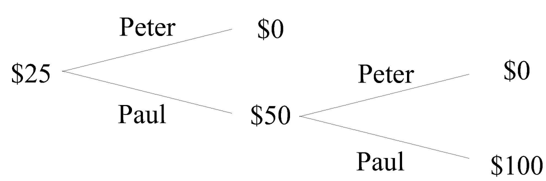
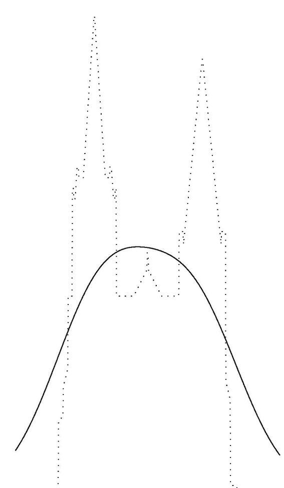
The diachronic Bayesian
(work in progress)

Vladimir Vovk





The Game-Theoretic Probability and Finance Project

Working Paper #64

First posted August 24, 2023. Last revised January 1, 2024.

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

Abstract

It is well known that a Bayesian probability forecast for the future observations should form a probability measure in order to satisfy a natural condition of coherence. The topic of this paper is the evolution of the Bayesian probability measure over 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

Consider a Bayesian forecaster predicting 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 [12, 14] and discussed in, e.g., [5, Sect. 4.1], that the Bayesian’s beliefs about the future observations should be encoded as a probability measure on the sequences of observations.

A fundamental role in de Finetti’s theory is played by the requirement of *coherence*: if the Bayesian’s beliefs do not form a probability measure, we can set a “Dutch book” against him, which is a system of bets leading to his sure loss. This is a property of consistency of the Bayesian’s beliefs, so we could call it internal coherence. We will also be interested in a stronger property, agreement with reality, in which sure loss is replaced by a substantial gain for an opponent who follows a strategy that can ever lose only a tiny amount. Both coherence and agreement with reality are defined in terms of betting.

Remark 1.1. The English word “coherence” seems to cover not only internal coherence but also agreement with reality (a kind of external coherence). For example, the earliest abstract use of “coherence” given in the Oxford English Dictionary (as of October 2023) is from Abraham Fraunce’s “The lawiers logike” (1588): “Where there is a greater cohærence and affinitie betweene the argument and the thing argued”. In this paper, however, I will use “coherence” in its meaning of internal coherence, which is standard in Bayesian statistics, except for a short discussion in Sect. 2.2.

Remark 1.2. One subtlety of de Finetti’s views is that the requirement of coherence only implies finite additivity and not countable additivity (see, e.g., [14, Sect. 18.3] and [5, Sect. 3.5.2]). In this paper we consider a finite case, in which the difference between finite and countable additivity disappears. (Starting from a finite case is standard in probability theory; see, e.g., [26, Chap. I] and [43, Chap. 1].)

Remark 1.3. 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, e.g., 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) [50, Sect. 7]. This view is sometimes called predictivism [52].

Remark 1.4. 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 these words as synonymous. I will never use more exotic words such as “prevision” [14, Sect. 3.1.2].

So we 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 [42, 11, 46], 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 [26, Sect. I.4]. This procedure for updating the Bayesian’s beliefs is known as “Bayesian conditioning” [4, 38, 40]. The principle that the new observations must be the only thing the Bayesian has learned is the *principle of total evidence* [39], and it is usually regarded as uncontroversial. Lewis [29] 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. While it may be convincing in coin-type situations, it is not in weather-type ones (cf. the first paragraph of this section). The main mathematical observation of Sect. 2 is that coherence-type requirements do not impose any restrictions on the forecasts at different times, so for discussing the diachronic aspects of Bayesian forecasting we need stronger requirements.

Section 3 proposes a testing protocol based on betting for the Bayesian’s predictions. This protocol is given in terms of observables and does not depend on Bayesian conditioning. A discussion of connections with the standard measure-theoretic picture follows in Sect. 4. The measure-theoretic picture will typically be an imaginary picture that does involve Bayesian conditioning (which may be happening deeply inside the imaginary picture, far from what we can observe). Section 5 adapts the testing protocols of the previous sections to predicting K steps ahead, which generalizes the case $K = 1$ considered earlier (in, e.g., [42, 11, 46]). Section 6 applies the testing protocol to making nearly optimal decisions, and Sect. 7 concludes. Appendixes A–C provide further information; the key one is Appendix A giving the proofs, and all the other appendixes should be omitted unless the reader has special reasons for reading them.

This paper has been inspired by the brief discussion of one-step-ahead prediction in [50, Sect. 7], and its title is adapted from [7] (being well-calibrated is an important aspect, namely the frequency aspect, of agreement with reality). Its other important source is Dawid’s super-strong prequential principle [11, Sect. 5.2]. My current understanding of this principle is that our testing protocol based on betting must agree with measure-theoretic probability, regardless of the imagined data-generating distribution. Technically, this is about the relation between the standard measure-theoretic and game-theoretic pictures discussed in Sects 3 and 4. Following [41, 42] I use “game-theoretic” to refer to being based on betting, and the kind of game theory involved here is the theory of perfect-information games rather than the probabilistic games studied in, e.g., economics (see, e.g., [42, Sect. 4.5]).

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 by $|a|$; 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:

- $\mathfrak{P}(A)$ is the set of all probability measures on A;
- if $P \in \mathfrak{P}(\mathbf{Y}^K)$ and $x \in \mathbf{Y}^k$ for $k \leq K$,
$$P(x) := P(x\mathbf{Y}^{K-k})$$
(however, when $x \in \mathbf{Y}^K$, I will still prefer $P(\{x\})$ to $P(x)$); in this paper we use this notation, and $P(x' \mid x)$ introduced next, for a finite $\mathbf{Y}$;
- if $P \in \mathfrak{P}(\mathbf{Y}^K)$, $x \in \mathbf{Y}^k$, and $x' \in \mathbf{Y}^{k'}$ for $k + k' \leq K$,
$$P(x' \mid x) := \frac{P(xx')}{P(x)};$$
- $\mathbf{Y}^{m:n}$ 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); this notation will be used mainly in Sect. 3.2.

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

1.2 Finite probability spaces

In my terminology I mainly follow [43, 44]. A *finite probability space* is a pair (Ω, P), where Ω is a finite set, implicitly equipped with the σ-algebra $\mathcal{F}$ of all subsets of Ω, and P is a probability measure on $(\Omega, \mathcal{F})$. Let us say that (Ω, P) is *positive* if the probability of each sample point is positive, $P(\{\omega\}) > 0$ for all $\omega \in \Omega$.

A *filtration* $(\mathcal{F}_n)$ in (Ω, P), where n ranges over a finite set of integers, is an increasing sequence of σ-algebras in Ω, $\mathcal{F}_{n_1} \subseteq \mathcal{F}_{n_2}$ when $n_1 \leq n_2$. We say that a sequence (Y_n) of random variables in (Ω, P) is *adapted* if Y_n is $\mathcal{F}_n$-measurable for all n.

1.3 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 [42]. The noun "Bayesian" will often be used as nearly synonymous with "Forecaster", and so the Bayesian is 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 *time 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 Forecaster 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., [42, Preface, point 2]).

Protocol 2.1 does not define a game, since we have not specified the players' goals, but we will often talk about the *plays* $P_1 y_1 \ldots P_N y_N$ proceeding according to the protocol's rules.

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" [30, Sect. 6.7].

2.1 Bayesian conditioning

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

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

Bayesian conditioning (as rule for updating beliefs) was criticised by Hacking in 1967 [21], although his criticism (ignoring the cost of thinking) is not the most important in the context of 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 [39].

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 [7, 10] 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 0/1 observation on 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). This is a common situation; to quote [20, Sect. 4], "in most cases of interest (e.g., the doctor's examination of the patient) it is unreasonable to suppose that, even in principle, there is a partition of possibilities over which probabilities and conditional probabilities could, in theory, be defined." 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, albeit in many cases this will be an unmanageable notion that is even 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 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. 5 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 [21] observation mentioned earlier.

Remark 2.4. In the main part of the paper we will only consider certain evidence and so will not discuss generalizations of Bayesian conditioning such as

Jeffrey's ([24, Chap. 11], [37]). See Appendix D for a brief discussion of uncertain evidence.

2.2 Weakness of coherence in the diachronic picture

The following proposition (proved in Sect. A.1 of Appendix A) shows that there is no reason to expect any connection between the forecasts P_n in Protocol 2.1 if we only assume a natural diachronic modification of coherence.

Proposition 2.5. *For any sequences of outcomes $y_1, \ldots, y_N \in \mathbf{Y}$ and any sequence of probability measures $P_n \in \mathfrak{P}(\mathbf{Y}^{N-n+1})$, $n = 1, \ldots, N$, there is a positive finite probability space (Ω, P) with filtration $\mathcal{F}_0, \ldots, \mathcal{F}_N$ and an adapted sequence of $\mathbf{Y}$-valued random elements $Y_1, \ldots, Y_n$ such that the event*

$$\forall n \in \{1, \ldots, N\} : P_n = P(\cdot \mid \mathcal{F}_{n-1}) \;\&\; Y_n = y_n$$

has a positive probability.

To interpret Proposition 2.5 in terms of coherence let us slightly modify this requirement. Remember that the Bayesian is incoherent if we can set a system of bets under which he always loses (a Dutch book). This notion is not applicable in our diachronic setting since we learn the full sequence $P_1, \ldots, P_N$ only at the very end of the forecasting session, when it is too late to bet. Let us say that the Bayesian (Forecaster in our protocol) is *dynamically incoherent* in a particular play if there is a way of betting against him that never leads to Sceptic's loss but, in this particular play, leads to Sceptic's gain. In other words, it's a gain that is not compensated by a potential loss; we will call it a *gratis gain*.

Proposition 2.5 says that Forecaster is never dynamically incoherent provided the bets are fair under P. This is true under any strategy for probability updating (or in the absence of such a strategy). There cannot be any diachronic inconsistency between P_n for different n leading to a gratis gain for Sceptic. In the following section we will see that such inconsistency can lead to an **almost** gratis gain for Sceptic. See also Remark 4.3 below.

3 Testing probability forecasts

Long-term prediction is much more complicated than one-step-ahead prediction that we have considered earlier [11, 42], and to have a clear understanding of the process we will use two pictures, which we call game-theoretic and measure-theoretic. The game-theoretic picture is based on betting (as in de Finetti [12]) and, in its more developed parts, on finance (such as prediction markets). In this paper 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., [23, Chap. 2] and [17]). Futures contracts is an old idea (see, e.g., [36]) 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. They will be described in Sect. 3.1.

The measure-theoretic picture is stochastic and assumes an overall probability measure used by Forecaster and Reality for generating their moves. This picture will be discussed in Sect. 4, where it will be shown to be equivalent to (albeit more complicated and less natural than) the game-theoretic picture.

3.1 Ideal futures contracts

In this section we complement the forecasting picture of the previous section by allowing Sceptic to bet against Forecaster. The kind of betting considered by de Finetti (see, e.g., [12]) is known as forward contracts in finance, but we will need a slight modification known as futures contracts.

Here we only discuss idealized futures contracts; this is 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 contract is settled at the end of step m; namely, its final price F_m^+ is announced by Reality. In the middle of step $n \in \{1, \ldots, m\}$, the current price F_n of Φ is announced by Forecaster, and Sceptic can then take any *position* $f_n \in \mathbb{R}$ 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 $f_n(F_{n+1} - F_n) < 0$). If $n = m$, at the end of the expiration step m (at *maturity*) Sceptic gains $f_m(F_m^+ - F_m)$. These gains keep accumulating as the play proceeds. (This picture will only be used in the interpretation of Protocols 3.1 and 3.2 in the main part of the paper.)

Let us say that Sceptic takes a *constant position* f at time $n < m$ if he maintain the same position f through steps $n, \ldots, m$. 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).

3.2 General testing protocol

The following extension of Protocol 2.1 describes a way of testing Forecaster's predictions. (It will be greatly simplified in Sect. 3.3.)

