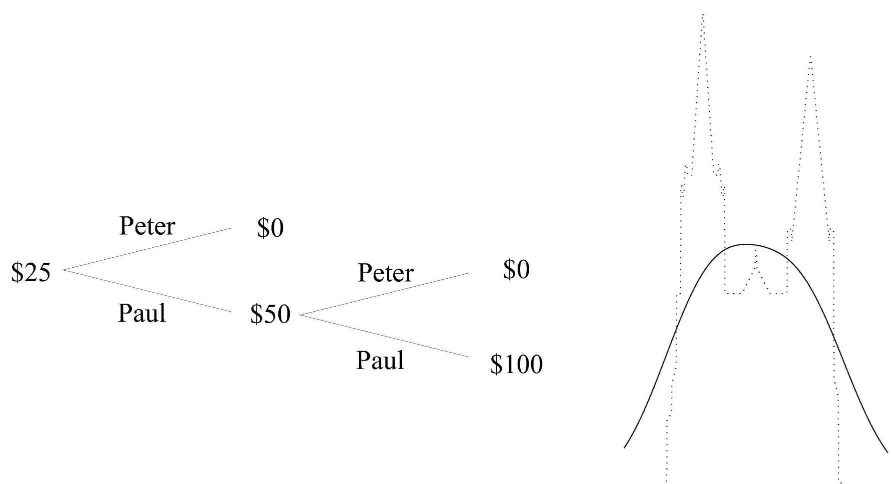
Asymptotic uniqueness in long-term prediction

Vladimir Vovk



The Game-Theoretic Probability and Finance Project

Working Paper #66

First posted December 4, 2023. Last revised April 5, 2024.

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

Abstract

This paper establishes the asymptotic uniqueness of long-term probability forecasts in the following form. Consider two forecasters who repeatedly issue probability forecasts for the infinite future. The main result of the paper says that either at least one of the two forecasters will be discredited or their forecasts will converge in total variation. This can be regarded as a game-theoretic version of the classical Blackwell–Dubins result getting rid of some of its limitations. This result is further strengthened along the lines of Richard Jeffrey's radical probabilism.

This paper has also been published as an arXiv report.

Contents

> ...scientific disagreements tend to disappear...when new data accumulate....
>
> Jeffreys (1938, p. 673); also in Jeffreys (1961, Sect. 1.9)

1 Introduction

This paper belongs to the general area of probabilistic prediction, and we will be interested in ways of testing long-term predictions and the asymptotic uniqueness of successful long-term predictions. The long-term nature of such predictions makes our results quite different from the usual results in conformal prediction (Vovk et al., 2022; Angelopoulos and Bates, 2023), prediction with expert advice (Cesa-Bianchi and Lugosi, 2006; Vovk, 1998), and game-theoretic probability (Shafer and Vovk, 2019), which are usually concerned with one-step-ahead prediction. In this paper "prediction" and "forecast" are used interchangeably.

To demonstrate the asymptotic uniqueness of successful long-term predictions, we consider two forecasters issuing such predictions. The asymptotic uniqueness holds if either the forecasts that they issue eventually become almost indistinguishable or we are able to demonstrate that at least one of the forecasters is inadequate. This paper establishes new results of this kind and reviews related known results.

Blackwell and Dubins (1962, Sect. 2) prove a classical result about the asymptotic uniqueness of long-term predictions. They consider two probability measures that agree in a certain sense (namely, one of them is absolutely continuous w.r. to the other). They then show that the predictions output by the two probability measures for the infinite future converge in total variation almost surely as time progresses. We will remove unnecessary restrictions and generalize this result.

The general phenomenon of convergence of opinions (i.e., forecasts for the future) for adequate forecasters will be referred to as *Jeffreys's law*, although originally this expression was used by Dawid (1984, Sect. 5.2) for his result about convergence of one-step-ahead predictions:

> I shall call this finding "Jeffreys's Law", after an admittedly distorted interpretation of Jeffreys (1938): "When a law has been applied to a large body of data without any systematic discrepancy being detected... the probability of a further inference from the law approaches certainty whether the law is true or not."

Another quote from the same paper (Jeffreys, 1938) is given as the epigraph to this paper.

We start in Sect. 2 from discussing ways of testing long-term predictions and applying those ways to deriving our first version of Jeffreys's law. This version is stated in terms of a testing protocol involving three players: Forecaster I, Forecaster II, and a new player, Sceptic, who performs testing. We will construct

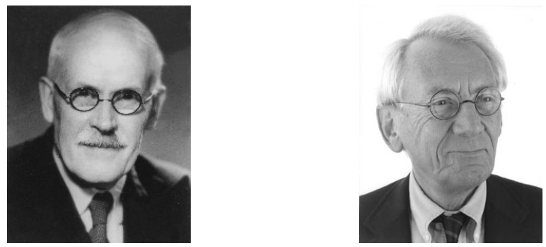


Figure 1: Jeffreys and Jeffrey. Left panel: Harold Jeffreys (1891–1989), British geophysicist. Right panel: Richard Jeffrey (1926–2002), American philosopher.

a strategy for Sceptic that discredits at least one of the two forecasters if their opinions do not converge. This removes limitations of Blackwell and Dubins's result, as discussed in Sect. 5.

One respect in which the result of Sect. 2 (and the Blackwell–Dubins original result) is restrictive is that it assumes that we observe data with certainty. Section 3 extends the result of Sect. 2 to Jeffrey's picture of radical probabilism, in which no evidence is certain.

Proofs of the results of Sects 2–3 are postponed to Sect. 4. All our proofs will be constructive and will exhibit simple strategies for Sceptic that enforce convergence of opinions.

Section 5 discusses related results in literature starting from the Blackwell–Dubins result (Sect. 5.1) and then going on to results about one-step-ahead prediction (Sect. 5.2). The most conspicuous difference between the two kinds of results, those for long-term and one-step-ahead prediction, is that the latter can be both asymptotic and small-sample, while the former are invariably asymptotic (to the best of my knowledge); making them non-asymptotic looks to me an important direction of further research. Section 6 concludes and lists some other directions of further research.

