Prequential probability: principles and properties

A. PHILIP DAWID[1]* and VLADIMIR G. VOVK[2]

[1]*Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, UK. e-mail: dawid@stats.ucl.ac.uk*
[2]*Department of Computer Science, Royal Holloway, University of London, Egham Hill, Egham, Surrey TW20 0EX, UK. e-mail: vovk@dcs.rhbnc.ac.uk*

Forcaster has to predict, sequentially, a string of uncertain quantities $(X_1, X_2, \ldots)$, whose values are determined and revealed, one by one, by Nature. Various criteria may be proposed to assess Forecaster's empirical performance. The *weak prequential principle* requires that such a criterion should depend on Forecaster's behaviour or strategy only through the actual forecasts issued. A wide variety of appealing criteria are shown to respect this principle. We further show that many such criteria also obey the *strong prequential principle*, which requires that, when both Nature and Forecaster make their choices in accordance with a common joint distribution **P** for $(X_1, X_2, \ldots)$, certain stochastic properties, underlying and justifying the criterion and inferences based on it, hold regardless of the detailed specification of **P**. In order to understand further this compliant behaviour, we introduce the *prequential framework*, a game-theoretic basis for probability theory in which it is impossible to violate the prequential principles, and we describe its connections with classical probability theory. In this framework, in order to show that some criterion for assessing Forecaster's empirical performance is valid, we have to exhibit a winning strategy for a third player, Statistician, in a certain perfect-information game. We demonstrate that many performance criteria can be formulated and are valid in this framework and, therefore, satisfy both prequential principles.

Keywords: farthingale; forecasting; goodness of fit; limit theorems of probability theory; martingale; perfect-information game; strong prequential principle; weak prequential principle

1. Introduction

The prequential approach to statistics (Dawid 1984, 1991, 1992a,b) is based on the idea that we can judge the quality of an inference method by converting it into a forecasting system and assessing the empirical success of the sequence of one-step-ahead forecasts that it implies. In this context, it is natural to impose what we here term the "weak prequential principle", which requires that any criterion for assessing the empirical validity of some Forecaster should depend only on the actual one-step-ahead forecasts that he/she issues, and the specific outcomes chosen by Nature, and not otherwise on Forecaster's strategy. We are particularly concerned with forecasts which have an interpretation in terms of a (perhaps

---

*To whom correspondence should be addressed.

partially specified) probability distribution, expressing Forecaster's uncertainty about the next observable. Then Forecaster might be thought of as having issued his forecasts in accordance with some overall joint distribution **P** for all the observables $(X_1, X_2, \ldots)$. The weak prequential principle then asserts the irrelevance of any aspect of **P**, beyond the sequence of probability forecasts actually made (which are obtained by conditioning on the actually realized past outcome values).

When we come to apply such a validity criterion, we typically want to examine stochastic measures, such as significance levels, which will generally depend on the full specification of **P**. However, if all we have observed is the string of forecasts issued by Forecaster, and the outcomes determined by Nature, we shall not be in a position fully to determine **P**. The "strong prequential principle" requires that the dependence of the stochastic properties of the criterion on **P** should vanish, at least asymptotically; when this applies, it licenses stochastic conclusions on the basis of the realized value of the validity criterion, even in the absence of full knowledge of **P**. In particular, meaningful stochastic inferences may then be made, given only the observed sequences of forecasts and outcomes, and no further information about **P**. It is a remarkable fact that many useful inferences do in fact have this strong property.

In this paper we first illustrate the above considerations for a variety of appealing criteria and then, in an attempt to understand this behaviour, introduce a new game-theoretic framework for probability theory, the "prequential framework", which is particularly suited to the study of such problems.

Section 2 introduces a variety of forecasting tasks, involving binary or uniformly bounded continuous variables. In Section 3 we consider several validity criteria, with various probabilistic motivations, which respect the weak prequential principle, while in Section 4 we demonstrate that these all respect the strong prequential principle also. Section 6 introduces the idea of a prequential game, which involves three players: Forecaster, Nature and Statistician. No underlying probabilistic model is required for this. In Section 7 the prequential framework is used to give definitions of, first, "full" ("almost certain") events and then "prequential probability". It is shown how results justifying the use of our performance criteria may be derived in this framework, with proofs often paralleling standard proofs of well-known probabilistic theorems ... except that now it is not necessary to assume even the existence of an underlying probability distribution. However, once the prequential result is obtained, the standard probabilistic result then follows and, because of its prequential provenance, must hold irrespective of the assumed underlying probability distribution. This then provides the desired explanation of the good behaviour of the criteria considered.

Two appendices provide sample proofs of required results within the prequential framework; specifically, they are related to the proofs of the strong law of large numbers and the law of the iterated logarithm.

2. Forecasting tasks

Every forecasting task involves at least two participants: Nature and Forecaster. Successively, for $i = 1, 2, \ldots$, Nature determines a value x_i for some uncertain quantity X_i. In this paper

we shall only consider the simplest cases where X_i take values in (i) a two-element set (the binary case) (we shall always take this set to be $\{0, 1\}$) or (ii) a closed interval of the real line $\mathbb{R}$ (we shall always take it to be $[-1, 1]$). Typical examples are as follows: $X_i = 1$ if the ith day is rainy and $X_i = 0$ if not (the first case); X_i is the maximum temperature on the ith day (in the second case, to fit the temperature into the interval $[-1, 1]$ we can take, say, 1000 °C as our unit of measurement).

Just before X_i is disclosed by Nature, Forecaster issues a forecast F_i for X_i. This can be a probability forecast, or something weaker, as the following examples demonstrate.

Example 1. *Probability forecasting of binary outcomes.* X_i takes values in $\{0, 1\}$; the allowed forecasts are $F_i = P_i \in [0, 1]$ (interpreted as the probability that $X_i = 1$).

Example 2. *Mean-value forecasting of continuous outcomes.* X_i takes values in $[-1, 1]$; the allowed forecasts are $F_i = M_i \in [-1, 1]$ (interpreted as Forecaster's expected value, at the time of issuing the forecast, of X_i).

Example 3. *Mean-value-and-variance forecasting of continuous outcomes.* X_i takes values in $[-1, 1]$; the allowed forecasts are pairs $F_i = (M_i, D_i)$ with $M_i \in [-1, 1]$ and $D_i \geqslant 0$ (with M_i interpreted as the expected value and D_i as the variance, at the time of issuing the forecast, of X_i).

Example 4. *Confidence-interval forecasting of continuous outcomes.* X_i takes values in $[-1, 1]$; the allowed forecasts are pairs $F_i = (L_i, U_i)$ such that $-1 \leqslant L_i \leqslant U_i \leqslant 1$, with the following interpretation: just before X_i is disclosed, the odds are at least 99 to 1 that $X_i \in [L_i, U_i]$.

Example 5. *Probability forecasting of continuous outcomes.* X_i takes values in $[-1, 1]$; the allowed forecasts F_i are arbitrary probability distributions P_i on $[-1, 1]$, with $P_i(E)$, $E \subseteq [-1, 1]$, interpreted as the forecast probability that $X_i \in E$. A slight variation in this forecasting task is where Forecaster is required to issue only *continuous* forecasts P_i (i.e. such that $P_i(\{x\}) = 0$ for any point $x \in [-1, 1]$).

3. Weak prequential principle

We wish to consider how good Forecaster is at forecasting Nature's outcomes. In some cases, each of Forecaster and Nature may use deterministic or stochastic strategies for deciding their moves (forecasts and outcomes, respectively). In others, these moves may be regarded as simply produced when required, with no underlying strategy (an example of this might be a weather forecaster's sequence of daily probabilities of precipitation in the next 24 h). Let f_i be Forecaster's issued value of his forecast F_i for X_i, and x_i the realized outcome of X_i. (In general, we denote uncertain quantities by capital letters and the specific values that they take by the corresponding small letters.) The *weak prequential principle* requires that any criterion for assessing the "agreement" between Forecaster and Nature should depend only on the

actual observed sequences $(f_1, f_2, \ldots)$ and $(x_1, x_2, \ldots)$, and not further on the strategies (if any) which might have produced these.

3.1. Violations

To clarify ideas, we start by exhibiting two criteria which do *not* respect the weak prequential principle.

3.1.1. Marginal probabilities

Suppose that the X_i are binary, and that Forecaster uses a *forecasting strategy*; this is a rule which determines his probability forecast for X_i:

$$F_i = P_i := \mathbf{P}(X_i = 1 | X_1, \ldots, X_{i-1}), \tag{1}$$

for any i and any values of the previous outcomes $(X_1, \ldots, X_{i-1})$. From all these conditional probabilities we can build up a full joint distribution $\mathbf{P}$ over the infinite sequence (X_i) and conversely, starting from any such distribution $\mathbf{P}$, we can define a strategy for Forecaster by (1) (at any rate, for outcome sequences which are not assigned zero probability). Thus, we can regard Forecaster's strategy as equivalent to his using the joint distribution $\mathbf{P}$ to model Nature's outcomes.

For this distribution, let π_i denote the *marginal* probability $\mathbf{P}(X_i = 1)$. Then we might impose, as our agreement criterion, the requirement that

$$\frac{1}{n} \sum_{i=1}^{n} (x_i - \pi_i) \to 0 \tag{2}$$

as $n \to \infty$.

