Merging of opinions in game-theoretic probability

Vladimir Vovk
vovk@cs.rhul.ac.uk
http://vovk.net

February 1, 2008

Abstract

This paper gives game-theoretic versions of several results on "merging of opinions" obtained in measure-theoretic probability and algorithmic randomness theory. An advantage of the game-theoretic versions over the measure-theoretic results is that they are pointwise, their advantage over the algorithmic randomness results is that they are non-asymptotic, but the most important advantage over both is that they are very constructive, giving explicit and efficient strategies for players in a game of prediction.

1 Introduction

The idea that the predictions made by two forecasters will become closer with increasing information goes back at least to de Finetti (1937, Chapter V); see also Savage 1954, §3.6. De Finetti's assumption was that both forecasters compute their predictions from exchangeable probability measures that are not too close to a power probability measure; in detail he considered only binary prediction from this point of view. Later it became quite popular in Bayesian statistics since it renders an element of objectivity to the subjective probability measures.

A similar phenomenon is also known in economics under the name of "Hotelling's law" (after Hotelling 1929) or the "principle of minimum differentiation".

The first general mathematical result about convergence of predictions made by successful forecasters appears to be Blackwell and Dubins's paper (1962). Blackwell and Dubins's result was about infinite-horizon forecasting, whereas in this paper we will be interested in one-step ahead forecasting. (In game-theoretic probability, Blackwell and Dubins's setting is much less natural, since there are no stochastic assumptions imposed on what happens outside the protocol.) An important paper in this direction was Kabanov et al. (1977) (see also Shiryaev 1996, Sect. VII.6, Jacod and Shiryaev 2003, and Greenwood and Shiryaev 1985), which generalized earlier results by Kakutani (1948) and by Hájek (1958) and Feldman (1958). The main disadvantage of the approach of these papers is that it is "bulk", stated in terms of absolute continuity and singularity of probability measures.

One way to obtain pointwise results about merging of opinions is to use the algorithmic theory of randomness. The first result of this kind was proved by Dawid (1985) (who refers to it as "Jeffreys's law" in Dawid 1984 and Dawid 2004). Dawid's result was for his version of von Mises's notion of randomness based on subsequence selection rules, and so his notion of merging was rather weak. A result based on the standard notion of randomness was obtained by Vovk (1987a) and later extended by Fujiwara (2007).

The algorithmic randomness approach has two major weaknesses. First, since it is based on the notion of computability, it imposes heavy restrictions on the types of measurable spaces it can deal with (typically one considers just finite or countable observation spaces Ω). Second, it is asymptotic in the sense that it never provides us with explicit inequalities. The von Mises-type notion of randomness does not assert anything at all about finite sequences of observations. But even the modern definitions (such as those due to Martin-Löf, Levin, and Schnorr) are based on a notion (the deficiency of randomness) that is defined only to within an additive constant and so can only be applied to finite sequences *en masse*.

The game-theoretic approach to probability was suggested in Vovk (1993) and developed in, e.g., Dawid and Vovk (1999), Shafer and Vovk (2001), Kumon et al. (2007). This approach makes it possible to use all flexibility of the algorithmic randomness approach without paying its high price; in particular, all our statements are either non-asymptotic or can be stated in a non-asymptotic manner.

2 Merging of opinions as criterion of success

Let Ω be a measurable space and $\mathcal{P}(\Omega)$ stand for the set of all probability measures on Ω; elements of Ω will be called *observations* and measurable subsets of Ω will be called *local events*. Suppose we have two forecasters at each step issuing probability forecasts for the next observation $\omega_n \in \Omega$ to be chosen by reality. The game-theoretic process of testing the forecasters' predictions can be represented in the following form.

Competitive testing protocol
Players: Reality, Forecaster I, Sceptic I, Forecaster II, Sceptic II
Protocol:
$\mathcal{K}_0^{\mathrm{I}} := 1.$
$\mathcal{K}_0^{\mathrm{II}} := 1.$
FOR $n = 1, 2, \dots$:
  Forecaster I announces $P_n^{\mathrm{I}} \in \mathcal{P}(\Omega)$.
  Forecaster II announces $P_n^{\mathrm{II}} \in \mathcal{P}(\Omega)$.
  Sceptic II announces $f_n^{\mathrm{II}} : \Omega \to [0, \infty]$ such that $\int f_n^{\mathrm{II}} \mathrm{d}P_n^{\mathrm{II}} = 1$.
  Sceptic I announces $f_n^{\mathrm{I}} : \Omega \to [0, \infty]$ such that $\int f_n^{\mathrm{I}} \mathrm{d}P_n^{\mathrm{I}} = 1$.
  Reality announces $\omega_n \in \Omega$.
  $\mathcal{K}_n^{\mathrm{I}} := \mathcal{K}_{n-1}^{\mathrm{I}} f_n^{\mathrm{I}}(\omega_n)$.

$\mathcal{K}_n^{\mathrm{II}} := \mathcal{K}_{n-1}^{\mathrm{II}} f_n^{\mathrm{II}}(\omega_n)$.
END FOR

The predictions output by Forecaster I are tested by Sceptic I, and the predictions output by Forecaster II are tested by Sceptic II. The Sceptics' success in detecting inadequacy of the Forecasters' predictions is measured by their *capital*, $\mathcal{K}_n^{\mathrm{I}}$ and $\mathcal{K}_n^{\mathrm{II}}$, respectively. The initial capital is 1 and the game is fair from the point of view of the Forecasters. The value of $\mathcal{K}_n^{\mathrm{I}}$ (resp. $\mathcal{K}_n^{\mathrm{II}}$) is interpreted as the degree to which Sceptic I (resp. Sceptic II) managed to discredit Forecaster I's (resp. Forecaster II's) predictions. The requirement that the Sceptics choose functions f_n^{I} and f_n^{II} taking nonnegative values reflects the restriction that they should never risk bankruptcy by gambling more than their current capital.

In our protocol we allow infinite values for the functions chosen by the Sceptics and, therefore, infinite values for their capital. We will use the convention $0\infty := 0$.

Let $\alpha \notin \{-1, 1\}$. For two probability measures P^{I} and P^{II} on Ω we define the *α-divergence* between them as

$$D^{(\alpha)}\left(P^{\mathrm{I}} \parallel P^{\mathrm{II}}\right) := \frac{4}{1-\alpha^2}\left(1 - \int_\Omega (\beta^{\mathrm{I}}(\omega))^{\frac{1-\alpha}{2}} (\beta^{\mathrm{II}}(\omega))^{\frac{1+\alpha}{2}} Q(\mathrm{d}\omega)\right) \tag{1}$$

(with the same convention $0\infty := 0$) where Q is any probability measure on Ω such that $P^{\mathrm{I}} \ll Q$ and $P^{\mathrm{II}} \ll Q$, β^{I} is any version of the density of P^{I} w.r.t. to Q and β^{II} is any version of the density of P^{II} w.r.t. to Q. (For example, one can set $Q := (P^{\mathrm{I}} + P^{\mathrm{II}})/2$; it is clear that the value of the integral does not depend on the choice of Q, β^{I} and β^{II}.) The expression (1) is always nonnegative: see, e.g., Amari and Nagaoka (2000). An important special case is the *Hellinger distance*, corresponding to $\alpha = 0$ and also given by the formula

$$D^{(0)}\left(P^{\mathrm{I}} \parallel P^{\mathrm{II}}\right) = 2\int_\Omega \left(\sqrt{\beta^{\mathrm{I}}(\omega)} - \sqrt{\beta^{\mathrm{II}}(\omega)}\right)^2 Q(\mathrm{d}\omega).$$

(Sometimes "Hellinger distance" refers to $\frac{1}{2}D^0(P^{\mathrm{I}}, P^{\mathrm{II}})$, as in Vovk 1987a, or to $\sqrt{D^0(P^{\mathrm{I}}, P^{\mathrm{II}})}$.)

For simplicity, in the main part of this section we will only consider the case where Forecaster II is "timid" on the given play of the game, in the sense that he does not deviate too much from Forecaster I. Formally, Forecaster II is *timid* if, for all n, $P_n^{\mathrm{I}} \ll P_n^{\mathrm{II}}$ (intuitively, if Forecaster II never declares a local event null unless it is already null according to Forecaster I). It should be remembered that the assumption of timidity is always imposed on the realized play of the game rather than on Forecaster II's strategy (in general, Forecaster II is not assumed to follow a strategy).

The following asymptotic result will be proved in Sect. 4 (its counterpart in the algorithmic theory of randomness has been recently proved by Fujiwara 2007, Theorem 3; in the special case $\alpha = 0$ it was obtained in Vovk 1987a). We will say that Sceptic I (resp. Sceptic II) *becomes infinitely rich* if $\lim_{n\to\infty} \mathcal{K}_n^{\mathrm{I}} = \infty$ (resp. $\lim_{n\to\infty} \mathcal{K}_n^{\mathrm{II}} = \infty$).

Theorem 1a *Let $\alpha \in (-1, 1)$. In the competitive testing protocol:*

1. *The Sceptics have a joint strategy guaranteeing that at least one of them will become infinitely rich if*
$$\sum_{n=1}^{\infty} D^{(\alpha)} \left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}} \right) = \infty \tag{2}$$
and Forecaster II is timid.

2. *Sceptic I has a strategy guaranteeing that he will become infinitely rich if*
$$\sum_{n=1}^{\infty} D^{(\alpha)} \left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}} \right) < \infty, \tag{3}$$
Sceptic II becomes infinitely rich, and Forecaster II is timid.

Sceptic I (resp. Sceptic II) becoming infinitely rich in this theorem can also be understood as $\limsup_{n\to\infty} \mathcal{K}_n^{\mathrm{I}} = \infty$ (resp. $\limsup_{n\to\infty} \mathcal{K}_n^{\mathrm{II}} = \infty$); in the next subsection we will see that this understanding of "infinitely rich" leads to an equivalent statement.

Remark Fujiwara (2007, Section 3.1) gives a simple example showing that Theorem 1a (namely, its Part 1) cannot be extended to the case $|\alpha| \geq 1$.

Before discussing the intuition behind Theorem 1a, it will be convenient to introduce some terminology (in part informal). We will say that Forecaster I (resp. Forecaster II) is *successful* (for a particular play of the game) if Sceptic I (resp. Sceptic II) does not become infinitely rich. We say that a Forecaster is *reliable* if we believe, even before the start of the game, that he will be successful. For example, the Forecaster might know the true stochastic mechanism producing the observations, or his predictions may be computed from a well-tested theory.

Part 1 of the theorem says that either (3) holds or at least one of the Sceptics becomes infinitely rich. Therefore, if both Forecasters are reliable, we expect their predictions to be close in the sense of (3).