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 $n = 2, \ldots, N$:
  Forecaster announces $P_n \in \mathfrak{P}(\mathbf{Y}^{N-n+1})$
  $\mathcal{K}_{n-1} := \mathcal{K}_{n-1}' + \sum_{x \in \mathbf{Y}^{1:N-n+1}} f_{n-1}(y_{n-1}x) P_n(x)$

$$-\sum_{x\in\mathbf{Y}^{2:N-n+2}} f_{n-1}(x)P_{n-1}(x) \quad (2)$$

Sceptic announces $f_n \in \mathbb{R}^{\mathbf{Y}^{1:N-n+1}}$
Reality announces $y_n \in \mathbf{Y}$
$$\mathcal{K}'_n := \mathcal{K}_{n-1} + f_n(y_n) - \sum_{y\in\mathbf{Y}} f_n(y)P_n(y). \quad (3)$$

Protocol 3.1 does not define $\mathcal{K}_N$, and we set $\mathcal{K}_N := \mathcal{K}'_N$. Sceptic's capital is not allowed to become negative (as soon as it does, the play is stopped and Sceptic loses).

We regard this protocol (and similar protocols below) as a way of testing Forecaster's predictions: a large $\mathcal{K}_N$ means lack of agreement with reality. When for a particular play $\mathcal{K}_N$ is large, we can regard it as an almost gratis gain. (This assumes that Sceptic's capital is measured in small monetary units, but we can always scale it down if the monetary units are not small.)

An advantage of Protocol 3.1 is that, even though it is stated for a finite time horizon N, it is relatively easy to modify to make the time horizon infinite, so that $n = 2, 3, \dots$ in the FOR loop. The simplified protocol of Sect. 3.3 will use the finiteness of the time horizon in a very essential way.

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 $m := |x|$, as discussed in Sect. 3.1. At each step n (but before observing y_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

$$F_n(x) := \begin{cases} P_n(\{x \setminus y_1\dots y_{n-1}\}) & \text{if } y_1\dots y_{n-1} \subset x \\ 0 & \text{if not.} \end{cases} \quad (4)$$

We assume, without loss of generality, that these prices indeed form a probability measure; otherwise, the market is not coherent and Sceptic can secure sure gain [14, 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 $n < N$ 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

$$\begin{aligned} F_n(x) = \sum_{x' \in \mathbf{Y}^N : x \subset x'} & F_n(x') \\ &= \sum_{x' \in \mathbf{Y}^N : x \subset x'} P_n(x' \setminus y_1 \dots y_{n-1}) = P_n(x \setminus y_1 \dots y_{n-1}) \end{aligned} \quad (5)$$

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 by $f_n(x)$ 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 (3). And after P_n is disclosed by Forecaster at the next step $n := n+1$, 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 (4) and (5), and the last expression agrees with (2).

3.3 Final simplification

We can rewrite Protocol 3.1 in other forms, getting rid of some of Forecaster's arbitrary and irrelevant choices. To compare protocols with the same allowed moves for Reality and Forecaster, we will use the notion of the *test martingale space* (TMS) 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's capital $\mathcal{K}_n$ (for all possible n) as function of Forecaster's and Reality's moves provided this function is nonnegative. The TMS of a given protocol is the set of all possible test martingales. We regard two protocols to be equivalent if they have the same TMS.

As already mentioned, the general testing protocol was formulated with a view towards an infinite time horizon, where N becomes ∞. In this subsection we introduce a much simpler protocol using an idea that only works for a finite time horizon. Namely, we simplify Protocol 3.1 radically by letting f_n take non-zero values only at the final futures contracts (those maturing at the end of step N).

Protocol 3.2.

$\mathcal{K}_0 := 1$

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

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

IF $n > 1$:

$$\begin{aligned}\mathcal{K}_{n-1} := \mathcal{K}_{n-2} &+ \textstyle\sum_{x\in\mathbf{Y}^{N-n+1}} f_{n-1}(y_{n-1}x)P_n(x)\\ &- \textstyle\sum_{x\in\mathbf{Y}^{N-n+2}} f_{n-1}(x)P_{n-1}(x)\end{aligned} \tag{6}$$

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

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

IF $n = N$:

$$\mathcal{K}_n := \mathcal{K}_{n-1} + f_n(y_n) - \textstyle\sum_{y\in\mathbf{Y}} f_n(y)P_n(\{y\}). \tag{7}$$

The interpretation of $\mathcal{K}_N$ is the same as for Protocol 3.1 (a large $\mathcal{K}_N$ evidences lack of agreement of the forecasts with reality, provided Sceptic is not allowed to go into debt).

Proposition 3.3. *Protocol 3.1 and Protocol 3.2 have identical TMS.*

Proposition 3.3, to be proved in Sect. A.2 of Appendix A, simplifies the market in futures contracts that we need: all the contracts now mature at the end of step N; we will call such futures contracts *final*.

Remark 3.4. The reason why the final futures contracts are sufficient is that a general futures contract $\Phi(x)$ is equivalent, to all intents and purposes, to the portfolio consisting of the final futures contracts $\Phi(x')$ for all $x' \supseteq x$.

Comparison with Bayesian conditioning

How does (6) compare with Bayesian conditioning, where no new information outside the protocol arrives and we just define P_n by (1)? In this case we can simplify the protocol by replacing Sceptic's moves f_n with $f'_n : \mathbf{Y} \to \mathbb{R}$ defined by

$$f'_n(y) := \sum_{x \in \mathbf{Y}^{N-n}} f_n(yx) P_n(x \mid y), \tag{8}$$

and then (6) 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 3.2 as

Protocol 3.5.

$\mathcal{K}_0 := 1$
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 (1)
  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., [42, Protocol 1.1]) except that Forecaster announces his forecasting strategy in advance. We can see that forecasting multiple steps ahead does not require any new methods under Bayesian conditioning: testing can proceed one step ahead.

Without any restrictions on Forecaster, we obtain, instead of Protocol 3.5, the following protocol equivalent to Protocols 3.1–3.2, in which we still use the notation (8).

Protocol 3.6.

$\mathcal{K}_0 := 1$
FOR $n = 1, \dots, N$:
  Forecaster announces $P_n \in \mathfrak{P}(\mathbf{Y}^{N-n+1})$
  IF $n > 1$:
    $\mathcal{K}_{n-1} := \mathcal{K}_{n-1} + \sum_{x \in \mathbf{Y}^{N-n+1}} f_{n-1}(y_{n-1}x)(P_n(x) - P_{n-1}(x \mid y_{n-1}))$
  Sceptic announces $f_n \in \mathbb{R}^{\mathbf{Y}}$
  Reality announces $y_n \in \mathbf{Y}$
  $\mathcal{K}_n := \mathcal{K}_{n-1} + f'_n(y_n) - \sum_{y \in \mathbf{Y}} f'_n(y) P_n(y)$.

Protocol 3.6 adds to Protocol 3.5 the possibility to update P_n is a way different from Bayesian conditioning and includes a term that describes betting on the difference between the actual forecast $P_n(x)$ and the Bayesian conditional probabilities $P_{n-1}(x \mid y_{n-1})$ computed from the previous forecast. The equivalence of Protocols 3.2 and 3.6 follows from the equality, for $n < N$, of the addend $f'_n(y_n)$ in the expression for $\mathcal{K}_n$ in Protocol 3.6 and the subtrahend $\sum_{x \in \mathbf{Y}^{N-n+1}} f_{n-1}(y_{n-1}x) P_{n-1}(x \mid y_{n-1})$ in the expression for $\mathcal{K}_{n-1}$ at the next step.

Merging Sceptic's opponents

If we are only interested in strategies for Sceptic (not in strategies for other players, as in [42, Preface, ideas 3 and 6]) we can simplify Protocol 3.2 further by merging Forecaster and Reality. We will refer to the combined player as Forecaster (rather than World, as in [41]); the reason for this will become clear in Sect. D.1 in Appendix D.

Protocol 3.7.

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

$$\mathcal{K}_{n-1} := \mathcal{K}_{n-2} + \textstyle\sum_{x \in \mathbf{Y}^N} F_{n-1}(x)(Q_n(x) - Q_{n-1}(x)) \tag{9}$$

  Sceptic announces $F_n \in \mathbb{R}^{\mathbf{Y}^N}$.

Protocol 3.7 uses the notation $\mathfrak{P}_n(\mathbf{Y}^N)$ for the set of all probability measures Q on $\mathbf{Y}^N$ satisfying $Q(x) = 1$ for some $x \in \mathbf{Y}^{n-1}$.

To embed Protocol 3.2 into Protocol 3.7, we should take as Q_n the extension of P_n to $\mathbf{Y}^N$, namely

$$Q_n(x) := \begin{cases} P_n(x \setminus (y_1 \dots y_{n-1})) & \text{if } (y_1 \dots y_{n-1}) \subseteq x \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

and we should take as F_n the extension of f_n to $\mathbf{Y}^N$, namely

$$F_n(x) := \begin{cases} f_n(x \setminus (y_1 \dots y_{n-1})) & \text{if } (y_1 \dots y_{n-1}) \subseteq x \\ u & \text{otherwise,} \end{cases}$$

where, e.g., $u := 0$ (but in fact the value of u does not matter as it is always multiplied by 0 in the embedded protocol, and we can use different us for different n and x).

Protocol 3.7 lasts for $N+1$ rather than N steps in order for $\mathcal{K}_N$ to be defined by (9). Sceptic's last move F_{N+1} is never used.

4 Measure-theoretic picture

As a sanity check, in this section we will see whether the testing procedures described in the previous sections conform to the standard measure-theoretic framework [26, 16] for probability. Our check will have two sides: the validity of the game-theoretic picture in the measure-theoretic framework, and the validity of a natural measure-theoretic picture in the game-theoretic framework. We will use a standard theorem of duality; in general, it can be said that the measure-theoretic and game-theoretic pictures are dual to each other in a certain sense.

4.1 Validity of the game-theoretic picture

Let (Ω, P) be a finite probability space equipped with a filtration $\mathcal{F}_n$, $n = 0, 1, \ldots$. 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 or 3.2; by step 1 in Protocol 3.1 I mean the statements preceding the FOR loop). 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, $\mathcal{F}_n$, $n = 1, 2, \ldots$, is the information available at the end of step n.

Let us concentrate on Protocol 3.1 (Protocol 3.2 is easier). In the measure-theoretic framework for it, we assume that $y_1, \ldots, y_N$ are realizations of an adapted $\mathbf{Y}$-valued process $Y_1, \ldots, Y_N$ (meaning, as usual, that Y_n is $\mathcal{F}_n$-measurable, $n = 1, \ldots, N$), that each P_n is computed from P as the conditional probability measure for Y_n given $\mathcal{F}_{n-1}$, and that Sceptic follows an adapted strategy (meaning that $f_n(x)$ is $\mathcal{F}_{n-1}$-measurable for each x).

Sceptic's capital $\mathcal{K}_n$ is then an adapted process. We have

$$P_n(\{x\}) = \mathbb{P}\left(\{Y_n \ldots Y_N = x\} \mid \mathcal{F}_{n-1}\right) \quad \text{a.s.,} \quad x \in \mathbf{Y}^{N-n+1}, \tag{11}$$

and so

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

Now the first increment (2) in Sceptic's capital is

$$\begin{aligned}\mathcal{K}_{n-1} - \mathcal{K}'_{n-1} = &\sum_{x \in \mathbf{Y}^{1:N-n+1}} f_{n-1}(Y_{n-1}x) P(\{x \subseteq Y_n \ldots Y_N\} \mid \mathcal{F}_{n-1}) \\ &- \sum_{x \in \mathbf{Y}^{2:N-n+2}} f_{n-1}(x) P(\{x \subseteq Y_{n-1} \ldots Y_N\} \mid \mathcal{F}_{n-2})\end{aligned}$$

and so we have

$$\mathbb{E}_P(\mathcal{K}_{n-1} - \mathcal{K}'_{n-1} \mid \mathcal{F}_{n-2}) = 0 \quad \text{a.s.,} \tag{12}$$

where $\mathbb{E}_P$ stands for the expected value under P. The second increment (3) is

$$\begin{aligned}\mathcal{K}'_n - \mathcal{K}_{n-1} &= f_n(Y_n) - \sum_y f_n(y) P(\{Y_n = y\} \mid \mathcal{F}_{n-1}) \\ &= f_n(Y_n) - \mathbb{E}_P(f_n(Y_n) \mid \mathcal{F}_{n-1}),\end{aligned}$$

which gives the analogue

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

of (12). Therefore, $\mathcal{K}_0, \mathcal{K}_1, \ldots, \mathcal{K}_{N-1}, \mathcal{K}'_N = \mathcal{K}_N$ is a measure-theoretic martingale w.r. to the filtration $(\mathcal{F}_n)$: $\mathbb{E}_P(\mathcal{K}_n \mid \mathcal{F}_{n-1}) = \mathcal{K}_{n-1}$, $n = 1, \ldots, N$.

