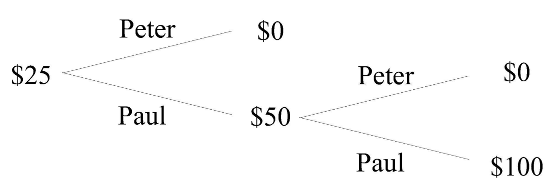
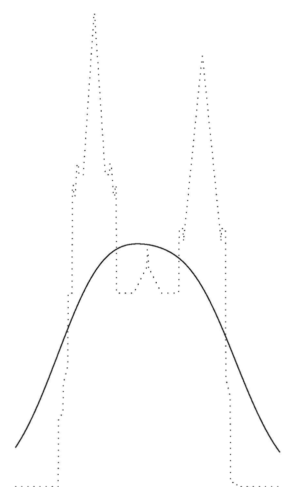
The diachronic Bayesian

Vladimir Vovk





The Game-Theoretic Probability and Finance Project

Working Paper #64

First posted August 24, 2023. Last revised September 21, 2023.

Project web site:
http://www.probabilityandfinance.com

Abstract

It is well known that a Bayesian's probability forecast for the future observations should form a probability measure in order to satisfy natural conditions of coherency. The topic of this paper is the evolution of the Bayesian's probability measure in time. We model the process of updating the Bayesian's beliefs in terms of prediction markets. The resulting picture is adapted to forecasting several steps ahead and making almost optimal decisions.

This paper has also been published as an arXiv report.

Contents

1 Introduction **1**
- 1.1 Notation 2
- 1.2 Dramatis personae 2

2 Basic prediction picture **3**

3 Raw testing protocol **4**

4 Equivalent protocols **6**
- 4.1 Martingale simplification 7
- 4.2 Final simplification 8

5 Super-strong prequential principle **10**

6 Predicting K steps ahead **11**

7 Bayesian decision making **12**
- 7.1 An optimality result for the Bayes decision strategy 12
- 7.2 Predicting $K < N$ steps ahead is essential in Theorem 7.5 15

8 Conclusion **15**

References **16**

A Ideal futures contracts **18**

B Mechanics of futures trading **18**

C Definitions and proofs **20**

1 Introduction

Consider a Bayesian forecasting finitely many discrete future observations. A standard example is where the observations are outcomes of coin tosses. Another standard example is where the observations are "dry" or "rain" for a number of consecutive days. Let us take the standard Bayesian position, due to de Finetti [8] and discussed in, e.g., [3, Sect. 4.1], that the Bayesian's beliefs about the future observations should be encoded as a probability measure on the sequences of observations.

Remark 1.1. One subtlety of de Finetti's views is that the coherence requirement only implies finite additivity and not countable additivity (see, e.g., [8, Sect. 18.3] and [3, Sect. 3.5.2]). In this paper we consider a finite case, in which the difference between finite and countable additivity disappears.

Remark 1.2. A fancier Bayesian picture is the one where, instead of one probability measure P over the future observations, the Bayesian's beliefs are modelled as a statistical model $\{P_\theta \mid \theta \in \Theta\}$ combined with a prior probability measure μ on Θ. The standard Bayesian point of view is that we should start from P and then, if this is more convenient mathematically, represent it as integral $P = \int P_\theta \mu(\mathrm{d}\theta)$. An example is the application of de Finetti's theorem to coin tossing, guaranteeing that any exchangeable probability measure P can be represented as a mixture $\int P_\theta \mu(\mathrm{d}\theta)$ of probability measures P_θ corresponding to independent and identically distributed observations. In Lindley's words, "We should be concentrating not on Greek letters but on the Roman letters" (i.e., not on θs, parameter values, but on the xs and ys, observables) [25, Sect. 7]. This view is sometimes called predictivism [26].

Remark 1.3. In this paper I will ignore any differences that are sometimes made between "forecast" and "prediction" (such as predictions being more categorical than forecasts) and will regard them as synonyms. I will never use more exotic words such as "prevision" [8, Sect. 3.1.2].

I will assume that at each point in time the Bayesian has a probability measure representing his beliefs for the future observations. But how does the Bayesian's probability measure change over time? And, if we are to use the betting interpretation of probability [23, 6, 24], how can we bet against the Bayesian's predictions encoded as probability measures? A standard simple answer is that we should include in our prediction picture **all** information that the Bayesian gets, and then we should condition on the new information in the usual sense of probability theory [15, Sect. I.4]. This procedure for updating the Bayesian's beliefs is known as "Bayesian conditioning" [2, 20, 22]. The principle that the new observations must be the only thing the Bayesian has learned is the *principle of total evidence* [21], and it is usually regarded as uncontroversial. Lewis [16] derives Bayesian conditioning (as updating rule) via his diachronic Dutch Book result, which implicitly relies on the principle of total evidence. In Sect. 2 we will discuss the narrowness of the principle of total evidence and, therefore, of Bayesian conditioning.

Section 3 proposes a testing protocol for the Bayesian's predictions, with discussions following in Sects 4 and 5. Section 6 adapts the testing protocols of the previous sections to predicting K steps ahead, which generalizes the case $K = 1$ considered earlier (in, e.g., [23, 6, 24]). Section 7 applied the testing protocol to making nearly optimal decisions, and Sect. 8 concludes. Appendixes A–C provide further information, with Appendix A defining the idealized prediction markets used in the main part of the paper.

This paper has been inspired by the brief discussion of one-step-ahead prediction in [25, Sect. 7], and its title is adapted from [4].

1.1 Notation

If a and b are finite sequences, we write $a \subseteq b$ to mean that a is a prefix of b, and we write $a \subset b$ to mean that $a \subseteq b$ and $a \neq b$. If $a \subseteq b$, $b \setminus a$ is the sequence obtained from b by crossing out its prefix a. The concatenation of a and b is written simply as ab; we use the same notation when a or b (or both) are elements; if B is a set of elements or finite sequences, aB stands for $\{ab \mid b \in B\}$. The length of a finite sequence a is denoted $|a|$; in particular, $|\Box| = 0$ for the empty sequence $\Box$.

If a and b are numbers, $a \wedge b := \min(a, b)$.

We will also use the following notation:

- $\mathbf{Y}^{m:n}$ stands for the set of all sequences of elements of $\mathbf{Y}$ of length between m and n inclusive (so that $\mathbf{Y}^{0:n}$ stands for the sequences of elements of $\mathbf{Y}$ of length at most n, and $\mathbf{Y}^{1:n}$ stands for the non-empty sequences of elements of $\mathbf{Y}$ of length at most n);
- $\mathfrak{P}(A)$ is the set of all probability measures on A;
- if $P \in \mathfrak{P}(\mathbf{Y}^K)$ and $x \in \mathbf{Y}^k$ for $k \leq K$,

$$P(x) := P(x\mathbf{Y}^{K-k}).$$

To reduce the number of required parentheses, the operator precedence for $:$ and $\wedge$ relative to addition and subtraction, $+/-$, is

$$\wedge, +/-, : .$$

1.2 Dramatis personae

These are the players in our prediction protocols (most of the protocols involve subsets of players).

- Reality (female): player who chooses sequential observations $y_1, y_2, \ldots,$ which are elements of the observation space $\mathbf{Y}$.
- Forecaster (male): player who issues probabilistic forecasts for the future observations.

- Sceptic (male): player who gambles against Forecaster's predictions. Informally, he is trying to discredit Forecaster.
- Decision Maker (female): player who makes decisions in light of Forecaster's predictions.

The players' sexes are defined in [23]. The noun "Bayesian" will often be used as nearly synonymous with "Forecaster", and so the Bayesian will be male.

2 Basic prediction picture

We are interested in the following sequential *Bayesian prediction protocol*.

Protocol 2.1.
FOR $n = 1, \ldots, N$:
Forecaster announces $P_n \in \mathfrak{P}(\mathbf{Y}^{N-n+1})$
Reality announces the actual observation $y_n \in \mathbf{Y}$.

In this paper we only consider the case of a finite horizon $N > 1$. At each step n, P_n is a prediction for the whole future $y_n y_{n+1} \ldots y_N$ (sequence of length $N - n + 1$). Earlier we have referred to this player as "Bayesian", in order to emphasize that his predictions are complete probability measures over the future observations (while in earlier work we often considered less complete predictions: see, e.g., [23, Preface], point 2).