Suppose we only know that Forecaster I is reliable; for concreteness, let us impose on Reality the requirement that $\mathcal{K}_n^{\mathrm{I}}$ should stay bounded. If the Sceptics invest a fraction (arbitrarily small) of their initial capital in strategies whose existence is guaranteed in Theorem 1a, we will have

$$\left(\limsup_{n\to\infty} \mathcal{K}_n^{\mathrm{II}} < \infty \right) \Longleftrightarrow \left(\sum_{n=1}^{\infty} D^{(\alpha)} \left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}} \right) < \infty \right).$$

(In the context of algorithmic randomness theory this equivalence is called a criterion of randomness in Vovk 1987a and Fujiwara 2007.)

In the following sections we will see several elaborations of Theorem 1a. One useful interpretation of our results is where Forecaster I computes his predictions using some well-tested stochastic theory, and we believe him to be reliable.

Forecaster II represents an alternative way of forecasting. We will be interested in the relation between the deviation of Forecaster II's predictions from Forecaster I's predictions and the degree of the former's success, as measured by Sceptic II's capital.

2.1 Equivalence of the two senses of becoming infinitely rich

Let us first simplify the competitive testing protocol:

TESTING PROTOCOL
Players: Reality, Forecaster, Sceptic
Protocol:
$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots$:
  Forecaster announces $P_n \in \mathcal{P}(\Omega)$.
  Sceptic announces $f_n : \Omega \to [0, \infty]$ such that $\int f_n \mathrm{d}P_n = 1$.
  Reality announces $\omega_n \in \Omega$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} f_n(\omega_n)$.
END FOR

Now we have only one Forecaster and one Sceptic. We again refer to $\mathcal{K}_n$ as Sceptic's *capital* at time n.

The proof of the following lemma will give an efficient procedure transforming a strategy for Sceptic into another strategy for Sceptic such that the second strategy makes Sceptic infinitely rich in the sense $\lim \mathcal{K}_n = \infty$ whenever the first strategy makes him infinitely rich in the sense $\limsup \mathcal{K}_n = \infty$. If $\mathcal{S}$ is a strategy for Sceptic, $P_1\omega_1P_2\omega_2\ldots$ is a sequence of moves by Forecaster and Reality, and $n \in \{1, 2, \ldots\}$, let $\mathcal{K}_n(\mathcal{S}, P_1\omega_1P_2\omega_2\ldots)$ be Sceptic's capital achieved when playing $\mathcal{S}$ against Forecaster and Reality playing $P_1\omega_1P_2\omega_2\ldots$.

Lemma 1 *For any strategy $\mathcal{S}$ for Sceptic there exists another strategy $\mathcal{S}'$ for Sceptic such that, for all $P_1\omega_1P_2\omega_2\ldots$,*

$$\limsup_{n\to\infty} \mathcal{K}_n(\mathcal{S}, P_1\omega_1P_2\omega_2\ldots) = \infty \Longrightarrow \lim_{n\to\infty} \mathcal{K}_n(\mathcal{S}', P_1\omega_1P_2\omega_2\ldots) = \infty. \quad (4)$$

Proof This proof will use the argument from Vovk and Shafer (2005) (the end of the proof of Theorem 3), which we learned from Sasha Shen; for another argument, see Shafer and Vovk (2001), Lemma 3.1.

Let $\mathcal{S}$ be any strategy for Sceptic. The transformed strategy $\mathcal{S}'$ works as follows. Start playing $\mathcal{S}$ until $\mathcal{K}_n$ exceeds 2 (play $\mathcal{S}$ forever if $\mathcal{K}_n$ never exceeds 2). As soon as this happens, set 1 aside and continue playing $\mathcal{S}$ with the initial active capital $\mathcal{K}_n - 1$ until the active capital exceeds 2. As soon as this happens, set another 1 aside (decreasing the active capital by this amount) and continue playing $\mathcal{S}$ until the active capital exceeds 2, etc.

Formally, if at the beginning of some step n the capital $\mathcal{K}_{n-1}$ attained by $\mathcal{S}'$ includes active capital $\mathcal{K}^{\mathrm{act}}_{n-1}$, with $\mathcal{K}_{n-1} - \mathcal{K}^{\mathrm{act}}_{n-1}$ set aside earlier, and $\mathcal{S}$

recommends move f_n, the move recommended by $\mathcal{S}'$ is

$$f'_n := \frac{\mathcal{K}^{\mathrm{act}}_{n-1}}{\mathcal{K}_{n-1}} f_n + \frac{\mathcal{K}_{n-1} - \mathcal{K}^{\mathrm{act}}_{n-1}}{\mathcal{K}_{n-1}},$$

so that Sceptic's capital becomes $\mathcal{K}^{\mathrm{act}}_{n-1} f_n(\omega_n) + (\mathcal{K}_{n-1} - \mathcal{K}^{\mathrm{act}}_{n-1})$ at the end of step n. Now (4) follows from the fact that Sceptic will set aside another 1 infinitely often when the antecedent of (4) is satisfied. ∎

2.2 General theorem about merging of opinions

In this subsection we drop the assumption that Forecaster II is timid; unfortunately, this will make the statement of the theorem more complicated (it is easy to see that Theorem 1a itself becomes false if the assumption of Sceptic II's timidity is removed). Even the competitive testing protocol has to be modified. Let us say that $(E^{\mathrm{I}}, E^{\mathrm{II}})$, where E^{I} and E^{II} are local events, is an *exceptional pair* for $P^{\mathrm{I}}, P^{\mathrm{II}} \in \mathcal{P}(\Omega)$ if $P^{\mathrm{I}}(E^{\mathrm{I}}) = 0$, $P^{\mathrm{II}}(E^{\mathrm{II}}) = 0$, and

$$P^{\mathrm{I}}(E) = 0 \Longleftrightarrow P^{\mathrm{II}}(E) = 0$$

for all $E \subseteq \Omega \setminus (E^{\mathrm{I}} \cup E^{\mathrm{II}})$.

Modified competitive testing protocol
Players: Reality, Forecaster I, Sceptic I, Forecaster II, Sceptic II
Protocol:
$\mathcal{K}^{\mathrm{I}}_0 := 1$.
$\mathcal{K}^{\mathrm{II}}_0 := 1$.
FOR $n = 1, 2, \dots$:
  Forecaster I announces $P^{\mathrm{I}}_n \in \mathcal{P}(\Omega)$.
  Forecaster II announces $P^{\mathrm{II}}_n \in \mathcal{P}(\Omega)$.
  Reality announces an exceptional pair $(E^{\mathrm{I}}_n, E^{\mathrm{II}}_n)$ for $P^{\mathrm{I}}_n, P^{\mathrm{II}}_n$.
  Sceptic II announces $f^{\mathrm{II}}_n : \Omega \to [0, \infty]$ such that $\int f^{\mathrm{II}}_n \mathrm{d}P^{\mathrm{II}}_n = 1$.
  Sceptic I announces $f^{\mathrm{I}}_n : \Omega \to [0, \infty]$ such that $\int f^{\mathrm{I}}_n \mathrm{d}P^{\mathrm{I}}_n = 1$.
  Reality announces $\omega_n \in \Omega$.
  $\mathcal{K}^{\mathrm{I}}_n := \mathcal{K}^{\mathrm{I}}_{n-1} f^{\mathrm{I}}_n(\omega_n)$.
  $\mathcal{K}^{\mathrm{II}}_n := \mathcal{K}^{\mathrm{II}}_{n-1} f^{\mathrm{II}}_n(\omega_n)$.
END FOR

The identity of the player who announces an exceptional pair does not matter as long as it is not one of the Sceptics. One way to chose $(E^{\mathrm{I}}, E^{\mathrm{II}})$ is to choose β^{I} and β^{II} first and then set $E^{\mathrm{I}} := \{\beta^{\mathrm{I}} = 0\}$ and $E^{\mathrm{II}} := \{\beta^{\mathrm{II}} = 0\}$.

Without the condition of timidity of Forecaster II, the condition of agreement (3) between the Forecasters has to be replaced by

$$\sum_{n=1}^{\infty} D^{(\alpha)}\left(P^{\mathrm{I}}_n \parallel P^{\mathrm{II}}_n\right) < \infty \quad \text{and} \quad \forall n : \omega_n \notin E^{\mathrm{I}}_n \cup E^{\mathrm{II}}_n. \tag{5}$$

Theorem 1b *Let $\alpha \in (-1, 1)$. In the modified competitive testing protocol:*

1. *The Sceptics have a joint strategy guaranteeing that at least one of them will become infinitely rich if (5) is violated.*

2. *Sceptic I has a strategy guaranteeing that he will become infinitely rich if (5) holds and Sceptic II becomes infinitely rich.*

3 Non-asymptotic version

In many cases it will be more convenient to use the following modification of the α-divergence (1) between two probability measures P^{I} and P^{II} on Ω:

$$D^{[\alpha]}\left(P^{\mathrm{I}} \parallel P^{\mathrm{II}}\right) := \frac{4}{\alpha^2-1} \ln \int_{\Omega} (\beta^{\mathrm{I}}(\omega))^{\frac{1-\alpha}{2}} (\beta^{\mathrm{II}}(\omega))^{\frac{1+\alpha}{2}} Q(\mathrm{d}\omega), \tag{6}$$

where the constant α is different from -1 and 1; $D^{[\alpha]}\left(P^{\mathrm{I}} \parallel P^{\mathrm{II}}\right)$ will also be referred to as α-divergence. The expression (6) is nonnegative: this follows from the fact that (1) is nonnegative. When P^{I} and P^{II} are close to each other (in the sense that

$$\int_{\Omega} (\beta^{\mathrm{I}}(\omega))^{\frac{1-\alpha}{2}} (\beta^{\mathrm{II}}(\omega))^{\frac{1+\alpha}{2}} Q(\mathrm{d}\omega), \tag{7}$$

called the *Hellinger integral of order* $\frac{1-\alpha}{2}$, is close to 1), the ratio of (6) to (1) is close to 1. In any case, the inequality $\ln x \le x - 1$ (for $x \ge 0$) implies that

$$\begin{aligned} |\alpha| < 1 &\Longrightarrow D^{(\alpha)}\left(P^{\mathrm{I}} \parallel P^{\mathrm{II}}\right) \le D^{[\alpha]}\left(P^{\mathrm{I}} \parallel P^{\mathrm{II}}\right) \\ |\alpha| > 1 &\Longrightarrow D^{(\alpha)}\left(P^{\mathrm{I}} \parallel P^{\mathrm{II}}\right) \ge D^{[\alpha]}\left(P^{\mathrm{I}} \parallel P^{\mathrm{II}}\right). \end{aligned} \tag{8}$$

In principle, it is possible that $D^{(\alpha)}\left(P^{\mathrm{I}} \parallel P^{\mathrm{II}}\right) = \infty$: this happens when the Hellinger integral in (6) is zero (for $|\alpha| < 1$) or infinity (for $|\alpha| > 1$).