4.2 Validity of the measure-theoretic picture

A *non-terminal situation* in Protocol 2.1 (and also in Protocols 3.1 and 3.2) is a tuple $(P_1, y_1, \dots, P_n)$ for some $n \in \{1, \dots, N\}$, where $y_i \in \mathbf{Y}$ and $P_i \in \mathfrak{P}(\mathbf{Y}^{N-i+1})$ for all i. Informally, this is a situation in which Sceptic makes a move. A *terminal situation* in those protocols is a tuple $(P_1, y_1, \dots, P_N, y_N)$, where again $y_i \in \mathbf{Y}$ and $P_i \in \mathfrak{P}(\mathbf{Y}^{N-i+1})$ for all i. Non-terminal situations and terminal situations are referred to collectively as *situations*. A strategy for Sceptic can be defined as a function mapping the non-terminal situations to an allowed move, such as mapping a situation $(P_1, y_1, \dots, P_n)$ to $f \in \mathbb{R}^{\mathbf{Y}^{1:N-n+1}}$ in the case of Protocol 3.1. For a fixed strategy for Sceptic his capital becomes a real-valued function of a situation; let us refer to such functions as *game-theoretic test martingales* provided they are nonnegative. (In Sect. 3.3 we called them simply test martingales, but in a moment we will discuss other kinds of test martingales.) Until the end of this section we will consider only Borel measurable game-theoretic martingales (corresponding to Borel measurable strategies for Sceptic).

More generally, a *game-theoretic process* is a Borel measurable real-valued function of a situation. A nonnegative game-theoretic process S is a *visible measure-theoretic test martingale* if, for any finite probability space (Ω, P) equipped with a filtration $(\mathcal{F}_n)$ and any adapted sequence of random variables $Y_1, \dots, Y_N$,

$$\begin{aligned} S_{n-1} &:= S(P_1, Y_1, \dots, P_n), \quad n = 1, \dots, N, \\ S_N &:= S(P_1, Y_1, \dots, P_N, Y_N) \end{aligned}$$

is a test martingale in the usual sense of $S_0 = 1$ and

$$\mathbb{E}_P(S_n \mid \mathcal{F}_{n-1}) = S_{n-1}, \quad n = 1, \dots, N, \tag{13}$$

where the P_i are defined by (11), which becomes

$$P_i(\{x\}) := P\left(\{Y_i \dots Y_N = x\} \mid \mathcal{F}_{i-1}\right), \quad x \in \mathbf{Y}^{N-i+1},$$

in our current notation (this definition does not depend on the choice of a version of the conditional probability). The adjective "visible" refers to the martingale (S_n) depending only on the players' moves in Protocol 2.1 (and not depending on the hidden aspects of the realized sample point $\omega \in \Omega$).

The following statement of agreement between the game-theoretic and measure-theoretic pictures will be proved in Sect. A.3.

Proposition 4.1. *A game-theoretic process is a game-theoretic test martingale if and only if it is a visible measure-theoretic test martingale.*

Proposition 4.1, however, has a weakness. Let us say that a game-theoretic process is a *game-theoretic test supermartingale* if it can be obtained as Sceptic's capital while he is allowed to discard part of his capital at each step (but is still not allowed to go into debt). For example, in the case of Protocol 3.2

this corresponds to replacing (6) and (7) by Sceptic's moves allowing him to choose $\mathcal{K}_{n-1}$ and $\mathcal{K}_n$, respectively, as any nonnegative number not exceeding the corresponding right-hand side. And a game-theoretic process is a *measure-theoretic test supermartingale* if it is defined in the same way as a measure-theoretic test martingale except that the "=" in (13) is replaced by "$\leq$". The notion of a game-theoretic test supermartingale is obviously redundant, in the sense of every game-theoretic test supermartingale being dominated by a game-theoretic test martingale. But the requirement (13) holding for any probability space might appear restrictive, and so it is less obvious that measure-theoretic test supermartingales are redundant in this sense. Therefore, in Sect. A.3 we will start from proving the following modification of Proposition 4.1.

Theorem 4.2. *A game-theoretic process is a game-theoretic test supermartingale if and only if it is a visible measure-theoretic test supermartingale.*

This theorem implies that every visible measure-theoretic test supermartingale is dominated by a visible measure-theoretic test martingale. We will also see it directly in Sect. A.3 that the game-theoretic version of this property implies the measure-theoretic version.

Remark 4.3. This section, and Sect. 2.2 above, illustrate the "hidden variable" account of belief change ([1, Chap. 4, note 14], [15, Theorem 2.1], [45, Sect. 1]), according to which coherent belief update is Bayesian conditioning in a bigger belief space.

5 Predicting K steps ahead

For a large N, the protocols considered in the previous sections are unrealistic in that Forecaster is asked to produce probability measures on huge sets such as $\mathbf{Y}^N$. Starting from this section, we will assume that all predictions made by Forecaster are only for the next $K < N$ observations, with $K \geq 1$, and we will sometimes refer to K as the *prediction horizon*. We are typically interested in the case $K \ll N$.

The Bayesian prediction protocol (Protocol 2.1) becomes:

Protocol 5.1.

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

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 general testing protocol (Protocol 3.1):

Protocol 5.2.

$\mathcal{K}_0 := 1$

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 $n = 2, \dots, N$:
Forecaster announces $P_n \in \mathfrak{P}(\mathbf{Y}^{K\wedge(N-n+1)})$
$\mathcal{K}_{n-1} := \mathcal{K}'_{n-1} + \sum_{x\in\mathbf{Y}^{1:(K-1)\wedge(N-n+1)}} f_{n-1}(y_{n-1}x)P_n(x)$
$\quad - \sum_{x\in\mathbf{Y}^{2:K\wedge(N-n+2)}} f_{n-1}(x)P_{n-1}(x)$
Sceptic announces $f_n \in \mathbb{R}^{\mathbf{Y}^{1:K\wedge(N-n+1)}}$
Reality announces $y_n \in \mathbf{Y}$
$\mathcal{K}'_n := \mathcal{K}_{n-1} + f_n(y_n) - \sum_{y\in\mathbf{Y}} f_n(y)P_n(y)$.

Remark 5.3. In the example of weather forecasting one week ahead (cf. Remark 2.2), the predictions in Protocol 5.1 are quite different from the predictions produced by a typical weather app. Weather apps produce marginal probabilities of rain whereas the probabilities in Protocol 5.1 are joint. Testing marginal probabilities would be much easier than the kind of testing exemplified by Protocol 5.2. See [49] for details of testing marginal probabilities.

For K steps ahead forecasting, the final testing protocol (Protocol 3.2) becomes:

Protocol 5.4.
$\mathcal{K}_0 := 1$
FOR $n = 1, \dots, N$:
Forecaster announces $P_n \in \mathfrak{P}(\mathbf{Y}^{K\wedge(N-n+1)})$
IF $n > 1$:
$\mathcal{K}_{n-1} := \mathcal{K}_{n-2} + \sum_{x\in\mathbf{Y}^{(K-1)\wedge(N-n+1)}} f_{n-1}(y_{n-1}x)P_n(x)$
$\quad - \sum_{x\in\mathbf{Y}^{K\wedge(N-n+2)}} f_{n-1}(x)P_{n-1}(x)$
Sceptic announces $f_n \in \mathbb{R}^{\mathbf{Y}^{K\wedge(N-n+1)}}$
Reality announces $y_n \in \mathbf{Y}$
IF $n = N$:
$\mathcal{K}_n := \mathcal{K}_{n-1} + f_n(y_n) - \sum_{y\in\mathbf{Y}} f_n(y)P_n(y)$.

6 Bayesian decision making

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

6.1 An optimality result for the Bayes decision strategy

Consider the following decision-making protocol.

Protocol 6.1.
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 $\mathbf{D}$ of permitted decisions. Before that, 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 order to weaken the conditions of our mathematical result (Theorem 6.5 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

$$\mathrm{Loss}_N := \sum_{n=1}^{N} \lambda_n(d_n, y_n \dots y_N) \in [0,N]. \tag{14}$$

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 6.1 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 6.1. The following protocol is a combination of Protocols 6.1 and 2.1.

Protocol 6.2.

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 6.2 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 $P_1, \dots, P_n$):

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

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 3.2.

Protocol 6.3.

$\mathcal{K}_0 := 1$

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

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

Forecaster announces $P_n \in \mathfrak{P}(\mathbf{Y}^{N-n+1})$
IF $n > 1$:
$$\mathcal{K}_{n-1} := \mathcal{K}_{n-2} + \sum_{x \in \mathbf{Y}^{N-n+1}} f_{n-1}(y_{n-1}x)P_n(x) - \sum_{x \in \mathbf{Y}^{N-n+2}} f_{n-1}(x)P_{n-1}(x)$$
Decision Maker announces $d_n \in \mathbf{D}$
Sceptic announces $f_n \in \mathbb{R}^{\mathbf{Y}^{N-n+1}}$
Reality announces $y_n \in \mathbf{Y}$
IF $n = N$:
$$\mathcal{K}_n := \mathcal{K}_{n-1} + f_n(y_n) - \sum_{y \in \mathbf{Y}} f_n(y)P_n(y).$$

In order to prove a law of large numbers for decision making showing that the Bayes strategy is indeed optimal in some sense, we need the following combination of Protocols 6.3 and 5.4. (We will see in Sect. 6.2 that this law of large numbers fails for Protocol 6.3.)

Protocol 6.4.
$\mathcal{K}_0 := 1$
FOR $n = 1, \ldots, N$:
Reality announces $\lambda_n : \mathbf{D} \times \mathbf{Y}^K \to [0, 1]$
Forecaster announces $P_n \in \mathfrak{P}(\mathbf{Y}^K)$
IF $n > 1$:
$$\mathcal{K}_{n-1} := \mathcal{K}_{n-2} + \sum_{x \in \mathbf{Y}^{(K-1)\wedge(N-n+1)}} f_{n-1}(y_{n-1}x)P_n(x) - \sum_{x \in \mathbf{Y}^{K\wedge(N-n+2)}} f_{n-1}(x)P_{n-1}(x) \qquad (16)$$
Decision Maker announces $d_n \in \mathbf{D}$
Sceptic announces $f_n \in \mathbb{R}^{\mathbf{Y}^{K\wedge(N-n+1)}}$
Reality announces $y_n \in \mathbf{Y}$
IF $n = N$:
$$\mathcal{K}_n := \mathcal{K}_{n-1} + f_n(y_n) - \sum_{y \in \mathbf{Y}} f_n(y)P_n(y).$$

Protocol 6.4 simplifies Forecaster's move $P_n \in \mathfrak{P}(\mathbf{Y}^{K\wedge(N-n+1)})$ in Protocol 5.4 to $P_n \in \mathfrak{P}(\mathbf{Y}^K)$ (but we will only use $P_n(x)$ for $x \in \mathbf{Y}^{0:K\wedge(N-n+1)}$).

We will continue to use the notation Loss_N introduced in (14), which is now modified to
$$\mathrm{Loss}_N := \sum_{n=1}^{N-K+1} \lambda_n(d_n, y_n \ldots y_{n+K-1}), \qquad (17)$$
but we will also be interested in Decision Maker's loss $\mathrm{Loss}_N(A)$ computed by replacing his actual decisions by decisions recommended by a decision strategy A:
$$\mathrm{Loss}_N(A) := \sum_{n=1}^{N-K+1} \lambda_n(d_n^A, y_n \ldots y_{n+K-1}),$$
where
$$d_n^A := A(\lambda_1, P_1, y_1, \lambda_2, P_2, \ldots, y_{n-1}, \lambda_n, P_n), \quad n = 1, \ldots, N-K+1;$$

we are only interested in strategies that are functions of the previous moves by the other players. Let us adapt the strategy (15) to Protocol 6.4:

$$d_n := d_n^A \in \arg\min_{d \in \mathbf{D}} \sum_{x \in \mathbf{Y}^K} \lambda_n(d, x) P_n(x), \tag{18}$$

with d_n^A chosen as the first element of the $\arg\min$ in a fixed linear order on $\mathbf{D}$ if there are ties among d.

