◇ ◇

- On impossibility of sequential algorithmic forecasting

- Vladimir V'yugin

- Institute for Information Transmission Problems
  Russian Academy of Sciences

⋄ Statistical forecasting ⋄

Let a 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 the future outcome ω_n.

Statistical method of defining forecasts p_n is based on some overal probability distribution P defined on all possible sequences of outcomes
$\omega_1\omega_2\ldots\omega_n$ for all $n = 1, 2, \ldots$.
The forecasts are defined as the conditional probabilities. In binary case we define
$p_n = P(\omega_n = 1|\omega_1\omega_2\ldots\omega_{n-1}) =$
$P(\omega_1\omega_2\ldots\omega_{n-1}1)/P(\omega_1\omega_2\ldots\omega_{n-1})$.
In practice, P is computable function.

⋄ Outside statistics ⋄

We consider more general approach.
A.P.Dawid's prequential principle says:
evaluation of a probability forecaster should depend only on his actual probability forecasts and the *observed* outcomes $\omega_1\omega_2\dots\omega_{n-1}$.
The additional information contained in a probability measure that has these probability forecasts as conditional probabilities should not enter in the evaluation.
Follow Dawid's *prequential framework* we do not consider numbers p_n as conditional probabilities generating by some overal probability distribution defined on all possible events.

⋄ Partial forecasting systems ⋄

Let $\omega = \omega_1 \dots \omega_n \dots$ be an infinite binary sequence. Forecasting system is a *partial* recursive function $p_n = \varphi(\omega_1 \dots \omega_{n-1})$.
We require only that $\varphi(\omega_1 \dots \omega_{n-1})$ is defined for all initial fragments of the sequence $\omega = \omega_1 \omega_2 \dots$.
Selection rule is a partial computable function δ on the set of all finite binary sequences taking values 0 and 1.
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

2) $n \in s$ just when $\delta(\omega_1 \omega_2 \dots \omega_{n-1}) = 1$.

◇ Algorithmic calibration ◇

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

1) $\varphi(\omega_1\omega_2\dots\omega_n)$ 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

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

Finally, $\omega_1\omega_2\dots\omega_n\dots$ is calibrable if some partial computable φ is calibrated for it; otherwise, $\omega_1\omega_2\dots\omega_n\dots$ is *noncalibrable*.

◇ Probabilistic machines ◇

Informally, probabilistic machine is defined by some algorithm, which when fed with an uniformly distributed finite or infinite sequence ω takes it sequentially bit by bit, processes it and produces an output sequence also sequentially bit by bit.

More correctly, probabilistic machine is a pair (L, F), where L is an uniform measure on the set of all binary sequences and $F(\omega) = \sup\{y \mid \exists x(x \subseteq \omega, \quad (x, y) \in \hat{F}\}$, where $\hat{F}$ is a recursively enumerable set such that

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

$\diamond$ Non-calibrable sequences $\diamond$

Theorem 1. *For any $\epsilon > 0$ there exists a probabilistic algorithm (L, F) which with probability $\geq 1 - \epsilon$ outputs an infinite binary sequence ω such that the following property holds: for each partial computable 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 $\omega_{n_1}, \omega_{n_2}, \ldots$ such that*

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

as $r \longrightarrow \infty$.
In other words, the probabilistic machine (L, F) generates with probability close to one an infinite noncalibrable sequence.

◇ History ◇

The first example of noncalibrabe sequence was presented by Oakes, Schervish, Dawid [1985]. Unfortunately, the methods used do not comply with prequential principle; they depend on assumptions about existence of the measure from which probability forecasts are derived as conditional probabilities.

The same drawback can be assigned to a method of generation noncalibrable sequence with probability arbitrary close to one proposed by V'yugin [1998].

$\diamond$ Relation to defensive forecasting $\diamond$

Vovk et al. [2004] showed that well calibrated forecasting is possible under very broad conditions. A typical example of selection rule used in wheather forecasting:

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

where f is a forecasting system and I_{p^*} is an interval around any point p^*.

Vovk's $K29$ algorithm defines a computable forecasting strategy $p_n = f(\omega_1\dots\omega_{n-1})$ such that for any sequence $\omega_1\omega_2\dots$

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

holds for any point p^* such that $\sum\limits_{i=1}^{n} I_{p^*}(p_i) \to \infty$. Here $I_{p^*}(p)$ is a smooth approximation to characteristic function of an arbitrary interval I_{p^*}.

◇ Relation to defensive forecasting ◇

We constructed "non-smooth" weights $\delta(p)$ such that it is possible to generate with probability close to one a sequence $\omega_1\omega_2\ldots$ for which

$$\frac{\sum_{i=1}^{n} \delta(p_i)(\omega_i - p_i)}{\sum_{i=1}^{n} \delta(p_i)} \not\to 0$$

and $\sum_{i=1}^{n} \delta(p_i) \to \infty$ hold for the sequence of forecasts

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

generated by any computable forecasting system f.