This paper and its predecessor (Vovk, 2024) were motivated by a conversation with A. Philip Dawid (Vovk and Shafer, 2023). Among topics of the conversation were one-step-ahead vs multi-step prediction (Vovk and Shafer, 2023, Sect. 7) and the Blackwell–Dubins theorem (Vovk and Shafer, 2023, Sect. 7 of the arXiv version).

Two people with similar surnames will play key roles in this paper, Jeffreys and Jeffrey. (See Fig. 1.) Harold Jeffreys (1891–1989) was a contemporary of Ronald Fisher who spent his professional life in Cambridge, England. His main speciality was geophysics, but he was also one of the founders of Bayesian statistics. Richard Jeffrey (1926–2002) was an influential American philosopher.

2 Jeffreys's law

Before we state Jeffreys's law we need to discuss ways of testing probability forecasts. Let us fix a finite *observation space* $\mathbf{Y}$ (equipped with the discrete σ-algebra). At each step Reality produces an *observation* $y_n \in \mathbf{Y}$, $n = 1, 2, \dots$, and Forecaster is trying to predict future observations by issuing a probability forecast $P_n \in \mathfrak{P}(\mathbf{Y}^\infty)$ ($\mathfrak{P}(\mathbf{Y}^\infty)$ standing for the family of all probability measures on $\mathbf{Y}^\infty$ equipped with the Borel σ-algebra). In the following testing protocol, we let $\mathbf{Y}^*$ stand for the set of all finite sequences of observations, $\mathbf{Y}^+$ stand for the set of all non-empty finite sequences of observations, $|x|$ stand for the length of $x \in \mathbf{Y}^*$, and $[x]$ stand for the set of all infinite continuations of a finite sequence $x \in \mathbf{Y}^*$ (in other words, $[x]$ is the set of all sequences in $\mathbf{Y}^\infty$ that have x as their prefix).

Protocol 1. Testing protocol:

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

$$\begin{aligned} \mathcal{K}_{n-1} := \mathcal{K}^-_{n-1} &+ \textstyle\sum_{x\in\mathbf{Y}^+} f_{n-1}(y_{n-1}x)P_n([x]) \\ &- \textstyle\sum_{x\in\mathbf{Y}^*:|x|>1} f_{n-1}(x)P_{n-1}([x]) \end{aligned} \tag{1}$$

  Sceptic announces $f_n \in \mathbb{R}^{\mathbf{Y}^+}$ such that
    $f_n(x) = 0$ for all but finitely many $x \in \mathbf{Y}^+$
  Reality announces $y_n \in \mathbf{Y}$

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

Protocol 1 is interpreted in terms of betting, as described in de Finetti (1937, Chap. 1) and, in our current terminology, Vovk (2024, Sect. 3). At each step n Forecaster announces a probability measure P_n for the infinite future $y_n, y_{n+1}, \dots$. The betting interpretation of P_n is that, for each non-empty finite sequence $x \in \mathbf{Y}^*$, $P_n([x])$ is the price of a ticket (the *x-ticket*) that pays $1_{\{x \subseteq (y_n, y_{n+1}, \dots)\}}$ (i.e., it pays 1 if and only if x is a prefix of the sequence $(y_n, y_{n+1}, \dots)$ of the future observations and pays nothing otherwise). Forecaster allows his opponent to buy any real number (positive, negative, or zero; not necessarily integer) of such tickets.

Testing the forecasts is performed by another player, Sceptic. At step n, for each non-empty $x \in \mathbf{Y}^*$, Sceptic announces the number $f_n(x)$ of x-tickets that he chooses to buy at this step. Therefore, his move is $f_n \in \mathbb{R}^{\mathbf{Y}^+}$, which means that $f_n : \mathbf{Y}^+ \to \mathbb{R}$. The numbers $f_n(x)$ are allowed to be different from zero only for finitely many x-tickets, and so the sums $\sum_x$ in the protocol are uncontroversial. At the end of each step Reality announces the actual observation $y_n \in \mathbf{Y}$, and the y-tickets for $y \in \mathbf{Y}$ are cashed in: the y_n-ticket pays 1, the other y-tickets do not pay anything, and the total cost of all the y-tickets is $\sum_{y\in\mathbf{Y}} f_n(y)P_n([y])$. The x-tickets for longer x are sold at the next step at the new prices (which accounts for the term $\sum_{x\in\mathbf{Y}^+} f_{n-1}(y_{n-1}x)P_n([x])$ in Protocol 1) and the loss due to their total cost is recorded also at the next step (which accounts for the term $\sum_{x\in\mathbf{Y}^*:|x|>1} f_{n-1}(x)P_{n-1}([x])$).

The interpretation of $\mathcal{K}_n$ is that it is Sceptic's capital at time n, but a large $\mathcal{K}_n$ only means that Forecaster's predictions $P_1, P_2, \dots$ have been discredited (so Sceptic is using "play money"). For this interpretation to be valid, Sceptic is never allowed to go into debt: as soon as $K_n < 0$, the game is stopped and Sceptic's attempt at discrediting Forecaster fails.

In the case of one-step-ahead prediction we do not need the capital update (1), and (2) is sufficient. On the other hand, for multi-step prediction, we do not need to have (2) (which is the standard capital update in game-theoretic probability) as a separate entry and can merge it into (1). The following protocol is a slightly simplified version of Protocol 1.

Protocol 2. Simplified testing protocol:

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

$$\mathcal{K}_{n-1} := \mathcal{K}_{n-2} + \textstyle\sum_{x\in\mathbf{Y}^*} f_{n-1}(y_{n-1}x)P_n([x]) - \textstyle\sum_{x\in\mathbf{Y}^+} f_{n-1}(x)P_{n-1}([x]) \tag{3}$$

  Sceptic announces $f_n \in \mathbb{R}^{\mathbf{Y}^+}$ such that
    $f_n(x) = 0$ for all but finitely many $x \in \mathbf{Y}^+$
  Reality announces $y_n \in \mathbf{Y}$.

In Protocol 2 we merge the steps (1) and (2) of Protocol 1 into one step (3).

The testing protocol that we use for stating Jeffreys's law involves one Sceptic playing simultaneously (but separately) against two forecasters who output probability forecasts $P_n^{\mathrm{I}}, P_n^{\mathrm{II}} \in \mathfrak{P}(\mathbf{Y}^\infty)$ at steps $n = 1, 2, \dots$.