Let us assume, for simplicity, that the set $\mathbf{Y}$ is finite (and equipped with the discrete σ-algebra); this will allow us to concentrate of conceptual issues avoiding technical difficulties and ambiguities (such as countable vs finite additivity). To exclude trivialities, we also assume $|\mathbf{Y}| > 1$.

In addition, we impose the requirement that $P_n(E) > 0$ unless $E = \emptyset$. This is a version of Lindley's "Cromwell's rule" [17, Sect. 6.7].

Protocol 2.1 goes beyond *Bayesian conditioning*, where we insist that, for each $n \geq 2$,

$$P_n(\{x\}) = P_{n-1}(x \mid y_{n-1}) := P_{n-1}(\{y_{n-1}x\})/P_{n-1}(y_{n-1}), \quad x \in \mathbf{Y}^{N-n+1}.$$

Bayesian conditioning (as rule for updating beliefs) was criticised by Hacking in 1967 [11], although his criticism (ignoring the cost of thinking) is largely irrelevant to this paper. What is more relevant to our picture is that at step n Forecaster can also learn other information apart from y_n (i.e., learn information outside the protocol); see Shafer [21].

Let us give an example where Bayesian conditioning, based on the principle of total evidence, is utterly unrealistic as an updating rule: we can't hope to have a comprehensive protocol including all the information a real-life Bayesian has access to. Consider the standard case [4, 5] of a weather forecaster who issues a probability for the rain on sequentially numbered days. The observations are the actual outcomes, say 0 or 1 (encoding a dry or rainy day). In the morning of day 1 the forecaster announces a joint probability for the future observations (for days $1, 2, \dots$) as his forecast, and in the morning of day 2 he announces a new forecast, for days $2, 3, \dots$. We can't assume that the observation of day 1 is all the extra information that he has in the morning of day 2: a serious weather forecaster, such as the UK Met Office, will have plenty of other information arriving from weather stations around the globe (and even from outer space). We will sometimes use the notation $\mathcal{F}_{n-1}$ (formally this is a σ-algebra) for the information available when making the prediction P_n at time n; this is an outlandish notion, difficult to imagine (while the moves in our protocols will be observable).

Remark 2.2. Obviously, we can't include the data arriving from weather stations around the globe in a realistic prediction protocol, but we can go further and argue that even our picture is unrealistic for a large time horizon N: e.g., the first probability measure $P_1 \in \mathfrak{P}(\mathbf{Y}^N)$ specifies $|\mathbf{Y}|^N - 1$ independent parameters, and this number grows exponentially in N even for $|\mathbf{Y}| = 2$. In Sect. 6 we consider a more realistic setting of forecasting K steps ahead (such as a week ahead for $K = 7$).

Remark 2.3. Another reason why we might want to consider a Bayesian violating Bayesian conditioning when updating his beliefs is that his computational resources might be limited: he might keep processing information already available at the previous steps obtaining new values for probabilities of the same events. This is Hacking's [11] observation mentioned earlier.

Remark 2.4. We will only consider certain evidence and so will not discuss generalizations of Bayesian conditioning such as Jeffrey's ([14, Chap. 11], [19]).

Long-term prediction is much more complicated than one-step-ahead prediction that we have considered earlier [6, 23], and to have a clear understanding of the process we will use two pictures (which we have referred to as game-theoretic and measure-theoretic). The first picture is financial and motivated by prediction markets: we complement the basic forecasting protocol with a "market" allowing a third player, Sceptic, to trade in futures contracts (these are the most standard financial derivatives; see, e.g., [13, Chap. 2] and [10]). Futures contracts is an old idea (see, e.g., [18]) that arose gradually in financial industry, but in our prediction protocols it will be a powerful way of reducing prediction multiple steps ahead to one-step-ahead prediction. This picture will be formalized and developed in Sects 3 and 4.

The second picture is stochastic and uses Dawid's super-strong prequential principle [6, Sect. 5.2]. It will be discussed in Sect. 5.

3 Raw testing protocol

The following extension of Protocol 2.1 describes a way of testing Forecaster's predictions.

Protocol 3.1.

$\mathcal{K}_0 := 1$

Forecaster announces $P_1 \in \mathfrak{P}(\mathbf{Y}^N)$
Sceptic announces $f_1 \in \mathbb{R}^{\mathbf{Y}^{1:N}}$
Reality announces $y_1 \in \mathbf{Y}$
$\mathcal{K}'_1 := \mathcal{K}_0 + f_1(y_1) - \sum_y f_1(y) P_1(y)$
FOR $n = 2, \dots, N$:
Forecaster announces $P_n \in \mathfrak{P}(\mathbf{Y}^{N-n+1})$

$$\mathcal{K}_{n-1} := \mathcal{K}'_{n-1} + \sum_{x \in \mathbf{Y}^{1:N-n+1}} f_{n-1}(y_{n-1}x) P_n(x) - \sum_{x \in \mathbf{Y}^{2:N-n+2}} f_{n-1}(x) P_{n-1}(x) \tag{1}$$

Sceptic announces $f_n \in \mathbb{R}^{\mathbf{Y}^{1:N-n+1}}$
Reality announces $y_n \in \mathbf{Y}$

$$\mathcal{K}'_n := \mathcal{K}_{n-1} + f_n(y_n) - \sum_{y \in \mathbf{Y}} f_n(y) P_n(y). \tag{2}$$

Protocol 3.1 does not define $\mathcal{K}_N$, and we set $\mathcal{K}_N := \mathcal{K}'_N$.

The financial interpretation of Protocol 3.1 is that we have a market of futures contracts $\Phi(x)$, $x \in \mathbf{Y}^{1:N}$, that pay

$$F_m^+(x) := 1_{\{y_1 \dots y_m = x\}}$$

at the end of step $m := |x|$. See Appendix A for the definition. At each step n Forecaster announces the prices for all the futures contracts

$$\Phi(x), \quad y_1 \dots y_{n-1} \subset x \in \mathbf{Y}^N,$$

in the form of a probability measure $P_n \in \mathfrak{P}(\mathbf{Y}^{N-n+1})$; namely, the price of $\Phi(x)$, $x \in \mathbf{Y}^N$, at step n is

$$F_n(x) := \begin{cases} P_n(\{x \setminus y_1 \dots y_{n-1}\}) & \text{if } y_1 \dots y_{n-1} \subset x \\ 0 & \text{if not.} \end{cases} \tag{3}$$

We assume, without loss of generality, that these prices indeed form a probability measure; otherwise, the market is not coherent and Sceptic can become arbitrarily rich [8, Chap. 3]. For example, if

$$\sum_{x \in \mathbf{Y}^N : y_1 \dots y_{n-1} \subset x} \Phi(x) < 1,$$

Sceptic can take constant position 1 in all these futures contracts $\Phi(x)$ at step n, which will bring him a sure gain of

$$1 - \sum_{x \in \mathbf{Y}^N : y_1 \dots y_{n-1} \subset x} \Phi(x)$$

at the end of step N. The gain can be scaled up arbitrarily.

In principle, at step n we need, in addition to the current prices $F_n(x)$ of $\Phi(x)$, $x \in \mathbf{Y}^N$, also the current prices $F_n(x)$ of $\Phi(x)$ for $x \in \mathbf{Y}^{n:N-1}$. However, coherence implies that we can assume, without loss of generality, that the only possibly non-zero values of F_n are

$$F_n(x) = \sum_{x' \in \mathbf{Y}^N : x \subset x'} F_n(x') = \sum_{x' \in \mathbf{Y}^N : x \subset x'} P_n(x') = P_n(x)$$

for all $x \in \mathbf{Y}^{n:N-1}$ such that $y_1 \dots y_{n-1} \subset x$. For example, if

$$F_n(x) < \sum_{x' \in \mathbf{Y}^N : x \subset x'} F_n(x')$$

for some $x \in \mathbf{Y}^{n:N-1}$ with $y_1 \dots y_{n-1} \subset x$, Sceptic can take constant position 1 in $\Phi(x)$ and constant position -1 in each $\Phi(x')$, $x \subset x' \in \mathbf{Y}^N$, which will bring him a sure gain of

$$-F_n(x) + \sum_{x' \in \mathbf{Y}^N : x \subset x'} F_n(x').$$

At step n Sceptic needs to take positions in all $\Phi(x)$, $y_1 \dots y_{n-1} \subset x \in \mathbf{Y}^{1:N}$. The position in $\Phi(y_1 \dots y_{n-1} x)$ is denoted $f_n(x)$ in Protocol 3.1.

