$

For any infinite sequence $\omega = \omega_1\omega_2\ldots$ and for any $k$ the overall probability (generated by this randomization) of the event

$$\left| \frac{1}{n} \sum_{i=1}^{n} I_k(\tilde{p}_i)\delta_k(\omega^{i-1})(\omega_i - \tilde{p}_i) \right| \leq \Delta$$

tends to one as $n \to \infty$,

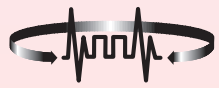
$\tilde{p}_i = f(\alpha : \omega^{i-1})$ is the random variable and $\omega^{i-1}$ is its parameter.

## Computable versions of these notions:

- partial recursive outcome-based checking rules $\{\delta_k\}$;
- *weakly computable* forecasting systems, i.e. such that

$$Pr_n\{\alpha : f(\alpha : \omega^{n-1}) < \frac{1}{2}\}$$

is a partial recursive function from $\omega^{n-1}$.



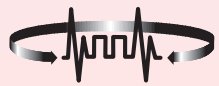
## Computable operation $F$

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

- $\hat{F}$ is recursively enumerable;
- $(x,\lambda) \in \hat{F}$ for each $x$;
- if $(x,y) \in \hat{F}$, $(x',y') \in \hat{F}$ and $x \sqsubseteq x'$ then $y \sqsubseteq y'$ or $y' \sqsubseteq y$ for all finite binary sequences $x, x', y, y'$.

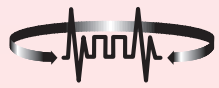
## A probabilistic algorithm is a pair $(L, F)$,

where $L(x) = L(\Gamma_x) = 2^{-l(x)}$ is the uniform measure on $\Omega$ and $F$ is a computable operation.
$L\{\omega : F(\omega) \in A\}$ is the probability of generating by means of $F$ a sequence from $A \subseteq \Omega$ given a uniformly distributed sequence $\omega$.



We show that the construction of the universal forecasting algorithm of Sandrony et al. is computationally non-efficient in a case when the class of all partial recursive outcome-based checking rules $\{\delta_k\}$ is used.

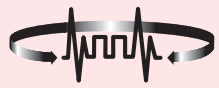
We construct *a probabilistic generator* of *non-learnable* sequences.



## The probabilistic generator of non-learnable sequences:

The generator outputs with probability close to one an infinite sequence $\omega$ such that for each weakly computable randomized forecasting system $\tilde{p}_i = f(\alpha; \omega^{i-1})$

- some computable outcome-based checking rule $\delta$ selects from $\omega$ an infinite subsequence $\omega_{i_1}, \omega_{i_2}, \ldots$ such that
- the sequence of forecasts $\tilde{p}_{i_1}, \tilde{p}_{i_2}, \ldots$ is not calibrated for $\omega_{i_1}, \omega_{i_2}, \ldots$ with the overall probability one.



## Theorem

*For any $\varepsilon > 0$ a probabilistic algorithm $(L, F)$ can be constructed, which with probability $\geq 1 - \varepsilon$ outputs an infinite binary sequence $\omega = \omega_1 \omega_2 \ldots$ such that for every partial weakly computable randomized forecasting system $f$ defined on all initial fragments of the sequence $\omega$ there exists a computable selection rule $\delta$ defined on all these fragments and such that for $\nu = 0$ or for $\nu = 1$ the overall probability of the event*

$$\limsup_{n\to\infty} \left| \frac{1}{n} \sum_{i=1}^{n} \delta(\omega^{i-1}) I_\nu(\tilde{p}_i)(\omega_i - \tilde{p}_i) \right| \geq 1/16$$

*equals one, where $I_0$ and $I_1$ are the characteristic functions of the intervals $[0, \frac{1}{2})$ and $[\frac{1}{2}, 1]$, $\tilde{p}_i = f(\omega^{i-1})$ is a random variable, $i = 1, 2, \ldots$, and the overall probability distribution is associated with $f$.*

The following theorem miscalibrates all partial defined computable deterministic forecasting systems.

**Theorem**

*For any $\varepsilon > 0$ a probabilistic algorithm $(L, F)$ can be constructed, which with probability $\geq 1 - \varepsilon$ outputs an infinite binary sequence $\omega = \omega_1 \omega_2 \ldots$ such that for every partial deterministic forecasting algorithm $f$ defined on all initial fragments of the sequence $\omega$ a computable outcome-based selection rule $\delta$ exists defined on all these fragments such that*

$$\limsup_{n \to \infty} \left| \frac{1}{n} \sum_{i=1}^{n} \delta(\omega^{i-1})(\omega_i - f(\omega^{i-1})) \right| \geq 1/8.$$