Remark The version (6) coincides, to within a constant factor and reparameterization, with Rényi's (1961) information gain, which in our context can be written as

$$D_{\alpha}\left(P^{\mathrm{I}}, P^{\mathrm{II}}\right) := \frac{1}{\alpha-1} \log \int_{\Omega} (\beta^{\mathrm{I}}(\omega))^{\alpha} (\beta^{\mathrm{II}}(\omega))^{1-\alpha} Q(\mathrm{d}\omega), \quad \alpha > 0, \ \alpha \ne 1, \tag{9}$$

log standing for the binary logarithm. However, we will never use the definition (9) in this paper; an important advantage of (6) is that, for any constants α_1 and α_2, the ratio of the divergences $D^{[\alpha_1]}(P^{\mathrm{I}} \parallel P^{\mathrm{II}})$ and $D^{[\alpha_2]}(P^{\mathrm{I}} \parallel P^{\mathrm{II}})$ (as well as the divergences $D^{(\alpha_1)}(P^{\mathrm{I}} \parallel P^{\mathrm{II}})$ and $D^{(\alpha_2)}(P^{\mathrm{I}} \parallel P^{\mathrm{II}})$) is close to 1 for P^{I} and P^{II} that are close to each other (in the sense of $\beta^{\mathrm{I}}/\beta^{\mathrm{II}} \approx 1$).

The main result of this section is the following non-asymptotic version of Theorems 1a and 1b.

Theorem 2 *In the competitive testing protocol:*

1. *For any $\alpha \in \mathbb{R}$, $\alpha \notin \{-1, 1\}$, the Sceptics have a joint strategy guaranteeing that, for all N,*

$$\frac{2}{1+\alpha} \ln \mathcal{K}_N^{\mathrm{I}} + \frac{2}{1-\alpha} \ln \mathcal{K}_N^{\mathrm{II}} = \sum_{n=1}^{N} D^{[\alpha]} \left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}} \right). \tag{10}$$

2. *For any $\alpha \in (-\infty, -1)$, Sceptic I has a strategy guaranteeing, for all N,*

$$\frac{2}{1+\alpha} \ln \mathcal{K}_N^{\mathrm{I}} + \frac{2}{1-\alpha} \ln \mathcal{K}_N^{\mathrm{II}} \le \sum_{n=1}^{N} D^{[\alpha]} \left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}} \right). \tag{11}$$

We regard (10) and (11) to be true if their left-hand side is an indefinite expression of the form $\infty - \infty$.

Part 1 of Theorem 2 is analogous to Part 1 of Theorem 1a. We will only be interested in the inequality "$\ge$" of (10) for $\alpha \in (-1, 1)$ (for such αs both coefficients $\frac{2}{1+\alpha}$ and $\frac{2}{1-\alpha}$ are positive). By (8) this inequality then implies

$$\frac{2}{1+\alpha} \ln \mathcal{K}_N^{\mathrm{I}} + \frac{2}{1-\alpha} \ln \mathcal{K}_N^{\mathrm{II}} \ge \sum_{n=1}^{N} D^{(\alpha)} \left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}} \right),$$

which is a precise quantification of Part 1 of Theorem 1a.

Part 2 of Theorem 2 greatly strengthens the inequality "$\le$" of (10) in the case $\alpha < -1$. Not only can this inequality be attained when the Sceptics collaborate with each other, but Sceptic I alone can enforce it, even if Sceptic II plays against him. It is close to being a quantification of Part 2 of Theorem 1a. There is, however, an essential difference between Part 2 of Theorem 2 and Part 2 of Theorem 1a: $\alpha < -1$ in the former and $\alpha \in (-1, 1)$ in the latter.

By (8), inequality (11) will continue to hold if $D^{[\alpha]}$ is replaced by $D^{(\alpha)}$. An important special case is where $\alpha = -3$ (considered in Vovk 1987a, Theorem 1); the (-3)-divergence becomes the χ^2 *distance*

$$D^{(-3)} \left(P^{\mathrm{I}} \parallel P^{\mathrm{II}} \right) = \frac{1}{2} \int_{\Omega} \frac{(\beta^{\mathrm{I}}(\omega) - \beta^{\mathrm{II}}(\omega))^2}{\beta^{\mathrm{II}}(\omega)} Q(\mathrm{d}\omega)$$

(in the notation of (1) and assuming $\beta^{\mathrm{II}} > 0$; there is no coefficient $\frac{1}{2}$ in Vovk 1987a).

In the rest of this section we will prove Theorem 2 mainly following Vovk (1987a) and Fujiwara (2007). There are, however, two important differences. First, our argument will be much more precise as compared to the $O(1)$ accuracy of the algorithmic theory of randomness. Second, we will pay careful attention to the "exceptional" cases where $\beta_n^{\mathrm{I}} = 0$ or $\beta_n^{\mathrm{II}} = 0$; this corresponds to getting rid of the assumption of local absolute continuity in measure-theoretic probability (accomplished by Pukelsheim 1986).

3.1 Proof of Part 1 of Theorem 2

Let Sceptic I play the strategy

$$f_n^{\mathrm{I}} := \frac{\left(\beta_n^{\mathrm{II}}/\beta_n^{\mathrm{I}}\right)^{\frac{1+\alpha}{2}}}{\int \left(\beta_n^{\mathrm{I}}\right)^{\frac{1-\alpha}{2}} \left(\beta_n^{\mathrm{II}}\right)^{\frac{1+\alpha}{2}} \mathrm{d}Q_n} = \left(\frac{\beta_n^{\mathrm{II}}}{\beta_n^{\mathrm{I}}}\right)^{\frac{1+\alpha}{2}} \exp\left(\frac{1-\alpha^2}{4} D^{[\alpha]}\left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right)\right) \tag{12}$$

(the Hellinger integral in the first denominator is just the normalizing constant) and Sceptic II play the strategy

$$f_n^{\mathrm{II}} := \frac{\left(\beta_n^{\mathrm{I}}/\beta_n^{\mathrm{II}}\right)^{\frac{1-\alpha}{2}}}{\int \left(\beta_n^{\mathrm{I}}\right)^{\frac{1-\alpha}{2}} \left(\beta_n^{\mathrm{II}}\right)^{\frac{1+\alpha}{2}} \mathrm{d}Q_n} = \left(\frac{\beta_n^{\mathrm{I}}}{\beta_n^{\mathrm{II}}}\right)^{\frac{1-\alpha}{2}} \exp\left(\frac{1-\alpha^2}{4} D^{[\alpha]}\left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right)\right). \tag{13}$$

From

$$\left(f_n^{\mathrm{I}}\right)^{\frac{2}{1+\alpha}} \left(f_n^{\mathrm{II}}\right)^{\frac{2}{1-\alpha}} = \exp\left(\frac{1-\alpha}{2} D^{[\alpha]}\left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right) + \frac{1+\alpha}{2} D^{[\alpha]}\left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right)\right) = \exp\left(D^{[\alpha]}\left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right)\right)$$

we now obtain (10).

Let us now look more carefully at the case where some of the denominators in (12) or (13) are zero and so the above argument is not applicable directly. If the Hellinger integral (7) at time n,

$$\int_{\Omega} \left(\beta_n^{\mathrm{I}}(\omega)\right)^{\frac{1-\alpha}{2}} \left(\beta_n^{\mathrm{II}}(\omega)\right)^{\frac{1+\alpha}{2}} Q(\mathrm{d}\omega), \tag{14}$$

is zero, the probability measures P_n^{I} and P_n^{II} are mutually singular. Choose a local event E such that $P_n^{\mathrm{I}}(E) = P_n^{\mathrm{II}}(\Omega \setminus E) = 0$. If the Sceptics choose

$$f_n^{\mathrm{I}}(\omega) := \begin{cases} \infty & \text{if } \omega \in E \\ 1 & \text{otherwise,} \end{cases} \qquad f_n^{\mathrm{II}}(\omega) := \begin{cases} 1 & \text{if } \omega \in E \\ \infty & \text{otherwise,} \end{cases}$$

(10) will be guaranteed to hold: both sides will be ∞.

Let us now suppose that the Hellinger integral (14) is non-zero. In (12) and (13), we interpret $0/0$ as 1 (and, of course, $t/0$ as ∞ for $t > 0$). As soon as the local event

$$\left(\beta_n^{\mathrm{I}} = 0 \;\&\; \beta_n^{\mathrm{II}} > 0\right) \text{ or } \left(\beta_n^{\mathrm{I}} > 0 \;\&\; \beta_n^{\mathrm{II}} = 0\right)$$

happens for the first time (if it ever happens), the Sceptics stop playing, in the sense of selecting $f^{\mathrm{I}}_N = f^{\mathrm{II}}_N := 1$ for all $N > n$. This will make sure that (10) always holds (in the sense of the convention introduced after the statement of the theorem).

3.2 Proof of Part 2 of Theorem 2

Fix $\alpha < -1$ and consider two strategies for Sceptic I: the one he played before, (12), and the strategy

$$f^{\mathrm{I}}_n = \frac{\beta^{\mathrm{II}}_n}{\beta^{\mathrm{I}}_n} f^{\mathrm{II}}_n. \tag{15}$$

Investing a fraction $c \in (0,1)$ of his initial capital of 1 in strategy (12) and investing the rest, $1-c$, in strategy (15), Sceptic I achieves a capital of

$$c \prod_{n=1}^N \left(\frac{\beta^{\mathrm{II}}_n}{\beta^{\mathrm{I}}_n}\right)^{\frac{1+\alpha}{2}} \exp\left(\frac{1-\alpha^2}{4} \sum_{n=1}^N D^{[\alpha]}\left(P^{\mathrm{I}}_n \parallel P^{\mathrm{II}}_n\right)\right) + (1-c)\mathcal{K}^{\mathrm{II}}_N \prod_{n=1}^N \frac{\beta^{\mathrm{II}}_n}{\beta^{\mathrm{I}}_n}. \tag{16}$$

To get rid of the likelihood ratio $x := \prod_{n=1}^N (\beta^{\mathrm{II}}_n / \beta^{\mathrm{I}}_n)$, we bound (16) from below by

$$\begin{align}
&\inf_{x>0} \left(c x^{\frac{1+\alpha}{2}} \exp\left(\frac{1-\alpha^2}{4} \sum_{n=1}^N D^{[\alpha]}\left(P^{\mathrm{I}}_n \parallel P^{\mathrm{II}}_n\right)\right) + (1-c)\mathcal{K}^{\mathrm{II}}_N x \right) \tag{17}\\
&= \left(\left(\frac{-1-\alpha}{2}\right)^{\frac{2}{1-\alpha}} + \left(\frac{2}{-1-\alpha}\right)^{-\frac{1+\alpha}{1-\alpha}} \right) c^{\frac{2}{1-\alpha}} (1-c)^{-\frac{1+\alpha}{1-\alpha}} \tag{18}\\
&\times \exp\left(\frac{1+\alpha}{2} \sum_{n=1}^N D^{[\alpha]}\left(P^{\mathrm{I}}_n \parallel P^{\mathrm{II}}_n\right)\right) \left(\mathcal{K}^{\mathrm{II}}_N\right)^{-\frac{1+\alpha}{1-\alpha}}. \tag{19}
\end{align}$$

To optimize this lower bound, we find c such that

$$c^{\frac{2}{1-\alpha}} (1-c)^{-\frac{1+\alpha}{1-\alpha}} \to \max,$$

which gives

$$c = \frac{2}{1-\alpha}.$$

For this value of c, the expression on line (18) equals

$$\left(\left(\frac{-1-\alpha}{2}\right)^{\frac{2}{1-\alpha}} + \left(\frac{2}{-1-\alpha}\right)^{-\frac{1+\alpha}{1-\alpha}} \right) \left(\frac{2}{1-\alpha}\right)^{\frac{2}{1-\alpha}} \left(\frac{-1-\alpha}{1-\alpha}\right)^{-\frac{1+\alpha}{1-\alpha}} = 1.$$

In combination with (19) this gives

$$\mathcal{K}^{\mathrm{I}}_N \ge \exp\left(\frac{1+\alpha}{2} \sum_{n=1}^N D^{[\alpha]}\left(P^{\mathrm{I}}_n \parallel P^{\mathrm{II}}_n\right)\right) \left(\mathcal{K}^{\mathrm{II}}_N\right)^{-\frac{1+\alpha}{1-\alpha}}.$$

Taking the logarithm of both sides and regrouping, we obtain (11).