After y_n is disclosed by Reality, the increment in Sceptic's capital (due to the futures contracts $\Phi(y_1 \dots y_{n-1} y)$) is

$$\begin{aligned}
\mathcal{K}'_n - \mathcal{K}_{n-1} &= \sum_{y \in \mathbf{Y}} f_n(y) \left(F_n^+(y_1 \dots y_{n-1} y) - F_n(y_1 \dots y_{n-1} y) \right) \\
&= f_n(y_n) - \sum_{y \in \mathbf{Y}} f_n(y) P_n(y),
\end{aligned}$$

which agrees with (2). And after P_n is disclosed by Forecaster, the increment in Sceptic's capital (due to the remaining futures contracts) is

$$\begin{aligned}
\mathcal{K}_{n-1} - \mathcal{K}'_{n-1} &= \sum_{x \in \mathbf{Y}^{2:N-n+2}} f_{n-1}(x) \big(F_n(y_1 \dots y_{n-2} x) - F_{n-1}(y_1 \dots y_{n-2} x) \big) \\
&= \sum_{x \in \mathbf{Y}^{1:N-n+1}} f_{n-1}(y_{n-1} x) F_n(y_1 \dots y_{n-1} x) \\
&\quad - \sum_{x \in \mathbf{Y}^{2:N-n+2}} f_{n-1}(x) F_{n-1}(y_1 \dots y_{n-2} x) \\
&= \sum_{x \in \mathbf{Y}^{1:N-n+1}} f_{n-1}(y_{n-1} x) P_n(x) - \sum_{x \in \mathbf{Y}^{2:N-n+2}} f_{n-1}(x) P_{n-1}(x),
\end{aligned}$$

where the last equality follows from (3) and the last expression agrees with (1).

4 Equivalent protocols

We can rewrite Protocol 3.1 in different forms, getting rid of some of Forecaster's arbitrary and irrelevant choices. To compare protocol with the same allowed moves for Reality and Forecaster, we will use the notion of the *test martingale space* (TSM) defined as follows. A strategy for Sceptic specifies his move as function of Forecaster's and Reality's previous moves, $d_n = d_n(P_1, y_1, \dots, P_n)$ in the case of Protocol 3.1 (and we do not impose any measurability conditions on strategies in this section). The corresponding *test martingale* is Sceptic capital $\mathcal{K}_n$ (for all possible n) as function of Forecaster's and Reality's moves provided this function is nonnegative. The TSM of a given protocol is the set of all possible test martingales. We regards two protocols to be equivalent if they have the same TMS.

The testing protocol introduced in the previous section was formulated with a view towards an infinite time horizon, where N becomes ∞, and this is continued in Sect. 4.1. The following Sect. 4.2 introduces a much simpler protocol using the finite horizon N in a more essential way.

4.1 Martingale simplification

First of all, we can simplify (2) by requiring that $\sum_{y\in\mathbf{Y}} f_n(y)P_n(y) = 0$. This way we will get rid of the subtrahend in (2). More generally, we have the following proposition.

Proposition 4.1. *We can assume, without loss of generality, that f_n is a* martingale difference *w.r. to P_n, in the sense that, for each $x \in \mathbf{Y}^{0:N-n}$,*

$$\sum_{y\in\mathbf{Y}} f_n(xy)P_n^x(y) = 0, \tag{4}$$

where P_n^x is the conditional probability measure

$$P_n^x(y) := P_n(y \mid x) := \frac{P_n(xy)}{P_n(x)}.$$

Formally, Protocol 3.1 and Protocol 4.2 below have identical TMS.

Proposition 4.1 gives us the representation of Protocol 3.1 as Protocol 4.2. And these protocols are equivalent in the sense of leading to the same capitals $\mathcal{K}_0, \mathcal{K}_1, \dots$ for Sceptic. Protocol 4.2 is simpler than Protocol 3.1 since the subtrahends in (1) and (2) are zero when f_n is a martingale difference w.r. to P_n for all n.

Protocol 4.2.

$\mathcal{K}_0 := 1$
Forecaster announces $P_1 \in \mathfrak{P}(\mathbf{Y}^N)$
Sceptic announces a martingale difference f_1 w.r. to P_1
Reality announces $y_1 \in \mathbf{Y}$
$\mathcal{K}'_1 := \mathcal{K}_0 + f_1(y_1)$
FOR $n = 2, \dots, N$:
  Forecaster announces $P_n \in \mathfrak{P}(\mathbf{Y}^{N-n+1})$
  $\mathcal{K}_{n-1} := \mathcal{K}'_{n-1} + \sum_{x\in\mathbf{Y}^{1:N-n+1}} f_{n-1}(y_{n-1}x)P_n(x)$ (5)
  Sceptic announces a martingale difference f_n w.r. to P_n
  Reality announces $y_n \in \mathbf{Y}$
  $\mathcal{K}'_n := \mathcal{K}_{n-1} + f_n(y_n).$ (6)

Proof of Proposition 4.1. The key observation is that, for any $x \in \mathbf{Y}^{2:N-n}$, adding any constant c to $f_n(x)$ and subtracting the same constant c from all

$f_n(xy)$, $y \in \mathbf{Y}$, will not change the increment in the capital $\mathcal{K}_n - \mathcal{K}'_n$ given by (1) at the next step. Namely, this operation, which we will denote $O(x, c)$, will change neither minuend nor subtrahend in (1). This way we can ensure (4) for each $x \in \mathbf{Y}^{2:N-n}$: first we apply $O(x, c)$ for suitable c to $x \in \mathbf{Y}^{N-n}$, then to $x \in \mathbf{Y}^{N-n}$, etc. Next, applying the operation $O(x, c)$ for any $x \in \mathbf{Y}$ does not change the increment in the capital $\mathcal{K}_n - \mathcal{K}_{n-1}$ given by (2) and by (1) at the next step. Indeed,

- the change in the sum in (2) and in the second sum in (1) at the next step will balance each other out, and
- the change in the term $f_n(y_n)$ in (2) and in the first sum in (1) at the next step will also balance each other out.

This operation can ensure (4) for each $x \in \mathbf{Y}$. Finally, adding the same constant to each $f_n(y)$, $y \in \mathbf{Y}$ will affect neither (2) nor (1) at the next step, and it will ensure (4) for $x = \Box$.

It remains to consider the last step, $n = N$. Sceptic then announces $f_n \in \mathbb{R}^{\mathbf{Y}}$, and adding a constant to f_n does not affect $\mathcal{K}'_N - \mathcal{K}_{N-1}$. $\Box$

How does (1) compare with Bayesian conditioning, where no new information outside the protocol arrives and we just set

$$P_n(\{x\}) := P_{n-1}^{y_{n-1}}(x) := \frac{P_{n-1}(y_{n-1}x)}{P_{n-1}(y_{n-1})} \tag{7}$$

for all $x \in \mathbf{Y}^{N-n+1}$? In this case line (5) becomes $\mathcal{K}_{n-1} := \mathcal{K}'_{n-1}$ and can be removed (if we drop the prime in (6)).

Instead of imposing a restriction on f_n, we could replace line (5) by

$$\mathcal{K}_{n-1} := \mathcal{K}'_{n-1} + \sum_{x \in \mathbf{Y}^{0:N-n}} \sum_{y \in \mathbf{Y}} P_n(x) f_{n-1}(y_{n-1}xy) \left(P_n^x(y) - P_{n-1}^{y_{n-1}x}(y) \right).$$

Informally, we assume that the expectation of $P_n^x(y)$ at the previous step is $P_{n-1}^{y_{n-1}x}(y)$; therefore, the protocol allows Sceptic, for each x and each possible y_{n-1}, to buy (at step $n-1$ after observing Forecaster's move) any number (positive, zero, or negative) of tickets with payoff

$$P_n^x(y) - P_{n-1}^{y_{n-1}x}(y).$$

4.2 Final simplification

Now let us simplify Protocol 3.1 more radically.

