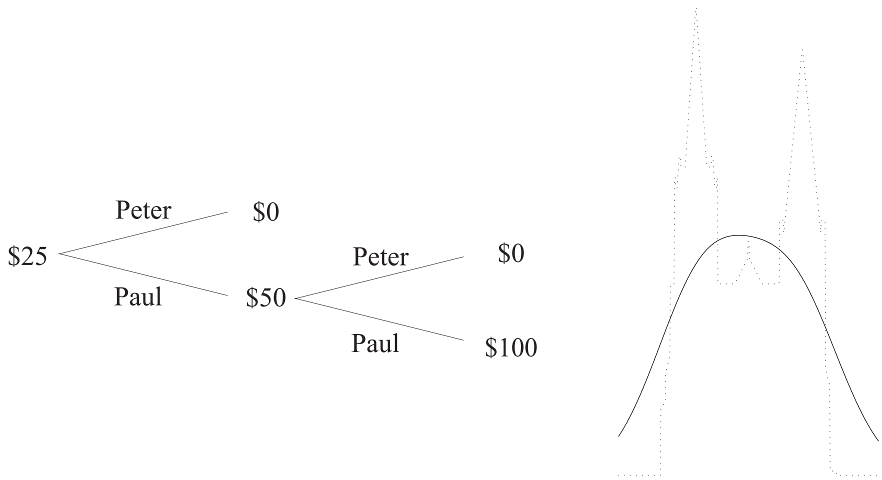
Prequential probability: game-theoretic = measure theoretic

Vladimir Vovk



The Game-Theoretic Probability and Finance Project

Working Paper #27

First posted January 26, 2009. Last revised May 16, 2009.

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

Abstract

This article continues study of the prequential framework for evaluating a probability forecaster. Testing the hypothesis that the sequence of forecasts issued by the forecaster is in agreement with the observed outcomes can be done using prequential notions of probability. It turns out that there are two natural notions of probability in the prequential framework: game-theoretic, whose idea goes back to von Mises and Ville, and measure-theoretic, whose idea goes back to Kolmogorov. The main result of this article is that, in the case of predicting binary outcomes, the two notions of probability in fact coincide on the analytic sets (in particular, on the Borel sets).

Contents

1 Background

The prequential framework for evaluating probability forecasters was introduced by A. P. Dawid in [4] and [5]. Suppose two players, Forecaster and Reality, interact according to the following protocol.

Binary prequential protocol

FOR $n = 1, 2, \dots$:
  Forecaster announces $p_n \in [0, 1]$.
  Reality announces $y_n \in \{0, 1\}$.
END FOR.

The interpretation is that p_n is Forecaster's subjective probability that $y_n = 1$ after having observed $y_1, \dots, y_{n-1}$ and taking account of all other relevant information available at the time of issuing the forecast. We will refer to p_n as *forecasts* and to y_n as *outcomes*. More generally, the outcomes take values in an arbitrary measurable space and the forecasts are probability distributions on that measurable space, but in this article we will restrict our attention to binary outcomes (as in [5]); this will be further discussed at the end of Section 7.

In general, the two players possess perfect information about each other's moves: Forecaster chooses p_1, Reality observes p_1 and chooses y_1, Forecaster observes y_1 and chooses p_1, etc. We might, however, be interested in "oblivious" strategies for a player, especially for Reality, who may generate her moves randomly according to a probability measure on $\{0, 1\}^\infty$ chosen in advance. On the other hand, the players may also react to events outside the protocol.

Dawid's *prequential principle* (see, e.g., [4, 5, 6]) says that when testing the adequacy of the forecaster in light of the outcomes y_n we should only use the forecasts p_n, not the forecasting strategy (if any) that Forecaster used to produce p_n. In this article we will be only interested in testing procedures that respect the prequential principle. In other words, we will be interested in testing the sequence

$$(p_1, y_1, p_2, y_2, \dots) \tag{1}$$

of forecast/outcome pairs (p_n, y_n) for agreement. This sequence may be infinite or finite.

There are two main ways to test sequences (1) for agreement, which we will call game-theoretic and measure-theoretic. For concreteness, suppose the sequence (1) does not satisfy

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (y_i - p_i) = 0 \tag{2}$$

(i.e., the sequence is not "unbiased in the large"; see, e.g., [6] for numerous other ways of testing probability forecasts). What do we mean when we say that violation of (2) evidences lack of agreement?

Two ways to answer this question correspond to two different approaches to the foundations of probability theory. One version of the game-theoretic answer is that we can gamble against the forecasts is such a way that, risking only one monetary unit, we can become infinitely rich when (2) is violated. The measure-theoretic answer is that, no matter what strategy Forecaster is using, the probability of (2) is one; therefore, if (2) is violated, an *a priori* specified event of probability zero (given by the negation of the formula (2)) has occurred. Both becoming infinitely rich and the occurrence of a pre-specified event of probability zero can be interpreted as lack of agreement between the forecasts and outcomes.

In fact, even the first answer can be expressed in terms of probability. The game-theoretic approach to the foundations of probability is as old as the standard measure-theoretic based on Kolmogorov's axioms ([11]; see [17] for the historical background). An imperfect version of the game-theoretic approach was championed by von Mises [14] and formalized, in different ways, by Wald [24] and Church [2]. Ville [21] gave an example demonstrating that von Mises's notion of a gambling strategy was too restrictive, and introduced a more general class of gambling strategies and a closely related notion of a martingale. However, the formal notion of game-theoretic probability was introduced only recently (see, e.g., [22], [6], or, for a much fuller treatment, [16]). In particular, an event has zero game-theoretic probability if and only if there is a gambling strategy that, risking at most one monetary unit, makes the player infinitely rich when the event happens.

The notion of game-theoretic probability makes the game-theoretic and measure-theoretic justifications of the testing procedure based on (2) look very similar: we just say that the probability (either game-theoretic or measure-theoretic) of (2) being violated is zero. The main result of this article says that the two notions of probability coincide on the analytic sets, and so the two approaches to testing probability forecasts are equivalent, in the prequential framework. The restriction to the analytic, and even Borel, sets is not a limitation in all practically interesting cases.