It remains to consider the exceptional cases. If the Hellinger integral (14) is infinite (this happens when P_n^{I} is not absolutely continuous w.r. to P_n^{II}), the right-hand side of (11) is also infinite, and there is nothing to prove. Suppose (14) is finite. When $\beta_n^{\mathrm{I}} = 0$, we will interpret the expressions (12) and (15) as ∞ (this will not affect $\int f_n^{\mathrm{I}} \mathrm{d}P_n^{\mathrm{I}}$; in principle, it is now possible that $\int f_n^{\mathrm{I}} \mathrm{d}P_n^{\mathrm{I}} < 1$, but this can only hurt Sceptic I). When $\beta_n^{\mathrm{I}} > 0$ & $\beta_n^{\mathrm{II}} = 0$, we, naturally, interpret (12) as ∞. In both cases Sceptic I's capital becomes infinite when $\beta_n^{\mathrm{I}} = 0$ or $\beta_n^{\mathrm{II}} = 0$, and (11) still holds.

4 Proof of Theorems 1a and 1b

Part 1 of Theorems 1a and 1b immediately follows from Part 1 of Theorem 2 (in the case of Theorem 1b, the Forecasters' moves f_n^{I} and f_n^{II} have to be slightly redefined by setting them to ∞ on E^{I} and E^{II}, respectively). Therefore, in this section we will only prove Part 2 of Theorems 1a and 1b. Instead of deducing this result from Part 2 of Theorem 2 (as was done in Vovk 1987a and Fujiwara 2007), we will prove it using methods of the theory of martingales and adapting the proof given in Shiryaev (1996, Theorem VII.6.4).

The following lemma from Fujiwara (2007) shows that all α-divergences, $\alpha \in (-1, 1)$, coincide to within a constant factor. We will be using the words "increasing" and "decreasing" in the wide sense (e.g., a constant function qualifies as both increasing and decreasing).

Lemma 2 *Let $P', P'' \in \mathcal{P}(\Omega)$. The function $(1-\alpha)D^{(\alpha)}(P', P'')$ is decreasing in $\alpha \in (-1, 1)$. The function $(1+\alpha)D^{(\alpha)}(P', P'')$ is increasing in $\alpha \in (-1, 1)$.*

Proof See the proof of Lemma 10 in Fujiwara (2007, Appendix B). ∎

Therefore, we could restrict ourselves only to the Hellinger distance. We will, however, consider the general case (as Fujiwara 2007).

We will be concerned with the following slight modification of the testing protocol.

Semimartingale protocol (multiplicative representation)
Players: Reality, Forecaster, Sceptic
Protocol:
$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \dots$:
  Forecaster announces $P_n \in \mathcal{P}(\Omega)$.
  Reality announces measurable $\xi_n : \Omega \to \mathbb{R}$.
  Sceptic announces $f_n : \Omega \to [0, \infty]$ such that $\int f_n \mathrm{d}P_n = 1$.
  Reality announces $\omega_n \in \Omega$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} f_n(\omega_n)$.
END FOR

The *martingale protocol* (resp. *submartingale protocol*, *supermartingale protocol*) differs from the semimartingale protocol in that Reality is required to ensure that, for all n, the function ξ_n is P_n-integrable and $\int \xi_n \mathrm{d}P_n$ is zero (resp. nonnegative, nonpositive).

An *event* is a property of the play $(P_n, \xi_n, f_n, \omega_n)_{n=1}^{\infty}$. We say that Sceptic *can force* an event E if he has a strategy guaranteeing that either E holds or $\mathcal{K}_n \to \infty$ as $n \to \infty$. Lemma 1 shows that replacing $\mathcal{K}_n \to \infty$ by $\limsup_n \mathcal{K}_n = \infty$ gives an equivalent definition.

Another representation of the semimartingale protocol is:

SEMIMARTINGALE PROTOCOL (ADDITIVE REPRESENTATION)
Players: Reality, Forecaster, Sceptic
Protocol:
$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \dots$:
  Forecaster announces $P_n \in \mathcal{P}(\Omega)$.
  Reality announces measurable $\xi_n : \Omega \to \mathbb{R}$.
  Sceptic announces $g_n : \Omega \to [-\mathcal{K}_{n-1}, \infty]$ such that $\int g_n \mathrm{d}P_n = 0$.
  Reality announces $\omega_n \in \Omega$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} + g_n(\omega_n)$.
END FOR

The correspondence between the two representations of the semimartingale protocol is given by $g_n = (f_n - 1)\mathcal{K}_{n-1}$. We will switch at will between the two representations.

If Sceptic follows a strategy $\mathcal{S}$, we will let $\mathcal{K}_n^{\mathcal{S}}$ stand for his capital at the end of step n (as a function of Forecaster's and Reality's moves). The additive representation makes it obvious that Sceptic's strategies can be mixed:

Lemma 3 *If $\mathcal{S}_1, \mathcal{S}_2, \dots$ is a sequence of strategies for Sceptic and $p_1, p_2, \dots$ is a sequence of positive weights summing to 1, there is a "master" strategy $\mathcal{S}$ for Sceptic ensuring*

$$\mathcal{K}_n^{\mathcal{S}} = \sum_{k=1}^{\infty} p_k \mathcal{K}_n^{\mathcal{S}_k}.$$

Proof It suffices for Sceptic to set $g_n := \sum_{k=1}^{\infty} p_k g_{k,n}$ at each step n, where $g_{k,n}$ is the move recommended by $\mathcal{S}_k$. ∎

Lemma 4 *In the martingale protocol, Sceptic can force*

$$\sum_{n=1}^{\infty} \int \xi_n^2 \mathrm{d}P_n < \infty \Longrightarrow \sup_N \sum_{n=1}^{N} \xi_n(\omega_n) < \infty.$$

Proof It is easy to see that for each $C > 0$ there is a strategy, say $\mathcal{S}_C$, for Sceptic leading to capital

$$\mathcal{K}_N^{\mathcal{S}_C} = \begin{cases} 1 + \frac{1}{C}\left(\left(\sum_{n=1}^{N} \xi_n(\omega_n)\right)^2 - \sum_{n=1}^{N} \int \xi_n^2 \mathrm{d}P_n\right) & \text{if } \sum_{n=1}^{N} \int \xi_n^2 \mathrm{d}P_n \le C \\ \mathcal{K}_{N-1}^{\mathcal{S}_C} & \text{otherwise,} \end{cases}$$

for all $N = 1, 2, \ldots$. It remains to mix all $\mathcal{S}_C$, $C = 1, 2, \ldots$, according to Lemma 3 (with arbitrary positive weights). ∎

Lemma 5 *In the submartingale protocol, Sceptic can force*

$$\sum_{n=1}^{\infty} \left(\int \xi_n \mathrm{d}P_n + \int \xi_n^2 \mathrm{d}P_n \right) < \infty \Longrightarrow \sup_N \sum_{n=1}^{N} \xi_n(\omega_n) < \infty.$$

Proof It suffices to apply Lemma 4 to $\tilde{\xi}_n := \xi_n - \int \xi_n \mathrm{d}P_n$ (notice that $\int \tilde{\xi}_n^2 \mathrm{d}P_n \le \int \xi_n^2 \mathrm{d}P_n$). ∎

Let $\mathbb{I}_E$ stand for the indicator of local event $E \subseteq \Omega$:

$$\mathbb{I}_E(\omega) := \begin{cases} 1 & \text{if } \omega \in E \\ 0 & \text{otherwise.} \end{cases}$$

The following lemma is a version of the Borel–Cantelli–Lévy lemma.

Lemma 6 *In the submartingale protocol, Sceptic can force*

$$\left(\forall n : \xi_n = \mathbb{I}_{E_n} \ \& \ \sum_{n=1}^{\infty} P_n(E_n) < \infty \right) \Longrightarrow (\omega_n \in E_n \textit{ for finitely many } n).$$

Proof This is a special case of Lemma 5. ∎

Our application of the Borel–Cantelli–Lévy lemma will be made possible by the following lemma, which we state using the notation introduced earlier in (1).