Protocol 4.3.

$\mathcal{K}_0 := 1$
FOR $n = 1, \dots, N$:
  Forecaster announces $P_n \in \mathfrak{P}(\mathbf{Y}^{N-n+1})$
  IF $n > 1$:

$$\mathcal{K}_{n-1} := \mathcal{K}_{n-2} + \textstyle\sum_{x\in\mathbf{Y}^{N-n+1}} f_{n-1}(y_{n-1}x)P_n(x) \\ \quad - \textstyle\sum_{x\in\mathbf{Y}^{N-n+2}} f_{n-1}(x)P_{n-1}(x) \tag{8}$$

Sceptic announces $f_n \in \mathbb{R}^{\mathbf{Y}^{N-n+1}}$
Reality announces $y_n \in \mathbf{Y}$
IF $n = N$:

$$\mathcal{K}_n := \mathcal{K}_{n-1} + f_n(y_n) - \textstyle\sum_{y\in\mathbf{Y}} f_n(y)P_n(y). \tag{9}$$

Proposition 4.4. *Protocol 3.1 and Protocol 4.3 have identical TMS.*

Proof. As discussed in the proof of Proposition 4.1, the operation $O(x, c)$ does not change the increment $\mathcal{K}_n - \mathcal{K}_{n-1}$ in Protocol 3.1 for any $x \in \mathbf{Y}^{1:N-n}$ (we discussed separately the cases $x \in \mathbf{Y}^{2:N-n}$ and $x \in \mathbf{Y}$). Therefore, we can assume, without loss of generality (i.e., without changing the TMS), that $f_n(x)$ is different from 0 only for $x \in \mathbf{Y}^{N-n+1}$, which implies that:

- we can ignore (2) for all steps n apart from $n = N$, and so (9) is performed only for $n = N$;
- we can ignore the bits "1 : " and "2 : " in (1), obtaining (8).

Protocol 4.3 also merges the four lines in Protocol 3.1 preceding the FOR loop into the loop. □

Proposition 4.4 simplifies the market in futures contracts that we need: all the contracts now mature at the end of step N.

In the case of Bayesian conditioning (7), we can simplify the protocol by replacing Sceptic's moves f_n by

$$f'_n(y) := \sum_{x\in\mathbf{Y}^{N-n}} f_n(yx)P_n(x \mid y),$$

in which case (8) becomes

$$\mathcal{K}_{n-1} := \mathcal{K}_{n-2} + f'_{n-1}(y_{n-1}) - \sum_{y\in\mathbf{Y}} f'_{n-1}(y)P_{n-1}(y).$$

Moving this command to the previous step, we can rewrite Protocol 4.3 as

Protocol 4.5.
$\mathcal{K}_0 := 1$
FOR $n = 1, \dots, N$:
IF $n = 1$:
Forecaster announces $P_1 \in \mathfrak{P}(\mathbf{Y}^N)$
ELSE:
Forecaster updates $P_{n-1} \in \mathfrak{P}(\mathbf{Y}^{N-n+2})$ to $P_n \in \mathfrak{P}(\mathbf{Y}^{N-n+1})$ by Bayesian conditioning (7)
Sceptic announces $f'_n \in \mathbb{R}^{\mathbf{Y}}$
Reality announces $y_n \in \mathbf{Y}$
$\mathcal{K}_n := \mathcal{K}_{n-1} + f'_n(y_n) - \sum_{y\in\mathbf{Y}} f'_n(y)P_n(y)$.

This is our standard one-step-ahead prediction protocol (cf., e.g., [23, Protocol 1.1]) except that Forecaster announces his forecasting strategy in advance.

5 Super-strong prequential principle

As a sanity check, in this section we will see whether our prediction protocols conform to the super-strong prequential principle introduced in [6, Sect. 5.2]. The super-strong prequential principle embeds those protocols in the standard measure-theoretic framework [15, 9] for probability. An informal statement of the super-strong prequential principle is:

> Suppose that Forecaster and Reality are colluding and compute/simulate their moves from a given probability measure P and that Sceptic follows a measurable strategy. Then Sceptic's capital is a martingale w.r. to P.

We will, at the same time, both formalize and check the super-strong prequential principle for our protocols. Let P be a probability measure on a measurable space $(\Omega, \mathcal{F})$ equipped with a filtration $\mathcal{F}_n$, $n = 0, 1, \dots$. Intuitively, we regard $\mathcal{F}_{n-1}$ as the information available to Forecaster and Sceptic at the beginning of step n in Protocols 3.1, 4.2, or 4.3; by step 1 I mean the statements preceding the FOR loop in the protocols. For concreteness, let us assume that all new information (including y_n, which is part of the new information) arrives at the end of step n and none arrives between the steps; therefore, $\mathcal{F}_n$, $n = 1, 2, \dots$, is the information available at the end of step n.

Let us concentrate on Protocol 3.1. Now $y_1, y_2, \dots$ become an adapted process (meaning that y_n is $\mathcal{F}_n$-measurable, $n = 1, 2, \dots$). Sceptic's capital $\mathcal{K}_n$ also becomes an adapted process. We have

$$P_n(\{x\}) = \mathbb{P}\left(\{y_n \dots y_N = x\} \mid \mathcal{F}_{n-1}\right) \quad \text{a.s.,} \quad x \in \mathbf{Y}^{N-n+1},$$

and so

$$P_n(x) = \mathbb{P}\left(\{x \subseteq y_n \dots y_N\} \mid \mathcal{F}_{n-1}\right) \quad \text{a.s.,} \quad x \in \mathbf{Y}^{1:N-n+1},$$

Now the first increment (1) in Sceptic's capital is

$$\begin{aligned}\mathcal{K}_{n-1} - \mathcal{K}'_{n-1} = &\sum_{x \in \mathbf{Y}^{1:N-n+1}} f_{n-1}(y_{n-1}x)\mathbb{P}(\{x \subseteq y_n \dots y_N\} \mid \mathcal{F}_{n-1}) \\ &- \sum_{x \in \mathbf{Y}^{2:N-n+2}} f_{n-1}(x)\mathbb{P}(\{x \subseteq y_{n-1} \dots y_N\} \mid \mathcal{F}_{n-2})\end{aligned}$$

and so we have

$$\mathbb{E}(\mathcal{K}_{n-1} - \mathcal{K}'_{n-1} \mid \mathcal{F}_{n-2}) = 0 \quad \text{a.s.} \tag{10}$$

The second increment (2) is

$$\mathcal{K}'_n - \mathcal{K}_{n-1} = f_n(y_n) - \sum_y f_n(y)\mathbb{P}(\{y_n = y\} \mid \mathcal{F}_{n-1}) = f_n(y_n) - \mathbb{E}(f_n \mid \mathcal{F}_{n-1}),$$

which gives the analogue

$$\mathbb{E}(\mathcal{K}'_n - \mathcal{K}_{n-1} \mid \mathcal{F}_{n-1}) = 0 \quad \text{a.s.}$$

of (10). Therefore, $\mathcal{K}_0, \mathcal{K}_1, \dots, \mathcal{K}_{N-1}, \mathcal{K}'_N$ is a martingale w.r. to the filtration $(\mathcal{F}_n)$.

6 Predicting K steps ahead

For a large N, the protocols considered in the previous sections are unrealistic in that Forecaster is asked to produce a probability measure on a huge set $\mathbf{Y}^N$. Starting from this section, we will assume that all predictions made by Forecaster are only for the next $K < N$ observations, with $K \geq 1$.

The Bayesian prediction protocol (Protocol 2.1) becomes:

Protocol 6.1.

FOR $n = 1, \ldots, N$:

Forecaster announces $P_n \in \mathfrak{P}(\mathbf{Y}^{K \wedge (N-n+1)})$

Reality announces the actual observation $y_n \in \mathbf{Y}$.

The previous theory applies, but now Sceptic is not allowed to bet more than K steps ahead. This gives the following modification of the raw testing protocol (Protocol 3.1):

Protocol 6.2.