Protocol 3. Double testing protocol:

$\mathcal{K}_0^{\mathrm{I}} = \mathcal{K}_0^{\mathrm{II}} := 1$
FOR $n = 1, 2, \dots$:
  Forecaster I announces $P_n^{\mathrm{I}} \in \mathfrak{P}(\mathbf{Y}^\infty)$
  Forecaster II announces $P_n^{\mathrm{II}} \in \mathfrak{P}(\mathbf{Y}^\infty)$
  IF $n > 1$:

$$\mathcal{K}_{n-1}^{\mathrm{I}} := \mathcal{K}_{n-2}^{\mathrm{I}} + \textstyle\sum_{x\in\mathbf{Y}^*} f_{n-1}^{\mathrm{I}}(y_{n-1}x)P_n^{\mathrm{I}}([x]) - \textstyle\sum_{x\in\mathbf{Y}^+} f_{n-1}^{\mathrm{I}}(x)P_{n-1}^{\mathrm{I}}([x])$$

$$\mathcal{K}_{n-1}^{\mathrm{II}} := \mathcal{K}_{n-2}^{\mathrm{II}} + \textstyle\sum_{x\in\mathbf{Y}^*} f_{n-1}^{\mathrm{II}}(y_{n-1}x)P_n^{\mathrm{II}}([x]) - \textstyle\sum_{x\in\mathbf{Y}^+} f_{n-1}^{\mathrm{II}}(x)P_{n-1}^{\mathrm{II}}([x])$$

  Sceptic announces $f_n^{\mathrm{I}}, f_n^{\mathrm{II}} \in \mathbb{R}^{\mathbf{Y}^+}$ such that
    $f_n^{\mathrm{I}}(x) = f_n^{\mathrm{II}}(x) = 0$ for all but finitely many $x \in \mathbf{Y}^+$
  Reality announces $y_n \in \mathbf{Y}$.

We will state Jeffrey's law using the total variation distance

$$\|P - Q\| := 2 \sup_E |P(E) - Q(E)| \in [0, 2]$$

between probability measures P and Q on $\mathbf{Y}^\infty$, where E ranges over the events in the common domain of P and Q.

Theorem 4. *Sceptic has a strategy in Protocol 3 that guarantees the disjunction of*

- $\|P_n^{\mathrm{I}} - P_n^{\mathrm{II}}\| \to 0$ *as* $n \to \infty$,
- $\mathcal{K}_n^{\mathrm{I}} \to \infty$ *as* $n \to \infty$,
- $\mathcal{K}_n^{\mathrm{II}} \to \infty$ *as* $n \to \infty$.

Theorem 4 can be interpreted as establishing a connection between the correspondence (see, e.g., Popper 1945) and convergence (see, e.g., Peirce 1877) theories of truth. The correspondence theory of truth (of the probability forecasts in this case) refers to agreement with reality, and our interpretation of a large $\mathcal{K}_n$ is lack of agreement with reality. The convergence theory of truth regards truth as the point at which different opinions converge. In Peirce's words, "the settlement of opinion is the sole end of inquiry" (Peirce, 1877). According to Theorem 4, a version of the correspondence theory implies a version of the convergence theory.

3 Agnostic probabilism

According to the idea of *radical probabilism*, put forward very clearly by Jeffrey (1968), empirical evidence is never certain. In particular, we never learn the outcomes $y_1, y_2, \dots$ for sure.

Jeffrey referred to the opposite point of view, in which we do observe the true y_n (in our current context), as dogmatic probabilism. We would like our mathematical results to cover dogmatic probabilism as a special case, and the title of this section, *agnostic probabilism*, refers to its prediction protocol eventually disclosing, or never disclosing, the true outcomes y_n, as the case may be for different n. In particular, the following protocol includes Protocol 3 as a special case.

Protocol 5. Double agnostic testing protocol:

$\mathcal{K}_0^{\mathrm{I}} = \mathcal{K}_0^{\mathrm{II}} := 1$
FOR $n = 1, 2, \dots$:
  Forecaster I announces $P_n^{\mathrm{I}} \in \mathfrak{P}(\mathbf{Y}^\infty)$
  Forecaster II announces $P_n^{\mathrm{II}} \in \mathfrak{P}(\mathbf{Y}^\infty)$
  IF $n > 1$:

$$\begin{aligned} \mathcal{K}_{n-1}^{\mathrm{I}} &:= \mathcal{K}_{n-2}^{\mathrm{I}} + \textstyle\sum_{x \in \mathbf{Y}^+} f_{n-1}^{\mathrm{I}}(x)(P_n^{\mathrm{I}}([x]) - P_{n-1}^{\mathrm{I}}([x])) \\ \mathcal{K}_{n-1}^{\mathrm{II}} &:= \mathcal{K}_{n-2}^{\mathrm{II}} + \textstyle\sum_{x \in \mathbf{Y}^+} f_{n-1}^{\mathrm{II}}(x)(P_n^{\mathrm{II}}([x]) - P_{n-1}^{\mathrm{II}}([x])) \end{aligned} \tag{4}$$

  Sceptic announces $f_n^{\mathrm{I}}, f_n^{\mathrm{II}} \in \mathbb{R}^{\mathbf{Y}^+}$ such that
    $f_n^{\mathrm{I}}(x) = f_n^{\mathrm{II}}(x) = 0$ for all but finitely many $x \in \mathbf{Y}^+$.

Protocol 5 does not include Reality as a separate player. Her role is played by the two forecasters: in order to model Reality of Protocol 3, they should choose P_n^{I} and P_n^{II} concentrated on $[(y_1, \dots, y_{n-1})]$ for some observations $y_1, \dots, y_{n-1} \in \mathbf{Y}$. Such a choice already implies a certain agreement between the forecasters; in particular, $P_n^{\mathrm{I}} - P_n^{\mathrm{II}} \to 0$ as $n \to \infty$ *weakly*, meaning that, for any $x \in \mathbf{Y}^*$,

$$P_n^{\mathrm{I}}([x]) - P_n^{\mathrm{II}}([x]) \to 0 \quad \text{as } n \to \infty.$$

More precisely, to embed Protocol 3 into Protocol 5, the two forecasters in the latter should choose the probability measures concentrated on $[(y_1, \ldots, y_{n-1})]$ and defined by

$$P_n^{\mathrm{I}}([y_1 \ldots y_{n-1}x]) := P_n^{\mathrm{I}}([x]),$$
$$P_n^{\mathrm{II}}([y_1 \ldots y_{n-1}x]) := P_n^{\mathrm{II}}([x]), \quad x \in \mathbf{Y}^*,$$

where the P_n^{I} and P_n^{II} on the right-hand sides are the predictions in the former.