Lemma 7 *For each $\alpha \in (-1, 1)$ there exists a constant $C = C(\alpha)$ such that, for all P^{I} and P^{II},*

$$P^{\mathrm{I}} \left\{ \beta^{\mathrm{I}} > e\beta^{\mathrm{II}} \right\} \le C D^{(\alpha)} \left(P^{\mathrm{I}} \parallel P^{\mathrm{II}} \right).$$

Proof Let E stand for the event $\left\{ \beta^{\mathrm{I}} > e\beta^{\mathrm{II}} \right\}$. Since

$$\begin{aligned}
D^{(\alpha)} \left(P^{\mathrm{I}} \parallel P^{\mathrm{II}} \right) &= \frac{4}{1-\alpha^2} \left(1 - \int_\Omega (\beta^{\mathrm{I}}(\omega))^{\frac{1-\alpha}{2}} (\beta^{\mathrm{II}}(\omega))^{\frac{1+\alpha}{2}} Q(\mathrm{d}\omega) \right) \\
&= \frac{4}{1-\alpha^2} \int_\Omega \frac{1-\alpha}{2} \beta^{\mathrm{I}}(\omega) + \frac{1+\alpha}{2} \beta^{\mathrm{II}}(\omega) - (\beta^{\mathrm{I}}(\omega))^{\frac{1-\alpha}{2}} (\beta^{\mathrm{II}}(\omega))^{\frac{1+\alpha}{2}} Q(\mathrm{d}\omega) \\
&\ge \frac{4}{1-\alpha^2} \int_E \frac{1-\alpha}{2} \beta^{\mathrm{I}}(\omega) + \frac{1+\alpha}{2} \beta^{\mathrm{II}}(\omega) - (\beta^{\mathrm{I}}(\omega))^{\frac{1-\alpha}{2}} (\beta^{\mathrm{II}}(\omega))^{\frac{1+\alpha}{2}} Q(\mathrm{d}\omega) \\
&\ge \frac{4}{1-\alpha^2} \int_E \frac{1-\alpha}{2} \beta^{\mathrm{I}}(\omega) + \frac{1+\alpha}{2e} \beta^{\mathrm{I}}(\omega) - \left(\beta^{\mathrm{I}}(\omega)\right)^{\frac{1-\alpha}{2}} \left(\frac{\beta^{\mathrm{I}}(\omega)}{e} \right)^{\frac{1+\alpha}{2}} Q(\mathrm{d}\omega) \\
&= \frac{4}{1-\alpha^2} \left(\frac{1-\alpha}{2} + \frac{1+\alpha}{2e} - e^{-\frac{1+\alpha}{2}} \right) P^{\mathrm{I}}(E)
\end{aligned}$$

(the first inequality follows from the fact that the geometric mean never exceeds the arithmetic mean, and the second inequality uses the fact that for each $\beta > 0$ the function

$$\frac{1-\alpha}{2}\beta + \frac{1+\alpha}{2}x - \beta^{\frac{1-\alpha}{2}} x^{\frac{1+\alpha}{2}}$$

is decreasing in $x \in [0, \beta/e]$, which can be checked by differentiation), we can set

$$C := \frac{1-\alpha^2}{4}\left(\frac{1-\alpha}{2} + \frac{1+\alpha}{2e} - e^{-\frac{1+\alpha}{2}}\right)^{-1} > 0$$

(the expression in the parentheses is a positive function of $\alpha \in (-1,1)$ since it takes value 0 at $\alpha = -1$ and $\alpha = 1$ and the function is strictly concave). ∎

We will also need the following elementary inequality, where $U : \mathbb{R} \to \mathbb{R}$ is the truncation function

$$U(x) := \begin{cases} x & \text{if } x \le 1 \\ 1 & \text{otherwise.} \end{cases}$$

Lemma 8 *For each $\gamma \in (0,1)$ there exists $B > 1$ such that, for all $x > 0$,*

$$xU(\ln x) + xU^2(\ln x) \le B(x-1) + \frac{B-1}{\gamma}\left(1 - x^{\gamma}\right). \tag{20}$$

Proof Let us first consider the case $x \le e$. To see that

$$x \ln x + x \ln^2 x \le B(x-1) + \frac{B-1}{\gamma}\left(1 - x^{\gamma}\right)$$

notice that the values and the first derivatives of the two sides of this inequality coincide when $x = 1$ (in fact, the coefficient $\frac{B-1}{\gamma}$ was chosen to match the derivatives) and that the inequality for the second derivatives,

$$\frac{3 + 2\ln x}{x} \le \frac{(B-1)(1-\gamma)}{x^{2-\gamma}},$$

holds when B is sufficiently large (namely, when $B \ge 5e^{1-\gamma}/(1-\gamma) + 1$).

In the case $x \ge e$, the inequality becomes

$$2x \le B(x-1) + \frac{B-1}{\gamma}\left(1 - x^{\gamma}\right);$$

since it is true for $x = e$ (by the previous paragraph), it suffices to make sure that the inequality between the derivatives of the two sides holds:

$$2 \le B - (B-1)x^{\gamma-1}.$$

This can be achieved by making $B \ge (2e - e^{\gamma})/(e - e^{\gamma}) > 2$. ∎

Now we have all we need to prove Part 2 of Theorem 1b. Let us first assume that the functions β_n^{I} and β_n^{II} are always positive and that $E_n^{\mathrm{I}} = E_n^{\mathrm{II}} = \emptyset$, for all n. Our goal is to prove that Sceptic I can force

$$\sum_{n=1}^{\infty} D^{(\alpha)}\left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right) < \infty \Longrightarrow \liminf_{n\to\infty} \mathcal{K}_n^{\mathrm{II}} < \infty.$$

Substituting

$$\gamma := \frac{1-\alpha}{2}, \quad x := \frac{\beta_n^{\mathrm{I}}}{\beta_n^{\mathrm{II}}}$$

in (20), multiplying by β_n^{II}, integrating over Q_n, and summing over $n = 1, 2, \ldots$, we obtain

$$\sum_{n=1}^{\infty} \int \beta_n^{\mathrm{I}} U\left(\ln \frac{\beta_n^{\mathrm{I}}}{\beta_n^{\mathrm{II}}}\right) + \beta_n^{\mathrm{I}} U^2\left(\ln \frac{\beta_n^{\mathrm{I}}}{\beta_n^{\mathrm{II}}}\right) \mathrm{d}Q_n \\ \le \frac{(B-1)(1+\alpha)}{2} \sum_{n=1}^{\infty} D^{(\alpha)}\left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right). \quad (21)$$

Let us combine this inequality with:

- Lemma 5 applied to Sceptic I and $\xi_n := U\left(\ln \frac{\beta_n^{\mathrm{I}}}{\beta_n^{\mathrm{II}}}\right)$. It is applicable because the inequality $xU(\ln x) \ge x - 1$, valid for all $x > 0$, implies

  $$\int \xi_n \mathrm{d}P_n^{\mathrm{I}} = \int \frac{\beta_n^{\mathrm{I}}}{\beta_n^{\mathrm{II}}} U\left(\ln \frac{\beta_n^{\mathrm{I}}}{\beta_n^{\mathrm{II}}}\right) \beta_n^{\mathrm{II}} \mathrm{d}Q_n \ge \int \left(\frac{\beta_n^{\mathrm{I}}}{\beta_n^{\mathrm{II}}} - 1\right) \beta_n^{\mathrm{II}} \mathrm{d}Q_n = 0$$

  (and ξ_n is P_n^{I}-integrable since $\xi_n \le 1$).
- Lemma 6 applied to Sceptic I and $E_n := \{\beta_n^{\mathrm{I}} > e\beta_n^{\mathrm{II}}\}$. It will be applicable by Lemma 7.

We can now see that it suffices to prove that Sceptic I can force

$$\left(\sup_N \sum_{n=1}^{N} U\left(\ln \frac{\beta_n^{\mathrm{I}}(\omega_n)}{\beta_n^{\mathrm{II}}(\omega_n)}\right) < \infty \;\&\; \beta_n^{\mathrm{I}}(\omega_n) \le e\beta_n^{\mathrm{II}}(\omega_n) \text{ from some } n \text{ on}\right) \\ \Longrightarrow \liminf_{n\to\infty} \mathcal{K}_n^{\mathrm{II}} < \infty. \quad (22)$$

Forcing (22) can be achieved by forcing

$$\sup_N \sum_{n=1}^{N} \ln \frac{\beta_n^{\mathrm{I}}(\omega_n)}{\beta_n^{\mathrm{II}}(\omega_n)} < \infty \Longrightarrow \liminf_{n\to\infty} \mathcal{K}_n^{\mathrm{II}} < \infty, \quad (23)$$

and the latter can be done with the strategy

$$f_n^{\mathrm{I}} := \frac{\beta_n^{\mathrm{II}}}{\beta_n^{\mathrm{I}}} f_n^{\mathrm{II}}.$$

It remains to get rid of the assumption $\beta_n^{\mathrm{I}} > 0$, $\beta_n^{\mathrm{II}} > 0$, $E_n^{\mathrm{I}} = E_n^{\mathrm{II}} = \emptyset$. Our argument so far remains valid if the observation space $\Omega = \Omega_n$ is allowed to depend on n (it can be chosen by Sceptic I's opponents at any time prior to his move). In particular, we can set $\Omega_n := \Omega \setminus (E_n^{\mathrm{I}} \cup E_n^{\mathrm{II}})$ and assume, without loss of generality, that $\beta_n^{\mathrm{I}} > 0$ and $\beta_n^{\mathrm{II}} > 0$ on Ω_n. This proves Part 2 of Theorem 1b.

To deduce Part 2 of Theorem 1a from Part 2 of Theorem 1b, notice that, for all n, $P_n^{\mathrm{I}}(E_n^{\mathrm{II}}) = 0$ (as $P_n^{\mathrm{I}} \ll P_n^{\mathrm{II}}$). Therefore, $P_n^{\mathrm{I}}(E_n^{\mathrm{I}} \cup E_n^{\mathrm{II}}) = 0$, and mixing (in the sense of Lemma 3) any of the strategies for Sceptic I whose existence is asserted in Part 2 of Theorem 1b with the strategy

$$f_n^{\mathrm{I}}(\omega) := \begin{cases} \infty & \text{if } \omega \in E_n^{\mathrm{I}} \cup E_n^{\mathrm{II}} \\ 1 & \text{otherwise} \end{cases}$$

we obtain a strategy for Sceptic I satisfying the condition of Part 2 of Theorem 1a.

Remark Part 2 of Theorem 2 gives a precise lower bound for the rate of growth of Sceptic I's capital in terms of the rate of growth of Sceptic II's capital and the cumulative divergence between the Forecasters' predictions. Unfortunately, our proof of Part 2 of Theorems 1a and 1b does not give such a bound; the step that prevents us from obtaining such a bound is replacing (22) with (23) (another such step would have been the use of Doob's martingale convergence theorem, as in Shiryaev 1996, but this proof avoids it). It remains an open problem whether such a bound exists.

5 The rate of growth of Sceptic II's capital in terms of the Kullback–Leibler divergence

Suppose Forecaster I is reliable and Sceptic I invests a part of his capital in strategies whose existence is asserted in Theorem 2. That theorem then gives rather precise bounds for the achievable rate of growth of $\mathcal{K}_n^{\mathrm{II}}$: Sceptic II can achieve

$$\ln \mathcal{K}_N^{\mathrm{II}} \geq \frac{1-\alpha}{2} \sum_{n=1}^{N} D^{[\alpha]}\left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right) - O(1) \tag{24}$$

but cannot achieve more than that:

$$\ln \mathcal{K}_N^{\mathrm{II}} \leq \frac{1-\alpha}{2} \sum_{n=1}^{N} D^{[\alpha]}\left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right) + O(1). \tag{25}$$

The problem with this is that (24) and (25) refer to different αs: $\alpha \in (-1, 1)$ in (24) and $\alpha < -1$ in (25). In this section we will derive similar (albeit cruder) bounds for the same α-divergence, namely for $\alpha = -1$, one of the two cases that have been excluded so far. The (-1)-*divergence*, or the *Kullback–Leibler divergence*, is defined by

$$D^{(-1)}\left(P^{\mathrm{I}} \parallel P^{\mathrm{II}}\right) := \int_{\Omega} \ln \frac{\beta^{\mathrm{I}}(\omega)}{\beta^{\mathrm{II}}(\omega)} P^{\mathrm{I}}(\mathrm{d}\omega),$$

using the notation of (1). Somewhat related results have been obtained in algorithmic randomness theory by Solomonoff (1978); see also Vovk (1989), Theorem 2.2.