For testing procedures based on events of probability zero (basically, on strong laws of probability theory, such as (2)), a special case of our result is sufficient: it is sufficient to know that a Borel set has zero game-theoretic probability if and only if it has zero measure-theoretic probability. Our full result is also applicable to events of merely low, not zero, probability. For example, we could reject the hypothesis of agreement if

$$\frac{1}{n}\left|\sum_{i=1}^{n}(y_i - p_i)\right| \geq C\sqrt{n} \qquad (3)$$

for prespecified large numbers C and n. Our result shows that this and similar procedures have equally strong game-theoretic and measure-theoretic justifications. Notice that in the case of (3) our decision to reject the hypothesis of agreement can be made after observing a finite sequence, $(p_1, y_1, \ldots, p_n, y_n)$.

2 This article

In the following two sections, 3 and 4, we formally introduce in the prequential framework the two notions of probability discussed in the previous section. The main result of this article is Theorem 1 in Section 5, asserting the coincidence of the two kinds of probability on the analytic sets. This result has several predecessors. In the situation where Forecaster's strategy is fixed, Ville ([21], Theorems 1 and 2 in Chapter IV) showed that a set E has game-theoretic probability zero if and only if it has measure-theoretic probability zero. (Ville stated this result in slightly different terms, without explicit use of game-theoretic probability.) This was generalized in [16] (Proposition 8.13) to the statement that game-theoretic and measure-theoretic probability coincide on the Borel sets. In the case of a finite-horizon protocol, a statement analogous to Theorem 1 was proved by Shafer in [15] (Proposition 12.7.4). The special case of Theorem 1 asserting the coincidence of game-theoretic and measure-theoretic probability on the open sets was first proved in [23] (Theorem 2).

In the same Section 5 we also prove that measure-theoretic probability never exceeds game-theoretic probability. This simple statement is true for all sets, not just analytic. The proof of the opposite inequality is given in Section 6. It relies on two fundamental results: Choquet's capacitability theorem [1] and Lévy's zero-one law in its game-theoretic version recently found in [19].

Some notation and definitions

The set of all natural (i.e., positive integer) numbers is denoted $\mathbb{N}$, $\mathbb{N} := \{1, 2, \ldots\}$. As always, $\mathbb{R}$ is the set of all real numbers.

Let $\Omega := \{0,1\}^\infty$ be the set of all infinite binary sequences and $\Omega^\diamond := \cup_{n=0}^\infty \{0,1\}^n$ be the set of all finite binary sequences. Set $\Pi := ([0,1] \times \{0,1\})^\infty$ and $\Pi^\diamond := \cup_{n=0}^\infty ([0,1] \times \{0,1\})^n$. The empty element (sequence of length zero) of both $\Omega^\diamond$ and $\Pi^\diamond$ will be denoted Λ. In our applications, the elements of Ω and $\Omega^\diamond$ will be sequences of outcomes (infinite or finite), and the elements of Π and $\Pi^\diamond$ will be sequences of forecasts and outcomes (infinite or finite). The set Π will sometimes be referred to as the *prequential space*.

For $x \in \Omega^\diamond$, let $\Gamma(x) \subseteq \Omega$ be the set of all infinite extensions of x that belong to Ω. Similarly, for $x \in \Pi^\diamond$, $\Gamma(x) \subseteq \Pi$ is the set of all infinite extensions of x that belong to Π. For each $\omega = (y_1, y_2, \ldots) \in \Omega$ and $n \in \mathbb{N} \cup \{0\}$, set $\omega^n := (y_1, \ldots, y_n)$. Similarly, for each $\pi = (p_1, y_1, p_2, y_2, \ldots) \in \Pi$ and $n \in \mathbb{N} \cup \{0\}$, set $\pi^n := (p_1, y_1, \ldots, p_n, y_n)$.

In some proofs and remarks we will be using the following notation, for $n \in \mathbb{N} \cup \{0\}$: $\Omega^n := \{0,1\}^n$ is the set of all finite binary sequences of length n; $\Omega^{\geq n} := \cup_{i=n}^\infty \Omega^i$ is the set of all finite binary sequences of length at least n; $\Pi^n := ([0,1] \times \{0,1\})^n$; $\Pi^{\geq n} := \cup_{i=n}^\infty ([0,1] \times \{0,1\})^i$.

3 Game-theoretic prequential probability

A *farthingale* is a function $V : \Pi^\diamond \to (-\infty, \infty]$ satisfying

$$V(p_1, y_1, \dots, p_{n-1}, y_{n-1}) \\ = (1 - p_n)V(p_1, y_1, \dots, p_{n-1}, y_{n-1}, p_n, 0) \\ + p_n V(p_1, y_1, \dots, p_{n-1}, y_{n-1}, p_n, 1) \quad (4)$$

for all $n \in \mathbb{N}$ and all $(p_1, y_1, p_2, y_2, \dots) \in \Pi$; the product 0∞ is defined to be 0. If we replace "=" by "$\geq$" in (4), we get the definition of a *superfarthingale*. These are prequential versions of the standard notions of martingale and supermartingale. We will be interested mainly in non-negative farthingales and superfarthingales.

The value of a farthingale can be interpreted as the capital of a gambler betting according to the odds announced by Forecaster. In the case of superfarthingales, the gambler is allowed to throw away part of his capital.

Game-theoretic probability can be introduced as either upper or lower probability; in this article the former is more convenient (and was used in the informal discussion of Section 1). A *prequential event* is a subset of Π. The *upper game-theoretic probability* of a prequential event E is

$$\mathbb{P}^{\text{game}}(E) := \inf \left\{ a \mid \exists V : V(\Lambda) = a \text{ and } \forall \pi \in E : \limsup_n V(\pi^n) \geq 1 \right\}, \quad (5)$$

where V ranges over the non-negative farthingales. It is clear that we will obtain the same notion of upper game-theoretic probability if we replace the $\geq$ in (5) by $>$, replace lim sup by sup or lim inf (we can always stop when 1 is reached), or allow V to range over the non-negative superfarthingales.

We will need the following property of countable subadditivity of game-theoretic probability.