Remark 6. Jeffrey traces the idea of radical probabilism back to Ramsey and de Finetti. In Jeffrey (1992, p. 66), he says, "this is his radical probabilism—Ramsey denies that our probable knowledge need be based on certainties". And in Jeffrey (1988, p. 2), he says, "De Finetti's probabilism is 'radical' in the sense of going all the way down to the roots: he sees probabilities as ultimate forms of judgment which need not be based on deeper all-or-none knowledge." (Although there is some tension between this interpretation of de Finetti's views and de Finetti's emphatic defence of Bayesian conditioning in de Finetti 2017, Sect. 4.5.3.)

Remark 7. Jeffrey often uses "radical probabilism" in the sense of our "agnostic probabilism": "Radical probabilism doesn't insist that probabilities be based on certainties" (Jeffrey, 1992, p. 11). However, in other places he appears to deny the existence of certain empirical evidence; e.g., in one of his earliest (1968) publications on radical probabilism he says, "Radical probabilism adds [to "probabilism"] the "nonfoundational" thought that there is no bedrock of certainty underlying our probabilistic judgments" (Jeffrey, 1992, pp. 44–45).

Example 8. Let us check that Theorem 4 does not hold under agnostic probabilism (without any further conditions). Suppose $P_n^{\mathrm{I}} = P^{\mathrm{I}}$ and $P_n^{\mathrm{II}} = P^{\mathrm{II}}$ do not depend on n. Then we have $\mathcal{K}_n^{\mathrm{I}} = \mathcal{K}_n^{\mathrm{II}} = 1$ for all n, and so we have no convergence of opinion for successful forecasters even if P^{I} and P^{II} are very different.

Theorem 9. *Sceptic has a strategy in Protocol 5 that guarantees* $\|P_n^{\mathrm{I}} - P_n^{\mathrm{II}}\| \to 0$ *as* $n \to \infty$ *whenever the following three conditions are satisfied:*

- $P_n^{\mathrm{I}} - P_n^{\mathrm{II}} \to 0$ *weakly as* $n \to \infty$*,*
- $\mathcal{K}_n^{\mathrm{I}} \not\to \infty$ *as* $n \to \infty$*, and*
- $\mathcal{K}_n^{\mathrm{II}} \not\to \infty$ *as* $n \to \infty$*.*

Theorem 9 includes Theorem 4 as a special case. According to Example 8, we need to impose some condition of agreement (which we want to make as weak as possible) between the two forecasters before we can claim that they agree in the strong sense of convergence in total variation. In Theorem 4 both forecasters weakly agree with Reality (and therefore, between themselves): for a fixed x, $P_n^{\mathrm{I}}([x])$ and $P_n^{\mathrm{II}}([x])$ of Protocol 5 become equal (namely, both equal to 0 or to 1) as soon as n exceeds the length of x. In Theorem 9 we require an even weaker agreement between forecasters (and do not require any agreement with Reality, who is not even a player).

In Sect. 2 we interpreted Jeffreys's law in the form of Theorem 4 as establishing a connection between the correspondence and convergence theories of truth. There is no absolute notion of truth under radical probabilism, and so this interpretation is not applicable to Theorem 9. Theorem 9 merely provides a means of boosting agreement between successful forecasters: agreement in the sense of weak convergence implies agreement in the sense of convergence in total variation.

4 Proofs

In this section we will prove Theorems 4 (in Sect. 4.1) and 9 (in Sect. 4.2).

4.1 Proof of Theorem 4

For simplicity, in the proofs in this section we assume that $P^{\mathrm{I}}_n([x]) > 0$ and $P^{\mathrm{II}}_n([x]) > 0$ for all $x \in \mathbf{Y}^*$ ("Cromwell's rule").

We will construct a strategy for Sceptic that guarantees the disjunction of

- $\|P^{\mathrm{I}}_n - P^{\mathrm{II}}_n\| \to 0$ as $n \to \infty$,
- $\sqrt{\mathcal{K}^{\mathrm{I}}_n \mathcal{K}^{\mathrm{II}}_n}$ is unbounded as $n \to \infty$.

This is sufficient since we can apply Proposition 11.2 in Shafer and Vovk (2019) to turn an unbounded capital $\mathcal{K}^{\mathrm{I}}_n$ or $\mathcal{K}^{\mathrm{II}}_n$ into $\mathcal{K}^{\mathrm{I}}_n \to \infty$ or $\mathcal{K}^{\mathrm{II}}_n \to \infty$.

As a second step, let us replace $\|P^{\mathrm{I}}_n - P^{\mathrm{II}}_n\| \to 0$ by $H(P^{\mathrm{I}}_n, P^{\mathrm{II}}_n) \to 1$, where H is the *Hellinger integral* (Shiryaev, 2016, Definition 3.9.3):

$$H(P,Q) := \int_{\mathbf{Y}^\infty} \sqrt{\mathrm{d}P \mathrm{d}Q} \in [0,1].$$

This can be done because of the standard connection between total variation distance and Hellinger integral (Shiryaev, 2016, Theorem 3.9.1):

$$2\,(1 - H(P,Q)) \le \|P - Q\| \le \sqrt{8\,(1 - H(P,Q))}. \tag{5}$$

We will also need an approximation to $H(P,Q)$, given in the following lemma, in terms of

$$H_m(P,Q) := \int_{\mathbf{Y}^m} \sqrt{(\mathrm{d}P|_m)(\mathrm{d}Q|_m)},$$

where $P|_m$ and $Q|_m$ stand for the restrictions of P and Q to the first m observations; in other words, $P|_m$ and $Q|_m$ are the probability measures on $\mathbf{Y}^m$ satisfying $(P|_m)(x) = P(x)$ and $(Q|_m)(x) = Q(x)$ for all $x \in \mathbf{Y}^m$.