Remark It might be tempting to set $\alpha < -1$ in (10). This will not lead to any useful bounds because of the possibility $\mathcal{K}_n^{\mathrm{I}} \to 0$.

For simplicity, we will again impose the assumption of "timidity" (in a much stronger sense than before) on Forecaster II: on the given play of the game, his predictions should stay within a constant factor of the reliable Forecaster I. More precisely, Forecaster II is *c-timid*, for a constant $c > 1$, if for all n the ratio $\beta_n^{\mathrm{II}}/\beta_n^{\mathrm{I}}$ (with $0/0$ interpreted as 1) is bounded above by c and bounded below by $1/c$. The value of the constant c is not disclosed to the players, and the strategies for the Sceptics constructed in this section will never depend on c.

Let x^+ stand for $\max(x, 0)$.

Theorem 3 *Let $c > 1$. There is a constant $C > 0$ depending only on c such that:*

1. *The Sceptics have a joint strategy in the competitive testing protocol that guarantees, starting from $N = 3$,*

$$\ln \mathcal{K}_N^{\mathrm{II}} \geq \sum_{n=1}^{N} D^{(-1)}\left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right) - C\sqrt{\frac{N}{\ln\ln N}}\left(\ln^+ \mathcal{K}_N^{\mathrm{I}} + \ln\ln N\right) \tag{26}$$

   on the plays where Forecaster II is c-timid.

2. *Sceptic I has a strategy that guarantees, starting from $N = 3$,*

$$\ln \mathcal{K}_N^{\mathrm{II}} \leq \sum_{n=1}^{N} D^{(-1)}\left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right) + C\sqrt{\frac{N}{\ln\ln N}}\left(\ln^+ \mathcal{K}_N^{\mathrm{I}} + \ln\ln N\right) \tag{27}$$

   on the plays where Forecaster II is c-timid.

Before proving Theorem 3, we will state a similar result for the case where the duration of the game, N, is known in advance. (Theorem 3 is applicable in this case as well, but it can be made more precise.)

Proposition 1 *Let $c > 1$. There is a constant $C > 0$ depending only on c such that:*

1. *For each $N \in \{2,3,\dots\}$, the Sceptics have a joint strategy that guarantees*

$$\ln \mathcal{K}_N^{\mathrm{II}} \geq \sum_{n=1}^{N} D^{(-1)}\left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right) - \left(\sqrt{N}-1\right)\left(\ln \mathcal{K}_N^{\mathrm{I}} + C\right) \tag{28}$$

on the plays where Forecaster II is c-timid.

2. *For each $N \in \{1,2,\dots\}$, Sceptic I has a strategy that guarantees*

$$\ln \mathcal{K}_N^{\mathrm{II}} \leq \sum_{n=1}^{N} D^{(-1)}\left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right) + \left(\sqrt{N}+1\right)\left(\ln \mathcal{K}_N^{\mathrm{I}} + C\right) \tag{29}$$

on the plays where Forecaster II is c-timid.

The intuition behind Theorem 3 and Proposition 1 is that Sceptic II can achieve the growth rate of his logarithmic capital close to the growth rate of the cumulative Kullback–Leibler divergence between P_n^{I} and P_n^{II}, but cannot achieve a better growth rate. Theorem 3 is related to the law of the iterated logarithm, in that it gives the accuracy of $O(\sqrt{N \ln\ln N})$ for a reliable Forecaster I, whereas Proposition 1 is related to the central limit theorem, in that it gives the accuracy of $O(\sqrt{N})$.

5.1 Proof of Part 1 of Theorem 3 and Proposition 1

Substituting $\alpha := -1 + 2\epsilon$, with $\epsilon \in (0,1)$, in (10) gives

$$\frac{1}{\epsilon} \ln \mathcal{K}_N^{\mathrm{I}} + \frac{1}{1-\epsilon} \ln \mathcal{K}_N^{\mathrm{II}} = \sum_{n=1}^{N} D^{[-1+2\epsilon]}\left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right),$$

which can be rewritten as

$$\ln \mathcal{K}_N^{\mathrm{II}} = -\frac{1}{\epsilon} \sum_{n=1}^{N} \ln \int \left(\frac{\beta_n^{\mathrm{II}}}{\beta_n^{\mathrm{I}}}\right)^{\epsilon} \mathrm{d}P_n^{\mathrm{I}} - \frac{1-\epsilon}{\epsilon} \ln \mathcal{K}_N^{\mathrm{I}}. \tag{30}$$

The inequality $e^x \leq 1 + x + \frac{x^2}{2} e^{x^+}$ gives

$$\begin{aligned}
\left(\frac{\beta_n^{\mathrm{II}}}{\beta_n^{\mathrm{I}}}\right)^{\epsilon} &= \exp\left(\epsilon \ln \frac{\beta_n^{\mathrm{II}}}{\beta_n^{\mathrm{I}}}\right) \\
&\leq 1 + \epsilon \ln \frac{\beta_n^{\mathrm{II}}}{\beta_n^{\mathrm{I}}} + \frac{1}{2}\epsilon^2 \ln^2 \frac{\beta_n^{\mathrm{II}}}{\beta_n^{\mathrm{I}}} \exp\left(\epsilon \ln^+ \frac{\beta_n^{\mathrm{II}}}{\beta_n^{\mathrm{I}}}\right) \\
&\leq 1 + \epsilon \ln \frac{\beta_n^{\mathrm{II}}}{\beta_n^{\mathrm{I}}} + \frac{1}{2}\epsilon^2 c^{\epsilon} \ln^2 c.
\end{aligned}$$

This further implies

$$
\begin{aligned}
-\frac{1}{\epsilon}\sum_{n=1}^{N}\ln\int\left(\frac{\beta_n^{\mathrm{II}}}{\beta_n^{\mathrm{I}}}\right)^{\epsilon}\mathrm{d}P_n^{\mathrm{I}}&\\
&\ge -\frac{1}{\epsilon}\sum_{n=1}^{N}\ln\left(1-\epsilon D^{(-1)}\left(P_n^{\mathrm{I}}\parallel P_n^{\mathrm{II}}\right)+\frac{1}{2}\epsilon^2 c^{\epsilon}\ln^2 c\right) \qquad (31)\\
&\ge -\frac{1}{\epsilon}\sum_{n=1}^{N}\left(-\epsilon D^{(-1)}\left(P_n^{\mathrm{I}}\parallel P_n^{\mathrm{II}}\right)+\frac{1}{2}\epsilon^2 c^{\epsilon}\ln^2 c\right) \qquad (32)\\
&= \sum_{n=1}^{N} D^{(-1)}\left(P_n^{\mathrm{I}}\parallel P_n^{\mathrm{II}}\right)-\frac{1}{2}N\epsilon c^{\epsilon}\ln^2 c
\end{aligned}
$$

(the transition from (31) to (32) uses the inequality $\ln x \le x-1$, valid for $x \ge 0$; the expression in the parentheses in (31) is nonnegative because of its provenance). Plugging the last inequality into (30), we obtain

$$
\ln\mathcal{K}_N^{\mathrm{II}} \ge \sum_{n=1}^{N} D^{(-1)}\left(P_n^{\mathrm{I}}\parallel P_n^{\mathrm{II}}\right)-\frac{1}{2}N\epsilon c^{\epsilon}\ln^2 c-\frac{1-\epsilon}{\epsilon}\ln\mathcal{K}_N^{\mathrm{I}}. \qquad (33)
$$

If N is known in advance (but c and $\mathcal{K}_N^{\mathrm{I}}$ are not), we can set $\epsilon := N^{-1/2}$ in (33), which gives

$$
\ln\mathcal{K}_N^{\mathrm{II}} \ge \sum_{n=1}^{N} D^{(-1)}\left(P_n^{\mathrm{I}}\parallel P_n^{\mathrm{II}}\right)-\frac{1}{2}\sqrt{N}c^{N^{-1/2}}\ln^2 c-\left(\sqrt{N}-1\right)\ln\mathcal{K}_N^{\mathrm{I}};
$$

this proves (28) with, e.g., $C := 2c\ln^2 c$ (in fact, we can take C arbitrarily close to $\frac{1}{2}\ln^2 c$ if we only want (28) to hold for sufficiently large N).

In the case of Theorem 3, where N is unknown, we will use the discrete form (Vovk 1987b, Theorem 1) of Ville's (1939) method of proving the law of the iterated logarithm (however, without worrying about constant factors). The method is based on the following simple corollary of Lemma 3. If $\mathcal{S}_2, \mathcal{S}_3, \dots$ is a sequence of strategies for Sceptic and $p_2, p_3, \dots$ is a sequence of positive weights summing to 1, there is a strategy $\mathcal{S}$ for Sceptic ensuring $\ln\mathcal{K}_N^{\mathcal{S}} \ge \ln\mathcal{K}_N^{\mathcal{S}_k}+\ln p_k$ for all $k = 2, 3, \dots$. In particular, taking $p_k \propto k^{-2}$, $k = 2, 3, \dots$, we obtain $\ln\mathcal{K}_N^{\mathcal{S}} \ge \ln\mathcal{K}_N^{\mathcal{S}_k} - 2\ln k$ for $k \ge 2$.

Let $\epsilon_2, \epsilon_3, \dots$ be a sequence of positive numbers, to be specified later on. Since for each $k = 2, 3, \dots$ the Sceptics can ensure (33) for $\epsilon := \epsilon_k$ (using simple strategies (12)–(13) depending only on the Forecasters' predictions), they can also ensure, for all $k = 2, 3, \dots$,

$$
\ln\mathcal{K}_N^{\mathrm{II}} + 2\ln k \ge \sum_{n=1}^{N} D^{(-1)}\left(P_n^{\mathrm{I}}\parallel P_n^{\mathrm{II}}\right) - C_1 N\epsilon_k - \frac{1-\epsilon_k}{\epsilon_k}\left(\ln\mathcal{K}_N^{\mathrm{I}}+2\ln k\right) \qquad (34)
$$

where C_1 (as well as C_2 to C_6 below) is a constant depending only on c. Weakening (34) to

$$\ln \mathcal{K}_N^{\mathrm{II}} \geq \sum_{n=1}^N D^{(-1)} \left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right) - C_1 N \epsilon_k - \frac{1}{\epsilon_k} \ln \mathcal{K}_N^{\mathrm{I}} - \frac{2 \ln k}{\epsilon_k}$$

and setting

$$k := \lceil \ln N \rceil, \quad \epsilon_k := \sqrt{\frac{\ln k}{e^k}} \tag{35}$$

(so that ϵ_k coincides with $\sqrt{\ln \ln N / N}$ to within a constant factor), we further obtain

$$\ln \mathcal{K}_N^{\mathrm{II}} \geq \sum_{n=1}^N D^{(-1)} \left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right) - C_2 \sqrt{N \ln \ln N} - C_3 \sqrt{\frac{N}{\ln \ln N}} \ln^+ \mathcal{K}_N^{\mathrm{I}}$$

for $N \geq 3$. This essentially coincides with (26).