$\mathcal{K}_0 := 1$

Forecaster announces $P_1 \in \mathfrak{P}(\mathbf{Y}^K)$

Sceptic announces $f_1 \in \mathbb{R}^{\mathbf{Y}^{1:K}}$

Reality announces $y_1 \in \mathbf{Y}$

$\mathcal{K}'_1 := \mathcal{K}_0 + f_1(y_1) - \sum_y f_1(y) P_1(y)$

FOR $n = 2, \ldots, N$:

Forecaster announces $P_n \in \mathfrak{P}(\mathbf{Y}^{K \wedge (N-n+1)})$

$$\mathcal{K}_{n-1} := \mathcal{K}'_{n-1} + \sum_{x \in \mathbf{Y}^{1:(K-1) \wedge (N-n+1)}} f_{n-1}(y_{n-1} x) P_n(x) - \sum_{x \in \mathbf{Y}^{2:K \wedge (N-n+2)}} f_{n-1}(x) P_{n-1}(x) \qquad (11)$$

Sceptic announces $f_n \in \mathbb{R}^{\mathbf{Y}^{1:K \wedge (N-n+1)}}$

Reality announces $y_n \in \mathbf{Y}$

$$\mathcal{K}'_n := \mathcal{K}_{n-1} + f_n(y_n) - \sum_{y \in \mathbf{Y}} f_n(y) P_n(y). \qquad (12)$$

The simplified testing protocol (Protocol 4.2) becomes:

Protocol 6.3.

$\mathcal{K}_0 := 1$

Forecaster announces $P_1 \in \mathfrak{P}(\mathbf{Y}^K)$

Sceptic announces a martingale difference f_1 w.r.t. to P_1

Reality announces $y_1 \in \mathbf{Y}$

$\mathcal{K}'_1 := \mathcal{K}_0 + f_1(y_1)$

FOR $n = 2, \ldots, N$:

Forecaster announces $P_n \in \mathfrak{P}(\mathbf{Y}^{K \wedge (N-n+1)})$

$\mathcal{K}_{n-1} := \mathcal{K}'_{n-1} + \sum_{x \in \mathbf{Y}^{1:(K-1) \wedge (N-n+1)}} f_{n-1}(y_{n-1} x) P_n(x)$

Sceptic announces a martingale difference f_n w.r.t. to P_n

Reality announces $y_n \in \mathbf{Y}$

$\mathcal{K}'_n := \mathcal{K}_{n-1} + f_n(y_n)$.

The martingale difference f_n is a function $f_n : \mathbf{Y}^{1:K \wedge (N-n+1)} \to \mathbb{R}$; its domain is determined by the domain of $P_n \in \mathfrak{P}(\mathbf{Y}^{K \wedge (N-n+1)})$. Similarly, $f_1 : \mathbf{Y}^{1:K} \to \mathbb{R}$.

Remark 6.4. In the example of weather forecasting one week ahead (cf. Remark 2.2), the predictions in Protocol 6.1 are quite different from the predictions produced by a typical weather app. Weather apps produce marginal probabilities of rain whereas the probabilities in Protocol 6.1 are joint. Testing marginal probabilities would be much easier than the kind of testing exemplified by Protocols 6.2 and 6.3.

7 Bayesian decision making

Why do we need long-term forecasts? One reason is that they facilitate nearly optimal decisions.

7.1 An optimality result for the Bayes decision strategy

Consider the following decision-making protocol.

Protocol 7.1.

FOR $n = 1, \dots, N$:
  Reality announces $\lambda_n : \mathbf{D} \times \mathbf{Y}^{N-n+1} \to [0, 1]$
  Decision Maker announces $d_n \in \mathbf{D}$
  Reality announces the actual observation $y_n \in \mathbf{Y}$.

At each step n Decision Maker is asked to choose a decision d_n from a finite set $\mathbf{D}$. Reality announces a loss function λ_n determining Decision Maker's loss

$$\lambda_n(d_n, y_n \dots y_N) \in [0, 1]$$

at this step. In applications the loss functions are usually given in advance, but we include them in the protocol in other to weaken the conditions of our mathematical result (Theorem 7.5 below). The loss functions are assumed bounded and scaled to the interval $[0, 1]$. The total loss can be computed only after the last step and equals

$$\text{Loss}_N := \sum_{n=1}^{N} \lambda_n(d_n, y_n \dots y_N) \in [0, N]. \tag{13}$$

Of course, Loss_N is a function of Reality's and Decision Maker's moves, but we will leave the arguments of Loss_N implicit.

A strategy for Decision Maker in Protocol 7.1 is a function giving a decision d_n at each step n as function of Reality's previous moves $y_1, \dots, y_{n-1}$ and $\lambda_1, \dots, \lambda_n$. It would be ideal to have a strategy A for Decision Maker that is provably either better than any other strategy B or approximately equally good, but this is clearly impossible; our decision making protocol is too poor for that.

As a first step towards the goal of designing an optimal (in some sense) strategy for Decision Maker, we add a new player, Forecaster, to Protocol 7.1. The following protocol is a combination of Protocols 7.1 and 2.1.

Protocol 7.2.

FOR $n = 1, \dots, N$:

Reality announces $\lambda_n : \mathbf{D} \times \mathbf{Y}^{N-n+1} \to [0,1]$

Forecaster announces $P_n \in \mathfrak{P}(\mathbf{Y}^{N-n+1})$

Decision Maker announces $d_n \in \mathbf{D}$

Reality announces the actual observation $y_n \in \mathbf{Y}$.

Protocol 7.2 allows us to design a plausible strategy ("Bayes strategy", or "Bayes optimal strategy") for Decision Maker (where d_n is now allowed to depend, additionally, on Forecaster's previous moves $P_1, \dots, P_n$):

$$d_n \in \arg\min_{d \in \mathbf{D}} \sum_{x \in \mathbf{Y}^{N-n+1}} \lambda_n(d, x) P_n(x). \tag{14}$$

However, we cannot prove anything about this strategy as we do not know anything about connections between the forecasts P_n and the actual observations y_n. Therefore, we add Sceptic to our protocol, as in Protocol 4.2.

Protocol 7.3.

$\mathcal{K}_0 := 1$

Reality announces $\lambda_1 : \mathbf{D} \times \mathbf{Y}^N \to [0,1]$

Forecaster announces $P_1 \in \mathfrak{P}(\mathbf{Y}^N)$

Decision Maker announces $d_1 \in \mathbf{D}$

Sceptic announces a martingale difference f_1 w.r. to P_1

Reality announces $y_1 \in \mathbf{Y}$

$\mathcal{K}'_1 := \mathcal{K}_0 + f_1(y_1)$

FOR $n = 2, \dots, N$:

Reality announces $\lambda_n : \mathbf{D} \times \mathbf{Y}^{N-n+1} \to [0,1]$

Forecaster announces $P_n \in \mathfrak{P}(\mathbf{Y}^{N-n+1})$

Decision Maker announces $d_n \in \mathbf{D}$

$\mathcal{K}_{n-1} := \mathcal{K}'_{n-1} + \sum_{x \in \mathbf{Y}^{1:N-n+1}} f_{n-1}(y_{n-1}x) P_n(x)$

Sceptic announces a martingale difference f_n w.r. to P_n

Reality announces $y_n \in \mathbf{Y}$

$\mathcal{K}'_n := \mathcal{K}_{n-1} + f_n(y_n)$.

In order to prove a law of large numbers for decision making we need the following combination of Protocols 7.3 and 6.3.

Protocol 7.4.

$\mathcal{K}_0 := 1$

Reality announces $\lambda_1 : \mathbf{D} \times \mathbf{Y}^K \to [0,1]$

Forecaster announces $P_1 \in \mathfrak{P}(\mathbf{Y}^N)$

Decision Maker announces $d_1 \in \mathbf{D}$

Sceptic announces a martingale difference f_1 w.r. to P_1

Reality announces $y_1 \in \mathbf{Y}$

$\mathcal{K}'_1 := \mathcal{K}_0 + f_1(y_1)$

FOR $n = 2, \dots, N$:

Reality announces $\lambda_n : \mathbf{D} \times \mathbf{Y}^{K \wedge (N-n+1)} \to [0,1]$

Forecaster announces $P_n \in \mathfrak{P}(\mathbf{Y}^{K\wedge(N-n+1)})$
Decision Maker announces $d_n \in \mathbf{D}$
$\mathcal{K}_{n-1} := \mathcal{K}'_{n-1} + \sum_{x\in\mathbf{Y}^{1:(K-1)\wedge(N-n+1)}} f_{n-1}(y_{n-1}x)P_n(x)$
Sceptic announces a martingale difference f_n w.r.t. to P_n
Reality announces $y_n \in \mathbf{Y}$
$\mathcal{K}'_n := \mathcal{K}_{n-1} + f_n(y_n)$.