If E is a property of Reality's, Forecaster's, and Decision Maker's moves in Protocol 6.4, we define the *upper game-theoretic probability* of E as the infimum of $\alpha > 0$ such that Sceptic has a strategy that guarantees $\mathcal{K}_n \geq 0$ for all n and that ensures $\alpha \mathcal{K}_n \geq 1$ whenever E happens. The following optimality result will be proved in Appendix A (Sect. A.5).

Theorem 6.5. *Let $\epsilon \in (0, 0.3)$. There is a strategy A for Decision Maker in Protocol 6.4 that guarantees*

$$\overline{\mathbb{P}} \left(\mathrm{Loss}_N(A) - \mathrm{Loss}_N \geq 2\sqrt{KN \ln \frac{1}{\epsilon}} \right) \leq \epsilon. \tag{19}$$

An alternative statement of Theorem 6.5 not using the notion of game-theoretic probability is that there exists a joint strategy for Decision Maker and Sceptic that achieves either

$$\mathrm{Loss}_N(A) - \mathrm{Loss}_N < 2\sqrt{KN \ln \frac{1}{\epsilon}}$$

or $\mathcal{K}_N \geq 1/\epsilon$. For a small ϵ and large N (as compared with $K \ln \frac{1}{\epsilon}$), this joint strategy demonstrates that A performs better than or similarly to the actual moves d_n unless Forecaster is discredited. This is a version of the law of large numbers that works only when $K \ll N$.

Remark 6.6. Notice that the strong law of large numbers for a fixed K (and with $N \to \infty$, as usual) is trivial: we can apply the standard one-step-ahead strong law of large numbers to each Kth observation (starting from observation 1, starting from observation 2,..., and finally starting from observation K). Theorem 6.5 is less trivial, but interestingly, it is based on the same idea. The argument used in the arXiv version 1 of this paper is different but leads to a weaker result (Theorem 7.5 in that version). See Remark C.7 for further details.

The strategy A in the statement of Theorem 6.5 can be chosen as the Bayes optimal strategy (18). Theorem 6.5 shows that, for any other strategy B for Sceptic, we have

$$\overline{\mathbb{P}} \left(\mathrm{Loss}_N(A) - \mathrm{Loss}_N(B) \geq 2\sqrt{KN \ln \frac{1}{\epsilon}} \right) \leq \epsilon; \tag{20}$$

we, however, prefer the stronger statement (19) allowing Forecaster to choose his moves on the fly. We can rewrite (20) as

$$\overline{\mathbb{P}} \left(\frac{1}{N} \mathrm{Loss}_N(A) - \frac{1}{N} \mathrm{Loss}_N(B) \geq \delta \right) \leq \exp\left(-\frac{\delta^2 N}{4K} \right)$$

for any $\delta \geq 2.2\sqrt{K/N}$. The restriction $\delta \geq 2.2\sqrt{K/N}$ is coming from the condition $\epsilon < 0.3$ in Theorem 6.5; without this restriction, we can still claim that

$$\overline{\mathbb{P}}\left(\frac{1}{N}\mathrm{Loss}_N(A) - \frac{1}{N}\mathrm{Loss}_N(B) \geq \delta\right) \leq 5\frac{K}{\delta^2 N}\exp\left(-\frac{\delta^2 N}{4K}\right)$$

(this follows from (48) with $\gamma := 1/\sqrt{2}$).

Remark 6.7. In Theorem 6.5 we compare Decision Maker's actual loss Loss_N with the loss she would have suffered following the strategy A defined by (18). Our interpretation of this theorem depends on the assumption that Reality's and Forecaster's moves are not affected by Decision Maker's moves.

6.2 Predicting $K < N$ steps ahead is essential for our statement of optimality

Theorem 6.5 is about predicting K steps ahead. How important is this restriction? Let us check that it may not be true that

$$\frac{1}{N}(\mathrm{Loss}_N(A) - \mathrm{Loss}_N) < \delta \tag{21}$$

with high probability in Protocol 6.3 for $\delta \ll 1$ if we use the definition of the cumulative loss given in (14) (there is little difference between (14) and (17) for $K \ll N$, but for $K = N$ the latter leads to vacuous statements for $\mathrm{Loss}_N(A) - \mathrm{Loss}_N$); as before, A stands for the Bayes optimal strategy. The intuition behind this demonstration 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 measure-theoretic probability) 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

$$\lambda_n(d_n, y_n \dots y_N) := \begin{cases} 0 & \text{if } d_n = y_N \\ 1 & \text{otherwise,} \end{cases} \tag{22}$$

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 (15) is $d_n^A := 0$. Let us compare it with the complementary strategy $B := 1 - A$ (or simply $B := 1$). We have

$$\frac{1}{N}(\mathrm{Loss}_N(A) - \mathrm{Loss}_N(B)) = \begin{cases} 1 & \text{with probability } 0.4 \\ -1 & \text{with probability } 0.6, \end{cases} \tag{23}$$

and so the inequality (21) is grossly violated with a significant probability.

Applying the idea leading to (23) on a smaller scale (to each Kth step instead of the last step), we obtain the following lower bound for Protocol 6.4.

Proposition 6.8. *For all N and $K < N/5$,*

$$\overline{\mathbb{P}}\left(\mathrm{Loss}_N(A) - \mathrm{Loss}_N \geq \sqrt{KN}\right) \geq \epsilon, \tag{24}$$

where A is the Bayes optimal strategy and ϵ is a universal positive constant.

The lower bound $\sqrt{KN}$ in (24) matches the upper bound in (19) (Theorem 6.5) as far as K and N are concerned. See Appendix C for related results (Propositions C.1 and C.6) in measure-theoretic probability.

Proposition 6.8 only concerns the optimality of the upper bound in (19) in K and N, but the next proposition shows that it is also close to being optimal in ϵ. In this proposition we use a slightly different definition of Loss_N: now, unlike in (17), we sum the losses of all decisions, including those of $d_{N-K+2}, \dots, d_N$ (they will be defined in a very natural way).

Proposition 6.9. *Suppose that N and K are such that N/K is an even integer. Then the Bayes optimal strategy A satisfies, for any $\epsilon > 0$,*

$$\overline{\mathbb{P}}\left(\mathrm{Loss}_N(A) - \mathrm{Loss}_N \geq \sqrt{KN \ln \frac{1}{\epsilon}}\right) \geq \epsilon^4/15 \tag{25}$$

provided

$$\sqrt{KN \ln \frac{1}{\epsilon}} \leq N/4. \tag{26}$$

The condition (26) is mild in this context; without it, the bound (25) appears useless. The substitution $\epsilon := \epsilon^4/15$ in Proposition 6.9 gives the following corollary, which shows that the upper bound in (19) is optimal if we ignore additive and multiplicative constants in the "regret term"

$$\sqrt{KN \ln \frac{1}{\epsilon}}.$$

Corollary 6.10. *Under the conditions of Proposition 6.9,*

$$\overline{\mathbb{P}}\left(\mathrm{Loss}_N(A) - \mathrm{Loss}_N \geq \frac{1}{2}\sqrt{KN \ln \frac{1}{15\epsilon}}\right) \geq \epsilon$$

provided the term $\sqrt{\dots}$ does not exceed $N/2$.

For proofs of Propositions 6.8 and 6.9, see Sects A.6 and A.7, respectively.

7 Conclusion

This paper has scratched the surface of the diachronic picture of realistic Bayesian forecasting not based on Bayesian conditioning. We discussed ways of testing such forecasts based on betting and their applications to Bayesian decision making.

Obvious directions of further research include, e.g., considering an infinite time horizon and more general observation spaces $\mathbf{Y}$. Another direction is to generalize our basic forecasting protocol: instead of assuming that the forecaster observes a new outcome y_n at each step, we could consider cases where beliefs are revised (perhaps because new information arrives from outside the protocol) without new outcomes becoming known, incorporating alternatives to Bayesian conditioning such as Jeffrey's rule; see Appendix D.

An important property of one-step-ahead prediction is that the predictions output by two successful forecasters must be in agreement with each other, at least asymptotically (see, e.g., [8, Sect. 5.2], [9, Theorem 7.1], and [42, Sect. 10.7]). This paper's setting can be used for extending this result to prediction multiple steps ahead; see [48].

Acknowledgments

My research has been partially supported by Mitie. Many thanks to Philip Dawid for advice on literature and useful discussions and to Ilia Nouretdinov for his input. Comments by the participants in the International Seminar on Selective Inference are gratefully appreciated. Following de Finetti's example [13, Sect. 32] of mixing science and politics, Слава Україні.

References

[1] Ernest W. Adams. *The Logic of Conditionals: An Application of Probability to Deductive Logic*. Springer, Dordrecht, 1975.

[2] Gunnar Andersson. The problem of the empirical basis in critical rationalism. In Jeremy Shearmur and Geoffrey Stokes, editors, *The Cambridge Companion to Popper*, chapter 5, pages 125–142. Cambridge University Press, New York, 2016.

[3] 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.

[4] 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.

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

[6] Thomas Cover. Universal portfolios. *Mathematical Finance*, 1:1–29, 1991.

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

[8] A. Philip Dawid. Statistical theory: the prequential approach (with discussion). *Journal of the Royal Statistical Society A*, 147:278–292, 1984.

[9] A. Philip Dawid. Calibration-based empirical probability (with discussion). *Annals of Statistics*, 13:1251–1285, 1985.

[10] 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.

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

[12] Bruno de Finetti. La prévision, ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7:1–68, 1937. English translation: Foresight: Its logical laws, its subjective sources. In [28] (both first and second editions).

[13] 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).

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

[15] Persi Diaconis and Sandy L. Zabell. Updating subjective probability. *Journal of the American Statistical Association*, 77:822–830, 1982.

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

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

[18] Paul Embrechts and Giovanni Puccetti. Bounds for functions of dependent risks. *Finance Stochastics*, 10:341–352, 2006.

[19] William Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley, New York, third edition, 1968.

[20] Michael Goldstein. The prevision of a prevision. *Journal of the American Statistical Association*, 78:817–819, 1983.

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

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

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

[24] Richard C. Jeffrey. *The Logic of Decision*. McGraw-Hill, New York, 1965. Second edition: 1983.

[25] Richard C. Jeffrey. Radical probabilism (prospectus for a user's manual). *Philosophical Issues*, 2:193–204, 1992.

[26] 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.

[27] Robert J. Kunsch and Daniel Rudolf. Optimal confidence for Monte Carlo integration of smooth functions. *Advances in Computational Mathematics*, 45:3095–3122, 2019.

[28] Henry E. Kyburg, Jr and Howard E. Smokler, editors. *Studies in Subjective Probability*. Krieger, Huntington, NY, second edition, 1980. First edition: Wiley, New York, 1964.

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

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

[31] G. D. Makarov. Estimates for the distribution function of the sum of two random variables with given marginal distributions. *Theory of Probability and its Applications*, 26:803–806, 1981.

[32] Jiří Matoušek and Bernd Gärtner. *Understanding and Using Linear Programming*. Springer, Berlin, 2007.

[33] Jiří Matoušek and Jan Vondrák. The probabilistic method. Available (in December 2023) on the web, 2008.

[34] Karl R. Popper. *The Logic of Scientific Discovery*. Hutchinson, London, 1950. German original: *Logik der Forschung*, Springer, Vienna, 1934.

[35] Giovanni Puccetti and Ludger Rüschendorf. Sharp bounds for sums of dependent risks. *Journal of Applied Probability*, 50:42–53, 2013.

[36] 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.

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

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

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

[40] 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.

[41] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's Only a Game!* Wiley, New York, 2001.

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

[43] Albert N. Shiryaev. *Probability-1*. Springer, New York, third edition, 2016.

[44] Albert N. Shiryaev. *Probability-2*. Springer, New York, third edition, 2019.

[45] Brian Skyrms. Coherence, probability and induction. *Philosophical Issues*, 2:215–226, 1992.

[46] 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.

[47] Vladimir Vovk. A game of prediction with expert advice. *Journal of Computer and System Sciences*, 56:153–173, 1998.