Lemma 10. *As $m \to \infty$, $H_m(P,Q) \to H(P,Q)$.*

Algorithm 1 Betting strategy for Sceptic for a fixed $\epsilon \in (0,1)$

1: **for** $n = 1, 2 \ldots$:
2: Observe P_n^{I} and P_n^{II} on step n of Protocol 3
3: **if** $H(P_n^{\mathrm{I}}, P_n^{\mathrm{II}}) < 1 - \epsilon$:
4: Find m such that $H_m(P_n^{\mathrm{I}}, P_n^{\mathrm{II}}) < 1 - \epsilon$
5: Buy a collection of x-tickets, $x \in \mathbf{Y}^m$, from Forecaster I that multiplies the current $\mathcal{K}_n^{\mathrm{I}}$ by $\dfrac{1}{H_m(P_n^{\mathrm{I}}, P_n^{\mathrm{II}})} \dfrac{\mathrm{d}\sqrt{P_n^{\mathrm{I}}|_m P_n^{\mathrm{II}}|_m}}{\mathrm{d}P_n^{\mathrm{I}}|_m}$
6: Buy a collection of x-tickets, $x \in \mathbf{Y}^m$, from Forecaster II that multiplies the current $\mathcal{K}_n^{\mathrm{II}}$ by $\dfrac{1}{H_m(P_n^{\mathrm{I}}, P_n^{\mathrm{II}})} \dfrac{\mathrm{d}\sqrt{P_n^{\mathrm{I}}|_m P_n^{\mathrm{II}}|_m}}{\mathrm{d}P_n^{\mathrm{II}}|_m}$
7: Skip the next m steps

Proof. Applying Doob's martingale convergence theorem (Shiryaev, 2019, Theorem 7.4.1) to $(\mathrm{d}P|_m)/(\mathrm{d}R|_m)$ and $(\mathrm{d}Q|_m)/(\mathrm{d}R|_m)$, where $R := (P+Q)/2$, we obtain their R-almost sure convergence as $m \to \infty$ to $\mathrm{d}P/\mathrm{d}R$ and $\mathrm{d}Q/\mathrm{d}R$, respectively (this step uses Carathéodory's theorem, as in the proof of Shiryaev 2019, Theorem 7.6.1). Therefore,

$$\sqrt{\frac{\mathrm{d}P|_m}{\mathrm{d}R|_m}\frac{\mathrm{d}Q|_m}{\mathrm{d}R|_m}} \to \sqrt{\frac{\mathrm{d}P}{\mathrm{d}R}\frac{\mathrm{d}Q}{\mathrm{d}R}}$$

R-almost surely and, since all these fractions take values in $[0,2]$, in L_1 w.r.t. R. $\square$

Fix temporarily an $\epsilon > 0$ (we will later mix over a sequence of $\epsilon \to 0$). Given ϵ, Sceptic can bet at the steps $n = 1, 2, \ldots$ as in Algorithm 1. The instructions in Algorithm 1 should be read in parallel with the description below.

In our interpretation of Protocol 3 we assumed that at each step Sceptic sells all tickets bought at the previous step and buys new tickets. For a given x-ticket, an important special case is where the new amount of x-tickets is equal to the old amount (and so we can assume that no trade in x-tickets takes place at this step). In particular, Sceptic can buy an x-ticket at any step n and hold it to maturity collecting $1_{\{(y_n,\ldots,y_{n+m-1})=x\}}$ at step $n+m$, where $m := |x|$.

Line 5 of Algorithm 1 instructs Sceptic to buy a collection of x-tickets multiplying his current capital by

$$\begin{aligned}\frac{1}{H_m(P_n^{\mathrm{I}}, P_n^{\mathrm{II}})} & \frac{\mathrm{d}\sqrt{P_n^{\mathrm{I}}|_m P_n^{\mathrm{II}}|_m}}{\mathrm{d}P_n^{\mathrm{I}}|_m} \\ &= \frac{1}{H_m(P_n^{\mathrm{I}}, P_n^{\mathrm{II}})} \frac{\sqrt{P_n^{\mathrm{I}}([(y_n,\ldots,y_{n+m-1})])P_n^{\mathrm{II}}([(y_n,\ldots,y_{n+m-1})])}}{P_n^{\mathrm{I}}([(y_n,\ldots,y_{n+m-1})])}. \quad (6)\end{aligned}$$

Let us check that it is indeed possible to turn an initial capital of 1 into (6). By definition, Sceptic can buy from Forecaster I a ticket paying $1_{\{x=(y_n,\ldots,y_{n+m-1})\}}$

(i.e., the x-ticket) for $P^{\mathrm{I}}_n(x)$, for any $x \in \mathbf{Y}^m$. Therefore, Sceptic can buy a ticket paying

$$\frac{\sqrt{P^{\mathrm{I}}_n([x])P^{\mathrm{II}}_n([x])}}{P^{\mathrm{I}}_n([x])} 1_{\{x=(y_n,\dots,y_{n+m-1})\}}$$

for $\sqrt{P^{\mathrm{I}}_n([x])P^{\mathrm{II}}_n([x])}$, for any $x \in \mathbf{Y}^m$. Buying such a ticket for each $x \in \mathbf{Y}^m$ results in a payoff of

$$\frac{\sqrt{P^{\mathrm{I}}_n([(y_n,\dots,y_{n+m-1})])P^{\mathrm{II}}_n([(y_n,\dots,y_{n+m-1})])}}{P^{\mathrm{I}}_n([(y_n,\dots,y_{n+m-1})])}$$

for the price of

$$\sum_{x\in\mathbf{Y}^m} \sqrt{P^{\mathrm{I}}_n([x])P^{\mathrm{II}}_n([x])} = H_m(P^{\mathrm{I}}_n, P^{\mathrm{II}}_n).$$

Therefore, we can indeed turn 1 into (6).

A similar argument applies to Forecaster II (line 6 of Algorithm 1).

When the condition in line 3 is satisfied, the geometric mean $\sqrt{\mathcal{K}^{\mathrm{I}}_n\mathcal{K}^{\mathrm{II}}_n}$ of Sceptic's capitals over the m steps in line 7 will be multiplied by $1/H_m(P^{\mathrm{I}}_n, P^{\mathrm{II}}_n)$, i.e., by more than $\frac{1}{1-\epsilon}$. Therefore, $\sqrt{\mathcal{K}^{\mathrm{I}}_n\mathcal{K}^{\mathrm{II}}_n}$ will be unbounded if the condition in line 3 is satisfied infinitely often.