Remark The strategies for Sceptic II constructed in our proof of Part 1 of Theorem 3 and Proposition 1 was somewhat complex, especially in the case of Theorem 3. In the spirit of Solomonoff (1978), we could take the simple "likelihood ratio" strategy $f_n^{\mathrm{II}} := \beta_n^{\mathrm{I}} / \beta_n^{\mathrm{II}}$. The expected value of $\ln f_n^{\mathrm{II}}$ with respect to P_n^{I} is $D^{(-1)}(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}})$, and according to standard results of game-theoretic probability Sceptic I can become infinitely rich unless $\ln \mathcal{K}_N^{\mathrm{II}} \approx \sum_{n=1}^N D^{(-1)} \left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right)$. The law of the iterated logarithm (see, e.g., Shafer and Vovk 2001, Chapter 5) will give a result similar to Part 1 of Theorem 3, and the weak law of large numbers (in the form of Proposition 6.1 in Shafer and Vovk 2001) or the central limit theorem (Shafer and Vovk 2001, Chapters 6–7) will give results similar to Part 1 of Proposition 1. An advantage of the proofs given in this subsection is that they show the dependence of Sceptic II's capital on Sceptic I's capital; it remains to be seen whether such explicit dependence can be achieved using limit theorems of game-theoretic probability.

5.2 Proof of Part 2 of Theorem 3 and Proposition 1

Substituting $\alpha := -1 - 2\epsilon$, with $\epsilon > 0$, in (11) gives

$$-\frac{1}{\epsilon} \ln \mathcal{K}_N^{\mathrm{I}} + \frac{1}{1+\epsilon} \ln \mathcal{K}_N^{\mathrm{II}} \leq \sum_{n=1}^N D^{[-1-2\epsilon]} \left(P_n^{\mathrm{I}} \parallel P_n^{\mathrm{II}}\right),$$

or, equivalently,

$$\ln \mathcal{K}_N^{\mathrm{II}} \leq \frac{1}{\epsilon} \sum_{n=1}^N \ln \int \left(\frac{\beta_n^{\mathrm{I}}}{\beta_n^{\mathrm{II}}}\right)^{\epsilon} \mathrm{d}P_n^{\mathrm{I}} + \frac{1+\epsilon}{\epsilon} \ln \mathcal{K}_N^{\mathrm{I}}. \tag{36}$$

Analogously to the transition from (30) to (33) but now using

$$\left(\frac{\beta_n^{\mathrm{I}}}{\beta_n^{\mathrm{II}}}\right)^{\epsilon} \leq 1 + \epsilon \ln \frac{\beta_n^{\mathrm{I}}}{\beta_n^{\mathrm{II}}} + \frac{1}{2} \epsilon^2 c^{\epsilon} \ln^2 c$$

and

$$\frac{1}{\epsilon}\sum_{n=1}^{N}\ln\int\left(\frac{\beta_n^{\mathrm{I}}}{\beta_n^{\mathrm{II}}}\right)^{\epsilon}\mathrm{d}P_n^{\mathrm{I}} \le \frac{1}{\epsilon}\sum_{n=1}^{N}\ln\left(1+\epsilon D^{(-1)}\left(P_n^{\mathrm{I}}\parallel P_n^{\mathrm{II}}\right)+\frac{1}{2}\epsilon^2 c^{\epsilon}\ln^2 c\right)$$
$$\le \sum_{n=1}^{N} D^{(-1)}\left(P_n^{\mathrm{I}}\parallel P_n^{\mathrm{II}}\right)+\frac{1}{2}N\epsilon c^{\epsilon}\ln^2 c,$$

we obtain

$$\ln\mathcal{K}_N^{\mathrm{II}} \le \sum_{n=1}^{N} D^{(-1)}\left(P_n^{\mathrm{I}}\parallel P_n^{\mathrm{II}}\right)+\frac{1}{2}N\epsilon c^{\epsilon}\ln^2 c+\frac{1+\epsilon}{\epsilon}\ln\mathcal{K}_N^{\mathrm{I}}.$$

If N is known in advance, setting $\epsilon := N^{-1/2}$ gives

$$\ln\mathcal{K}_N^{\mathrm{II}} \le \sum_{n=1}^{N} D^{(-1)}\left(P_n^{\mathrm{I}}\parallel P_n^{\mathrm{II}}\right)+\left(C_4+\ln\mathcal{K}_N^{\mathrm{I}}\right)\left(\sqrt{N}+1\right)$$

and so completes the proof of Proposition 1. (We can take $C := c\ln^2 c$ in (29), or, if we are interested in (29) holding from some N on, $C \approx \frac{1}{2}\ln^2 c$.)

As for Theorem 3, we now have

$$\ln\mathcal{K}_N^{\mathrm{II}} \le \sum_{n=1}^{N} D^{(-1)}\left(P_n^{\mathrm{I}}\parallel P_n^{\mathrm{II}}\right)+\frac{1}{2}N\epsilon_k c^{\epsilon_k}\ln^2 c+\frac{1+\epsilon_k}{\epsilon_k}\left(\ln\mathcal{K}_N^{\mathrm{I}}+2\ln k\right).$$

Setting, as before, (35), we now obtain

$$\ln\mathcal{K}_N^{\mathrm{II}} \le \sum_{n=1}^{N} D^{(-1)}\left(P_n^{\mathrm{I}}\parallel P_n^{\mathrm{II}}\right)$$
$$+C_5\sqrt{N\ln\ln N}+C_6\left(1+\sqrt{\frac{N}{\ln\ln N}}\right)\left(\ln^+\mathcal{K}_N^{\mathrm{I}}+2\ln\ln N\right),$$

which completes the proof of Theorem 3.

6 Criteria of absolute continuity and singularity

A simple measure-theoretic counterpart of the competitive testing protocol is the measurable space Ω^∞ equipped with two probability measures, $\mathbb{P}^{\mathrm{I}},\mathbb{P}^{\mathrm{II}}\in\mathcal{P}(\Omega^\infty)$. The generic element of Ω^∞ will be denoted $\omega_1\omega_2\ldots$. Let P_n^{I} (resp. P_n^{II}) be a regular conditional distribution of ω_n given $\omega_1\ldots\omega_{n-1}$ w.r. to the probability measure $\mathbb{P}^{\mathrm{I}}$ (resp. $\mathbb{P}^{\mathrm{II}}$). (For regular conditional distributions to exist it suffices to assume that Ω is a Borel space: see, e.g., Shiryaev 1996, Theorem II.7.5.) The strategies of the two Forecasters are fixed: Forecaster I is playing P_n^{I} and

Forecaster II is playing P_n^{II}; therefore, they cease to be active players in the game.

We will consider the filtration $(\mathcal{F}_n)_{n=0}^{\infty}$ where each $\mathcal{F}_n$ is generated by $\omega_1, \ldots, \omega_n$ and sometimes write $\mathcal{F}$ for $\mathcal{F}_\infty$. By a *normalized nonnegative measure-theoretic martingale* w.r. to a probability measure $\mathbb{P}$ on Ω^∞ we will mean a martingale $(\xi_n)_{n=0}^{\infty}$ (see, e.g., Shiryaev 1996, Chapter VII) w.r. to $\mathbb{P}$ and $(\mathcal{F}_n)_{n=0}^{\infty}$ such that $\xi_0 = 1$ and $\xi_n \geq 0$ for all $n = 1, 2, \ldots$; we will allow ξ_n to take value ∞ (of course, with probability zero). We will usually write $\xi(\omega_1, \ldots, \omega_n)$ for $\xi_n(\omega_1, \omega_2, \ldots)$ and regard ξ as a function on the set Ω^* of all finite sequences of elements of Ω. A *normalized nonnegative game-theoretic martingale* w.r. to $\mathbb{P}$ (either $\mathbb{P}^{\mathrm{I}}$ or $\mathbb{P}^{\mathrm{II}}$) is Sceptic's (either Sceptic I's or Sceptic II's, respectively) capital represented as a function of Reality's moves $\omega_1 \ldots \omega_n$.

To state the connection between measure-theoretic and game-theoretic notions of normalized nonnegative martingale (in the current measure-theoretic framework), let us say that two *processes* (i.e., measurable functions on Ω^*) are *equivalent* (w.r. to $\mathbb{P}$) if they coincide $\mathbb{P}$-almost surely. It is customary in measure-theoretic probability to identify equivalent processes (although this practice does not carry over to the game-theoretic framework). It is easy to see that the normalized nonnegative measure-theoretic martingales are the closure of the measurable normalized nonnegative game-theoretic martingales w.r. to this relation of equivalence.

The following proposition is a special case of the infinitary version of Ville's theorem; for a proof see, e.g., Shafer and Vovk (2001), Proposition 8.14.

Proposition 2 *Let $E \in \mathcal{F}$ and $\mathbb{P}$ be either $\mathbb{P}^{\mathrm{I}}$ or $\mathbb{P}^{\mathrm{II}}$.*

1. *If a normalized nonnegative measure-theoretic martingale diverges to infinity when E happens, then $\mathbb{P}\, E = 0$.*

2. *If $\mathbb{P}\, E = 0$, then there is a normalized nonnegative measure-theoretic martingale that diverges to infinity if E happens.*

This proposition is also true for the measurable normalized nonnegative game-theoretic martingales and establishes the connection between the game-theoretic and measure-theoretic notions of a "null event": for example, an event (in the measure-theoretic sense, i.e., a measurable subset of Ω^∞) E satisfies $\mathbb{P}^{\mathrm{I}}(E) = 0$ if and only if Sceptic I has a measurable strategy that makes him infinitely rich on E.

The following is a special case of the Kabanov–Liptser–Shiryaev (1977) criterion of absolute continuity and singularity. For simplicity we assume that $\mathbb{P}^{\mathrm{I}}\Big|_{\mathcal{F}_n} \ll \mathbb{P}^{\mathrm{II}}\Big|_{\mathcal{F}_n}$ for all n (this is the standard assumption of local absolute continuity, which simplifies measure-theoretic results: see, e.g., Jacod and Shiryaev 2003, Sect. IV.2c). This will ensure the timidity of Forecaster II (at least after changing the regular conditional distributions on a set of probability zero, both under $\mathbb{P}^{\mathrm{I}}$ and $\mathbb{P}^{\mathrm{II}}$).