Lemma 1. *For any sequence $E_1, E_2, \dots$ of prequential events,*

$$\mathbb{P}^{\text{game}}\left(\cup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}^{\text{game}}(E_i).$$

In particular, if $\mathbb{P}^{\text{game}}(E_i) = 0$ for all i, then $\mathbb{P}^{\text{game}}(\cup_{i=1}^{\infty} E_i) = 0$.

Proof. It suffices to notice that the sum of a sequence of non-negative farthingales is again a non-negative farthingale. □

4 Measure-theoretic prequential probability

A *forecasting system* is a function $\phi : \Omega^\diamond \to [0, 1]$. Let Φ be the set of all forecasting systems. For each $\phi \in \Phi$ there exists a unique probability measure $\mathbb{P}_\phi$ on Ω (equipped with the Borel σ-algebra) such that, for each $x \in \Omega^\diamond$,

$\mathbb{P}_\phi(\Gamma(x1)) = \phi(x)\,\mathbb{P}_\phi(\Gamma(x))$. (In other words, such that $\phi(x)$ is a version of the conditional probability, according to $\mathbb{P}_\phi$, that x will be followed by 1.) The notion of a forecasting system is close to that of a probability measure on Ω: the correspondence $\phi \mapsto \mathbb{P}_\phi$ becomes an isomorphism if we only consider forecasting systems taking values in the open interval $(0,1)$ and probability measures taking positive values on the sets $\Gamma(x)$, $x \in \Omega^\diamond$.

For each sequence $(y_1,\dots,y_n) \in \Omega^\diamond$ and each forecasting system $\phi \in \Phi$, let

$$(y_1,\dots,y_n)^\phi := (\phi(\Lambda), y_1, \phi(y_1), y_2, \dots, \phi(y_1,\dots,y_{n-1}), y_n) \in \Pi^\diamond.$$

Similarly, for each $(y_1, y_2, \dots) \in \Omega$ and each $\phi \in \Phi$,

$$(y_1, y_2, \dots)^\phi := (\phi(\Lambda), y_1, \phi(y_1), y_2, \phi(y_1, y_2), y_3, \dots) \in \Pi.$$

We can apply the idea of measure-theoretic probability to prequential events as follows, in the spirit of [9], Section 10.2. For each forecasting system ϕ and prequential event $E \subseteq \Pi$, define

$$\mathbb{P}^\phi(E) := \mathbb{P}_\phi\left\{\omega \in \Omega \mid \omega^\phi \in E\right\} = \mathbb{P}_\phi(E^\phi),$$

where $E^\phi := \left\{\omega \in \Omega \mid \omega^\phi \in E\right\}$ and $\mathbb{P}_\phi(A)$ is understood, in general, as the outer measure of A, i.e., as $\inf_B \mathbb{P}_\phi(B)$, B ranging over the Borel sets containing A. The convention about using the outer measure is important only for our proofs, not for the statement of the main result: according to Luzin's theorem (see, e.g., [10], Theorem 21.10), every analytic set is universally measurable, and E^ϕ is analytic whenever E is. Now we define the *upper measure-theoretic probability* of E as

$$\mathbb{P}^{\mathrm{meas}}(E) := \sup_\phi \mathbb{P}^\phi(E). \tag{6}$$

Remark 1. Our definition (6) is not fully adequate from the intuitive point of view: even if we are willing to assume that Forecaster follows some forecasting strategy (which is a non-trivial assumption: cf. the discussion in [5], pp. 1255–1256), why should this forecasting strategy depend only on the past outcomes? For example, a meteorologist forecasting rain might have data about temperatures, winds, etc. (See [5], Section 9, for further discussion.) A more satisfactory definition would involve a supremum over all probability spaces equipped with a filtration and for each such a probability space a further supremum over all forecasting systems adapted to the corresponding filtration (with a natural more general definition of a forecasting system). Our definition (6) is the simplest one mathematically and leads to the strongest inequality $\mathbb{P}^{\mathrm{game}}(E) \le \mathbb{P}^{\mathrm{meas}}(E)$ (for the analytic sets), which is the non-trivial part of our main result, Theorem 1.

5 Main result

Now we are have all ingredients needed to state our main result.

Theorem 1. *For all analytic sets $E \subseteq \Pi$, $\mathbb{P}^{\mathrm{game}}(E) = \mathbb{P}^{\mathrm{meas}}(E)$.*

Intuitively, this theorem establishes the equivalence between the purely prequential and Bayesian viewpoints in the framework of probability forecasting. The definition of measure-theoretic probability is Bayesian, in that Forecaster is modeled as a coherent subjectivist Bayesian having a joint probability distribution over the sequences of outcomes (cf. [3], Section 1); we represent this joint probability distribution as a forecasting system. Rejecting his forecasts is the same as rejecting all forecasting systems that could have produced those forecasts: cf. the $\sup_\phi$ in (6). The definition of game-theoretic probability is purely prequential, in that it does not postulate any joint probability distribution behind the forecasts; the latter are used for testing directly.

Remark 2. As discussed in the previous section (Remark 1), our Bayesian forecaster is somewhat naive: he conditions only on the observed outcomes. It would be easy (but would complicate the exposition) to allow Reality to issue a *signal* s_n, taking one of a finite number of values, before Forecaster chooses his forecast p_n. Allowing both farthingales and forecasting systems to depend on the signals, one could still prove that $\mathbb{P}^{\mathrm{game}}(E) = \mathbb{P}^{\mathrm{meas}}(E)$ for all analytic $E \subseteq \Pi$ following the proof of Theorem 1.

In this section we will only prove the inequality $\geq$ in Theorem 1. It turns out that this inequality holds for all sets E, not necessarily analytic.

Theorem 2. *For all sets $E \subseteq \Pi$, $\mathbb{P}^{\mathrm{game}}(E) \geq \mathbb{P}^{\mathrm{meas}}(E)$.*

The simple proof of Theorem 2 will follow from Ville's inequality ([21], p. 100; in modern probability textbooks this result is often included among "Doob's inequalities": see, e.g., [20], Theorem VII.3.1.III).

Let ϕ be a forecasting system. A *martingale* w.r. to ϕ is a function $V : \Omega^\diamond \to (-\infty, \infty]$ satisfying

$$V(x) = (1 - \phi(x))V(x, 0) + \phi(x)V(x, 1)$$

for all $x \in \Omega^\diamond$ (with the same convention $0\infty := 0$).