[48] Vladimir Vovk. Convergence of opinions. Technical Report arXiv:2312.02033 [math.ST], arXiv.org e-Print archive, December 2023.

[49] Vladimir Vovk. Logic of subjective probability. Technical Report arXiv:2309.01173 [cs.AI], arXiv.org e-Print archive, September 2023.

[50] Vladimir Vovk and Glenn Shafer. A conversation with A. Philip Dawid. *Statistical Science*, 2023. See "Future Papers" on the journal web site. An extended version is published as arXiv:2312.00632.

[51] Vladimir Vovk and Ruodu Wang. Combining p-values via averaging. *Biometrika*, 107:791–808, 2020.

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

A Proofs

A.1 Proof of Proposition 2.5

Let us fix such sequences of probability measures $P_n \in \mathfrak{P}(\mathbf{Y}^{N-n+1})$ and outcomes $y_n \in \mathbf{Y}$. As a first step, define Ω as $\mathbf{Y}^N$, P as P_1, $Y_n(\omega)$ as the nth element ω_n of $\omega \in \Omega$, and let the σ-algebra $\mathcal{F}_n$ be generated by $Y_1, \dots, Y_n$.

Next modify the finite probability space (Ω, P) and filtration $(\mathcal{F}_n)$ as follows. Split each sample point $y_1\omega_2 \dots \omega_N$ that starts from y_1 into two sample points, $y_1'\omega_2 \dots \omega_N$ and $y_1''\omega_2 \dots \omega_N$, and make the sets $\{y_1'\} \times \Omega^{N-1}$ and $\{y_1''\} \times \Omega^{N-1}$ $\mathcal{F}_n$-measurable for $n \geq 1$. Split the old value $c := P_1(\{y_1\} \times \Omega^{N-1})$ into $P(\{y_1'\} \times \Omega^{N-1}) := \epsilon c$, for a sufficiently small $\epsilon > 0$, and $P(\{y_1''\} \times \Omega^{N-1}) := (1 - \epsilon)c$. Without changing $P(\{\omega_1 \dots \omega_N\})$ for $\omega_1 \notin \{y_1', y_1''\}$, set

$$\frac{P(\{y_1'\omega_2 \dots \omega_N\})}{P(\{y_1'\} \times \Omega^{N-1})} := P_2(\{\omega_2 \dots \omega_N\}), \quad \omega_2, \dots, \omega_N \in \Omega,$$

and define

$$\frac{P(\{y_1''\omega_2 \dots \omega_N\})}{P(\{y_1'\} \times \Omega^{N-1})}, \quad \omega_2, \dots, \omega_N \in \Omega,$$

in such a way that we have an agreement with P_1:

$$\forall \omega_2, \dots, \omega_N \in \Omega : \frac{P(\{y_1'\omega_2 \dots \omega_N, y_1''\omega_2 \dots \omega_N\})}{P(\{y_1', y_1''\} \times \Omega^{N-1})} = \frac{P_1(\{y_1\omega_2 \dots \omega_N\})}{P_1(\{y_1\} \times \Omega^{N-1})};$$

this is possible for a sufficiently small $\epsilon > 0$.

Apply the same procedure to the probability subspace of (Ω, P) consisting of the sample points $y_1'\omega_2 \dots \omega_N$, thereby splitting y_2 into y_2' and y_2''. Continue by splitting y_3, y_4, etc.

A.2 Proof of Proposition 3.3

Consider step $n < N$ of Protocol 3.1. Let $O(x, c)$, where $x \in \mathbf{Y}^{1:N-n}$ and $c \in \mathbb{R}$, be the operation that adds the constant c to $f_n(x)$ and subtracts the same constant c from all $f_n(xy)$, $y \in \mathbf{Y}$. The key observation used in our simplifications of Protocol 3.1 is that, for any $x \in \mathbf{Y}^{1:N-n}$ and $c \in \mathbb{R}$, $O(x, c)$ does not change the increment in the capital $\mathcal{K}_n - \mathcal{K}_{n-1}$. Let us check this property. If $x \in \mathbf{Y}^{2:N-n}$, $O(x, c)$ will not affect (3) whatsoever, and it will change neither minuend nor subtrahend in (2) at the next step (there is a next step since $n < N$). And if $x \in \mathbf{Y}$, applying the operation $O(x, c)$ does not change the increment in the capital $\mathcal{K}_n - \mathcal{K}_{n-1}$ given by (3) and then by (2) at the next step since

- the change in the sum in (3) and in the second sum in (2) at the next step will balance each other out, and
- the change in the term $f_n(y_n)$ in (3) and in the first sum in (2) at the next step will also balance each other out (this is relevant only when $x = y_n$).

Applying $O(x, c)$ repeatedly to the xs in the order of increasing length, 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 (3) for all steps n apart from $n = N$, and so (7) is performed only for $n = N$;
- we can ignore the bits "1 : " and "2 : " in (2), obtaining (6).

Protocol 3.2 also merges the four lines in Protocol 3.1 preceding the FOR loop into the loop.

A.3 Proof of Theorem 4.2 and Proposition 4.1

We start from Theorem 4.2. The proof uses the duality theorem of linear programming as described in [32, Sect. 6.2] (this textbook gives a general and detailed description of the dual problem). Let us consider in detail only the first step in Protocol 2.1, when we move from prediction $P := P_1$ to prediction $Q := P_2$. We regard P as fixed (so that our argument if conditional on P) and use the notation $P(y)$, where $y \in \mathbf{Y}$, and $P(x \mid y)$, where $y \in \mathbf{Y}$ and $x \in \mathbf{Y}^{N-1}$, as usual. We also use $Q_r(x \mid y)$ for the various $Q = P_2$ possible after observing y. We will be able to apply the standard duality theorem since r ranges over a finite set; remember that we consider a finite probability space.

In Sect. 4.1 we saw that every game-theoretic test martingale is a visible measure-theoretic test martingale, and this implies that every game-theoretic test supermartingale is a visible measure-theoretic test supermartingale; therefore, we will be only interested in the opposite direction. Let $S_{y,r}$ be the values at step 2 of a visible measure-theoretic test supermartingale. Our goal is to show that they are the values at step 2 of a game-theoretic test supermartingale.

The primary (measure-theoretic) linear programming problem involves variables $X_{y,r} \geq 0$ subject to the constraints

$$\sum_r X_{y,r} = 1 \tag{27}$$

for all y and

$$\sum_r X_{y,r} Q_r(x \mid y) = P(x \mid y) \tag{28}$$

for all y and x. The interpretation is that $X_{y,r}$ is the conditional probability of Q_r after observing y. The relevant optimization problem is

$$\sum_y P(y) \sum_r S_{y,r} X_{y,r} \to \max. \tag{29}$$

By the choice of S, the max value is at most 1.

Following the recipe given in [32, Sect. 6.2], we find the dual (game-theoretic) problem as

$$\sum_{y,x} P(x \mid y) Y_{y,x} + \sum_y Y_y \to \min \tag{30}$$

subject to

$$\sum_x Q_r(x \mid y) Y_{y,x} + Y_y \geq P(y) S_{y,r} \tag{31}$$

for all y, r. The dual variables $Y_{y,x}$ and Y_y are unconstrained. Let us replace them with new variables Z defined by

$$\begin{aligned} Y_{y,x} &= P(y) Z_{y,x} \\ Y_y &= P(y) Z_y. \end{aligned}$$

Instead of the optimization problem (30)–(31), we obtain

$$\sum_{y,x} P(yx) Z_{y,x} + \sum_y Z_y P(y) \to \min \tag{32}$$

subject to

$$\sum_x Q_r(x \mid y) Z_{y,x} + Z_y \geq S_{y,r}, \tag{33}$$

with the same value, at most 1. We can see that the optimization problem (32)–(33) agrees with the interpretation in terms of futures contracts.

Before proving Proposition 4.1 let us make a detour and check that every visible measure-theoretic test supermartingale is dominated by a visible measure-theoretic test martingale. First we make $S = (S_{y,r})$ admissible replacing each $S_{y,r}$ by the left-hand side of (33). The expression being maximized in (29) becomes

$$\begin{aligned} \sum_y P(y) \sum_r X_{y,r} S_{y,r} &= \sum_y P(y) \sum_r X_{y,r} \left(\sum_x Q_r(x \mid y) Z_{y,x} + Z_y \right) \\ &= \sum_y P(y) \sum_x Z_{y,x} P(x \mid y) + \sum_y P(y) Z_y \\ &= \sum_{y,x} P(y,x) Z_{y,x} + \sum_y P(y) Z_y, \end{aligned}$$

where the second equality uses (27) and (28). The last expression is very natural, and does not depend at all on the primary variables $X_{y,r}$, which shows that $S_{y,r}$ are the values at step 2 of a visible measure-theoretic test martingale.

Finally, if $S = (S_{y,r})$ satisfies the measure-theoretic martingale property, it will coincide with its dominating measure-theoretic test martingale and coincide with the corresponding game-theoretic test martingale, which completes the proof of Proposition 4.1.

A.4 Game-theoretic probability

In the proof of Theorem 6.5 in Sect. A.5 we will need some basic definitions and results in game-theoretic probability given in this subsection; see [42] for further information. We will let $\mathbb{E}_n$ denote the game-theoretic expectation (to be defined momentarily) at the point in Protocol 6.4 right after Decision Maker announcing her move d_n (let us call this point the *checkpoint*), which in our current context can be defined as follows. 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 $n + K - 1 \geq N$),

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

More generally, if f depends on other future moves (by Reality and other players), $\mathbb{E}_n f$ is the initial capital (if it exists) starting from which Sceptic can attain the final value of f 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.

The *game-theoretic sample space* Ω consists of all possible sequences of moves

$$\omega := (\lambda_1, P_1, d_1, y_1, \dots \lambda_N, P_N, d_N, y_N)$$

by non-Sceptic players in Protocol 6.4. A *nonnegative variable* X is a function $X : \Omega \to [0, \infty)$. The *upper expectation* of X is defined as

$$\overline{\mathbb{E}}(X) := \inf \{\alpha > 0 \mid \exists \text{ strategy for Sceptic } \forall \omega \in \Omega : \alpha \mathcal{K}_N(\omega) \geq X(\omega)\},$$

where ω are the non-Sceptic player's moves and $\mathcal{K}_N$ is regarded as function of ω. In words, $\overline{\mathbb{E}}(X)$ is the smallest (in the sense of inf) initial capital that Sceptic can turn into $X(\omega)$. An *event* is a set $E \subseteq \Omega$. The *upper probability* $\overline{\mathbb{P}}(E)$ of an event E is defined to be $\overline{\mathbb{E}}(1_E)$.

Lemma A.1. *For any bounded nonnegative variable X,*

$$\overline{\mathbb{E}}(X) \leq \int_0^\infty \overline{\mathbb{P}}(X \geq u) \, \mathrm{d}u. \tag{34}$$

Proof. Set $f(u) := \overline{\mathbb{P}}(X \geq u)$; then $f : [0, \infty) \to [0, 1]$ is a decreasing function. Replace the ∞ in (34) by C for some upper bound C for X. For each $k = 0, \dots, \lceil C/\epsilon \rceil$, fix a strategy for Sceptic that turns $f(k\epsilon) + \epsilon$ into $1_{\{X \geq k\epsilon\}}$ or more. Multiplying by ϵ and then summing over the k, we obtain a strategy that turns

$$\sum_{k=0}^{\lceil C/\epsilon \rceil} \epsilon (f(k\epsilon) + \epsilon) \tag{35}$$

into at least

$$\sum_{k=0}^{\lceil C/\epsilon \rceil} \epsilon 1_{\{X(\omega) \geq k\epsilon\}} \geq X(\omega).$$

It remains to notice that (35) tends to $\int_0^C f(u) \, \mathrm{d}u$ as $\epsilon \to 0$. □

A.5 Proof of Theorem 6.5

This subsection uses the definitions and results from game-theoretic probability given in Sect. A.4. The reader familiar with measure-theoretic probability who encounters game-theoretic probability for the first time might prefer to read Appendix C first as a gentle introduction to this subsection.

We will also need the following lemma, which is widely used in robust risk aggregation.