We will continue to use the notation Loss_N introduced in (13), but we will also be interested in Decision Maker's loss $\text{Loss}_N(A)$ computed by replacing his actual decisions by decisions recommended by a decision strategy A:

$$\text{Loss}_N(A) := \sum_{n=1}^{N} \lambda_n(d_n^A, y_n \dots y_{(n+K-1)\wedge N}),$$

where

$$d_n^A := A(\lambda_1, P_1, y_1, \lambda_2, P_2, \dots, y_{n-1}\lambda_n, P_n), \quad n = 1, 2, \dots;$$

we are only interested in strategies that are functions of the previous moves by the other players. Let us adapt the strategy (14) to Protocol 7.4:

$$d_n := d_n^0 \in \arg\min_{d\in\mathbf{D}} \sum_{x\in\mathbf{Y}^{K\wedge(N-n+1)}} \lambda_n(d,x)P_n(x), \tag{15}$$

with d_n^0 chosen as the first element of the arg min in a fixed linear order on $\mathbf{D}$ if there are ties among d. The following optimality result will be proved in Appendix C.

Theorem 7.5. *Let $\epsilon > 0$. There is a strategy A for Decision Maker in Protocol 7.4 that guarantees*

$$\overline{\mathbb{P}}\left(\frac{1}{N}\left(\text{Loss}_N(A) - \text{Loss}_N\right) \geq \epsilon\right) \leq \exp\left(-\frac{N\epsilon^2}{8K^2}\right). \tag{16}$$

This statement of Theorem 7.5 uses the notion of game-theoretic probability $\overline{\mathbb{P}}$ defined in, e.g., [23]. An alternative statement of Theorem 7.5 is that there exists a joint strategy for Decision Maker and Sceptic that achieves either

$$\frac{1}{N}\left(\text{Loss}_N(A) - \text{Loss}_N\right) < \epsilon$$

or $\mathcal{K}_N \geq \exp(N\epsilon^2/(8K^2))$. For a small ϵ and large N (as compared with $K^2\epsilon^{-2}$), this joint strategy demonstrates that A performs better than or within ϵ of the actual moves d_n unless Forecaster is discredited. This is a version of the law of large numbers that works only when $K \ll \sqrt{N}$.

The strategy A in the statement of Theorem 7.5 can be chosen as (15). Theorem 7.5 shows that, for any other strategy B for Sceptic, we have

$$\overline{\mathbb{P}}\left(\frac{1}{N}\left(\text{Loss}_N(A) - \text{Loss}_N(B)\right) \geq \epsilon\right) \leq \exp\left(-\frac{N\epsilon^2}{8K^2}\right) \tag{17}$$

instead of (16); we, however, prefer a stronger statement allowing Forecaster to choose his moves on the fly.

Remark 7.6. In Theorem 7.5 we compare Decision Maker's actual loss Loss_N with the loss she would have suffered following the strategy A defined by (15). Our interpretation of this theorem depends on the assumption that Reality's and Forecaster's moves are not affected by Decision Maker's moves.

7.2 Predicting $K < N$ steps ahead is essential in Theorem 7.5

Theorem 7.5 is about predicting K steps ahead. How important is this restriction? Let us check that it ceases to hold when $K = N$ in Protocol 7.4. The intuition is that at each step Decision Maker is asked to predict the last outcome, and this creates heavy dependence between losses at different steps that ruins the law of large numbers.

Set $\mathbf{D} := \mathbf{Y} := \{0, 1\}$, and suppose (in the spirit of the super-prequential principle) that all players know and comply with a probability measure $P \in \mathfrak{P}(\{0,1\}^N)$ governing Reality. The loss functions output by Reality are

$$\lambda_n(d_n, y_n \dots y_N) := \begin{cases} 0 & \text{if } d_n = y_N \\ 1 & \text{otherwise,} \end{cases}$$

and the true probability measure P is such that $P(\{y_N = 1\}) = 0.4$ (so that $y_N = 0$ is slightly likelier than $y_N = 1$).

The Bayes optimal strategy A given by (14) is $d_n^A := 0$. Let us compare it with the complementary strategy $B := 1 - A$ (or simply $B := 1$). We have

$$\frac{1}{N}(\mathrm{Loss}_N(A) - \mathrm{Loss}_N(B)) = \begin{cases} 1 & \text{with probability } 0.4 \\ -1 & \text{with probability } 0.6, \end{cases}$$

and so (17) (with P in place of $\overline{\mathbb{P}}$) is grossly violated.

8 Conclusion

The bound given by the right-hand side of (16) in Theorem 7.5 is unlikely to be tight, and obtaining tighter bounds is the most obvious direction of further research. It is also interesting to prove similar results for other limit theorems of probability theory, such as the strong law of large numbers or central limit theorem.

In this paper we took the standard synchronic Bayesian picture for granted and used a standard Dutch Book argument to show that Sceptic can achieve a sure gain if Forecaster' forecast (in the form of futures prices) at a given time does not form a probability measure. However, the sure gain only happens at the end of step N, and at the intermediate steps Sceptic's capital can go down. Can we limit the risk of Sceptic's Dutch Book at the intermediate steps?

Acknowledgments

My research has been partially supported by Mitie. Following de Finetti's example [7, Sect. 32] of mixing science and politics, Слава Україні.

References

[1] Kenneth J. Arrow, Robert Forsythe, Michael Gorham, Robert Hahn, Robin Hanson, John O. Ledyard, Saul Levmore, Robert Litan, Paul Milgrom, Forrest D. Nelson, George R. Neumann, Marco Ottaviani, Thomas C. Schelling, Robert J. Shiller, Vernon L. Smith, Erik Snowberg, Cass R. Sunstein, Paul C. Tetlock, Philip E. Tetlock, Hal R. Varian, Justin Wolfers, and Eric Zitzewitz. The promise of prediction markets. *Science*, 320:877–878, 2008.

[2] Thomas Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418, 1763.

[3] José M. Bernardo and Adrian F. M. Smith. *Bayesian Theory*. Wiley, Chichester, 2000.

[4] A. Philip Dawid. The well-calibrated Bayesian (with discussion). *Journal of the American Statistical Association*, 77:605–613, 1982.

[5] A. Philip Dawid. Probability forecasting. In Samuel Kotz, N. Balakrishnan, Campbell B. Read, Brani Vidakovic, and Norman L. Johnson, editors, *Encyclopedia of Statistical Sciences*, volume 10, pages 6445–6452. Wiley, Hoboken, NJ, second edition, 2006.

[6] A. Philip Dawid and Vladimir Vovk. Prequential probability: Principles and properties. *Bernoulli*, 5:125–162, 1999.

[7] Bruno de Finetti. Probabilism: A critical essay on the theory of probability and on the value of science. *Erkenntnis*, 31:169–223, 1989. Translated by Maria Concetta Di Maio, Maria Carla Galavotti, and Richard C. Jeffrey from "Probabilismo" (1931).

[8] Bruno de Finetti. *Theory of Probability*. Wiley, Chichester, 2017. First published in 1970 as *Teoria delle probabilità*.

[9] Joseph L. Doob. *Stochastic Processes*. Wiley, New York, 1953.

[10] Darrell Duffie. *Futures Markets*. Prentice-Hall, Englewood Cliffs, NJ, 1989.

[11] Ian Hacking. Slightly more realistic personal probability. *Philosophy of Science*, 34:311–325, 1967.

[12] Larry Harris. *Trading and Exchanges: Market Microstructure for Practitioners*. Oxford University Press, Oxford, 2003.