Proposition 1 ([21])**.** *If ϕ is a forecasting system, V is a non-negative martingale w.r. to ϕ, and $C > 0$,*

$$\mathbb{P}_\phi \left\{ \omega \in \Omega \mid \sup_n V(\omega^n) \geq C \right\} \leq \frac{V(\Lambda)}{C}.$$

If V is a farthingale, the function $V^\phi : \Omega^\diamond \to (-\infty, \infty]$ defined by

$$V^\phi(x) := V\left(x^\phi\right), \quad x \in \Omega^\diamond,$$

is a martingale w.r. to ϕ. It is important that this statement does not require measurability of the farthingale V; even if V is not measurable, V^ϕ is always measurable, like any other function on $\Omega^\diamond$ (which is why there was no need to include the requirement of measurability in our definition of a martingale).

Proof of Theorem 2. Let $E \subseteq \Pi$. It suffices to prove that $\mathbb{P}_\phi(E^\phi) \leq V(\Lambda)$ for any forecasting system ϕ and any non-negative farthingale V satisfying $\limsup_n V(\pi^n) \geq 1$ for all $\pi \in E$. Fix such ϕ and V. Then V^ϕ is a non-negative martingale w.r.t. to ϕ satisfying $\limsup_n V^\phi(\omega^n) \geq 1$ for all $\omega \in E^\phi$. Applying Proposition 1 to V^ϕ, we can see that indeed $\mathbb{P}_\phi(E^\phi) \leq V^\phi(\Lambda) = V(\Lambda)$. □

6 Proof of the inequality $\leq$ in Theorem 1

We start from proving a special case of Theorem 1.

Lemma 2. *If $E \subseteq \Pi$ is a compact set,* $\mathbb{P}^{\text{meas}}(E) = \mathbb{P}^{\text{game}}(E)$.

Proof. Fix a compact prequential event $E \subseteq \Pi$. (Of course, "compact" is the same thing as "closed" in this context.) Represent E as the intersection $E = \cap_{i=1}^\infty E_i$ of a nested sequence $E_1 \supseteq E_2 \supseteq \cdots$ of closed sets such that

$$\forall \pi \in \Pi : \pi \in E_i \Longrightarrow \Gamma(\pi^i) \subseteq E_i \tag{7}$$

is satisfied for all i. Informally, E_i is a property of the first i forecasts and outcomes. For each $i = 1, 2, \ldots$, define a superfarthingale W_i by setting

$$W_i(x) := \begin{cases} 1 & \text{if } \Gamma(x) \subseteq E_i \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

for all $x \in \Pi^{\geq i}$ and then proceeding inductively as follows. If $W_i(x)$ is already defined for $x \in \Pi^n$, $n = i, i-1, \ldots, 1$, define $W_i(x)$, for each $x \in \Pi^{n-1}$, by

$$W_i(x) := \sup_{p \in [0,1]} \bigl((1-p)W_i(x,p,0) + pW_i(x,p,1)\bigr). \tag{9}$$

It is clear that $W_1 \geq W_2 \geq \cdots$.

Let us check that $W_i(x)$ is upper semicontinuous as a function of $x \in \Pi^\diamond$. By (8) this is true for $x \in \Pi^{\geq i}$. Suppose this is true for $x \in \Pi^n$, $n \in \{i, i-1, \ldots, 2\}$, and let us prove that it is true for $x \in \Pi^{n-1}$, using the inductive definition (9). It is clear that $f(x,p) := (1-p)W_i(x,p,0) + pW_i(x,p,1)$ is upper semicontinuous as function of $p \in [0,1]$ and $x \in \Pi^{n-1}$. It is well known that $\sup_p f(x,p)$ is upper semicontinuous whenever f is upper semicontinuous and x and p range over compact sets (see, e.g., [7], Theorem I.2(d)). A simple proof of a slightly more general fact will be given below in Lemma 3. Therefore, $W_i(x) = \sup_{p\in[0,1]} f(x,p)$ is an upper semicontinuous function of $x \in \Pi^{n-1}$.

An important implication of the upper semicontinuity of W_i and the compactness of $[0,1]$ is that the supremum in (9) is attained: it is easy to check that an upper semicontinuous function attains its supremum over a compact set (cf. [8], Problem 3.12.23(g)). For each $i = 1, 2, \ldots$, we can now define a forecasting system ϕ_i as follows. For each $x \in \Omega^n$, $n = 0, 1, \ldots, i-1$, choose $\phi_i(x)$ such that

$$(1-\phi_i(x))W_i(x^{\phi_i},\phi_i(x),0)+\phi_i(x)W_i(x^{\phi_i},\phi_i(x),1)$$
$$=\sup_p\bigl((1-p)W_i(x^{\phi_i},p,0)+pW_i(x^{\phi_i},p,1)\bigr)=W_i(x^{\phi_i})$$

(this is an inductive definition; in particular, x^{ϕ_i} is already defined at the time of defining $\phi_i(x)$). For $x \in \Omega^{\geq i}$, set, for example, $\phi_i(x) := 0$. The important property of ϕ_i is that $W_i^{\phi_i}$ is a martingale w.r. to ϕ_i, and so $\mathbb{P}^{\phi_i}(E_i) = W_i(\Lambda)$.