In this case, Forecaster's actually issued one-step-ahead probability forecasts would be given by

$$f_i = p_i := \mathbf{P}(X_i = 1 | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}), \tag{3}$$

where $x_1, x_2, \ldots$ are the actual outcomes chosen by Nature. It is clear that, given only the sequences (x_i) and (p_i), but not the underlying distribution $\mathbf{P}$ used to construct these forecasts, it is impossible to determine the marginal probabilities (π_i), and this can readily be confirmed by constructing different $\mathbf{P}$, having different marginal probabilities (π_i), which yield identical forecasts (p_i) for some data sequence (x_i). Consequently the criterion (2) does not depend only on the observed sequences (p_i) and (x_i), but further on the strategy $\mathbf{P}$ which Forecaster uses; it thus violates the weak prequential principle.

3.1.2. Significance level

Consider a problem of probability forecasting of (say) continuous outcomes, where again a strategy for Forecaster may be expressed in terms of his underlying joint probability model $\mathbf{P}$

for the (X_i). Let us take a test statistic T, and assess the agreement beteween Forecaster and Nature by means of the *observed significance level*

$$\alpha := \mathbf{P}(T \geqslant t), \tag{4}$$

t being the value of T for the specific sequence (x_i). Again it is clear that, although quoting the *value t* of the test statistic (the "nominal inference" (Dawid 1983)) would respect the weak prequential principle, the associated significance level α (the "stochastic inference") will typically depend further on $\mathbf{P}$ and thus would not do so.

   We shall nevertheless see later that, for suitable test statistics, the dependence of α on $\mathbf{P}$ may be asymptotically negligible. Both types of behaviour are exhibited in Example 6 below.

   Our further examples will be formulated directly in terms of the sequences of forecasts and outcomes, so that they will automatically respect the weak prequential principle.

3.2. Probability forecasting of binary outcomes

Let $f_i = p_i$ be the issued probability of the outcome 1 for the binary variable X_i, and x_i the realized value of X_i.

3.2.1. Calibration

We might require that

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - p_i) \to 0 \tag{5}$$

as $n \to \infty$. This is the "simple calibration requirement" (Dawid 1982, 1986).

   An important generalization of requirement (5) is that we might average, not over all i, but over some "predictable" sequence i_k; if $(i_1, i_2, \ldots)$ is an increasing sequence of predictable finite stopping times, we can require that

$$\frac{1}{n}\sum_{k=1}^{n}(x_{i_k} - p_{i_k}) \to 0 \tag{6}$$

as $n \to \infty$. We might then regard the (p_i) as being poor probability assessments if (5), or (6) for a suitable subsequence or subsequences, were to fail. Dawid (1985) showed that two sequences of forecasts which both satisfy (6) for sufficiently many subsequences must be asymptotically equivalent.

3.2.2. Central limit theorem

Another criterion (Seillier-Moiseiwitsch and Dawid 1993) is motivated by the central limit theorem. Choose a large constant $B > 0$, and wait until the first n such that

$$\sum_{i=1}^{n} p_i(1 - p_i) \geqslant B. \tag{7}$$

Then we could require that, for this n,

$$B^{-1/2} \sum_{i=1}^{n} (x_i - p_i) \tag{8}$$

be "not large" as judged by the distribution $N(0, 1)$. We could quote, as a measure of disagreement, the "observed significance level" obtained by referring (8) to $N(0, 1)$, a small value counting against the validity of the issued forecasts. Once again, it is possible to apply such a criterion to predictable sequences i_k and further, in this case, to treat values calculated from disjoint subsequences as independent, thus allowing "omnibus" reference of a sum of squares of such values to a χ^2 distribution (Seillier-Moiseiwitsch and Dawid 1993).

3.2.3. Law of the iterated logarithm

Yet another criterion is related to the law of the iterated logarithm. We might require that *either*

$$\sum_{i=1}^{\infty} p_i(1 - p_i) < \infty \tag{9}$$

or

$$\limsup_{n \to \infty} \sum_{i=1}^{n} (x_i - p_i) \bigg/ \left[2 \left\{ \sum_{i=1}^{n} p_i(1 - p_i) \right\} \ln \ln \left\{ \sum_{i=1}^{n} p_i(1 - p_i) \right\} \right]^{1/2} = 1. \tag{10}$$

In other words, we require (10) to hold so long as the denominator of its left-hand side tends to infinity, and we would thus reject the validity of the forecasts (p_i) if both (9) and (10) were to fail. Again, one could extend consideration to predictable subsequences.

3.3. Mean-value forecasting of continuous outcomes

Now let $f_i = m_i$ be the issued mean-value forecast for the real quantity X_i, with realized value x_i.

3.3.1. Strong law of large numbers

For this forecasting task we might require that (in accordance with the strong law of large numbers)

$$\frac{1}{n} \sum_{i=1}^{n} (x_i - m_i) \to 0 \ (n \to \infty). \tag{11}$$

3.4. Mean-value-and-variance forecasting of continuous outcomes

Now $f_i = (m_i, d_i)$. Here natural requirements (in addition to (11)) are the following.

3.4.1. Central limit theorem

Select a large number $B > 0$, wait until

$$\sum_{i=1}^{n} d_i \geqslant B$$

and refer

$$B^{-1/2} \sum_{i=1}^{n} (x_i - m_i) \tag{12}$$

to $N(0, 1)$.

3.4.2. Law of the iterated logarithm

Require that *either*

$$\sum_{i=1}^{\infty} d_i < \infty \tag{13}$$

or

$$\limsup_{n \to \infty} \sum_{i=1}^{n} (x_i - m_i) \Big/ \left\{ 2 \left(\sum_{i=1}^{n} d_i \right) \ln \ln \left(\sum_{i=1}^{n} d_i \right) \right\}^{1/2} = 1. \tag{14}$$

3.5. Confidence-interval forecasting of continuous outcomes

If Forecaster claims that just before observing each X_i the odds in favour of $X_i \in [L_i, U_i]$ are at least 99:1, we could require that

$$\liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}(x_i \in [l_i, u_i]) \geqslant 0.99, \tag{15}$$

where $\mathbf{I}(E)$ is defined to be 1 if event E happens and 0 if not. We could similarly require that, for each increasing sequence i_k of predictable finite stopping times,

$$\liminf_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{I}(x_{i_k} \in [l_{i_k}, u_{i_k}]) \geqslant 0.99. \tag{16}$$

3.6. Probability forecasting of continuous outcomes

In the rest of this section, $f_i = p_i$ is the issued probability distribution for X_i.

3.6.1. Strong law of large numbers

Let Z_i, $i = 1, 2, \ldots$, be arbitrary uniformly bounded measurable functions of $(X_1, \ldots, X_i)$. Put

$$z_i = Z_i(x_1, \ldots, x_i),$$

$$m_i = \mathrm{E}_{Y \sim p_i} Z_i(x_1, \ldots, x_{i-1}, Y);$$

these are the "realized" values of Z_i and of the one-step-ahead expectations of Z_i (we assume that the latter exist). Then require

$$\frac{1}{n} \sum_{i=1}^{n} (z_i - m_i) \to 0 \ (n \to \infty). \tag{17}$$

3.6.2. Central limit theorem; law of the iterated logarithm

In a similar way we could also introduce the one-step-ahead variances

$$d_i = \mathrm{var}_{Y \sim p_i} Z_i(x_1, \ldots, x_{i-1}, Y)$$

and formulate criteria motivated by the central limit theorem and the law of the iterated logarithm.

3.7. Prequential decision theory

Fix a sequence of loss functions $\lambda_i(x, a)$, x ranging over the outcome space $[-1, 1]$ and a over some fixed *decision space* $\mathscr{A}$. Forecaster's loss at stage i, if he takes decision a_i and Nature then produces x_i, is $\lambda_i(x_i, a_i)$. We assume that the (λ_i) are uniformly bounded. For every possible forecast p (a distribution over $[-1, 1]$), take $B_i(p)$ to be the Bayes act under $X \sim p$ for loss function λ_i, i.e.

$$B_i(p) = \arg \min_{a \in \mathscr{A}} \int \lambda_i(x, a) p \, \mathrm{d}x$$

(assuming that this arg min exists). We could consider $B_i(p)$ to be the decision that Forecaster would take at stage i, if he issues forecast distribution p for X_i. We can then assess the quality of the forecasts (p_i) by the cumulative loss of their implied decisions:

$$\sum_{i=1}^{n} \lambda_i(x_i, B_i(p_i)).$$

We thus might reject (p_i) if some other forecaster is able to issue forecasts (q_i) such that

$$\lim_{n\to\infty}\left(\sum_{i=1}^{n}\lambda_i(x_i,\,B_i(p_i)) - \sum_{i=1}^{n}\lambda_i(x_i,\,B_i(q_i))\right) = \infty. \tag{18}$$

More generally, we might reject the (p_i) if some other decision maker, with the same information as Forecaster, is able to select actions a_i (a_i selected before x_i is disclosed) such that

$$\lim_{n\to\infty}\left(\sum_{i=1}^{n}\lambda_i(x_i,\,B_i(p_i)) - \sum_{i=1}^{n}\lambda_i(x_i,\,a_i)\right) = \infty. \tag{19}$$

3.8. Cumulative probability integral transform

Forecaster is now required to assess a probability distribution P_i for real X_i, having continuous cumulative distribution function. Let p_i be the issued distribution. Define $u_i = \mathrm{prob}_{Y\sim p_i}(Y \leqslant x_i)$. Then we might assess (in various possible ways) the (u_i) for agreement with the hypothesis that they were obtained by random sampling from $U[0, 1]$ (the uniform probability distribution in $[0, 1]$). For example, we could formally apply the Kolmogorov–Smirnov test for uniformity or, as a test for independence, the uniform conditional test (Cox and Lewis 1966). A wide range of other tests could be similarly performed.