Lemma A.2. *For any $C > 0$, any $\alpha \in (0, C/K)$, and any $x_1, \ldots, x_K \in \mathbb{R}$,*

$$\sum_{k=1}^{K} g(x_k) \geq 1_{\{\sum_{k=1}^{K} x_k \geq C\}}, \tag{36}$$

where g is the continuous function

$$g(x) := \begin{cases} 0 & \text{if } x < C/K - \alpha \\ \frac{x-(C/K-\alpha)}{K\alpha} & \text{if } C/K - \alpha \leq x \leq C/K + (K-1)\alpha \\ 1 & \text{if } x > C/K + (K-1)\alpha. \end{cases} \tag{37}$$

Proof. We argue indirectly. Suppose there is a set of numbers $x_1, \ldots, x_K$ for which (36) holds with "$<$" in place of "$\geq$", and fix such a set. If $x_i < C/K - \alpha$ and $x_j > C/K + (K-1)\alpha$, we can replace x_i by $x_i + t$ and x_j by $x_j - t$, where $t > 0$ is the smallest number such that $x_i + t = C/K - \alpha$ or $x_j - t = C/K + (K-1)\alpha$; therefore, we can assume, without loss of generality, that there is no such pair (i, j). In this case, $x_k \leq C/K + (K-1)\alpha$ for all k, but perhaps $x_j < C/K - \alpha$ for some j. It remains to apply Jensen's inequality: as the average of x_k is at least C/K, the average of $f(x_k)$ is at least $f(C/K) = 1/K$. □

See the proof of Theorem 4.2 in [18] for another proof of Lemma A.2, and see Appendix C for further information about robust risk aggregation.

Set $Q := \lfloor N/K \rfloor$. Let us first assume that $N = QK + K - 1$; later we will get rid of this assumption (it will be easy as $N = QK + K - 1$ is, in a sense, the worst case).

To get a handle on the difference $\mathrm{Loss}_N(A) - \mathrm{Loss}_N$ in Protocol 6.4, we first consider its increment

$$\lambda(d_i^A, y_i \ldots y_{i+K-1}) - \lambda(d_i, y_i \ldots y_{i+K-1}) \tag{38}$$

on step $i \leq N - K + 1$, where d_i^A is the prediction output by the strategy A defined by (18). By the choice of d_i^A, the difference (38) is a supermartingale difference, meaning that its $\mathbb{E}_i$ expectation is nonpositive. Namely,

$$\mathbb{E}_i\big(\lambda(d_i^A, y_i \ldots y_{i+K-1}) - \lambda(d_i, y_i \ldots y_{i+K-1})\big) \\ = \sum_{x \in \mathbf{Y}^K} \big(\lambda(d_i^A, x) - \lambda(d_i, x)\big) P_i(x) \leq 0.$$

For each $k \in \{1, \dots, K\}$, we consider the process

$$L_n^k = \mathbb{E}_n \sum_{i \in \{k, k+K, \dots, k+(Q-1)K\}} \Biggl(\lambda(d_i^A, y_i \dots y_{i+K-1}) - \lambda(d_i, y_i \dots y_{i+K-1}) \\ + \sum_{x \in \mathbf{Y}^K} \bigl(\lambda(d_i, x) - \lambda(d_i^A, x) \bigr) P_i(x) \Biggr); \quad (39)$$

considering only every Kth step in the sum simplifies the analysis and, more importantly, makes the result stronger (cf. Remark 6.6). This process starts from zero, and it is a game-theoretic martingale (namely, $L_n^k = \mathcal{K}_{n-1}$ for some strategy for Sceptic), as the following explicit expression shows:

$$\begin{aligned} L_n^k := & \sum_{i \in \{k, k+K, \dots, k+(q-1)K\}} \Biggl(\lambda(d_i^A, y_i \dots y_{i+K-1}) - \lambda(d_i, y_i \dots y_{i+K-1}) \\ & \qquad + \sum_{x \in \mathbf{Y}^K} \bigl(\lambda(d_i, x) - \lambda(d_i^A, x) \bigr) P_i(x) \Biggr) \\ & + \sum_{x \in \mathbf{Y}^{K-j}} \bigl(\lambda(d_{k+qK}^A, y_{k+qK} \dots y_{n-1} x) - \lambda(d_{k+qK}, y_{k+qK} \dots y_{n-1} x) \bigr) P_n(x) \\ & \qquad + \sum_{x \in \mathbf{Y}^K} \bigl(\lambda(d_{k+qK}, x) - \lambda(d_{k+qK}^A, x) \bigr) P_{k+qK}(x) \end{aligned} \quad (40)$$

where q and $j \in \{0, \dots, K-1\}$ are the integers from the representation $n = k + qK + j$, and we are only interested in $n \le QK$. The first sum (i.e., the sum $\sum_{i \in \{k, k+K, \dots, k+(q-1)K\}}$) in (40) involves the terms (38) (for $i \equiv k \pmod{K}$) that are determined by the checkpoint on step n. The rest of the expression in (40) involves the term (38) that is partially determined, which corresponds to $i = k + qK$. And we do not have terms corresponding to $i > k + qK$ since at the checkpoint on step n the expectation of the expression in the outer parentheses in (39) is still 0 for such i.

To check that (40) is indeed a game-theoretic martingale, it suffices to notice that

$$L_n^k - L_{n-1}^k = \sum_{x \in \mathbf{Y}^{K-j}} f_{n-1}(y_{n-1} x) P_n(x) - \sum_{x \in \mathbf{Y}^{K-j+1}} f_{n-1}(x) P_{n-1}(x),$$

where

$$f_{n-1}(x) := \lambda(d_{k+qK}^A, y_{k+qK} \dots y_{n-2} x) - \lambda(d_{k+qK}, y_{k+qK} \dots y_{n-2} x),$$

has the same form as the capital increment in (16). This assumes that n is not one of the borderline values $k + qK$, which case should be considered separately.

If we only consider the values of the game-theoretic martingale (40) at steps $k + qK$, $q = 0, 1, \dots, Q$,

$$L^k_{k+qK} := \sum_{i\in\{k,k+K,\dots,k+(q-1)K\}} \Bigl(\lambda(d^A_i, y_i \dots y_{i+K-1}) - \lambda(d_i, y_i \dots y_{i+K-1}) + \sum_{x\in\mathbf{Y}^K} \bigl(\lambda(d_i,x) - \lambda(d^A_i,x)\bigr) P_i(x)\Bigr), \quad q = 0,1,\dots,Q, \tag{41}$$

its increments will be bounded by 2 in absolute value, and we can apply the game-theoretic Hoeffding inequality [42, Corollary 3.8 for Protocol 3.5] to it. However, a tighter inequality is obtained when we apply the one-sided version of the game-theoretic Hoeffding inequality [42, Corollary 3.8 for Protocol 3.7] to the process (41) with the sum over $x \in \mathbf{Y}^K$ removed. This process is a game-theoretic supermartingale whose increments are bounded by 1 in absolute value, and the one-sided Hoeffding inequality gives

$$\overline{\mathbb{P}}\left(X_k \ge U\right) \le \exp\left(-\frac{U^2}{2Q}\right) \le \exp\left(-U^2\frac{K}{2N}\right), \tag{42}$$

where $U \ge 0$ and

$$X_k := \sum_{i\in\{k,k+K,\dots,k+(Q-1)K\}} \bigl(\lambda(d^A_i, y_i \dots y_{i+K-1}) - \lambda(d_i, y_i \dots y_{i+K-1})\bigr);$$

we assume that the game-theoretic supermartingale is constant after $k+(Q-1)K$ (the last i in the range of summation in (41)).

Applying (42) and Lemmas A.1 and A.2 (see below for details) gives, for any $C > 0$,

$$\overline{\mathbb{P}}(X_1 + \dots + X_K \ge C) \le \sum_{k=1}^K \overline{\mathbb{E}}(g(X_k)) \le \sum_{k=1}^K \int_0^\infty \overline{\mathbb{P}}(g(X_k) \ge u)\,\mathrm{d}u \tag{43}$$

$$\le \sum_{k=1}^K \int_0^\infty \overline{\mathbb{P}}\left(X_k \ge \gamma\frac{C}{K} + (1-\gamma)Cu\right)\mathrm{d}u \tag{44}$$

$$\le \sum_{k=1}^K \int_0^\infty \exp\left(-\left(\gamma\frac{C}{K} + (1-\gamma)Cu\right)^2 \frac{K}{2N}\right)\mathrm{d}u \tag{45}$$

$$= \frac{\sqrt{KN}}{(1-\gamma)C} \int_{\frac{\gamma C}{\sqrt{KN}}}^\infty \exp\left(-v^2/2\right)\mathrm{d}v \tag{46}$$

$$= \frac{\sqrt{KN}}{(1-\gamma)C}\sqrt{2\pi}\bar{\Phi}\left(\frac{\gamma C}{\sqrt{KN}}\right) < \frac{KN}{\gamma(1-\gamma)C^2}\exp\left(-\frac{\gamma^2C^2}{2KN}\right). \tag{47}$$

The first and second equalities in (43) follow from Lemmas A.2 and A.1, respectively. The inequality (44) follows from the definition of g in (37) with $\alpha := (1-\gamma)C/K$. Indeed, we can assume, without loss of generality, $u > 0$, and then $g(X) \ge u$ implies

$$\frac{X - (C/K - \alpha)}{K\alpha} \ge u,$$

which is equivalent to

$$X \geq \gamma \frac{C}{K} + (1-\gamma)Cu.$$

The inequality (45) follows from Hoeffding's inequality (42). The equality (46) follows by the substitution

$$v := \frac{\gamma C}{\sqrt{KN}} + (1-\gamma)C\sqrt{\frac{K}{N}}u.$$

The equality in (47) introduces the notation $\bar{\Phi} := 1 - \Phi$ for the survival function of the standard Gaussian distribution. And the last inequality in the chain follows by applying the standard upper bound [19, Lemma VII.1.2] on $\bar{\Phi}$.

We can rewrite the inequality between the extreme terms in the chain (43)–(47) as

$$\overline{\mathbb{P}}\left(\mathrm{Loss}_N(A) - \mathrm{Loss}_N \geq C\right) \leq \frac{KN}{\gamma(1-\gamma)C^2} \exp\left(-\frac{\gamma^2 C^2}{2KN}\right). \tag{48}$$

Comparing this with (19), we can see that we need to solve the inequality

$$\frac{KN}{\gamma(1-\gamma)C^2} \exp\left(-\frac{\gamma^2 C^2}{2KN}\right) \leq \epsilon. \tag{49}$$

Ignoring the part before the exp and replacing "$\leq$" by "$=$", we obtain the solution

$$C = \frac{\sqrt{2KN \ln \frac{1}{\epsilon}}}{\gamma},$$

which motivates the substitution

$$C := \frac{\sqrt{2KN \ln \frac{1}{\epsilon} x}}{\gamma} \tag{50}$$

in (49). After this substitution, (49) simplifies to

$$\epsilon^{x-1} \leq 2\frac{1-\gamma}{\gamma} x \ln \frac{1}{\epsilon}. \tag{51}$$

Setting $x := 2\gamma^2$ in (50) gives an expression that matches the corresponding expression in (19).

The condition $x > 1$ (required for (51) to hold as $\epsilon \to 0$) narrows down the range of γ from $\gamma \in (0,1)$ to $\gamma \in (2^{-1/2}, 1)$. Setting, e.g., $\gamma := 0.8$ ensures that (51) holds for all $\epsilon \in (0, 0.32)$.

It remains to consider the case $N < QK + K - 1$. If the final value of L^k_{k+qK} (corresponding to $q = Q$) is undefined (because $k + qK + K - 1 > N$), we set it equal to its previous value (for $q = Q - 1$).

A.6 Proof of Proposition 6.8

Similarly to (22), let us set $\mathbf{D} := \mathbf{Y} := \{0, 1\}$ and

$$\lambda_n(d_n, y_n \dots y_{n+K-1}) := \begin{cases} 1 & \text{if } d_n \neq y_{\lceil n/K \rceil K} \\ 0 & \text{otherwise.} \end{cases} \tag{52}$$

We are only interested in $n \le N - K + 1$ (see (17)), which implies $n + K - 1 \le N$ and $\lceil n/K \rceil K \le N$; therefore, (52) is well-defined. Now the true probability measure P is such that $y_n = 1$ with probability $1/2$ independently for different n (and now we will rely on our tie-breaking convention). As in Sect. 6.2, the players comply with P. Let B be the decision strategy that always outputs 1; notice that A always outputs 0 (assuming that the linear order on $\mathbf{D}$ is $0 < 1$).