[13] John C. Hull. *Options, Futures, and Other Derivatives*. Pearson, Hoboken, NJ, eleventh edition, 2021.

[14] Richard Jeffrey. *The Logic of Decision*. McGraw-Hill, New York, 1965.

[15] Andrei N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933. English translation: *Foundations of the Theory of Probability*. Chelsea, New York, 1950. Third edition of the Russian translation: *Основные понятия теории вероятностей*, ФАЗИС, Moscow, 1998.

[16] David Lewis. Why conditionalize? In *Papers in Metaphysics and Epistemology*, chapter 23, pages 403–407. Cambridge University Press, Cambridge, 1999.

[17] Dennis V. Lindley. *Making Decisions*. Wiley, London, second edition, 1985.

[18] Ulrike Schaede. Forwards and futures in Tokugawa-period Japan: A new perspective on the Dōjima rice market. *Journal of Banking and Finance*, 13:487–513, 1989.

[19] Glenn Shafer. Jeffrey's rule of conditioning. *Philosophy of Science*, 48:337–362, 1981.

[20] Glenn Shafer. Bayes's two arguments for the rule of conditioning. *Annals of Statistics*, 10:1075–1089, 1982.

[21] Glenn Shafer. Conditional probability (with discussion). *International Statistical Review*, 53:261–277, 1985.

[22] Glenn Shafer. The notion of event in probability and causality: Situating myself relative to Bruno de Finetti. *International Journal of Approximate Reasoning*, 141:171–178, 2022.

[23] Glenn Shafer and Vladimir Vovk. *Game-Theoretic Foundations for Probability and Finance*. Wiley, Hoboken, NJ, 2019.

[24] Vladimir Vovk. A logic of probability, with application to the foundations of statistics (with discussion). *Journal of the Royal Statistical Society B*, 55:317–351, 1993.

[25] Vladimir Vovk and Glenn Shafer. A conversation with A. Philip Dawid. *Statistical Science*, 2023. See "Future Papers" on the journal web site.

[26] Sergio Wechsler. Exchangeability and predictivism. *Erkenntnis*, 38:343–350, 1993.

A Ideal futures contracts

In this appendix we will only discuss idealized futures contracts; these are all we need in this paper. Real futures contracts will be briefly discussed in Appendix B. Our terminology will be slightly adapted to our needs (for example, the unit of time will be a step rather than, e.g., a day, and the trader will be called Sceptic).

A futures contract Φ has an *expiration step* m. The price of the contract is settled at the end of step m; namely, its final price F_m^+ is announced by Reality. In the middle of step $n \in \{1, \dots, m\}$, the current price F_n of Φ is announced by Forecaster, and Sceptic can take any *position* f_n in Φ. If $n < m$, Sceptic gains capital $f_n(F_{n+1} - F_n)$ at the next step (which actually means losing capital if $F_{n+1} < F_n$). At the end of the expiration step m (at *maturity*) Sceptic gains $f_m(F_m^+ - F_m)$.

Let us say that Sceptic takes a *constant position* f at time $n < m$ if he maintain the same position f through steps $n, \dots, m$. This leads to gaining capital $f(F_m^+ - F_n)$ at maturity. This mode of using futures contracts emulates forward contracts (which are similar to futures contracts but not exchange-traded).

B Mechanics of futures trading

The previous section gives an idealized picture of futures trading. The main elements of simplification in it are:

- the interest rate is assumed to be zero
- the positions and futures prices are assumed to take any real values
- there is no difference between the selling and buying prices (no bid/ask spread)
- there are no other transaction costs.

In this paper we are only interested in binary futures contracts (where the outcome is 0 or 1). However, the most popular market mechanism, described in this section, works for general futures contracts, which are not restricted to the binary case.

A good reference for traditional futures markets is [10]. While some of the physical details of trading described in it might be obsolete, the general principles are still applicable. Another good reference is [12].

By far the most popular platform for prediction markets is the Iowa Electronic Markets (IEM). The IEM was created in 1988 and has always been a small-scale operation; the development of prediction markets has been greatly hindered by the US anti-gambling regulation [1]. The IEM was created by academics, and its role is mainly educational; in particular, it has a great help system explaining the market microstructure (which I often follow in this section).

It received two no-action letters, in 1992 and 1993, from the US Commodity Futures Trading Commission (CFTC) reducing the chance of legal action against it. Its competitors sometimes have better bid/ask spreads, but their positions are less secure; e.g., Intrade (1999–2013) is now defunct and PredictIt (launched in 2014) had their CFTC no-action letter withdrawn in 2022.

A futures contract is a contract that pays a specified amount F_m at a specified future time, called the *expiration time* m (it was the expiration step in the previous section). The amount is uncertain now but becomes well-defined at the expiration time. An example of m and F_m is "6 November 2024" and "Democratic Nominee's share of the two-party popular vote in the 2024 US Presidential election" in US dollars (this is, essentially, one of the futures contracts traded at the IEM in August 2023). At each time the market participants can hold any number of the futures contracts (positive, zero, or negative), which is known as their *positions* in the futures contracts. They can also submit orders to change their positions. The main kinds of orders are *market orders* and *limit orders*. A limit order specifies the number of futures contracts to buy or sell at a given price (known as the *bid price* for orders to buy and the *ask price* for orders to sell); it may also specify the time when the order expires.

At the core of a futures market is the *order book* listing the outstanding limit orders. The prices specified in those orders are

$$B_{n_B} < B_{n_B-1} < \cdots < B_1 < A_1 < A_2 < \cdots < A_{n_A}, \tag{18}$$

where n_B is the number of different bid prices in the currently active limit orders to buy and n_A is the number of different ask prices in the currently active limit orders to sell. The prices in the list (18) are sorted in the ascending order, and the difference $A_1 - B_1$ is known as the *bid/ask spread*. With each price level x is associated the total number $N(x)$ of futures contracts that the market participants with active limit orders wish to trade (to buy if $x = B_n$ for some n and to sell if $x = A_n$ for some n; $N(x) = 0$ for all other x). The order book consists of the prices (18) and the numbers $N(x)$ of futures contracts offered at each price level x (within each price level x older orders appear before newer orders). It consists of a *bid queue* (the data related to the bid prices) and an *ask queue* (the data related to the ask prices).

A market order is simpler than a limit order and only specifies the number of futures contracts to buy or sell. When a new market order is submitted by a market participant MP, it is matched with the order book immediately and a trade is performed. Namely, if the order is to sell N contracts, the bid queue is traversed from the top (i.e., from B_1) until the required number of contracts to buy is found: we find the smallest k such that $N(A_1) + \cdots + N(A_k) \geq N$ (all the N contracts are bought if there is no such k) and arrange a trade with MP selling all his futures contracts to the market participants with active limit orders with the prices in $\{A_1, \ldots, A_k\}$; for the price A_k only the oldest orders are fulfilled (perhaps partially). The procedure for market orders to buy is analogous.

When a new limit order is submitted by a market participant, it is simply added to the order book. We can assume that the limit orders to buy specify prices below A_1 and the limit orders to sell specify prices above B_1 (otherwise, a market order can be submitted). When a limit order in the order book expires, it is, of course, removed from it.

An important element of futures markets is the system of *margins*. Typically market participants have positions in several futures contracts and other securities, and the total values of their portfolios can go up or down. To reduce the chance of the exchange losing money, they are required to maintain margin accounts at specified levels. If a margin account falls below the specified level as result of changing market prices, a *margin call* is issued requiring the account to be replenished.

In the IEM, short positions are formally prohibited, which allows it to avoid imposing margin requirements. But it is still easy to emulate short positions (e.g., a short position in the vote share for the Democratic Nominee can be modelled as a long position in the vote share for the Republican Nominee).

A natural question is how a futures market is started; namely how to make the order book non-empty. In the IEM, the markets participants are allowed to buy *fixed price bundles* for a given price. For example, such a bundle might contain the vote share for the Democratic Nominee and the vote share for the Republican Nominee, with a fixed price of \$1 (the sum of the two vote shares is 1, and so the final pay-off of the bundle is known to be \$1).