Since the set Φ of all forecasting systems is compact in the product topology, the sequence ϕ_i has a convergent subsequence ϕ_{i_k}, $k = 1, 2, \ldots$; let $\phi := \lim_{k\to\infty} \phi_{i_k}$. We assume, without loss of generality, $i_1 < i_2 < \cdots$. Set

$$c := \inf_i W_i(\Lambda) = \lim_{i\to\infty} W_i(\Lambda).$$

Fix an arbitrarily small $\epsilon > 0$. Let us prove that $\mathbb{P}_\phi(E^\phi) \geq c - \epsilon$. Let $K \in \mathbb{N}$. The restriction of $\mathbb{P}_{\phi_{i_k}}$ to Ω^{i_K} (more formally, the probability measure assigning weight $\mathbb{P}_{\phi_{i_k}}(\Gamma(x))$ to each singleton $\{x\}$, $x \in \Omega^{i_K}$) comes within ϵ of the restriction of $\mathbb{P}_\phi$ to Ω^{i_K} in total variation distance from some k on; let the total variation distance be at most ϵ for all $k \geq K' \geq K$. Let $k \geq K'$. Since $\mathbb{P}_{\phi_{i_k}}(E_{i_k}^{\phi_{i_k}}) \geq c$, it is also true that $\mathbb{P}_{\phi_{i_k}}(E_{i_K}^{\phi_{i_k}}) \geq c$; therefore, it is true that $\mathbb{P}_\phi(E_{i_K}^{\phi_{i_k}}) \geq c - \epsilon$. By Fatou's lemma, we now obtain

$$\mathbb{P}_\phi\left(\limsup_k E_{i_K}^{\phi_{i_k}}\right) \geq \limsup_{k\to\infty} \mathbb{P}_\phi(E_{i_K}^{\phi_{i_k}}) \geq c - \epsilon. \tag{10}$$

Let us check that

$$\limsup_k E_{i_K}^{\phi_{i_k}} \subseteq E_{i_K}^\phi. \tag{11}$$

Indeed, let $\omega \notin E_{i_K}^\phi$, i.e., $\omega^\phi \notin E_{i_K}$. Since $\phi_{i_k} \to \phi$ in the product topology and the set E_{i_K} is closed, $\omega^{\phi_{i_k}} \notin E_{i_K}$ from some k on. This means that $\omega \in E_{i_K}^{\phi_{i_k}}$ for only finitely many k, i.e., $\omega \notin \limsup_k E_{i_K}^{\phi_{i_k}}$.

From (10) and (11) we can see that $\mathbb{P}_\phi(E_{i_K}^\phi) \geq c - \epsilon$, for all $K \in \mathbb{N}$. This implies $\mathbb{P}_\phi(E^\phi) \geq c - \epsilon$. Since this holds for all ϵ, $\mathbb{P}_\phi(E^\phi) \geq c$.

The rest of the proof is easy: since

$$\mathbb{P}^{\text{game}}(E) \leq c \leq \mathbb{P}_\phi(E^\phi) \leq \mathbb{P}^{\text{meas}}(E) \leq \mathbb{P}^{\text{game}}(E)$$

(the last inequality following from Theorem 2), we have

$$\mathbb{P}^{\text{game}}(E) = c = \mathbb{P}_\phi(E^\phi) = \mathbb{P}^{\text{meas}}(E). \qquad \square$$

In the proof of Lemma 2 we referred to the following simple result.

Lemma 3. *Suppose X and Y are topological spaces and Y is compact. If a function $f : X \times Y \to \mathbb{R}$ is upper semicontinuous, then the function $x \in X \mapsto g(x) := \sup_{y\in Y} f(x, y)$ is also upper semicontinuous.*

Proof. For any $c \in \mathbb{R}$, we are required to show that the set $G := \{x \mid \sup_y f(x,y) < c\}$ is open. Let $x \in G$. For any $y \in Y$ there exists a neighborhood O'_y of x and a neighborhood O''_y of y such that, for some $\epsilon > 0$, $f(x',y') < c - \epsilon$ for all $x' \in O'_y$ and all $y' \in O''_y$. By the compactness of Y, there is a finite family $O''_{y_1}, \ldots, O''_{y_K}$ that covers Y. The intersection of $O'_{y_1}, \ldots, O'_{y_K}$ will contain x and will be a subset of G. Therefore, G is indeed open.

The argument in [7], proof of Theorem I.2(d), is even simpler, but it assumes that X is compact (which is, however, sufficient for the purpose of Lemma 2). □

The idea of the proof of Theorem 1 is to extend Lemma 2 to the analytic sets using Choquet's capacitability theorem (stated below). Remember that a function γ (such as $\mathbb{P}^{\mathrm{game}}$ or $\mathbb{P}^{\mathrm{meas}}$) mapping the power set of a topological space X (such as Π) to $[0,\infty)$ is a *capacity* if:

- for any subsets A and B of X,

$$A \subseteq B \Longrightarrow \gamma(A) \le \gamma(B); \tag{12}$$

- for any nested increasing sequence $A_1 \subseteq A_2 \subseteq \cdots$ of arbitrary subsets of X,

$$\gamma\left(\cup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \gamma(A_i); \tag{13}$$

- for any nested decreasing sequence $K_1 \supseteq K_2 \supseteq \cdots$ of compact sets in X,

$$\gamma\left(\cap_{i=1}^{\infty} K_i\right) = \lim_{i\to\infty} \gamma(K_i). \tag{14}$$

Condition (14) is sometimes replaced by a different condition which is equivalent to (14) for compact metrizable spaces X: cf. [10], Definition 30.1.

It turns out that both $\mathbb{P}^{\mathrm{game}}$ and $\mathbb{P}^{\mathrm{meas}}$ are capacities. We start from $\mathbb{P}^{\mathrm{game}}$.

Theorem 3. *The set function* $\mathbb{P}^{\mathrm{game}}$ *is a capacity.*

It is obvious that $\mathbb{P}^{\mathrm{game}}$ satisfies condition (12). The following two statements establish conditions (13) and (14). Condition (14) is easier to check: it can be extracted from the proof of Lemma 2.

Lemma 4. *If* $K_1 \supseteq K_2 \supseteq \cdots$ *is a nested sequence of compact sets in* Π*,*

$$\mathbb{P}^{\mathrm{game}}\left(\cap_{i=1}^{\infty} K_i\right) = \lim_{i\to\infty} \mathbb{P}^{\mathrm{game}}(K_i). \tag{15}$$

Proof. We will use the equality $\mathbb{P}^{\mathrm{game}}(E) = \lim_{i\to\infty} \mathbb{P}^{\mathrm{game}}(E_i)$, in the notation of the proof of Lemma 2. This equality follows from

$$\mathbb{P}^{\mathrm{game}}(E) = c = \lim_{i\to\infty} W_i(\Lambda) \ge \lim_{i\to\infty} \mathbb{P}^{\mathrm{game}}(E_i)$$

(the opposite inequality is obvious).