Let $\mathcal{K}^{\mathrm{I}}_n(\epsilon)$ and $\mathcal{K}^{\mathrm{II}}_n(\epsilon)$ be the capital processes that result from using Algorithm 1 with parameter ϵ in Protocol 3. Then

$$\mathcal{K}^{\mathrm{I}}_n := \sum_{j=1}^{\infty} 2^{-j}\mathcal{K}^{\mathrm{I}}_n(2^{-j}) \quad \text{and} \quad \mathcal{K}^{\mathrm{II}}_n := \sum_{j=1}^{\infty} 2^{-j}\mathcal{K}^{\mathrm{II}}_n(2^{-j})$$

are also capital processes for a valid strategy for Sceptic, and $\sqrt{\mathcal{K}^{\mathrm{I}}_n\mathcal{K}^{\mathrm{II}}_n}$ will be unbounded unless $H(P^{\mathrm{I}}_n, P^{\mathrm{II}}_n) \to 1$. This completes the proof.

Remark 11. The proof given in this section relies on the method of combining probability measures P^{I} and P^{II} known as "geometric pooling"; see, e.g., Pettigrew and Weisberg (2024) for a recent review.

4.2 Proof of Theorem 9

We follow the proof of Theorem 4 in Sect. 4.1 modifying it slightly. We still use Algorithm 1, but now P^{I}_n and P^{II}_n have a different meaning: both of them are predictions for the whole sequence of observations $y_1, y_2, \dots$ rather than only for the future observations $y_n, y_{n+1}, \dots$ (there is no clear-cut division between past and future under radical probabilism).

It will be convenient to say that in (4) in Protocol 5 Sceptic invests in the x-tickets at step $n-1$, with his initial investment being $P^{\mathrm{I}}_{n-1}([x])$ and his payoff being $P^{\mathrm{I}}_n([x])$ for each x-ticket. Since Sceptic can hold the same position $f^{\mathrm{I}}_n(x)$ over a number of steps, he can invest in the x-tickets at step n receiving a payoff at step $N > n$, with the initial investment being $P^{\mathrm{I}}_n([x])$ and the final payoff being $P^{\mathrm{I}}_N([x])$.

Now we modify the argument leading to the possibility to increase Sceptic's capital (6)-fold (where (6) is the equation number). Sceptic can invest $P_n^{\mathrm{I}}(x)$ in the x-ticket at step n, and his payoff at step $N > n$ will be $P_N^{\mathrm{I}}(x)$, for any $x \in \mathbf{Y}^m$. Therefore, Sceptic can invest $\sqrt{P_n^{\mathrm{I}}([x])P_n^{\mathrm{II}}([x])}$ for a payoff of

$$\frac{\sqrt{P_n^{\mathrm{I}}([x])P_n^{\mathrm{II}}([x])}}{P_n^{\mathrm{I}}([x])} P_N^{\mathrm{I}}(x).$$

Investing in each $x \in \mathbf{Y}^m$ results in a payoff of

$$\sum_{x\in\mathbf{Y}^m} \frac{\sqrt{P_n^{\mathrm{I}}([x])P_n^{\mathrm{II}}([x])}}{P_n^{\mathrm{I}}([x])} P_N^{\mathrm{I}}(x)$$

for the initial investment of

$$\sum_{x\in\mathbf{Y}^m} \sqrt{P_n^{\mathrm{I}}([x])P_n^{\mathrm{II}}([x])} = H_m(P_n^{\mathrm{I}}, P_n^{\mathrm{II}}).$$

Therefore, starting from step 5 of Algorithm 1 the capital $\mathcal{K}_n^{\mathrm{I}}$ can be multiplied by

$$\frac{1}{H_m(P_n^{\mathrm{I}}, P_n^{\mathrm{II}})} \sum_{x\in\mathbf{Y}^m} \frac{\sqrt{P_n^{\mathrm{I}}([x])P_n^{\mathrm{II}}([x])}}{P_n^{\mathrm{I}}([x])} P_N^{\mathrm{I}}([x]) \tag{7}$$

by step N (later we will make N large, definitely $N > m$), and starting from step 6 the capital $\mathcal{K}_n^{\mathrm{II}}$ will be multiplied by

$$\frac{1}{H_m(P_n^{\mathrm{I}}, P_n^{\mathrm{II}})} \sum_{x\in\mathbf{Y}^m} \frac{\sqrt{P_n^{\mathrm{I}}([x])P_n^{\mathrm{II}}([x])}}{P_n^{\mathrm{II}}([x])} P_N^{\mathrm{II}}([x]) \tag{8}$$

by step N.

Using the weak convergence $P_N^{\mathrm{I}} - P_N^{\mathrm{II}} \to 0$ and the uniform continuity of (7) as function of $P_N^{\mathrm{I}}|_m$ and of (8) as function of $P_N^{\mathrm{II}}|_m$, we can replace, asymptotically, $P_N^{\mathrm{I}}|_m$ and $P_N^{\mathrm{II}}|_m$ by the same probability measure P on $\mathbf{Y}^m$, and so the geometric mean of (7) and (8) can be bounded below, for a sufficiently large N, by the last term of the chain

$$\begin{aligned}
&\frac{1}{H_m(P_n^{\mathrm{I}}, P_n^{\mathrm{II}})} \sqrt{\left(\sum_{x\in\mathbf{Y}^m} \frac{\sqrt{P_n^{\mathrm{I}}([x])P_n^{\mathrm{II}}([x])}}{P_n^{\mathrm{I}}([x])} P([x])\right)\left(\sum_{x\in\mathbf{Y}^m} \frac{\sqrt{P_n^{\mathrm{I}}([x])P_n^{\mathrm{II}}([x])}}{P_n^{\mathrm{II}}([x])} P([x])\right)} \\
&\geq \frac{1}{H_m(P_n^{\mathrm{I}}, P_n^{\mathrm{II}})} \sum_{x\in\mathbf{Y}^m} \sqrt{\frac{\sqrt{P_n^{\mathrm{I}}([x])P_n^{\mathrm{II}}([x])}}{P_n^{\mathrm{I}}([x])} \frac{\sqrt{P_n^{\mathrm{I}}([x])P_n^{\mathrm{II}}([x])}}{P_n^{\mathrm{II}}([x])}} P([x]) \\
&= \frac{1}{H_m(P_n^{\mathrm{I}}, P_n^{\mathrm{II}})} > \frac{1}{1-\epsilon},
\end{aligned}$$

where the "$\geq$" follows from the concavity of the geometric mean function $(u, v) \in [0, \infty)^2 \mapsto \sqrt{uv}$ and Jensen's inequality (Ferguson, 1967, Lemma 2.8.1). Therefore, we can wait until the geometric mean of the capitals increases $\frac{1}{1-\epsilon}$-fold. The proof is completed as before, by mixing over ϵ. Notice that in our construction Sceptic's capital cannot become negative.