3.9. Farthingale

Fix a sequence of quantities

$$S_i = S_i(P_1,\,X_1,\,\ldots,\,P_i,\,X_i), \qquad i = 0,\,1,\,\ldots, \tag{20}$$

such that all $S_i \geqslant 0$, $S_0 = 1$, and, for all $i \geqslant 1$, and fixed $P_1,\,X_1,\,P_2,\,X_2,\,\ldots,$

$$\mathrm{E}_{Y\sim P_i}S_i(P_1,\,X_1,\,\ldots,\,P_{i-1},\,X_{i-1},\,P_i,\,Y) = S_{i-1}(P_1,\,X_1,\,\ldots,\,P_{i-1},\,X_{i-1}). \tag{21}$$

We shall call a sequence satisfying (21) a *farthingale*. It is similar to a martingale, except that the distribution used for taking the expectation in (21) is specified internally, rather than being constructed from an external global distribution for the (X_i).

We can interpret S_i as the accumulated capital, after game i, of Gambler playing a sequence of games against Forecaster, each of which is regarded as fair by Forecaster at the time it is offered. Gambler starts with £1 and never risks going into debt. If Forecaster is "successful" in describing Nature's outcomes, it should not be possible for Gambler to win an unlimited amount. Accordingly, we might require, as a validity criterion, that the realized sequence

$$s_i = S_i(p_1,\,x_1,\,\ldots,\,p_i,\,x_i)$$

be bounded.

As an elaboration, for finite as well as infinite outcome sequences, we might use

$\alpha = (\sup_i s_i)^{-1}$ as a measure of mismatch between forecasts and outcomes, small values of α being "significant".

As a particular case, suppose all p_i are required to have positive densities (which will be denoted by the same symbols). Let **Q** be some fixed strategy whereby Gambler produces a positive density q_i for X_i after seeing $p_1, x_1, \ldots, p_{i-1}, x_{i-1}, p_i$. Putting $S_0 = 1$ and

$$S_i(p_1, x_1, \ldots, p_i, x_i) = S_{i-1}(p_1, x_1, \ldots, p_{i-1}, x_{i-1}) \frac{q_i(x_i)}{p_i(x_i)}, \ i \geq 1,$$

where q_i is calculated from $p_1, x_1, \ldots, p_{i-1}, x_{i-1}, p_i$, we obtain a farthingale. Actually, we can obtain all positive farthingales with initial value 1 this way.

Note that, if q_i depends only on $(x_1, \ldots, x_{i-1})$, so that **Q** can be considered as a joint distribution for the (X_i), and if Forecaster likewise produces his forecasts (p_i) as appropriate conditional probabilities from some joint distribution **P**, then S_i is just the likelihood ratio for testing **Q** against **P** on the basis of data $(x_1, \ldots, x_i)$ and thus reflects the relative performance of Gambler's **Q** as against Forecaster's **P**. If this likelihood ratio were to tend to infinity, then Forecaster should be discredited as being worse at forecasting than Gambler. Even in our more general context, the intuitive force of this argument continues to apply.

Dawid (1985, Section 13.2) proposed, as an agreement criterion, the boundedness of all computable farthingales; Vovk (1993a) suggested a non-algorithmic approach similar to that of this paper. In Sections 6 and 7 below we shall see how the idea underlying the farthingale can be made the cornerstone of a general approach to probability.

4. Probabilistic behaviour

Some motivation and justification for the criteria introduced in the previous section can be given, if we are willing to make the assumption that, for some "underlying" joint distribution **P** for $(X_1, X_2, \ldots)$, Nature simulates her outcome values from **P**, and Forecaster uses the same "true" distribution **P** to construct his forecasts. In such a case, we should surely regard Forecaster as doing as well as possible. Under such a model, various stochastic properties can be derived relating these "ideal" forecasts to observed outcomes. These describe how our criteria should behave in such ideal circumstances and can thus be regarded as validating their use for assessing empirical performance in general.

We assume the usual regularity conditions that ensure the existence of the relevant regular conditional probabilties.

4.1. Probability forecasting of binary outcomes

If **P** is the underlying probability distribution, then the "ideal" probability forecast for the binary quantity X_i is

$$P_i = \mathrm{E}_{\mathbf{P}}(X_i | X_1, \ldots, X_{i-1}). \tag{22}$$

By the usual martingale strong law of large numbers (see, for example, Stout (1974, Theorem 3.3.1)), with **P** probability 1

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - P_i) \to 0 \tag{23}$$

(Dawid 1982). This is true no matter what **P** is and can thus be regarded as a justification for insisting on the calibration criterion (5) as a necessary requirement for "valid" forecasts. Similarly, the stochastic analogue of (6) holds with **P** probability 1, thus justifying its use.

The martingale central limit theorem of Billingsley (1986, Theorem 35.9) implies that the random variable

$$B^{-1/2}\sum_{i=1}^{\tau_B}(X_i - P_i), \tag{24}$$

where τ_B is the smallest n satisfying

$$\sum_{i=1}^{n} P_i(1 - P_i) \geqslant B, \tag{25}$$

should weakly converge to $N(0, 1)$ as $B \to \infty$, provided that

$$\sum_{i=1}^{\infty} P_i(1 - P_i) = \infty \tag{26}$$

with **P** probability 1.

Without invoking condition (26), which is not quite "prequential" (it depends on the behaviour of non-realized forecasts), we can still assert that, for any interval (a, b),

$$\liminf_{B \to \infty} \mathbf{P}\left(B^{-1/2}\sum_{i=1}^{\tau_B}(X_i - P_i) \in (a, b) \text{ or } \tau_B = \infty \right) \geqslant \Phi(b) - \Phi(a), \tag{27}$$

where $\Phi(\cdot)$ is the standard normal distribution function; the interpretation of $\sum_{i=1}^{\tau_B}$ for $\tau_B = \infty$ is obviously not essential to the meaning of (27). It is easy to see that (27) implies the weak convergence of (24) to $N(0, 1)$ when (26) holds almost surely. These considerations then justify the use of criterion (8), judged against a standard normal reference distribution.

The law of the iterated logarithm for the case of binary outcomes asserts that, with **P** probability 1, *either*

$$\sum_{i=1}^{\infty} D_i < \infty \tag{28}$$

or

$$\limsup_{n \to \infty} \sum_{i=1}^{n}(X_i - M_i) \bigg/ \left\{ 2\left(\sum_{i=1}^{n} D_i\right) \ln \ln \left(\sum_{i=1}^{n} D_i\right) \right\}^{1/2} = 1, \tag{29}$$

where $M_i = P_i$ and $D_i = P_i(1 - P_i)$. This follows from the general result of Stout (1970, Theorems 1 and 2).

4.2. Mean-value forecasting of continuous outcomes

Here we have, as "ideal mean-value forecast",

$$M_i = E_{\mathbf{P}}(X_i | X_1, \ldots, X_{i-1}). \tag{30}$$

By the martingale variant of Kolmogorov's strong law of large numbers (Stout 1974, Theorem 3.3.1), with $\mathbf{P}$ probability 1,

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - M_i) \to 0 \ (n \to \infty), \tag{31}$$

which again justifies the use of criterion (11).

4.3. Mean-value-and-variance forecasting of continuous outcomes

In the context of mean-value-and-variance forecasting of continuous outcomes, we have ideal forecasts

$$
\begin{aligned}
M_i &= E_{\mathbf{P}}(X_i | X_1, \ldots, X_{i-1}), \\
D_i &= \operatorname{var}_{\mathbf{P}}(X_i | X_1, \ldots, X_{i-1}).
\end{aligned}
\tag{32}
$$

For justification of the criteria in Section 3.4, it again suffices to refer to the appropriate central limit theorem (Billingsley 1986), and its prequential variant analogous to (27), and also to the general law of the iterated logarithm: with $\mathbf{P}$ probability 1, either (28) or (29) holds (Stout 1970).

4.4. Confidence-interval forecasting of continuous outcomes

Let us assume that Forecaster uses some measurable strategy for selecting (L_i, U_i) as a function of the past outcomes $X_1, \ldots, X_{i-1}$. We can interpret an ideal Forecaster's claim that the odds in favour of $X_i \in [L_i, U_i]$ are at least 99:1 as the following assertion about the underlying probability distribution $\mathbf{P}$:

$$\mathbf{P}(X_i \in [L_i, U_i] | X_1, \ldots, X_{i-1}) \geqslant 0.99, \tag{33}$$

for all i. The martingale analogue of Kolmogorov's strong law of large numbers implies that

$$\frac{1}{n} \sum_{i=1}^{n} \{\mathbf{I}(X_i \in [L_i, U_i]) - \mathbf{P}(X_i \in [L_i, U_i] | X_1 \ldots, X_{i-1})\} \to 0$$

with **P** probability 1, which in its turn implies that

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}(X_i \in [L_i U_i]) \geq 0.99 \tag{34}$$

with **P** probability 1, thus justifying requirement (15). Analogously, the stochastic analogue of (16) also holds with **P** probability 1.