Represent each K_n in the form $K_n = \cap_{i=1}^{\infty} E_i$ where $E_1 \supseteq E_2 \supseteq \cdots$ and each E_i satisfies (7); we will write $K_{n,i}$ in place of E_i. Without loss of generality we will assume that $K_{1,i} \supseteq K_{2,i} \supseteq \cdots$ for all i. Then the set $K := \cap_{i=1}^{\infty} K_i$ can be represented as $K = \cap_{i=1}^{\infty} K_{i,i}$, and so (15) follows from

$$\begin{aligned}\mathbb{P}^{\text{game}}(K) = \mathbb{P}^{\text{game}}\left(\cap_{i=1}^{\infty} K_{i,i}\right) = \lim_{i\to\infty} \mathbb{P}^{\text{game}}(K_{i,i}) = \lim_{n\to\infty}\lim_{i\to\infty} \mathbb{P}^{\text{game}}(K_{n,i}) \\ = \lim_{n\to\infty} \mathbb{P}^{\text{game}}\left(\cap_{i=1}^{\infty} K_{n,i}\right) = \lim_{n\to\infty} \mathbb{P}^{\text{game}}(K_n). \quad \square\end{aligned}$$

To check condition (13) for $\mathbb{P}^{\text{game}}$, we will need the game-theoretic version, proved in [19], of Lévy's zero-one law ([13], Section 41). For each $x \in \Pi^{\diamond}$, define the *conditional upper game-theoretic probability* of $E \subseteq \Pi$ by

$$\mathbb{P}^{\text{game}}(E \mid x) := \inf\left\{a \mid \exists V : V(x) = a \text{ and } \forall \pi \in E \cap \Gamma(x) : \limsup_{n} V(\pi^n) \geq 1\right\},$$

where V ranges over the non-negative (super)farthingales.

Proposition 2 ([19])**.** *Let $E \subseteq \Pi$. For almost all $\pi \in E$,*

$$\mathbb{P}^{\text{game}}(E \mid \pi^n) \to 1 \tag{16}$$

as $n \to \infty$. (In other words, there exists a prequential event N such that $\mathbb{P}^{\text{game}}(N) = 0$ and (16) holds for all $\pi \in E \setminus N$.)

Proof. It suffices to construct a non-negative farthingale V starting from 1 that tends to ∞ on the sequences $\pi \in E$ for which (16) is not true. Without loss of generality we replace "for which (16) is not true" by

$$\liminf_{n\to\infty} \mathbb{P}^{\text{game}}(E \mid \pi^n) < a,$$

where $a \in (0, 1)$ is a given rational number (see Lemma 1).

Let π be any sequence in Π; we will define $V(\pi^n)$ by induction for $n = 1, 2, \ldots$ (intuitively, we will describe a gambling strategy with capital process V). Start with 1 monetary unit: $V(\Lambda) := 1$. Keep setting $V(\pi^n) := 1$, $n = 1, 2, \ldots$, until $\mathbb{P}^{\text{game}}(E \mid \pi^n) < a$ (if this never happens, $V(\pi^n)$ will be 1 for all n). Let N_1 be the first n when this happens: $\mathbb{P}^{\text{game}}(E \mid \pi^{N_1}) < a$ but $\mathbb{P}^{\text{game}}(E \mid \pi^n) \geq a$ for all $n < N_1$. Choose a non-negative farthingale S_1 starting at π^{N_1} from 1, $S_1(\pi^{N_1}) = 1$, whose upper limit exceeds $1/a$ on all extensions of π^{N_1} in E. Keep setting $V(\pi^n) := S_1(\pi^n)$, $n = N_1, N_1 + 1, \ldots$, until $S_1(\pi^n)$ reaches a value $s_1 > 1/a$. After that keep setting $V(\pi^n) := V(\pi^{n-1})$ until $\mathbb{P}^{\text{game}}(E \mid \pi^n) < a$. Let N_2 be the first n when this happens. Choose a non-negative farthingale S_2 starting at π^{N_2} from s_1, $S_2(\pi^{N_2}) = s_1$, whose upper limit exceeds s_1/a on all extensions of π^{N_2} in E. Keep setting $V(\pi^n) := S_2(\pi^n)$, $n = N_2, N_2 + 1, \ldots$, until $S_2(\pi^n)$ reaches a value $s_2 > s_1(1/a) > (1/a)^2$. After that keep setting $V(\pi^n) := V(\pi^{n-1})$ until $\mathbb{P}^{\text{game}}(E \mid \pi^n) < a$. Let N_3 be the first n when this happens. Choose a non-negative farthingale S_3 starting at π^{N_3} from s_2 whose upper limit exceeds s_2/a on all extensions of π^{N_3} in E. Keep setting $V(\pi^n) := S_3(\pi^n)$, $n = N_3, N_3+1, \ldots$, until S_3 reaches a value $s_3 > s_2(1/a) > (1/a)^3$. And so on. □

Lemma 5. *If $A_1 \subseteq A_2 \subseteq \cdots \subseteq \Pi$ is a nested sequence of prequential events,*

$$\mathbb{P}^{\mathrm{game}}\left(\cup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mathbb{P}^{\mathrm{game}}(A_i). \tag{17}$$

Proof. Let $A_1, A_2, \ldots$ be a nested increasing sequence of prequential events. The non-trivial inequality in (17) is $\leq$. For each A_i the process

$$S_i(x) := \mathbb{P}^{\mathrm{game}}(A_i \mid x)$$

is a non-negative superfarthingale (see Lemma 6 below). By Proposition 2, $\limsup_n S_i(\pi^n) \geq 1$ for almost all $\pi \in A_i$. The sequence S_i is increasing, $S_1 \leq S_2 \leq \cdots$, so the limit $S := \lim_{i\to\infty} S_i = \sup_i S_i$ exists and is a non-negative superfarthingale such that $S(\Lambda) = \lim_{i\to\infty} \mathbb{P}^{\mathrm{game}}(A_i)$ and $\limsup_n S(\pi^n) \geq 1$ for almost all $\pi \in \cup_i A_i$ (by Lemma 1). We can get rid of "almost" by adding to S a non-negative farthingale V that starts at $V(\Lambda) < \epsilon$, for an arbitrarily small $\epsilon > 0$, and satisfies $\limsup_n V(\pi^n) \geq 1$ for all $\pi \in \cup_i A_i$ violating $\limsup_n S(\pi^n) \geq 1$. □