5 Comparison with known results

In this section we will discuss two kinds of Jeffreys's law: for predicting the infinite future (Sect. 5.1) and for one-step-ahead prediction (Sect. 5.2).

5.1 Blackwell–Dubins result

The main topic of this subsection is limitations of Blackwell and Dubins's classical result (1962, Sect. 2) and how they are overcome by this paper's results.

From the purely mathematical point of view, one limitation of Blackwell and Dubins's result is that they consider two probability measures, Q and P (which correspond to our Forecasters I and II) such that Q is absolutely continuous w.r. to P, denoted $Q \ll P$. This requirement means that $Q(E) = 0$ for any event E such that $P(E) = 0$. Let us write P^{I} and P^{II} for Q and P, respectively.

We regard P^{I} as a base forecasting strategy; the corresponding nth forecast P_n^{I} is the conditional probability of the future observations $y_n, y_{n+1}, \dots$ given the past observations $y_1, \dots, y_{n-1}$, with the observations $y_1, y_2, \dots$ generated from P^{I}. Then P^{II} is an alternative forecasting strategy producing, in a similar manner, $P_1^{\mathrm{II}}, P_2^{\mathrm{II}}, \dots$. The condition $P^{\mathrm{I}} \ll P^{\mathrm{II}}$ can be interpreted as P^{II} being at least as adaptive as P^{I}. Blackwell and Dubins's result says that P_n^{I} and P_n^{II} converge in total variation with P^{I}-probability 1 whenever P^{I} agrees with P^{II}, in the sense of P^{II} being at least as adaptive as P^{I}. (This ignores technical issues surrounding the existence of conditional distributions, which are less acute in our current context of a finite observation space $\mathbf{Y}$.)

What if the two forecasting strategies do not agree (we have neither $P^{\mathrm{I}} \ll P^{\mathrm{II}}$ nor $P^{\mathrm{II}} \ll P^{\mathrm{I}}$)? For example, suppose that they agree perfectly everywhere apart from two events, E^{I} and E^{II}, for which

$$P^{\mathrm{I}}(E^{\mathrm{I}}) = P^{\mathrm{II}}(E^{\mathrm{II}}) = 10^{-6}, \quad P^{\mathrm{II}}(E^{\mathrm{I}}) = P^{\mathrm{I}}(E^{\mathrm{II}}) = 0.$$

For this case the Blackwell–Dubins result does not say anything. To prevent the possibility of waiting until one or both events become settled, suppose further that, for all n,

$$P^{\mathrm{I}}(E^{\mathrm{I}} \mid y_1, \dots, y_n), P^{\mathrm{II}}(E^{\mathrm{I}} \mid y_1, \dots, y_n), P^{\mathrm{I}}(E^{\mathrm{II}} \mid y_1, \dots, y_n), P^{\mathrm{II}}(E^{\mathrm{II}} \mid y_1, \dots, y_n) \in (0,1).$$

In our context, the condition of absolute continuity becomes not only restrictive but also less natural. Namely, in our framework, there are no *a priori* connections between the probability forecasts $P_1^{\mathrm{I}}, P_2^{\mathrm{I}}, \dots$ and $P_1^{\mathrm{II}}, P_2^{\mathrm{II}}, \dots$ output at different steps, and so it is possible to have $P_1^{\mathrm{I}} \ll P_1^{\mathrm{II}}$ followed by $P_2^{\mathrm{I}} \perp P_2^{\mathrm{II}}$ ($\perp$ meaning mutual singularity) followed by $P_3^{\mathrm{I}} \ll P_3^{\mathrm{II}}$, etc.

The condition of absolute continuity is very natural (or even unavoidable) under the Bayesian interpretation of Blackwell and Dubins's result (both Blackwell and Dubins were Bayesians (DeGroot, 1986, pp. 43–44 and p. 48)). Its Bayesian interpretation is that Forecaster I believes that the forecasts issued by the two forecasters will converge in total variation. Bayesian theory is based on personal probability, and for a Bayesian interpretation it is essential to indicate

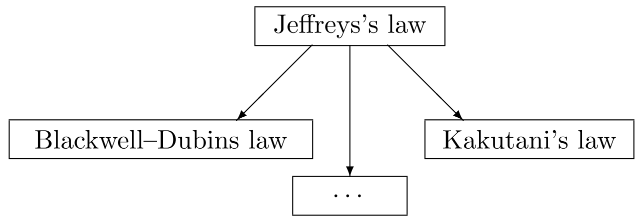


Figure 2: Two known special cases of Jeffreys's law

the person whose beliefs we are talking about. It is difficult to see who can be such a person without the requirement of absolute continuity.

Our interpretation of Theorem 9 was in terms of boosting: weak convergence can be boosted to convergence in total probability. Blackwell and Dubins's result can also be interpreted in these terms: agreement between two forecasting strategies in the sense of $P^{\mathrm{I}} \ll P^{\mathrm{II}}$ can be boosted to convergence in total variation P^{I}-almost surely.

The game-theoretic version of Blackwell and Dubins's result proved in Sect. 2 of this paper not only removes some of the unnecessary restrictions but is also more constructive. We have an explicit strategy for Sceptic that discredits at least one of the forecasters by successful betting against his forecasts unless the predictions that the forecasters output converge in total variation. While Blackwell and Dubins's result under its Bayesian interpretation only concerns a Bayesian's beliefs, our result establishes connections with idealized reality.