4.5. Probability forecasting of continuous outcomes

Now each ideal forecast P_i, $i = 1, 2, \ldots$, is the conditional distribution (under **P**) of the random variable X_i, given past outcomes:

$$P_i(E) = \mathbf{P}(X_i \in E | X_1, \ldots, X_{i-1}), \tag{35}$$

E ranging over the measurable subsets of $[-1, 1]$.

If now Z_i, $i = 1, 2, \ldots$, are arbitrary uniformly bounded measurable functions of $(X_1, \ldots, X_i)$, and $M_i = \mathrm{E}(Z_i | X_1, \ldots, X_{i-1})$, then we again have the strong law of large numbers:

$$\frac{1}{n}\sum_{i=1}^{n}(Z_i - M_i) \to 0 \ (n \to \infty) \tag{36}$$

with **P** probability 1; similarly the appropriate central limit theorem and law of the iterated logarithm are again special cases of the classical results mentioned above.

4.6. Prequential decision theory

Recall that the (λ_i) are uniformly bounded. For any predictable sequence (A_i) of actions $((A_i)$ being predictable may be interpreted as A_i being chosen by some measurable strategy given $X_1, \ldots, X_{i-1}$),

$$\sum_{i=1}^{n} \lambda_i(X_i, B_i(P_i)) - \sum_{i=1}^{n} \lambda_i(X_i, A_i)$$

is, under suitable measurability assumptions, a **P** supermartingale with bounded increments, and so, with **P** probability 1,

$$\liminf_{n\to\infty} \left(\sum_{i=1}^{n} \lambda_i(X_i, B_i(P_i)) - \sum_{i=1}^{n} \lambda_i(X_i, A_i)\right) < \infty \tag{37}$$

(Shiryayev 1984, Corollary to Theorem VII.5.1). This justifies rejection of the issued probability forecasts (p_i) if another decision maker, with the same information, can force (19).

As a special case of the above, we can require that $A_i = B_i(Q_i)$, where Q_i is the probability forecast for X_i derived by Second Forecaster, who calculates his forecasts as

appropriate conditional probabilities based on some joint distribution $\mathbf{Q}$. In this case we must have

$$\liminf_{n\to\infty}\left(\sum_{i=1}^{n}\lambda_i(X_i,\,B_i(P_i))-\sum_{i=1}^{n}\lambda_i(X_i,\,B_i(Q_i))\right)<\infty \tag{38}$$

with $\mathbf{P}$ probability 1.

4.7. Cumulative probability integral transform

Suppose that the conditional distributions P_i are continuous, and define $U_i = U(P_i, X_i)$, where $U(p,\,x) = \text{prob}_{Y\sim p}\{Y\leqslant x\}$. Then, under $\mathbf{P}$, the (U_i) are independent random variables distributed exactly as $U[0,\,1]$ (Rosenblatt 1952). This then justifies, as a criterion of empirical validity of the forecasts, the application to the (u_i) (where $u_i = U(p_i, x_i)$ is the realized value of U_i) of any suitable test for this independent uniform distributional model.

4.8. Farthingale

Assuming that all S_i are measurable, (S_i) is a non-negative martingale under $\mathbf{P}$, and so, for all $c > 1$,

$$\mathbf{P}\{\sup S_i < c\}\geqslant 1-\frac{1}{c} \tag{39}$$

(this is a special case of Doob's inequality (see, for example, Shiryayev (1984))); this implies that

$$\mathbf{P}\{S_i\text{ bounded}\} = 1. \tag{40}$$

4.9. Marginal probabilities

Finally let us consider the criterion (2), which does not respect the weak prequential principle. General forms of the strong law of large numbers may be invoked to argue that, under some conditions on $\mathbf{P}$, we should have, with $\mathbf{P}$ probability 1,

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\pi_i)\to 0 \tag{41}$$

as $n\to\infty$. However, the generality of such a result is considerably less than in the cases considered above. For example, if under $\mathbf{P}$ the (X_i) are modelled as exchangeable, being independent Bernoulli trials with probability parameter Θ, where Θ is an unobserved quantity with a uniform distribution over $[0,\,1]$, then we have $\pi_i\equiv\frac{1}{2}$, but the $\mathbf{P}$-almost sure limit of $(1/n)\sum_{i=1}^{n}X_i$ is the random variable Θ, rather than the constant $\frac{1}{2}$ as would be required by (41). This stands in contrast with property (23), which does still hold in this exchangeable case.

5. Strong prequential principle

It is noteworthy that each of the probabilistic properties listed in Sections 4.1–4.8, which may be regarded as providing "stochastic justification" for the corresponding criteria in Section 3, holds for *all* **P**. It is only required that the *same* joint distribution **P** should govern both Nature's (randomized) strategy for determining outcomes, and Forecaster's strategy for producing probability forecasts—a requirement that may be regarded as an expression of the "null hypothesis" that we do indeed have agreement between Forecaster and Nature. This robustness is surprising, since the relevant "stochastic inferences" are constructed using, in a fundamental way, the distribution **P**, and thus the strategies of the players; nevertheless, when those strategies agree, they can be essentially ignored. That is to say, the stochastic inferences considered above, while seemingly in violation of the weak prequential principle (since they are defined in terms of the strategies used, not just the sequences of outcomes), in fact turn out to be in accordance with that principle (at least asymptotically). A stochastic inference with this robustness property, and its associated agreement criterion, will be said to respect the *strong prequential principle*.

5.1. Collusion

Another way of expressing the strong prequential principle is as follows. Suppose that Nature, instead of being oblivious to the choices made by Forecaster, in fact observes them and reacts appropriately. Specifically, if Forecaster has issued a probability forecast P_i for X_i, then Nature will produce x_i by simulating from P_i; if Forecaster gives a mean-and-variance forecast (M_i, D_i) for X_i, Nature will simulate from some distribution with that mean and variance, etc. Any such strategy for Nature may be called *collusive*. (There is just one collusive strategy for probability forecasting, but typically many when the forecasts are less fully specified.)

   If Nature is colluding with Forecaster, she is endeavouring, in so far as is possible, to bring about agreement between outcomes and forecasts. Hence the "null hypothesis", to be addressed by some proposed agreement criterion, may now be formulated as the assertion that Nature is being collusive, and a stochastic justification for the criterion should be based on performance under this hypothesis.

   However, collusion does not of itself determine a full distribution for the (X_i), under which such performance could be assessed, since Forecaster is still free to choose his forecasts (F_i) as he desires. Suppose that he in fact does so by using some fixed deterministic strategy for constructing F_i in the light of previous outcomes and forecasts, and suppose that Nature uses a collusive strategy. It is easy to see that this pair of strategies now determines Nature's joint distribution **P** for simulating $(x_1, x_2, \ldots)$, and that Forecaster's forecasts are then consistent with **P**. Conversely, any common **P**, underlying both Nature's and Forecaster's choices, can be interpreted as arising this way. Consequently, the strong prequential principle may be rephrased as requiring that, so long as Nature plays a collusive strategy, the desired stochastic inference will be probabilistically justified, no matter what Forecaster's strategy may be.

5.2. Super-strong prequential principle

So far we have assumed that the only past information available, to both Nature and Forecaster, is the set of previous forecasts and outcomes. More generally, we could allow both players to have access to (the same) additional information, e.g. temperature information when forecasting rain. This set-up can be formally expressed in terms of a filtered space $(\Omega, \mathscr{F}, (\mathscr{F}_i))$, with the outcome sequence X_i being adapted and the forecast sequence F_i being predictable. We consider a probability distribution **P** defined on $(\Omega, \mathscr{F})$. At every stage i, Forecaster issues his forecast for X_i in accordance with the relevant conditional distribution $\mathbf{P}(\cdot|\mathscr{F}_{i-1})$ given all observed past information, and Nature simulates X_i from the same distribution. All the results of Sections 4.1–4.8 can be easily extended to this more general setting, with the proofs essentially unchanged.

   We could further strengthen the strong prequential principle to require validity of stochastic inferences in this more general setting (still requiring that Nature and Forecaster use the same, otherwise arbitrary, underlying distribution **P** or, equivalently, that Forecaster use any **P**, while Nature operates a collusive strategy). This we shall term the *super-strong prequential principle*.

   An application of the above general framework (wherein the database is extended to include the results of extraneous randomization) allows Forecaster to use a stochastic strategy which, given past outcomes $(x_1, \ldots, x_{i-1})$ and forecasts $(f_1, \ldots, f_{i-1})$, determines, not a fixed forecast F_i, but rather a distribution Π_i for F_i. After observing past outcomes and forecasts, Forecaster first randomly generates F_i from Π_i, and then Nature, playing a collusive strategy, randomly generates X_i in accordance with F_i. The super-strong prequential principle requires validity of stochastic inferences in this randomized set-up.

5.3. Counter-example

We see from Section 4 above that all the criteria introduced in Sections 3.2–3.9 satisfy the strong (indeed, the super-strong) prequential principle.

   To clarify ideas, we now exhibit an example for which the strong prequential principle is violated.