Lemma 6. *For any prequential event E, the function $x \in \Pi^\diamond \mapsto \mathbb{P}^{\mathrm{game}}(E \mid x)$ is a superfarthingale.*

Proof. Suppose there are $x \in \Pi^\diamond$ and $p \in [0,1]$ such that

$$\mathbb{P}^{\mathrm{game}}(E \mid x) < (1-p)\,\mathbb{P}^{\mathrm{game}}(E \mid x, p, 0) + p\,\mathbb{P}^{\mathrm{game}}(E \mid x, p, 1).$$

Then there exists a non-negative farthingale V with $\limsup_n V(\pi^n) \geq 1$ for all $\pi \in E \cap \Gamma(x)$ that satisfies

$$V(x) < (1-p)\,\mathbb{P}^{\mathrm{game}}(E \mid x, p, 0) + p\,\mathbb{P}^{\mathrm{game}}(E \mid x, p, 1)$$

and, therefore,

$$(1-p)V(x,p,0) + pV(x,p,1) < (1-p)\,\mathbb{P}^{\mathrm{game}}(E \mid x, p, 0) + p\,\mathbb{P}^{\mathrm{game}}(E \mid x, p, 1).$$

The last inequality implies that there exists $j \in \{0,1\}$ such that $V(x,p,j) < \mathbb{P}^{\mathrm{game}}(E \mid x, p, j)$, which is impossible. □

This completes the proof of Theorem 3. Let us now check that measure-theoretic probability is also a capacity.

Lemma 7. *The set function $\mathbb{P}^{\mathrm{meas}}$ is a capacity.*

Proof. Property (12) is obvious for $\mathbb{P}^{\text{meas}}$. Property (14) follows from Lemmas 2 and 4.

Let us now check the remaining property (13), with $\mathbb{P}^{\text{meas}}$ as γ. Suppose there exists an increasing sequence $A_1 \subseteq A_2 \subseteq \cdots \subseteq X$ of prequential events such that

$$\mathbb{P}^{\text{meas}}\left(\cup_{i=1}^{\infty} A_i\right) > \lim_{i\to\infty} \mathbb{P}^{\text{meas}}(A_i).$$

Let ϕ be a forecasting system satisfying

$$\mathbb{P}^{\phi}\left(\cup_{i=1}^{\infty} A_i\right) > \lim_{i\to\infty} \mathbb{P}^{\text{meas}}(A_i).$$

Then ϕ will satisfy $\mathbb{P}^{\phi}\left(\cup_{i=1}^{\infty} A_i\right) > \lim_{i\to\infty} \mathbb{P}^{\phi}(A_i)$, which is equivalent to the obviously wrong $\mathbb{P}_{\phi}\left(\cup_{i=1}^{\infty} A_i^{\phi}\right) > \lim_{i\to\infty} \mathbb{P}_{\phi}(A_i^{\phi})$. □

In combination with Choquet's capacitability theorem, Theorem 3 and Lemma 7 allow us to finish the proof of Theorem 1.

Choquet's Capacitability Theorem ([1]). *If X is a compact metrizable space, γ is a capacity on X, and $E \subseteq X$ is an analytic set,*

$$\gamma(E) = \sup\left\{\gamma(K) \mid K \textit{ is compact}, K \subseteq E\right\}.$$

For a proof of Choquet's theorem, see, e.g., [10], Theorem 30.13.

Proof of Theorem 1. Combining Choquet's capacitability theorem (applied to the compact metrizable space Π), Lemma 2, Theorem 3, and Lemma 7, we obtain

$$\mathbb{P}^{\text{game}}(E) = \sup_{K\subseteq E} \mathbb{P}^{\text{game}}(K) = \sup_{K\subseteq E} \mathbb{P}^{\text{meas}}(K) = \mathbb{P}^{\text{meas}}(E),$$

K ranging over the compact sets. □

Remark 3. The fact that game-theoretic probability and measure-theoretic probability are capacities has allowed us to prove their coincidence on the analytic sets, and it might be useful for other purposes as well. In general, neither of these capacities is *strongly subadditive*, in the sense of satisfying

$$\gamma(A \cup B) + \gamma(A \cap B) \leq \gamma(A) + \gamma(B)$$

for all prequential events A and B. To demonstrate this it suffices, in view of Theorem 1, to find analytic sets A and B that violate

$$\mathbb{P}^{\text{game}}(A \cup B) + \mathbb{P}^{\text{game}}(A \cap B) \leq \mathbb{P}^{\text{game}}(A) + \mathbb{P}^{\text{game}}(B). \tag{18}$$

We can define $\mathbb{P}^{\text{game}}(E)$ for subsets of Π^n by (5) with $\limsup_n$ omitted. This is an example of subsets A and B of Π^2 for which (18) is violated:

$$A = \left\{\left(0, 0, \frac{1}{2}, 0\right), \left(\frac{1}{2}, 0, 0, 0\right)\right\}, \tag{19}$$

$$B = \left\{ \left(0, 0, \frac{1}{2}, 0\right), \left(\frac{1}{2}, 1, 0, 0\right) \right\}. \tag{20}$$

For these subsets we have

$$\mathbb{P}^{\text{game}}(A \cup B) + \mathbb{P}^{\text{game}}(A \cap B) = 1 + \frac{1}{2} > \frac{1}{2} + \frac{1}{2} = \mathbb{P}^{\text{game}}(A) + \mathbb{P}^{\text{game}}(B).$$

To obtain an example of subsets A and B of the full prequential space Π for which (18) is violated, it suffices to add $00\ldots$ at the end of each element of the sets A and B defined by (19) and (20).

7 Application to the limit theorems of probability theory

The *lower game-theoretic probability* of a prequential event E is defined to be $1 - \mathbb{P}^{\text{game}}(\Pi \setminus E)$. Similarly, the *lower measure-theoretic probability* of a prequential event E is defined to be $1 - \mathbb{P}^{\text{meas}}(\Pi \setminus E)$.