5.2 One-step-ahead prediction

Figure 2 represents two known special cases of Jeffreys's law. Blackwell and Dubins's result is an instance of one of them; more generally, Fig. 2 refers to the *Blackwell–Dubins law* as an asymptotic qualitative statement about convergence between forecasts for the infinite future (or, in the case of radical probabilism, infinite past, present, and future). Theorem 4 and its generalization Theorem 9 are also results of this kind. Despite overcoming some limitations of Blackwell and Dubins's result, they are still asymptotic and qualitative.

A very different approach to Jeffreys's law was pioneered by Kakutani (1948), with a crucial step made by Kabanov et al. (1977). Their results also show that, provided $P^{\mathrm{I}} \ll P^{\mathrm{II}}$, the one-step-ahead predictions computed from P^{I} and P^{II} converge, but the nature of these results very different. Their important advantage is that they are quantitative (albeit with interesting qualitative implications); the price to pay, however, is that they only cover one-step ahead prediction.

The results by Kakutani (1948) and Kabanov et al. (1977) are measure-theoretic and have similar disadvantages to those of the Blackwell–Dubins result discussed in Sect. 5.1, but these disadvantages were eliminated in the game-theoretic versions described in Shafer and Vovk (2019, Sect. 10.7) (and developed in the references given in Shafer and Vovk 2019, Sect. 10.9). One such result is that in Protocol 3 Sceptic can ensure

$$\ln \mathcal{K}_n^{\mathrm{I}} + \ln \mathcal{K}_n^{\mathrm{II}} \geq \frac{1}{4} \sum_{i=1}^{n} \left\| (P_i^{\mathrm{I}}|_1) - (P_i^{\mathrm{II}}|_1) \right\|^2 \tag{9}$$

for all n, where $P_i^{\mathrm{I}}|_1$ and $P_i^{\mathrm{II}}|_1$ are the restrictions of P_i^{I} and P_i^{II}, respectively, to $\mathbf{Y}$, as defined earlier (i.e., they are the one-step-ahead restrictions). (To obtain (9), combine Proposition 10.17 in Shafer and Vovk (2019) with the standard bound for Hellinger distance, which is essentially the right-hand side of (5) (Shiryaev, 2016, Theorem 3.9.1, (23)).)

It is clear that (9) implies Theorem 4 with P_n^{I} and P_n^{II} replaced by $P_n^{\mathrm{I}}|_1$ and $P_n^{\mathrm{II}}|_1$, respectively. However, (9) is an explicit lower bound rather than merely an asymptotic result. Figure 2 refers to the class of such quantitative one-step-ahead results as *Kakutani's law*. The ellipsis represents new special cases of Jeffreys's law, those yet to be discovered.

6 Conclusion

A big disadvantage of Theorems 4 and 9, inherited from their measure-theoretic prototype (Blackwell and Dubins, 1962), is that they are merely asymptotic; it is not obvious how to make them quantitative. The versions given in Shafer and Vovk (2019, Sect. 10.7) are much more precise, but they cover only one-step-ahead forecasting. Bridging the gap between these two very different kinds of results (depicted in Fig. 2) looks to me an interesting direction of further research.

Other possible directions of further research are:

- Generalizing our testing protocols and results to the case of an infinite observation space $\mathbf{Y}$.

- Are there situations where a wider set of permitted moves for Sceptic would be useful? In Protocols 1–3 and 5 we considered Sceptic's moves f such that $f(x) = 0$ apart from finitely many x. More generally, we could permit f such that $f(x) \neq 0$ for arbitrarily long x, but this would require careful analysis of convergence of the resulting infinite series in expressions for Sceptic's capital.

Acknowledgments

Many thanks to A. Philip Dawid for his questions and comments. Research on this paper has been partially supported by Mitie.

References

Anastasios N. Angelopoulos and Stephen Bates. *Conformal Prediction: a Gentle Introduction*, volume 16:4 of *Foundations and Trends in Machine learning*. Springer, NOW, 2023.

David Blackwell and Lester Dubins. Merging of opinions with increasing information. *Annals of Mathematical Statistics*, 33:882–886, 1962.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, Cambridge, 2006.

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

Bruno de Finetti. La prévision, ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7:1–68, 1937. English translation: Foresight: Its logical laws, its subjective sources. In Kyburg and Smokler (1980) (both first and second editions).

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

Morris H. DeGroot. A conversation with David Blackwell. *Statistical Science*, 1:40–53, 1986.

Thomas S. Ferguson. *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York, 1967.

Richard C. Jeffrey. Probable knowledge. In Imre Lakatos, editor, *The Problem of Inductive Logic*, pages 166–190. Elsevier, 1968. This paper is included as Chap. 3 (pp. 30–43) of Jeffrey (1992).

Richard C. Jeffrey. De Finetti's radical probabilism. Richard C. Jeffrey Papers archive at the University of Pittsburgh. Accessed online in April 2024, 1988.

Richard C. Jeffrey. *Probability and the Art of Judgment*. Cambridge University Press, Cambridge, 1992.

Harold Jeffreys. Science, logic and philosophy. *Nature*, 141:672–676 and 716–719, 1938.

Harold Jeffreys. *Theory of Probability*. Oxford University Press, Oxford, third edition, 1961.

Yury M. Kabanov, Robert Sh. Liptser, and Albert N. Shiryaev. To the question of absolute continuity and singularity of probability measures. *Mathematics of the USSR—Sbornik*, 33:203–221, 1977.

Shizuo Kakutani. On equivalence of infinite product measures. *Annals of Mathematics, Second Series*, 49:214–224, 1948.

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

Charles Sanders Peirce. The fixation of belief. *Popular Science Monthly*, 12: 1–15, November 1877.

Richard Pettigrew and Jonathan Weisberg. Geometric pooling: A user's guide. *British Journal for the Philosophy of Science*, 2024. To appear.

Karl R. Popper. *The Open Society and Its Enemies*. Routledge, London, 1945. In 2 volumes. Fifth edition (revised): 1966.

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

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

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

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

Vladimir Vovk. The diachronic Bayesian. Technical Report arXiv:2308.13069 [stat.ME], arXiv.org e-Print archive, March 2024.

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

Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, Cham, second edition, 2022.