Example 6. Consider a special case of probability forecasting of continuous outcomes. For simplicity, we shall here relax the technical restriction that the values be confined to $[-1, 1]$ or any other finite interval (similar examples could be constructed incorporating this restriction, or in other contexts we consider, such as probability forecasting of binary outcomes).

   Let us consider as our agreement criterion the observed significance level (4), based on the statistic $T \equiv X_n$ for a fixed $n > 1$. Although the actual value of the statistic depends only on the observed outcomes and is thus in accord with the weak prequential principle, the significance level explicitly involves the distribution **P**, i.e. the strategy used by Forecaster.

To verify that this does not respect the strong prequential principle, consider two different strategies for Forecaster, corresponding to the following two joint distributions.

(i) Under $\mathbf{P}_1$, the X_i are modelled as following a stationary normal autoregressive model:

$$X_1 \sim N(0, \tfrac{4}{3}),$$

$$X_i \sim N(0.5X_{i-1}, 1) \qquad (i > 1).$$

(ii) Under $\mathbf{P}_2$, the X_i are modelled as independent, each having distribution $N(2, 1)$, except for $X_1 \sim N(0, \tfrac{4}{3})$.

The marginal distribution of X_n ($n > 1$) under $\mathbf{P}_1$ is $N(0, \tfrac{4}{3})$, while under $\mathbf{P}_2$ it is $N(2, 1)$.

Suppose that Nature produces data $x_i \equiv 4$, all i. Then, for each i both strategies $\mathbf{P}_1$ and $\mathbf{P}_2$ produce the identical sequence of probability forecasts: $N(0, \tfrac{4}{3})$ for $i = 1$, and $N(2, 1)$ for $i > 1$. If the strong prequential principle were satisfied, our assessment of agreement with the data would be the same for both $\mathbf{P}_1$ and $\mathbf{P}_2$. However, using the criterion above, we find that $\alpha = \mathrm{prob}\{N(0, 1) > 12^{1/2}\} = 0.00027$ for case (i), while $\alpha = \mathrm{prob}\{N(0, 1) > 2\} = 0.023$ for case (ii). Since these differ, although the sequences of outcomes and forecasts do not, this criterion does not respect the strong prequential principle.

It is interesting to note that, had we used instead the statistic $T \equiv \sum_{i=1}^{n} X_i$, we would have obtained asymptotically identical observed significance levels in the two cases above, even though the null distributions of T are quite different. This can be verified directly. The reason is that, in each case, this statistic is essentially equivalent to that considered in (12), and the result then follows as in Section 4.3. This phenomenon extends to general stationary autoregressive moving-average models. □

Appendix 1 describes a range of plausible agreement criteria, all respecting the weak prequential principle. Whereas some of them also respect the strong prequential principle, still others do not.

These counter-examples notwithstanding, we have seen that, perhaps surprisingly, it is indeed possible to construct numerous appealing inferences and agreement criteria which *do* satisfy the strong, as well as the weak, prequential principle. The purpose of the remainder of this paper is to seek to understand more fully why this should be, and when.

6. Prequential games

The main reason why there are so many inferences satisfying the strong prequential principle seems to be that these inferences can be expressed within the *prequential framework*, a new game-theoretic approach to probability theory which has the weak and strong prequential principles 'built in'. In the prequential framework, each forecasting task is represented as a perfect-information game (with the goal of the game to be specified) involving, besides Nature and Forecaster, one more player, Statistician. Statistician plays against the combined forces of Forecaster and Nature. The role of Statistician will be to demonstrate that our

inferences are valid. We shall discuss what this means in more detail in the next section. Here we describe the means that Statistician should have to accomplish his task.

We shall use the term "game" somewhat loosely, often to denote a partially specified game in which, for example, the rules of play are specified, but not the starting situation or the criterion for determining who has won. In describing our games, we shall often say that one player pays some amount to another player. We mean not money but infinitely divisible points. We let £n stand for "n points".

A typical prequential game can be represented as in Figure 1. In addition to the moves (F_i) of Forecaster and (X_i) of Nature, we have moves (H_i) for Statistician. The players, playing in the order indicated by the arrows, are allowed to see each other's past moves (the situation of complete information). The rules of an individual game will determine the allowable moves for each player. After each stage, when all three players have moved, Forecaster pays a certain amount, depending on those moves, to Statistician; again, the rules of a particular game will specify how this amount is to be calculated. Finally, we shall need a rule for determining who has won the game (this may not be determined until an infinite number of stages of the game).

We now describe in more detail some specific prequential games relevant for our purposes.

6.1. Probability forecasting of binary outcomes

The simplest prequential game corresponds to X_i binary. Forecaster's moves are $F_i = P_i \in [0, 1]$; Statistician's moves are $H_i \in \mathbb{R}$. The interpretation of the moves is as follows. Before X_i is disclosed by Nature, Forecaster selects $P_i \in [0, 1]$, interpreted as his probability that $X_i = 1$, or equivalently as his expectation ("fair price") for X_i. We take this to mean (cf. de Finetti (1974, Section 3.3)) that Forecaster allows Statistician to buy, before X_i is disclosed, any number H_i (possibly negative or non-integer) of tickets, each of which will pay £X_i; the price is £P_i for a ticket.

Define

$$\mathcal{K}_i = \mathcal{K}_{i-1} + H_i(X_i - P_i), \qquad i \geqslant 1, \tag{42}$$

$\mathcal{K}_0$ being some given initial value. We interpret $\mathcal{K}_0$ as Statistician's initial capital; then $\mathcal{K}_i$ is Statistician's capital at stage i (we shall sometimes refer to $\mathcal{K}_i$ as the *capital process*). Note that, given any fixed strategy whereby Statistician chooses H_i as a function of previous moves, $(\mathcal{K}_i)$ will be a farthingale in the sense of (21).
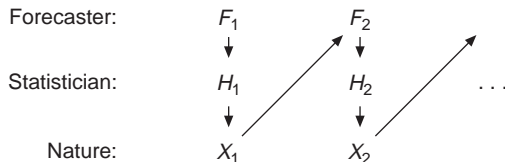

Fig. 1. Typical prequential game; the arrows indicate the order of moves of the three players.

We shall denote this game by $\mathscr{G}^{\mathrm{bin}}(\mathscr{K}_0)$. (To complete the description of the game we still need to specify a rule for determining the winner; we shall do this in the next section.)

6.2. Mean-value forecasting of continuous outcomes

This game $\mathscr{G}^{\mathrm{mean}}(\mathscr{K}_0)$ is a straightforward generalization of $\mathscr{G}^{\mathrm{bin}}(\mathscr{K}_0)$. The only difference is that now Nature is allowed to select Her move X_i from $[-1, 1]$ and Forecaster is also allowed to select his move $F_i = M_i$ from $[-1, 1]$; again, M_i is interpreted as Forecaster's "fair price" for X_i, before its value is disclosed by Nature. The tickets offered by Forecaster will be called E tickets. The capital process $\mathscr{K}_i$ is defined (cf. (42)) by

$$\mathscr{K}_i = \mathscr{K}_{i-1} + H_i(X_i - M_i), \qquad i \geqslant 1.$$

6.3. Mean-value-and-variance forecasting of continuous outcomes

Now Forecaster's moves are $F_i = (M_i, D_i)$, where $M_i \in [-1, 1]$ and $D_i \geqslant 0$; Statistician's moves are $H_i = (V_i, W_i)$, with $V_i, W_i \in \mathbb{R}$. Instead of (42), we have now

$$\mathscr{K}_i = \mathscr{K}_{i-1} + V_i(X_i - M_i) + W_i\{(X_i - M_i)^2 - D_i\}. \tag{43}$$

Intuitively, before Nature discloses X_i, Forecaster, besides the E tickets, is offering for sale also var tickets, which oblige Forecaster to pay some "fine" if his forecast M_i is bad. A var ticket obliges Forecaster to pay $£(X_i - M_i)^2$ to Statistician; D_i is the price for a var ticket. After Forecaster's move, Statistician decides how many E tickets (V_i) and var tickets (W_i) he would like to buy; (43) shows Statistician's capital at each stage.

This game will be denoted by $\mathscr{G}^{\mathrm{mean\&var}}(\mathscr{K}_0)$.

6.4. Confidence-interval forecasting of continuous outcomes

Now Nature's moves are $X_i \in [-1, 1]$ and Forecaster's moves are (L_i, U_i) with $-1 \leqslant L_i \leqslant U_i \leqslant 1$; Statistician's moves are $H_i \geqslant 0$; (42) is replaced by

$$\mathscr{K}_i = \mathscr{K}_{i-1} + H_i\{\mathbf{I}(X_i \notin [L_i, U_i]) - 0.01\}. \tag{44}$$

Intuitively, before Nature discloses X_i, Forecaster offers tickets which will pay £1 in case $X_i \notin [L_i, U_i]$ for the price of £0.01. The requirement $H_i \geqslant 0$ means that Forecaster does not assert that the odds in favour of $X_i \in [L_i, U_i]$ are exactly 99:1, but that they are at least 99:1. Our notation for this game will be $\mathscr{G}^{99\%}(\mathscr{K}_0)$.