The N steps is now split into $\lceil N/K \rceil$ blocks of K steps (except, possibly, the last step), $n \in \{1, \dots, K\}$, $n \in \{K+1, \dots, 2K\}$, etc. Within each block, A suffers the same loss at each step, and B suffers the same loss at each step. By the central limit theorem, the probability is at least ϵ (a universal positive constant) that A performs worse than B in at least $\sqrt{N/K} + 1$ more blocks than vice versa. In such cases

$$\mathrm{Loss}_N(A) - \mathrm{Loss}_N(B) \ge K\sqrt{N/K} = \sqrt{KN}.$$

This gives (24) with P in place of $\overline{\mathbb{P}}$. By Ville's inequality, we can replace the probability measure P by the upper game-theoretic probability $\overline{\mathbb{P}}$.

A.7 Proof of Proposition 6.9

We will obtain Proposition 6.9 by applying a lower bound for large deviations in the form of [33, Proposition 7.3.2] to the argument of the previous subsection, as follows (see below for some explanations):

$$\begin{aligned} &\overline{\mathbb{P}}\left(\mathrm{Loss}_N(A) - \mathrm{Loss}_N(B) \ge \sqrt{KN \ln \frac{1}{\epsilon}}\right) \\ &= \overline{\mathbb{P}}\left(\frac{1}{K}\mathrm{Loss}_N(A) \ge \frac{N}{2K} + \frac{1}{2}\sqrt{\frac{N}{K} \ln \frac{1}{\epsilon}}\right) && (53) \\ &= \overline{\mathbb{P}}\left(X \ge \frac{n}{2} + t\right) \ge \frac{1}{15}\exp\left(-16t^2/n\right) = \epsilon^4/15. && (54) \end{aligned}$$

The first equality, (53), follows from

$$\mathrm{Loss}_N(A) + \mathrm{Loss}_N(B) = N \tag{55}$$

(which allows us to eliminate $\mathrm{Loss}_N(B)$) and obvious transformations. The first equality in (54) introduces the notation

$$X := \frac{1}{K}\mathrm{Loss}_N(A), \quad n := \frac{N}{K}, \quad t := \frac{1}{2}\sqrt{\frac{N}{K} \ln \frac{1}{\epsilon}},$$

which is the notation used in [33, Proposition 7.3.2]. The inequality "$\ge$" in (54) is identical to [33, Proposition 7.3.2].

Remark A.3. Kunsch and Rudolf [27, Lemma 3] slightly improve the constants in [33, Proposition 7.3.2], and using their result we can improve the bound $\epsilon^4/15$ in (54) to $\epsilon^3/5$. The last statement can be rewritten in the form

$$\overline{\mathbb{P}}\left(\mathrm{Loss}_N(A) - \mathrm{Loss}_N(B) \ge \sqrt{\frac{1}{3}KN\ln\frac{5}{\epsilon}}\right) \ge \epsilon \tag{56}$$

for any $\epsilon > 0$.

Remark A.4. Let us check informally what the optimal counterparts of the constants $1/3$ and 5 in (56) would be in the domain of applicability of the central limit theorem. We have for the probability measure P of Sect. A.6:

$$P\left(\mathrm{Loss}_N(A) - \mathrm{Loss}_N(B) \ge \bar{\Phi}^{-1}(\epsilon)\sqrt{KN}\right) \approx \epsilon,$$

where $\bar{\Phi}^{-1}(\epsilon)$ is the upper ϵ-quantile of the standard Gaussian distribution. This follows from the variance of

$$\mathrm{Loss}_N(A) - \mathrm{Loss}_N(B) = 2\mathrm{Loss}_N(A) - N$$

(cf. (55)) being approximately KN. This gives the ideal approximate equality

$$P\left(\mathrm{Loss}_N(A) - \mathrm{Loss}_N(B) \ge \sqrt{2KN\ln\frac{1}{\epsilon}}\right) \approx \epsilon$$

in place of (56). It is interesting that this is exactly what we get from (50) when we make $\gamma \approx 1$ and $x \approx 1$ (it is clear that we can make $\gamma \in (0,1)$ and $x > 1$ as close to 1 as we want at the price of restricting ϵ to a narrower range $(0, \epsilon^*)$).

B Mechanics of futures trading

Section 3.1 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 (although we are only interested in positive prices for futures contracts);
- 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 appendix, works for general futures contracts, which are not restricted to the binary case.

A good reference for traditional futures markets is [17]. While some of the physical details of trading described in it might be obsolete, the general principles are still applicable. Another good reference is [22].

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 [3]. 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 main part of the paper). The amount is uncertain at the time of trading but becomes well-defined at the expiration time, when trading ceases. 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 types of futures contracts traded at the IEM in August 2023 (of the "vote share" variety; the other main variety is "winner takes all"). Let us fix m and F_m. 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

$$B_{n_B} < B_{n_B-1} < \cdots < B_1 < A_1 < A_2 < \cdots < A_{n_A}, \tag{57}$$

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 (57) 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 (57) 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 orders to buy is found: we find the smallest k such that $N(B_1)+\cdots+N(B_k) \geq N$ (all the $N(B_1)+\cdots+N(B_{n_B})$ contracts requested in the bid queue 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 $\{B_1, \ldots, B_k\}$; for the price B_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 types of futures contracts (corresponding to different m and F_m) 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 (i.e., negative) 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 market 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 Measure-theoretic martingale law of large numbers

Our discussion of Bayesian decision theory in Sect. 6 was based on a law of large numbers for predicting K steps ahead. This law of large numbers may also present an independent interest, and the purpose of this appendix is to give clean self-contained measure-theoretic statements of its various versions. In this appendix we consider general probability spaces $(\Omega, \mathcal{F}, P)$, not necessarily finite.

A *filtration* $(\mathcal{F}_n)$, $n = 0, 1, \dots, N$, in a general probability space $(\Omega, \mathcal{F}, P)$ is still an increasing sequence of σ-algebras, $\mathcal{F}_0 \subseteq \cdots \subseteq \mathcal{F}_N$. A sequence $Y_1, \dots, Y_N$ of random variables in $(\Omega, \mathcal{F}, P)$ is *adapted* if Y_n is $\mathcal{F}_n$-measurable for $n = 1, \dots, N$. We usually assume $|Y_n| \le 1$ for agreement with the assumption $\lambda_n \in [0, 1]$ that we made in Sect. 6 about the loss functions: Y_n corresponds to a difference between two values of such a loss function λ_n.

Interestingly, we can get nearly optimal results by using the primitive idea of decomposing forecasting K steps ahead into K processes of forecasting one step ahead, as in Remark 6.6. This gives us the following proposition (analogous to Theorem 6.5).

Proposition C.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space equipped with a filtration $(\mathcal{F}_n)$, $n = 0, 1, \dots, N$. Fix a prediction horizon $K \in \{1, \dots, N\}$. Let $Y_1, \dots, Y_N$ be an adapted sequence of random variables in $(\Omega, \mathcal{F}, P)$ bounded by 1 in absolute value, $|Y_n| \le 1$ for $n = 1, \dots, N$. Then we have, for any $\epsilon \in (0, 0.7)$,*

$$P\left(\left|\sum_{n=K}^{N} (Y_n - \mathbb{E}_P(Y_n \mid \mathcal{F}_{n-K}))\right| \ge 4\sqrt{KN \ln \frac{1}{\epsilon}}\right) \le \epsilon. \tag{58}$$

Proof. In this proof we will need one result from robust risk aggregation (this theory was originated by Kolmogorov [31]; it is briefly described in [51, Remark 2] and then widely used in that paper). Namely, we will need the following special case of Theorem 4.2 of Embrechts and Puccetti [18].

Suppose nonnegative random variables X_k, $k = 1, \dots, K$, satisfy

$$\mathbb{P}(X_k \ge x) = \exp(-ax^2) \tag{59}$$

for all $x \ge 0$, where a is a positive constant. The value E of the optimization problem

$$\mathbb{P}(X_1 + \cdots + X_K \ge C) \to \max \tag{60}$$

(the max, or at least sup, being over all joint distributions for $(X_1, \dots, X_K)$ with the given marginals) does not exceed

$$E := \inf_{t < C/K} \frac{K \int_t^{C-(K-1)t} \exp(-ax^2)\,\mathrm{d}x}{C - Kt}. \tag{61}$$

We can extend the statement in the previous paragraph to a wider class of random variables X_k, $k = 1, \dots, K$. Namely, it suffices to assume that they satisfy

$$\mathbb{P}(X_k \ge x) \le \exp(-ax^2) \tag{62}$$

for all $x \ge 0$, instead of (59). We will apply the statement to the random variables X_k given by

$$X_k := \sum_{n \in \{k+K, k+2K, \dots, k+\lfloor N/K \rfloor K\}} (Y_n - \mathbb{E}_P(Y_n \mid \mathcal{F}_{n-K}))\,.$$

By Hoeffding's inequality, for any $C > 0$ and any $k \in \{0, \dots, K-1\}$,

$$P\left(X_k \ge C\right) \le \exp\left(-C^2/(2\lfloor N/K \rfloor)\right) \le \exp\left(-C^2/(2N/K)\right),$$

where the non-existent terms in the sum (those corresponding to $n > N$ if any) are interpreted as 0. Therefore, (62) holds with

$$a := \frac{K}{2N}. \tag{63}$$

Let us set $t := \frac{C}{2K}$ in (61) (this is the middle of the range of t). This gives the upper bound

$$\frac{2K}{C} \int_{\frac{C}{2K}}^{\infty} \exp(-ax^2)\,\mathrm{d}x$$

for E, which can be rewritten (see below for an explanation) as

$$\frac{2K}{C} \frac{1}{\sqrt{2a}} \int_{\sqrt{2a}\frac{C}{2K}}^{\infty} \exp(-y^2/2)\,\mathrm{d}y = \frac{2K}{C} \frac{\sqrt{2\pi}}{\sqrt{2a}} \bar{\Phi}\left(\sqrt{2a}\frac{C}{2K}\right) \tag{64}$$

$$= \frac{2\sqrt{2\pi}\sqrt{KN}}{C} \bar{\Phi}\left(\frac{C}{2\sqrt{KN}}\right) < \frac{4KN}{C^2} \exp\left(-\frac{C^2}{8KN}\right). \tag{65}$$

The first expression in (64) is obtained by the substitution $y := \sqrt{2a}x$, the equality in (64) uses the notation $\bar{\Phi}$ for the survival function of the standard Gaussian distribution, the following equality (the one in (65)) is obtained by plugging in (63), and the final inequality in (65) follows from the usual upper bound for $\bar{\Phi}$ [19, Lemma VII.1.2].

To find a suitable solution to the inequality

$$\frac{4KN}{C^2} \exp\left(-\frac{C^2}{8KN}\right) \le \frac{\epsilon}{2},$$

we plug in $C = \sqrt{8KN \ln \frac{1}{\epsilon} x}$ (intuitively, $x \approx 1$) obtaining, after simplification,

$$\epsilon^{x-1} \le x \ln \frac{1}{\epsilon}.$$

Assuming $\epsilon < 0.7$, we can set $x := 2$. $\square$

Remark C.2. In the proof of Proposition C.1 we did not make any attempt to optimize the coefficient 4 in (58). However, the same argument shows that 4 can be replaced by a number as close to $\sqrt{2}$ as we wish if we narrow down the permitted range of ϵ (leaving the lower end of the range at 0, of course).

Remark C.3. Since the bound E in (61) plays an important role in this appendix (and implicitly in Appendix A.5), it is reassuring to know that in many interesting cases E actually coincides with the value of the optimization problem (60). This is shown in Theorem 2.3 by Puccetti and Rüschendorf [35]. (And the restatement of Embrechts and Puccetti's result in [35, Sect. 1] is particularly convenient.) One of the cases [35, Sect. 3] in which E is the value of the optimization problem is where the probability density function of X_k is monotonically decreasing over its domain $[0,\infty)$. This condition, however, is only satisfied for $x \ge 1/\sqrt{2a}$ (the last condition becomes $x \ge \sqrt{N/K}$ for the value of a, given in (63), that we will be interested in).