The game-theoretic strong law of large numbers (see, e.g., [16], Section 3.3) implies that (2) holds with lower game-theoretic probability one. The standard martingale strong law of large numbers implies that (2) holds with lower measure-theoretic probability one. Our Theorem 1 establishes the equivalence between these two statements. Similarly, Theorem 1 establishes the equivalence between the game-theoretic law of the iterated logarithm for binary outcomes (a special case of Theorems 5.1 and 5.2 in [16]) and the martingale law of the iterated logarithm for binary outcomes in measure-theoretic probability theory.

Transition from game-theoretic to measure-theoretic laws of probability, corresponding to the inequality $\geq$ in Theorem 1, depends only on Ville's inequality, and so can be easily done for a wide variety of prediction protocols (see, e.g., [16], Section 8.1). Transition in the opposite direction, corresponding to the inequality $\leq$, is more difficult, and its feasibility has been demonstrated only in a very limited number of cases.

In an important respect Theorem 1 is only an existence result. For example, in combination with the standard martingale strong law of large numbers in measure-theoretic probability theory it implies the game-theoretic strong law of large numbers for binary outcomes, but the resulting farthingale is very complex. The corresponding strategy for the gambler (or Skeptic, in the terminology of [16]) is also very complex. This contrasts with the simple and efficient gambling strategies designed in game-theoretic probability: see, e.g., [16], Section 3.2, and [12].

It would be interesting to design efficient general procedures producing simple gambling strategies witnessing that $\mathbb{P}^{\text{game}}(E) = 0$ for natural classes of prequential events satisfying $\mathbb{P}^{\text{meas}}(E) = 0$. For example, such a procedure might be applicable to all prequential events satisfying $\mathbb{P}^{\text{meas}}(E) = 0$ and situated at a given low level of the Borel hierarchy. This would allow an automatic procedure of transition from measure-theoretic to constructive game-theoretic laws of probability: e.g., the set of sequences (1) violating the strong law of large numbers (2) is in the class Σ_3^0 of the Borel hierarchy, and the set of sequences violating the law of the iterated logarithm is in Δ_4^0.

In this article we have only considered the case where the outcomes y_n are restricted to the binary outcome space $Y := \{0,1\}$. It is easy to extend our results to the case where Y is any finite set and Forecaster outputs probability measures on Y, interpreted as his probability forecasts for y_n. It remains an open problem whether it is possible to modify our definitions in a natural way so that the equivalence between game-theoretic and measure-theoretic probability extends to a wide classes of outcome spaces and prequential events; this would require imposing suitable measurability or topological conditions on the farthingales (or superfarthingales) used in the definition (5) of game-theoretic probability.

Acknowledgements

This article has greatly benefitted from conversations with Philip Dawid, Glenn Shafer, and Alexander Shen. Its main result answers, within the prequential framework, a question that has been asked independently by several people, including Shafer and, more recently, Shen. I am grateful to Alexander Kechris for his advice about capacities. This work was supported in part by EPSRC (grant EP/F002998/1).

References

[1] Gustave Choquet. Theory of capacities. *Annales de l'institut Fourier*, 5:131–295, 1954.

[2] Alonzo Church. On the concept of a random sequence. *Bulletin of American Mathematical Society*, 46:130–135, 1940.

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

[4] A. Philip Dawid. Statistical theory: the prequential approach. *Journal of the Royal Statistical Society A*, 147:278–292, 1984.

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

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

[7] Claude Dellacherie. *Ensembles analytiques, capacités, mesures de Hausdorff*, volume 295 of *Lecture Notes in Mathematics*. Springer, Berlin, 1972.

[8] Ryszard Engelking. *General Topology*. Heldermann, Berlin, second edition, 1989.

[9] Peter J. Huber. *Robust Statistics.* Wiley, New York, 1981.

[10] Alexander S. Kechris. *Classical Descriptive Set Theory.* Springer, New York, 1995.

[11] Andrei N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer, Berlin, 1933. English translation: *Foundations of the Theory of Probability.* Chelsea, New York, 1950.

[12] Masayuki Kumon and Akimichi Takemura. On a simple strategy weakly forcing the strong law of large numbers in the bounded forecasting game. *Annals of the Institute of Statistical Mathematics*, 60:801–812, 2008.

[13] Paul Lévy. *Théorie de l'addition des variables aléatoires.* Gauthier-Villars, Paris, 1937. Second edition: 1954.

[14] Richard von Mises. Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 5:52–99, 1919.

[15] Glenn Shafer. *The Art of Causal Conjecture.* MIT Press, Cambridge, MA, 1996.

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

[17] Glenn Shafer and Vladimir Vovk. The origins and legacy of Kolmogorov's *Grundbegriffe*. The Game-Theoretic Probability and Finance project, `http://probabilityandfinance.com`, Working Paper 4, October 2005. Part of this technical report (covering the *Grundbegriffe* and the period before its publication) appeared as [18].

[18] Glenn Shafer and Vladimir Vovk. The sources of Kolmogorov's *Grundbegriffe*. *Statistical Science*, 21:70–98, 2006.

[19] Glenn Shafer, Vladimir Vovk, and Akimichi Takemura. Lévy's zero-one law in game-theoretic probability. The Game-Theoretic Probability and Finance project, Working Paper 29, `http://probabilityandfinance.com`, `http://arxiv.org/abs/0905.0254`, May 2009.

[20] Albert N. Shiryaev. *Probability.* Springer, New York, second edition, 1996. Third Russian edition published in 2004.

[21] Jean Ville. *Etude critique de la notion de collectif.* Gauthier-Villars, Paris, 1939.

[22] 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.

[23] Vladimir Vovk and Alexander Shen. Prequential randomness. In Yoav Freund, László Györfi, György Turán, and Thomas Zeugmann, editors, *Proceedings of the Nineteenth International Conference on Algorithmic Learning Theory*, volume 5254 of *Lecture Notes in Artificial Intelligence*, pages 154–168, Berlin, 2008. Springer. Journal version submitted to the Special Issue of *Theoretical Computer Science* devoted to ALT 2008.

[24] Abraham Wald. Die Widerspruchfreiheit des Kollectivbegriffes der Wahrscheinlichkeitsrechnung. *Ergebnisse eines Mathematischen Kolloquiums*, 8:38–72, 1937.