6.5. Probability forecasting of continuous outcomes

The game $\mathscr{G}^{\mathrm{prob}}(\mathscr{K}_0)$ is, like $\mathscr{G}^{\mathrm{mean}}(\mathscr{K}_0)$, generalization of $\mathscr{G}^{\mathrm{bin}}(\mathscr{K}_0)$. The difference from $\mathscr{G}^{\mathrm{bin}}(\mathscr{K}_0)$ is that Nature's moves X_i are now allowed to belong to the whole of $[-1, 1]$,

Forecaster's moves $F_i = P_i$ are probability distributions in $[-1, 1]$, Statistician's moves H_i are Borel functions on $[-1, 1]$, and (42) is replaced by

$$\mathscr{K}_i = \mathscr{K}_{i-1} + H_i(X_i) - \int H_i \, dP_i. \tag{45}$$

The interpretation is as follows. Forecaster's move P_i means that he is willing to sell Statistician any of the following tickets: for any Borel function $H: [-1, 1] \to \mathbb{R}$, the H ticket obliges Forecaster to pay £$H(X_i)$ to Statistician after X_i is disclosed; the price for the H ticket is given by its mean value $\int H \, dP_i$ with respect to the forecast distribution P_i.

The variant of this game where the forecasts P_i are required to be continuous distributions in $[-1, 1]$ will be denoted by $\mathscr{G}^{\mathrm{prob}'}(\mathscr{K}_0)$; we need this variant to examine the cumulative probability integral transform.

For prequential decision theory we shall need one more player, Second Forecaster; just before X_i is disclosed, Second Forecaster announces his forecast Q_i (a probability distribution in $[-1, 1]$) for X_i. This game will be denoted by $\mathscr{G}^{\mathrm{prob+SF}}(\mathscr{K}_0)$. When considering (37) instead of (38), we shall need a more powerful player, Decision Maker, who can directly choose actions A_i rather than probability distributions Q_i; the corresponding game will be denoted by $\mathscr{G}^{\mathrm{prob+DM}}(\mathscr{K}_0)$.

By *forecasting task* we shall mean a game with $\mathscr{K}_0$ unspecified (so far we have used this notion informally); to denote a forecasting task we shall simply drop '$(\mathscr{K}_0)$' from our notation for the corresponding game.

7. Prequential framework

Statistician is a mathematician whose role is to deduce various probabilistic consequences from Forecaster's assertions about the world. Thus he might wish to demonstrate that the calibration property (5) holds, but this cannot be expected in complete generality; rather, only under the "null hypothesis" that Forecaster's forecasts provide a valid model of Nature's outcomes. In previous sections this null hypothesis of agreement between Forecaster and Nature was interpreted in terms of the existence of a common (albeit unspecified) joint distribution **P** underlying both Nature's and Forecaster's strategies. Here we take a different approach.

The tickets offered to Statistician by Forecaster in the games described in the previous section are priced "fairly", from the point of view of Forecaster. If Forecaster's forecasts are "good" (in agreement with Nature), Statistician should not be able to make much money by betting at these fair prices. Conversely, if the forecasts are bad, Statistician should be able to profit from their mismatch with Nature's outcomes. Thus, for any specific realization of the game, we can regard it as evidence that Forecaster was *wrong* (i.e. out of agreement with Nature) when Statistician, starting from £1 (or any other positive capital), managed greatly to increase his capital, without ever risking getting into debt (and, the greater the increase in his capital, the more convincing the evidence).

Statistician could thus prove that the calibration property (5) holds, whenever Nature and Forecaster are in agreement, if he could ensure that, *whatever* Nature and Forecaster

might do, *either* (5) will hold *or* he will be able to demonstrate (by substantially increasing his capital) that Forecaster and Nature are not in agreement. If he can find a strategy with this property, then he will have achieved his goal. This idea is the basis of the next definition.

7.1. Full events

Let $\mathscr{G}$ be any of the above forecasting tasks and let E be some *substantive event*, i.e. a property of the sequence of moves $F_1 X_1 F_2 X_2 \ldots$ of Forecaster and Nature alone. In our applications, the statement "$(f_1 x_1 f_2 x_2 \ldots) \in E$" will express some desirable property, such as (5), which Statistician claims should hold if Forecaster's forecasts are empirically valid.

Given any substantive event E, we can complete the description of a prequential game as follows. We say that Statistician *wins game* $\mathscr{G}_E$ if, in game $\mathscr{G}(1)$, (a) $\mathscr{K}_i \geqslant 0$, for all i and (b) either E occurs, or $\mathscr{K}_i \to \infty$.

The disjunction in (b) corresponds to Statistician having demonstrated *either* that Nature has colluded with Forecaster to bring about the occurrence of E *or* that there was a mismatch between forecasts and outcomes, which has allowed Statistician to amass an infinite fortune (while never going into debt). We are particularly interested in events E for which Statistician can force a win in game $\mathscr{G}_E$. Then, whenever E does not occur, we have evidence of non-collusion between Nature and Forecaster, with such a mismatch between forecasts and outcomes that Statistician makes an infinite fortune from a sequence of "fair bets". We can thus regard such an event E as a criterion of agreement between forecasts and outcomes. We shall say that event E is *full* (and its complement E^c is *null*) in forecasting task $\mathscr{G}$ if Statistician has a winning strategy in game $\mathscr{G}_E$.

Our programme now is to show, for various interesting events E, that Statistician has a winning strategy in $\mathscr{G}_E$ (and even, wherever possible, to *exhibit* such a strategy) and thus to demonstrate that E is full. We can then regard Forecaster as discredited if E fails to occur (so that Statistician, playing his winning strategy, makes an infinite amount of money).

The following theorem asserts the existence of Statistician's winning strategies for many of the events considered above.

Theorem 1.

(a) *(Compare* (31)*.) Event*

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - M_i) \to 0 \qquad (n \to \infty)$$

is full in forecasting task $\mathscr{G}^{\text{mean}}$.

(b) *(Compare* (28) *and* (29)*.) Event* either

$$\sum_{i=1}^{\infty} D_i < \infty$$

or

$$\limsup_{n\to\infty} \sum_{i=1}^{n}(X_i - M_i) \bigg/ \left\{ 2\left(\sum_{i=1}^{n} D_i\right) \ln\ln\left(\sum_{i=1}^{n} D_i\right) \right\}^{1/2} = 1$$

is full in forecasting task $\mathscr{G}^{\mathrm{mean\&var}}$.

(c) *(Compare* (34).*) Event*

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\mathbf{I}(X_i \in [L_i, U_i]) \geqslant 0.99$$

is full in forecasting task $\mathscr{G}^{99\%}$.

(d) *(Compare* (37).*) Event*

$$\liminf_{n\to\infty} \left(\sum_{i=1}^{n}\lambda_i(X_i, B_i(P_i)) - \sum_{i=1}^{n}\lambda_i(X_i, A_i)\right) < \infty$$

is full in forecasting task $\mathscr{G}^{\mathrm{prob+DM}}$.

(e) *(Compare* (40).*) Let* (S_i) *be non-negative functions, as in* (20), *satisfying* $S_0 = 1$ *and* (21). *Then event*

$$\sup_{i} S_i(P_1, X_1, \ldots, P_i, X_i) < \infty$$

is full in forecasting task $\mathscr{G}^{\mathrm{prob}}$.

For all the above assertions, proofs in our prequential framework can be closely modelled on already existing proofs of their probabilistic counterparts. Two sample proofs are given in the appendices. Appendix 2 proves a prequential analogue of Doob's convergence theorem for non-negative martingales and briefly describes other ideas that we need to deduce from it the "prequential strong law of large numbers", (a) above. Appendix 3 proves the "prequential law of the iterated logarithm" (upper half), (b) above; it is further shown how the proof given there also implies (c) and (a). The proofs are, in essence, simple modifications of well-established martingale-based proofs.

It is important that we can explicitly construct winning strategies for Statistician which satisfy many desiderata: are measurable, computable, etc. (in the case of items (d) and (e), provided that the loss functions and the functions S_i satisfy these desiderata).

Theorem 1 (a) implies that event (36) is full in forecasting task $\mathscr{G}^{\mathrm{prob}}$ (in some sense game $\mathscr{G}^{\mathrm{prob}}(1)$ is easier for Statistician to play than $\mathscr{G}^{\mathrm{mean}}(1)$; in $\mathscr{G}^{\mathrm{prob}}(1)$, Forecaster is required to give more specific forecasts). Analogously, Theorem 1 (d) implies that event (38) is full in the forecasting task $\mathscr{G}^{\mathrm{prob+SF}}$. It is also clear that the events (23) and "(28) or (29)" (with $M_i = P_i$, $D_i = P_i(1 - P_i)$) are full in $\mathscr{G}^{\mathrm{bin}}$.

7.1.1. Connection with the conventional notion of almost sure event

Let E be any full event. Fix any joint distribution $\mathbf{P}$ for $(X_1, X_2, \ldots)$. Let Forecaster use either the strategy that calculates its forecasts from $\mathbf{P}$ (see (22), (30), (32), and (35)) or a

measurable strategy that "agrees" with **P** (as in (33)), and let Nature use the collusive randomized strategy in which X_i is simulated from the distribution function $F_i(t) = \mathbf{P}(X_i \leq t | X_1 = x_1, \ldots, X_{i-1} = x_{i-1})$. Then $(X_1, X_2, \ldots)$ is simulated from **P**. Consider the behaviour of Statistician's capital process $(\mathscr{K}_i)$ (with $\mathscr{K}_0 = 1$) under his measurable winning strategy. Since every non-negative supermartingale is bounded with probability 1 (we already used a similar fact in Section 4.8) and $(\mathscr{K}_i)$ is now a non-negative **P** supermartingale (even a **P** martingale in all cases except (33)), the event "$\mathscr{K}_i \to \infty$" has **P** probability 0; therefore, by Theorem 1(b) of the requirement for a full event, E occurs with **P** probability 1. This argument holds for any **P**.