Remark C.4. In the proof of Proposition C.1 we set $t := \frac{C}{2K}$ in (61). In the arXiv version 2 of this paper, we used two other choices, $t \to \frac{C}{K}$ and $t := 0$, which led to weaker results (if we ignore the coefficient in front of the $\sqrt{}$ in (58)). Namely, the former choice is equivalent to using Bonferroni's inequality (as noticed by Embrechts and Puccetti [35, Remark 4.1(i)]), and the latter choice gives a worse dependence of ϵ, namely ϵ^{-2} in place of $\ln \frac{1}{\epsilon}$.

Let us state Proposition C.1 in a cruder way. Now we consider a sequence of probability spaces $(\Omega_N, \mathcal{F}_N, P_N)$, $N = 1, 2, \dots$, each equipped with a filtration $(\mathcal{F}_{N,n})$, $n = 0, 1, \dots, N$. Fix a sequence $K_N \in \{1, \dots, N\}$, $N = 1, 2, \dots$, of prediction horizons. Let, for each N, $Y_{N,1}, \dots, Y_{N,N}$ be an adapted sequence of random variables in $(\Omega_N, \mathcal{F}_N, P_N)$ bounded by 1 in absolute value, $|Y_{N,n}| \le 1$ for $n = 1, \dots, N$. When I say that a relation $R_N(O(X_N))$ involving $O(X_N)$ (such as (66) below) holds in probability, I mean that for any $\epsilon > 0$ there exists $C > 0$ such that $P_N(R_N(CX_N)) \ge 1 - \epsilon$ from some N on.[1] According to (58),

$$\left| \sum_{n=K_N}^{N} \left(Y_{N,n} - \mathbb{E}_{P_N}(Y_{N,n} \mid \mathcal{F}_{N,n-K_N}) \right) \right| = O\left(\sqrt{K_N N} \right) \tag{66}$$

in probability. An even cruder form of (66) (and of Proposition C.1) is the following corollary.

Corollary C.5. *Let $(\Omega_N, \mathcal{F}_N, P_N)$, $N = 1, 2, \dots$, be a sequence of probability spaces $(\Omega_N, \mathcal{F}_N, P_N)$ each equipped with a filtration $(\mathcal{F}_{N,n})$, $n = 0, 1, \dots, N$. Suppose the sequence $K_N \in \{1, \dots, N\}$, $N = 1, 2, \dots$, of prediction horizons satisfies $K_N = o(N)$. Let, for each N, $Y_{N,1}, \dots, Y_{N,N}$ be an adapted sequence of random variables in $(\Omega_N, \mathcal{F}_N, P_N)$ bounded by 1 in absolute value. Then*

$$\left| \frac{1}{N - K_N + 1} \sum_{n=K_N}^{N} \left(Y_{N,n} - \mathbb{E}_{P_N}(Y_{N,n} \mid \mathcal{F}_{N,n-K_N}) \right) \right| \to 0 \quad (N \to \infty) \tag{67}$$

holds in probability.

Remember that when we say that random variables ξ_N in probability spaces $(\Omega_N, \mathcal{F}_N, P_N)$ converge to 0 in probability, as in (67), we mean that, for any $\delta > 0$, $P_N(|\xi_N| > \delta) \to 0$ as $N \to \infty$.

The following proposition (analogous to Proposition 6.8) is an inverse to (66). To make it slightly stronger, we state it for finite probability spaces.

[1] Of course, this definition makes an intuitive sense only when the statement $R_N(x)$ becomes weaker as x increases.

Proposition C.6. *There exist $\epsilon > 0$, a sequence of finite probability spaces (Ω_N, P_N), $N = 1, 2, \dots$, each equipped with a filtration $(\mathcal{F}_{N,n})$, $n = 0, 1, \dots, N$, and, for each N, an adapted sequence $Y_{N,1}, \dots, Y_{N,N}$ of random variables in (Ω_N, P_N) bounded by 1 in absolute values, $|Y_{N,n}| \le 1$ for $n = 1, \dots, N$, such that, for any sequence $K_N \in \{1, \dots, \lfloor N/5 \rfloor\}$, $N = 5, 6, \dots$, and for all $N \ge 5$, we have*

$$P_N\left(\sum_{n=K_N}^{N} (Y_{N,n} - \mathbb{E}_{P_N}(Y_{N,n} \mid \mathcal{F}_{N,n-K_N})) \ge \sqrt{K_N N}\right) \ge \epsilon.$$

Proof. Fix independent $\{-1, 1\}$-valued variables $X_1, \dots, X_{\lceil N/K_N \rceil}$ in (Ω_N, P_N) taking values ± 1 with equal probabilities, and set

$$Y_{N,n} := X_{\lceil n/K_N \rceil}, \quad n = 1, \dots, N.$$

Therefore, the N steps are split into $\lceil N/K_N \rceil$ blocks of length K_N (with a possible exception of the last block, which may be shorter), and $Y_{N,n}$ is constant within each block. By the central limit theorem, the probability is at least ϵ (a universal positive constant) that $Y_{N,n} = 1$ in at least $\sqrt{N/K_N} + 1$ more blocks than $Y_{N,n} = -1$. In such cases

$$\sum_{n=K_N}^{N} (Y_{N,n} - \mathbb{E}_{P_N}(Y_{N,n} \mid \mathcal{F}_{N,n-K_N})) = \sum_{n=K_N}^{N} Y_{N,n} \ge K_N \sqrt{N/K_N} = \sqrt{K_N N},$$

where each $\mathcal{F}_{N,n}$ is generated by $Y_{N,1}, \dots, Y_{N,n}$. $\square$

Remark C.7. One inefficient approach to the K-steps ahead martingale law of large numbers (used in the arXiv version 1 of this paper and already alluded to in Remark 6.6) is to apply Hoeffding's inequality to the martingale difference

$$X_n := \sum_{i=n}^{(n+K_N-1)\wedge N} \left(\mathbb{E}_{P_N}(Y_{N,i} \mid \mathcal{F}_{N,n}) - \mathbb{E}_{P_N}(Y_{N,i} \mid \mathcal{F}_{N,n-1})\right),$$

whose increments are bounded by $2K_N$ in absolute value. It is a martingale difference in the sense $\mathbb{E}(X_n \mid \mathcal{F}_{N,n-1}) = 0$, $n = 1, \dots, N$, and it satisfies

$$\begin{aligned}
\sum_{n=1}^{N} X_n &= \sum_{n=K_N}^{N} (Y_{N,n} - \mathbb{E}_{P_N}(Y_{N,n} \mid \mathcal{F}_{N,n-K_N})) \\
&\quad + \sum_{n=N+1}^{N+K_N-1} (\mathbb{E}_{P_N}(Y_{N,n} \mid \mathcal{F}_{N,N}) - \mathbb{E}_{P_N}(Y_{N,n} \mid \mathcal{F}_{N,n-K_N})) \\
&\approx \sum_{n=K_N}^{N} (Y_{N,n} - \mathbb{E}_{P_N}(Y_{N,n} \mid \mathcal{F}_{N,n-K_N}))
\end{aligned}$$

(where the $\approx$ assumes $K_N \ll N$ and ignores borderline effects). This argument, however, requires $K_N = o(N^{1/2})$.

D Radical probabilism

Our testing protocols, such as Protocol 3.2, assume that we learn the observations y_n with full certainty. According to Jeffrey's doctrine of radical probabilism [25], we do not learn anything for certain; at best, we learn that the nth observation is y_n with a high probability (similar statements have been made by many philosophers of science; see, e.g., [2]). In this section we will discuss two modifications of Protocol 3.2 allowing uncertain evidence.

Remark D.1. The uncertainty of observations is a recurring topic for many other philosophers: see, e.g., Popper's discussion of "basic statements" in [34, Chap. 5] (where he also refers to Reininger's and Neurath's similar ideas).

D.1 Additive picture

A straightforward modification of Protocol 3.7 making evidence uncertain is the following one.

Protocol D.2.

$\mathcal{K}_0 := 1$
FOR $n = 1, 2, \dots$:
  Forecaster announces $Q_n \in \mathfrak{P}(\mathbf{Y}^N)$
  IF $n > 1$:
    $$\mathcal{K}_{n-1} := \mathcal{K}_{n-2} + \sum_{x \in \mathbf{Y}^N} F_{n-1}(x)(Q_n(x) - Q_{n-1}(x)) \tag{68}$$
  Sceptic announces $F_n \in \mathbb{R}^{\mathbf{Y}^N}$.

Whereas the main loop in Protocols 3.2 and 3.7 is over finite ranges of n, in Protocol D.2 the loop is infinite since we do not learn any of $y_1, \dots, y_N$ with certainty. Even though in Protocol D.2 y_n are never disclosed explicitly, they may be disclosed implicitly via Q_n: cf. (10). The capital updating rule (68) is very natural: namely, a possible interpretation of this rule is that Q_{n-1} is the expectation of Q_n (cf. [20, Theorem in Sect. 3]).

D.2 Multiplicative picture

Protocol D.2 and all the protocols discussed in the main part of the paper are similar to Protocol 3.1 in that Sceptic's capital is updated by adding various terms. This subsection introduces a multiplicative protocol, in which Sceptic's capital will be updated by multiplication. Both multiplicative and additive protocols are ubiquitous in game-theoretic probability (although the difference between them is rarely pointed out). This is the multiplicative version of Protocol D.2:

Protocol D.3.

$\mathcal{K}_0 := 1$
FOR $n = 1, 2, \dots$:
  Forecaster announces $Q_n \in \mathfrak{P}(\mathbf{Y}^N)$
  IF $n > 1$:

$$\mathcal{K}_{n-1} := \mathcal{K}_{n-2} \sum_{x\in\mathbf{Y}^N} \frac{Q_n(x)}{Q_{n-1}(x)} G_{n-1}(x) \tag{69}$$

Sceptic announces $G_n \in \mathfrak{P}(\mathbf{Y}^N)$.

To see the equivalence of the two protocols, notice that (69) is equivalent to

$$\begin{aligned}
\mathcal{K}_{n-1} - \mathcal{K}_{n-2} &= \left(\sum_{x\in\mathbf{Y}^N} \frac{Q_n(x)}{Q_{n-1}(x)} G_{n-1}(x) - 1 \right) \mathcal{K}_{n-2} \\
&= \left(\sum_{x\in\mathbf{Y}^N} (Q_n(x) - Q_{n-1}(x)) \frac{G_{n-1}(x)}{Q_{n-1}(x)} \right) \mathcal{K}_{n-2}.
\end{aligned}$$

This establishes the one-to-one correspondence

$$F_{n-1}(x) = \frac{G_{n-1}(x)}{Q_{n-1}(x)} \mathcal{K}_{n-2}$$

between F_{n-1} in (68) and G_{n-1} in (69). (This correspondence presupposes that $\mathcal{K}_{n-2} > 0$.)

Protocol D.3 is a special case of Cover's protocol modelling investment into $|\mathbf{Y}|^N$ securities (see, e.g., [6] or [47, Example 9]). As in Sect. 3, we have a market in securities $\Phi(x)$, $x \in \mathbf{Y}^N$, but they are never settled; at best $\Phi(x)$ may approach 0 or 1. For each security $\Phi(x)$ the protocol gives its price $Q_n(x)$ at time n. The prices are normalized in that $Q_n(x)$ sum to 1 over x; therefore, $Q_n(x)$ are actually the market shares. The capital update rule (69) involves the *price relative* $Q_n(x)/Q_{n-1}(x)$ (as used in [6]). At each step Sceptic decides on the distribution G_n of his current capital $\mathcal{K}_{n-1}$ among the securities $\Phi(x)$.

D.3 Radical probabilism and reality

The additive picture, and especially the multiplicative one, shed new light on the protocols in the main part of the paper. The latter cover the case where Q_n, $n = 1, \ldots, N$, is concentrated on $[x] \subseteq \mathbf{Y}^N$ (the set of all continuations of x) for some $x \in \mathbf{Y}$. This corresponds to the difference between futures contracts and other securities, such as stocks. Sooner or later, reality forces itself on a futures contract, but stock prices can be forever arbitrary (in our ideal picture).

It would be interesting to establish conditions under which this paper's results can be extended to the more general and simpler protocol of this appendix.