Corollary 1 *In the measure-theoretic competitive testing protocol:*

1. *For any $\alpha \in (-1,1)$, $\mathbb{P}^{\mathrm{I}} \ll \mathbb{P}^{\mathrm{II}}$ if and only if (3) holds $\mathbb{P}^{\mathrm{I}}$-almost surely.*

2. *For any $\alpha \in (-1,1)$, $\mathbb{P}^{\mathrm{I}} \perp \mathbb{P}^{\mathrm{II}}$ if and only if (2) holds $\mathbb{P}^{\mathrm{I}}$-almost surely.*

Proof We start from Part "if" of Part 1. Suppose (3) holds $\mathbb{P}^{\mathrm{I}}$-almost surely and E is an event such that $\mathbb{P}^{\mathrm{II}}(E) = 0$; our goal is to prove that $\mathbb{P}^{\mathrm{I}}(E) = 0$. Let Sceptic II play a measurable strategy that makes him infinitely rich on the event E, and let Sceptic I play the half-and-half mixture (in the sense of Lemma 3) of the following two strategies: one of the measurable strategies whose existence is guaranteed in Part 2 of Theorem 1a (cf. the remark following this proof) and a measurable strategy that makes him infinitely rich when the event (3) fails to happen. It is easy to check that Sceptic I is guaranteed to become infinitely rich on the event E, no matter whether (3) holds or not. Therefore, indeed $\mathbb{P}^{\mathrm{I}}(E) = 0$.

Next we prove Part "only if" of Part 1. Fix measurable strategies $\mathcal{S}^{\mathrm{I}}$ and $\mathcal{S}^{\mathrm{II}}$ for the Sceptics that make at least one of them infinitely rich when the event (3) fails to happen; Part 1 of Theorem 1a guarantees that such strategies exist. Sceptic II will play strategy $\mathcal{S}^{\mathrm{II}}$ whereas Sceptic I will play a mixture of $\mathcal{S}^{\mathrm{I}}$ and another strategy. Let E be the event that Sceptic II becomes infinitely rich. Since $\mathbb{P}^{\mathrm{II}}(E) = 0$ and $\mathbb{P}^{\mathrm{I}} \ll \mathbb{P}^{\mathrm{II}}$, we have $\mathbb{P}^{\mathrm{I}}(E) = 0$ and so Sceptic I has a measurable strategy that makes him infinitely rich on E; let him play the half-and-half mixture of this strategy and $\mathcal{S}^{\mathrm{I}}$. This strategy for Sceptic I will guarantee his becoming infinitely rich whenever (3) fails to happen.

The proof of Part "if" of Part 2 again relies on Part 1 of Theorem 1a. Define $\mathcal{S}^{\mathrm{I}}$ and $\mathcal{S}^{\mathrm{II}}$ as before, let Sceptic II play $\mathcal{S}^{\mathrm{II}}$, and let Sceptic I play the half-and-half mixture of $\mathcal{S}^{\mathrm{I}}$ and a measurable strategy that makes him infinitely rich when (2) fails to happen. It is clear that one of the Sceptics becomes infinitely rich no matter what happens. Therefore, $\mathbb{P}^{\mathrm{I}}$ and $\mathbb{P}^{\mathrm{II}}$ are mutually singular (e.g., take E as the event that Sceptic I becomes infinitely rich; then $\mathbb{P}^{\mathrm{I}}(E) = 0$ and $\mathbb{P}^{\mathrm{II}}(\Omega^{\infty} \setminus E) = 0$).

It remains to prove Part "only if" of Part 2. Let E be an event such that $\mathbb{P}^{\mathrm{II}}(E) = 0$ and $\mathbb{P}^{\mathrm{I}}(E) = 1$. Let Sceptic II play a measurable strategy that makes him infinitely rich on E and let Sceptic I play the half-and-half mixture of the following two strategies: one of the measurable strategies whose existence is guaranteed in Part 2 of Theorem 1a and a measurable strategy that makes him infinitely rich when the event E fails to happen. Now if the event (2) fails to happen, Sceptic I is guaranteed to become infinitely rich, regardless of whether E happens. This completes the proof. ∎

Remark Notice that to deduce Corollary 1 we need slightly more than stated in Theorem 1a: namely, we need measurable strategies for the Sceptics (in the case of Part 1) or Sceptic I (in the case of Part 2). It is easy to see that the strategies constructed in Theorem 1a satisfy this property, but it would have been awkward to include their measurability in the statement. The strategies are not only measurable but also computable, which is a much stronger property. Therefore, it would have been more natural to include the strategies' computability in the statement of Theorem 1a; however, this would significantly complicate the exposition, especially that there are several popular non-equivalent definitions of computability, even for real-valued functions of real variable. Following Shafer and Vovk (2001), we dropped any references to the properties of regularity of the constructed strategies for the Sceptics in the formal statements of our results.

To discuss the intuition behind Corollary 1, let us assume that Forecaster I is reliable in the following measure-theoretic sense: we do not expect an event E to happen if it is chosen in advance and satisfies $\mathbb{P}^{\mathrm{I}}(E) = 0$. Then $\mathbb{P}^{\mathrm{I}} \ll \mathbb{P}^{\mathrm{II}}$ means that Forecaster II is "automatically reliable": if E is chosen in advance and satisfies $\mathbb{P}^{\mathrm{II}}(E) = 0$, we also do not expect it to happen. Part 1 of Corollary 1 says that Forecaster II is automatically reliable if and only if the Forecasters' predictions are close in the sense of (3) holding almost surely w.r. to the reliable probability measure $\mathbb{P}^{\mathrm{I}}$.

On the other hand, $\mathbb{P}^{\mathrm{I}} \perp \mathbb{P}^{\mathrm{II}}$ means that Forecaster II is "automatically unreliable": we can choose in advance an event E which we expect to happen ($\mathbb{P}^{\mathrm{I}}(E) = 1$) but which will falsify the probability measure $\mathbb{P}^{\mathrm{II}}$ ($\mathbb{P}^{\mathrm{II}}(E) = 0$). Part 2 of Corollary 1 says that Forecaster II is automatically unreliable if and only if the Forecasters' predictions are far apart in the sense of (2) holding almost surely w.r. to the reliable probability measure $\mathbb{P}^{\mathrm{I}}$.

The game-theoretic Theorem 1a has several advantages over Corollary 1. First of all, Theorem 1a is "pointwise": it carries information about specific plays of the game. It also has the flexibility provided by the game-theoretic framework in general:

- the Forecasters can react to the Sceptics' moves (and Reality can react to both the Forecasters and the Sceptics);
- all players can react to various events outside the protocol;
- game-theoretic results about merging of opinions, unlike the standard measure-theoretic results about absolute continuity and singularity, do not depend on the (vast) parts of the space Ω^* that are never reached or even approached by Reality.

Acknowledgments

I am grateful to Akio Fujiwara and Phil Dawid for encouragement and useful discussions.

References

Amari, S., and Nagaoka, H. (2000). *Methods of information geometry*. Providence, RI: American Mathematical Society. Japanese original (1993): Tokyo: Iwanami Shoten.

Blackwell, D., and Dubins, L. (1962). Merging of opinions with increasing information. *Annals of Mathematical Statistics*, *33*, 882–886.

Dawid, A. P. (1984). Statistical theory: the prequential approach. *Journal of the Royal Statistical Society* A, *147*, 278–292.

Dawid, A. P. (1985). Calibration-based empirical probability (with discussion). *Annals of Statistics*, *13*, 1251–1285.

Dawid, A. P. (2004). Probability, causality and the empirical world: a Bayes–de Finetti–Popper–Borel synthesis. *Statistical Science*, *19*, 44–57.

Dawid, A. P., Vovk, V. (1999). Prequential probability: principles and properties. *Bernoulli*, *5*, 125–162.

de Finetti, B. (1937). La prévision, ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, *7*, 1–68. English translation (1980): Foresight: its logical laws, its subjective sources. In H. E. Kyburg, Jr, and H. E. Smokler (Ed.) *Studies in subjective probability* (pp. 53–118). Huntington, NY: Krieger. Second edition. First edition (1964): New York: Wiley.

Feldman, J. (1958). Equivalence and perpendicularity of Gaussian measures. *Pacific Journal of Mathematics*, *8*, 699–708.

Fujiwara, A. (2007). Randomness criteria in terms of α-divergences. Manuscript.

Greenwood, P. E., Shiryaev, A. N. (1985). *Contiguity and the statistical invariance principle*. New York: Gordon & Breach.

Hájek, J. (1958). On a property of normal distribution of an arbitrary stochastic process (in Russian). *Czechoslovak Mathematical Journal*, *8*:610–618.

Hotelling, H. (1929). Stability in competition. *Economic Journal*, *39*, 41–57.

Jacod, J., Shiryaev, A. N. (2003). *Limit theorems for stochastic processes*. Berlin: Springer. Second edition.

Kabanov, Y. M., Liptser, R. S., and Shiryaev, A. N. (1977). On the question of absolute continuity and singularity of probability measures. *Mathematics of the USSR—Sbornik*, *33*, 203–221.

Kakutani, S. (1948). On equivalence of infinite product measures. *Annals of Mathematics*, *49*, 214–224.

Kumon, M., Takemura, A., Takeuchi, K. (2007). Game-theoretic versions of strong law of large numbers for unbounded variables. *Stochastics*, to appear.

Pukelsheim, F. (1986). Predictable criteria for absolute continuity and singularity of two probability measures. *Statistics and Decisions*, *4*, 227-236.

Rényi, A. (1961). On measures of information and entropy. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 547–561). Berkeley, CA: University of California Press.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley. Second edition (1972): New York: Dover.

Shafer, G., Vovk, V. (2001). *Probability and finance: it's only a game!* New York: Wiley.

Shiryaev, A. N. (1996). *Probability*. New York: Springer. Second edition. Third Russian edition published in 2004.

Solomonoff, R. J. (1978). Complexity-based induction systems: comparisons and convergence theorems. *IEEE Transactions on Information Theory*, *IT-24*, 422–432,

Ville, J. (1939). *Etude critique de la notion de collectif*. Paris: Gauthier-Villars.

Vovk, V. (1987a). On a randomness criterion. *Soviet Mathematics Doklady*, *35*, 656–660.

Vovk, V. (1987b). The law of the iterated logarithm for random Kolmogorov, or chaotic, sequences. *Theory of Probability and its Applications*, *32*, 413–425.

Vovk, V. (1989). Prediction of stochastic sequences. *Problems of Information Transmission*, *25*, 285–296.

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

Vovk, V., Shafer, G. (2005). Good randomized sequential probability forecasting is always possible. *Journal of the Royal Statistical Society* B, *67*, 747–763.