To see now that each of the agreement criteria mentioned in Theorem 1 must satisfy the strong prequential principle (as already confirmed directly in Section 4), it is enough to combine the above argument with the results of Theorem 1. In fact, an essentially identical argument demonstrates that the super-strong prequential principle will also be satisfied in such cases, thus allowing Forecaster to take into account additional information, beyond the revealed history of the game.

7.2. Prequential probability

The above development has allowed us to introduce a game-theoretic analogue of the concept of an event's having probability 0 or 1. We now extend this framework to allow measurement of non-extreme probabilities.

Up to now, Statistician was effectively playing against Forecaster (with Nature as a neutral player), and Forecaster had infinite capital. Suppose now that Statistician starts with capital $\mathscr{K}_0 \in [0, 1]$, Forecaster with capital $1 - \mathscr{K}_0$. After i stages, Statistician's capital becomes $\mathscr{K}_i$. Either player loses when his capital becomes negative. Let E be some event. We wish to define what it means for E to be "reasonably certain" (unless the forecasts are bad).

Say that Statistician *wins game* $\mathscr{G}_E(\mathscr{K}_0)$ if (a) his capital $\mathscr{K}_i$ reaches above 1 before reaching below 0 or (b) his capital $\mathscr{K}_i$ is always in [0, 1], and E happens. If, for $\mathscr{K}_0$ small, Statistician wins the game, then if E has not occurred, we must be in situation (a); and then the large increase in Statistician's capital provides strong (although less than categorical) evidence that Forecaster was wrong (out of agreement with Nature).

If Statistician has and uses a *winning strategy* for $\mathscr{G}_E(\mathscr{K}_0)$, then (a) his capital can never be negative (at least if he stops as soon as it exceeds 1) and (b) whenever E fails, he will have made more than £1 out of £$\mathscr{K}_0$, thus casting doubt on the agreement between Forecaster and Nature. The smaller $\mathscr{K}_0$, the greater the doubt. In other words, if E is an event for which Statistician can find a winning strategy in $\mathscr{G}_E(\mathscr{K}_0)$ with $\mathscr{K}_0$ small, then, whenever E fails, an outcome occurs (large increase in Statistician's capital) that is not very credible under the hypothesis that Forecaster's forecasts and Nature's outcomes are in agreement. Intuitively, the complement of E has "low probability", and thus E has "high probability", under this hypothesis; the smaller we can make $\mathscr{K}_0$, the higher is the "probability" of E. This motivates the following definition.

Let

$$\mathscr{T}(E) = \{1 - \mathscr{K}_0 : \text{Statistician can force a win in } \mathscr{G}_E(\mathscr{K}_0)\}$$

be the range of Forecaster's initial capital for which Statistician has a winning strategy. Clearly,

$$(a \in \mathscr{T}(E), b < a) \Rightarrow b \in \mathscr{T}(E).$$

We define the *prequential probability* of an event E to be

$$\mathbf{PP}(E) = \sup \mathscr{T}(E),$$

with $\sup \varnothing = 0$. This notion generalizes that of a full event; it can be shown that, for all our forecasting tasks, an event E is full if and only if $\mathbf{PP}(E) = 1$ (provided that we allow Statistician to buy infinitely many tickets or, in the case of $\mathscr{G}^{\mathrm{prob}}$, to buy tickets with potentially infinite pay-off, with natural conventions for operations with infinities).

We could generalize further and define "prequential expectation" (cf. Vovk 1993b, Section 5). However, we shall not discuss this here.

Now we can state the theorem corresponding to the remaining agreement criteria.

Theorem 2.

(a) *In forecasting tasks $\mathscr{G}^{\mathrm{mean\&var}}$ and $\mathscr{G}^{\mathrm{bin}}$,*

$$\mathbf{PP}\left(B^{-1/2} \sum_{i=1}^{\tau_B} (X_i - M_i) \in (a, b) \text{ or } \tau_B = \infty \right) \to \Phi(b) - \Phi(a)$$

as $B \to \infty$, where (a, b) is any interval of the extended real line and

$$\tau_B = \min\left(n : \sum_{i=1}^{n} D_i \geqslant B \right).$$

(b) *Let the cumulative probability integral transform CPIT be the transformation that maps every forecast or outcome sequence $(p_1, x_1, p_2, x_2, \ldots)$ into the corresponding sequence*

$$(\mathrm{prob}_{Y \sim p_1}\{Y \leqslant x_1\}, \mathrm{prob}_{Y \sim p_2}\{Y \leqslant x_2\}, \ldots).$$

Let E be any Borel set in $[0, 1]^\infty$. In forecasting task $\mathscr{G}^{\mathrm{prob}'}$,

$$\mathbf{PP}\{\mathrm{CPIT}(P_1, X_1, P_2, X_2, \ldots) \in E\} = \mathbf{U}(E),$$

$\mathbf{U} = (U[0, 1])^\infty$ *being the uniform distribution in $[0, 1]^\infty$.*

(c) *In forecasting task $\mathscr{G}^{\mathrm{prob}}$,*

$$\mathbf{PP}(\sup_i S_i(P_1, X_1, \ldots, P_i, X_i) < c) \geqslant 1 - \frac{1}{c}, \text{ for } c > 1,$$

provided that the functions S_i satisfy $S_i \geqslant 0$, $S_0 = 1$ and (21).

Proof *(outline)*. The proof of the central limit theorem (a) for forecasting task $\mathscr{G}^{\mathrm{bin}}$ can be

found in Vovk (1993a, Theorem 5), which can be routinely extended to cover $\mathscr{G}^{\text{mean\&var}}$. Once again, the proof is a simple modification of a standard proof (due to Lindeberg) from conventional probability theory. Proof of (c) is trivial. For both (a) and (c), the proofs involve explicit construction of winning strategies for Statistician.

In the case of (b) our construction is also in some sense explicit. Suppose that we have $\epsilon \in (0, 1)$ and a set $E \subseteq [0, 1]^\infty$ such that $\mathbf{U}(E) > 1 - \epsilon$. Then there exists (Vovk 1993a, Theorem 4) a sequence (S_i) of functions $S_i: [0, 1]^i \to [0, \infty[$, $i = 0, 1, \ldots$, such that $S_0 < \epsilon$, $\liminf_{i \to \infty} S_i(t_1, \ldots, t_i) > 1$ when $(t_1, t_2, \ldots) \notin E$, and, for all i and $(t_1, \ldots, t_i) \in [0, 1]^i$,

$$\mathrm{E}_{T \sim U[0,1]}\, S_{i+1}(t_1, \ldots, t_i, T) = S_i(t_1, \ldots, t_i)$$

(the martingale property). Exhibiting such a martingale S_i can be interpreted as a "proof" that $\mathbf{U}(E) > 1 - \epsilon$. Theorem 2 (b) is "constructive" in the following sense: given a martingale S "proving" that $\mathbf{U}(E) > 1 - \epsilon$, one can easily find Statistician's strategy in $\mathscr{G}^{\text{prob}'}(\epsilon)$ which witnesses that

$$\mathbf{PP}\{\mathrm{CPIT}(P_1, X_1, P_2, X_2, \ldots) \in E\} \geqslant 1 - \epsilon. \qquad \square$$

Remark. In the terminology of Shafer (1996) and Vovk (1993b), our "prequential probability" is "lower probability", and the "global probability" of Vovk (1993a) is "upper probability" (in our current framework, we can define the upper probability of event E as $1 - \mathbf{PP}(E^c)$). Lower probabilities are analogous to Shafer's (1976) belief functions. Another language for introducing "prequential probability" is provided by Shafer's (1996) martingale trees.

7.2.1. Connection with the conventional notion of probability

For discussing connections between prequential probability and the conventional notion of probability, it is convenient to change slightly the definition of $\mathbf{PP}(E)$, adding the requirement that Statistician's winning strategy should be measurable. Theorem 2 continues to hold under this changed definition of $\mathbf{PP}$ (in the case of (b) and (c), provided that E and S_i are measurable).

Consider a distribution $\mathbf{P}$ for $(X_1, X_2, \ldots)$ and Forecaster's measurable strategy $\mathscr{S}$ that forms its forecasts in accordance with $\mathbf{P}$. It is easy to see that

$$\mathbf{PP}(E^*) \leqslant \mathbf{P}(E), \tag{46}$$

for any Borel set E, E^* standing for the event that $(X_1, X_2, \ldots) \in E$ and all P_i are selected in accordance with $\mathscr{S}$. Indeed, for any fixed $\epsilon > 0$ satisfying $\mathbf{PP}(E^*) > 1 - \epsilon$, let $(\mathscr{K}_i)$ be the capital process of Statistician's measurable winning strategy starting from £ϵ. Then $(\mathscr{K}_i)$ is a non-negative $\mathbf{P}$ supermartingale with $\mathscr{K}_0 = \epsilon$. So (a) of the definition of Statistician winning holds with $\mathbf{P}$ probability at most ϵ (by Doob's inequality). Since event E is bound to happen when (a) fails, $\mathbf{P}(E) \geqslant 1 - \epsilon$.