C Definitions and proofs

This appendix uses some basic definitions in game-theoretic probability [23]. Let $\mathbb{E}_n$ denote the game-theoretic expectation at the point in Protocol 7.4 right after Decision Maker announcing her move d_n (let us call it the *checkpoint*). If $f = f(y_n \dots y_{(n+K-1)\wedge N})$ is a function of the K consecutive moves by Reality starting from y_n (and ending with y_N if $N - n \geq K$),

$$\mathbb{E}_n f := \sum_{x \in \mathbf{Y}^{K \wedge (N-n+1)}} f(x) P_n(x).$$

More generally, if f depends on other future moves (by Reality and other players), $\mathbb{E}_n f$ is the value (if it exists) starting from which Sceptic can attain $f(y_n, \dots, y_N, \dots)$ at the end of step N. If f also depends on the moves preceding the step n checkpoint, $\mathbb{E}_n f$ is found separately for each set of these preceding moves.

Proof of Theorem 7.5

To get a handle on the difference $\mathrm{Loss}_N(A) - \mathrm{Loss}_N$ in Protocol 7.4, we first consider its increment

$$\lambda(d_i^0, y_i \dots y_{(i+K-1)\wedge N}) - \lambda(d_i, y_i \dots y_{(i+K-1)\wedge N}) \tag{19}$$

on step i, where d_i^0 is the prediction output by the strategy A defined by (15). By the choice of d_i^0, the difference (19) is a supermartingale difference, meaning that its $\mathbb{E}_i$ expectation is nonpositive. Namely,

$$\mathbb{E}_i\bigl(\lambda(d_i^0, y_i \dots y_{(i+K-1)\wedge N}) - \lambda(d_i, y_i \dots y_{(i+K-1)\wedge N})\bigr) \\ = \sum_{x \in \mathbf{Y}^{K\wedge(N-i+1)}} \bigl(\lambda(d_i^0, x) - \lambda(d_i, x)\bigr) P_i(x).$$

Let us set

$$L_n = \mathbb{E}_n \sum_{i=1}^{N} \Biggl(\lambda(d_i^0, y_i \dots y_{(i+K-1)\wedge N}) - \lambda(d_i, y_i \dots y_{(i+K-1)\wedge N}) \\ + \sum_{x \in \mathbf{Y}^{K\wedge(N-i+1)}} \bigl(\lambda(d_i, x) - \lambda(d_i^0, x)\bigr) P_i(x)\Biggr);$$

this is a game-theoretic martingale starting from zero (namely, $L_n = \mathcal{K}_{n-1}$ for some strategy for Sceptic). An explicit expression for this game-theoretic martingale is

$$L_n := \sum_{i=1}^{n-K} \Biggl(\lambda(d_i^0, y_i \dots y_{i+K-1}) - \lambda(d_i, y_i \dots y_{i+K-1}) \\ + \sum_{x \in \mathbf{Y}^K} \bigl(\lambda(d_i, x) - \lambda(d_i^0, x)\bigr) P_i(x)\Biggr) \\ + \sum_{i=n-K+1}^{n-1} \Biggl(\sum_{x \in \mathbf{Y}^{(K-n+i)\wedge(N-n+1)}} \bigl(\lambda(d_i^0, y_i \dots y_{n-1}x) \\ - \lambda(d_i, y_i \dots y_{n-1}x)\bigr) P_n(x) \\ + \sum_{x \in \mathbf{Y}^{K\wedge(N-i+1)}} \bigl(\lambda(d_i, x) - \lambda(d_i^0, x)\bigr) P_i(x)\Biggr). \quad (20)$$

The first sum (i.e., the sum $\sum_{i=1}^{n-K}$) in (20) involves the terms (19) that are determined by the checkpoint on step n. The second sum (i.e., the sum $\sum_{i=n-K+1}^{n-1}$) in (20) involves the terms (19) that are partially determined. And we do not have a term corresponding to the sum $\sum_{i=n}^{N}$ since at the checkpoint on step n the expectation of (19) is still 0 for such i.

We have constructed a game-theoretic martingale, (20), whose final value is a function of the difference $\mathrm{Loss}_N(A) - \mathrm{Loss}_N$ between the optimal loss and the actual loss (apart from the last step, which should be considered separately). It is easy to see that the differences $L_n - L_{n-1}$ are bounded by $2K$ in absolute value: the expression (20) for L_n and a similar expression for L_{n-1} differ only in addends corresponding to the steps $i \in \{n-K, \dots, n-1\}$, and the difference is in the weighted averages w.r.t. to P_n vs P_{n-1}. The expressions being averaged are between -1 and 1, and this gives the bound of $2K$. It remains to apply Corollary 3.8 (game-theoretic Hoeffding inequality) in [23].

As a sanity check, let me spell out the difference between L_n and L_{n-1}. Let us use the notation λ_i for the addends in (20),

$$L_n = \sum_{i=1}^{n-K} \lambda_i + \sum_{i=n-K+1}^{n-1} \lambda_i$$

and let us use the notation λ_i^- for the addends in the analogous expression for L_{n-1} when they are different from λ_i,

$$L_{n-1} = \sum_{i=1}^{n-K-1} \lambda_i + \sum_{i=n-K}^{n-2} \lambda_i^- .$$

The first difference between L_n and L_{n-1} arises when $i = n - K$:

$$\begin{aligned}\lambda_{n-K} - \lambda_{n-K}^- &= \lambda(d_{n-K}^0, y_{n-K} \dots y_{n-1}) - \lambda(d_{n-K}, y_{n-K} \dots y_{n-1}) \\ &\quad - \sum_{y \in \mathbf{Y}} \bigl(\lambda(d_{n-K}^0, y_{n-K} \dots y_{n-2} y) - \lambda(d_{n-K}, y_{n-K} \dots y_{n-2} y)\bigr) P_{n-1}(y)\end{aligned}$$

(each of the addends λ_{n-K} and λ_{n-K}^- also contains a sum $\sum_{x \in \mathbf{Y}^K}$, but these two sums cancel out). We can see that it is both a martingale difference and bounded by 2 in absolute value: indeed, it consists of the components

$$\lambda(d_{n-K}^0, y_{n-K} \dots y_{n-1}) - \sum_{y \in \mathbf{Y}} \lambda(d_{n-K}^0, y_{n-K} \dots y_{n-2} y) P_{n-1}(y)$$

and

$$\lambda(d_{n-K}, y_{n-K} \dots y_{n-1}) - \sum_{y \in \mathbf{Y}} \lambda(d_{n-K}, y_{n-K} \dots y_{n-2} y) P_{n-1}(y)$$

each of which is a martingale difference bounded by 1 in absolute value; cf. (12).

The next source of the difference between L_n and L_{n-1} is the terms corresponding to $i = n-K+1, \dots, n-2$ (this step should be ignored when $K \in \{1, 2\}$, as there are no such terms):

$$\begin{aligned}\lambda_i - \lambda_i^- &= \sum_{x \in \mathbf{Y}^{(K-n+i) \wedge (N-n+1)}} \bigl(\lambda(d_i^0, y_i \dots y_{n-1} x) - \lambda(d_i, y_i \dots y_{n-1} x)\bigr) P_n(x) \\ &\quad - \sum_{x \in \mathbf{Y}^{(K-n+i+1) \wedge (N-n+2)}} \bigl(\lambda(d_i^0, y_i \dots y_{n-2} x) - \lambda(d_i, y_i \dots y_{n-2} x)\bigr) P_{n-1}(x)\end{aligned}$$

(these two addends, λ_i and λ_i^-, also contain identical sums, which cancel out). Again we have a martingale difference bounded by 2 in absolute value; cf. (11).

The final source of the difference between L_n and L_{n-1} corresponds to $i = n - 1$ and can be written as

$$\lambda_{n-1} = \sum_{x \in \mathbf{Y}^{(K-1)\wedge(N-n+1)}} \big(\lambda(d^0_{n-1}, y_{n-1}x) - \lambda(d_{n-1}, y_{n-1}x)\big) P_n(x) \\ + \sum_{x \in \mathbf{Y}^{K\wedge(N-n+2)}} \big(\lambda(d_{n-1}, x) - \lambda(d^0_{n-1}, x)\big) P_{n-1}(x);$$

we can again use (11) to see that this is a martingale difference bounded by 2 in absolute value.