This observation shows that Theorem 2 (a)–(c) imply their conventional counterparts given in Section 4.

Remark. We prefer not to impose the requirement of measurability on the players' strategies because (i) on the theoretical side, this requirement is likely to complicate more advanced parts of our theory needlessly and (ii) from the applied point of view, it is too weak (useful strategies must be at least computable). (Of course, the assumption of measurability of strategies is very useful when discussing the relation of our approach to the conventional approach.) On the other hand, measurable (in some sense) events E play a special role in our approach: Martin's (1975, 1990) theorem asserts that the games $\mathscr{G}_E$ and $\mathscr{G}_E(\mathscr{K}_0)$ are determined (in the sense that either Statistician or the team Forecaster and Nature has a winning strategy) when E is pseudo-Borel (which is the same as Borel when the players only can select their moves from a countable set).

8. Discussion

Satisfaction of the weak prequential principle is an appealing requirement to impose on an assessment criterion for agreement between forecasts and outcomes, since it focuses on the actual realized sequences, obviating the need to consider how these might have been produced. Of itself, however, it is clearly not sufficient. For example, the trivial criterion which always asserts that Forecaster has been successful, or the equally vacuous criterion which always asserts that Forecaster has failed, are perfectly acceptable from the point of view of the weak prequential principle alone. Some further justification for any proposed criterion is required. Such justification might be supplied by appealing to stochastic properties of the criterion, e.g. almost sure behaviour or an approximate significance level, which apply under the assumption that Forecaster is successful.

This last assumption can be formalized in various ways. In Section 4 we regarded it as asserting that Forecaster calculates his forecasts using the "true" distribution P from which Nature generates the outcomes. Since, however, we do not want the unknown true distribution to be relevant to the definition of "success", we introduced the strong prequential principle, requiring that dependence of the relevant stochastic properties on P be (exactly or asymptotically) absent. We saw that many appealing criteria respect both principles.

In Section 6 we described an alternative, completely prequential game-theoretic framework, as the basis of a new approach to probability. The weak prequential principle is automatically satisfied; there is no way to use anything beyond the forecasts actually made by Forecaster (there is nothing more; we make no assumptions on the way the forecasts are generated). In Section 7 we reinterpreted in this setting the assumption that Forecaster is "successful" and showed how this could be used to explain why, in the more standard approach, many results turn out to be independent of the underlying distribution, thus validating a wide variety of assessment criteria as satisfying both prequential principles.

It might be argued that our approach to probability via prequential games is merely using another language to express already well-known results. There is some truth in this, but it is a language which is particularly well suited to bring out aspects which are often under-emphasized, namely (a) the minimum specification of the problem necessary for the required results and (b) the universal applicability of the results, irrespective of the assumed underlying probability distribution.

Most important, the prequential results are available and have essentially the same intuitive interpretation, even when we do not wish to assume the existence of an underlying distribution. For example, we can justify assessing a weather forecaster's probability of precipitation forecasts against the calibration criterion (5), in the knowledge that Statistician has a strategy such that, if (5) fails, he will be able to discredit Forecaster by making a fortune from him. This entirely non-probabilistic argument stands on its own merits.

Another technical point, not taken up here, is the possibility of refining the idea of a "null event" and distinguishing between different categories of null event, in terms of the rate at which Statistician's fortune can grow when the event happens. This does not appear possible within the conventional approach. (Some work in this direction has been done within the algorithmic approach to probability theory (see, for example, Loveland (1969), Schnorr (1970, 1971a,b) and Vovk (1987)).)

An interesting feature of the prequential framework is that it is based, not on measure theory, but on the theory of perfect-information games. In it, the usual task of proving something about averages is replaced by the task of constructing strategies which are guaranteed to achieve some goal. We hope that we have succeeded in convincing the reader that this approach is intuitive and fruitful, and that it could be valuable to develop a more general theory of prequential games such as those considered in Section 6 (in this paper we have not even given a fully general definition of a prequential game).

Appendix 1: Hill's criteria

In this appendix we present an interesting family of agreement criteria introduced by Hill (1982). We consider the probability forecasting game for continuous outcomes (with or without the restriction $X_i \in [-1, 1]$). Let B be some subset of $\mathbb{R}^\infty$. We denote by $\mathscr{P}(B)$ the set of all sequences $(P_1, P_2, \ldots)$ of probability forecasts such that the product distribution $P_1 \times P_2 \times \ldots$ gives probability 1 to B. Let $\hat{B}$ denote the set of pairs $(\mathbf{f}, \mathbf{x})$ of sequences of moves by Forecaster and Nature respectively such that

$$\mathbf{f} \in \mathscr{P}(B) \Rightarrow \mathbf{x} \in B,$$

i.e. *either* $\mathbf{x} \in B$ *or* $\mathbf{f} \notin \mathscr{P}(B)$.

We might consider, as a possible agreement criterion between Forecaster and Nature, the requirement that $(\mathbf{f}, \mathbf{x}) \in \hat{B}$. We may term this *Hill's B criterion*. Note that, by definition, this will be satisfied with probability 1 whenever Nature generates her outcomes from a product distribution, and Forecaster constructs his forecasts from the same distribution.

For any joint distribution $\mathbf{P}$, consider the event

$$B(\mathbf{P}) := \text{``}\{(P_1, P_2, \ldots), (X_1, X_2, \ldots)\} \in \hat{B}\text{''},$$

where the outcomes $X_1, X_2, \ldots$ and the forecasts $P_1, P_2, \ldots$ are generated from $\mathbf{P}$ (see (35)). Then Hill's B criterion will respect the strong prequential principle if, for any probability distribution $\mathbf{P}$ over $\mathbb{R}^\infty$, $B(\mathbf{P})$ is $\mathbf{P}$-almost certain. As remarked above, this is satisfied for all product measures $\mathbf{P}$.

As shown by Hill (1982), for a wide class of sets B, Hill's B criterion satisfies the strong prequential principle; namely Hill shows that it is satisfied for

$$B = \left\{ (x_1, x_2, \ldots) \in \mathbb{R}^\infty \colon \sum_{i=1}^\infty x_i \text{ converges} \right\},$$

$$B = \liminf_{n \to \infty} \{ (x_1, x_2, \ldots) \in \mathbb{R}^\infty \colon x_n \in A_n \},$$

$$B = \limsup_{n \to \infty} \{ (x_1, x_2, \ldots) \in \mathbb{R}^\infty \colon x_n \in A_n \},$$

where $A_1, A_2, \ldots$ is a sequence of Borel sets in $\mathbb{R}$. Therefore, the following counter-example (Hill 1982) is of interest.

Theorem 3. *Hill's B criterion violates the strong prequential principle when*

$$B = \left\{ (x_1, x_2, \ldots) \colon \limsup_{n \to \infty} \left[\sum_{i=1}^n x_i \bigg/ \left\{ n \ln \ln \left(\frac{n}{2} \right) \right\}^{1/2} \right] = 1 \right\}.$$

Proof. Define $\pi^{(+)}$ to be the discrete uniform distribution over $\{0, 1\}$, and $\pi^{(-)}$ that over $\{-1, 0\}$. Let $X_i = S_i - S_{i-1}$, where the (S_i) are independently drawn from $\pi^{(+)}$. The distribution, **M** say, of the sequence $(P_1, P_2, \ldots)$ of one-step-ahead forecasts is such that the terms are independent, with $P_i = \pi^{(+)}$ or $\pi^{(-)}$, each with probability $\frac{1}{2}$. Our goal is to prove that event $B(\mathbf{P})$ is not almost certain. Since clearly $(X_1, X_2, \ldots) \in B$ is always false, we are only required to prove that $(P_1, P_2, \ldots) \in \mathscr{P}(B)$ with positive probability. In fact, the following stronger property holds.

Lemma 1. *Let $(P_1, P_2, \ldots)$ have joint distribution* **M**. *Then*

$$(P_1, P_2, \ldots) \in \mathscr{P}(B)$$

almost surely.

Proof. Consider the following stochastic scenario: first a sequence $(P_1, P_2, \ldots)$ is generated from **M**, and then a sequence of outcomes $(X_1, X_2, \ldots)$ is generated from the product distribution $P_1 \times P_2 \times \ldots$ in $\mathbb{R}^\infty$. It is clear that $X_1, X_2, \ldots$ are independent random variables taking values $-1, 0, 1$ with probabilities $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$, respectively. Therefore, by the usual law of the iterated logarithm (see (29) above),

$$\limsup_{n \to \infty} \left[\sum_{i=1}^n X_i \bigg/ \left\{ n \ln \ln \left(\frac{n}{2} \right) \right\}^{1/2} \right] = 1$$

almost surely, and so this equality holds almost surely under **M**-almost all product distributions $P_1 \times P_2 \times \ldots$. $\qquad\square$

Appendix 2: Doob's convergence theorem

A proof of a prequential assertion can often be obtained by modifying some of the usually many proofs of its conventional counterpart. Examination of the proofs of the classical versions of the assertions in Section 4 shows that most of them rely on "martingale methods", and martingales are associated with games and gambling (Ville 1939; Doob 1953) (see also Stout (1974) or Shiryayev (1984)). It is thus often possible to extract a "game-theoretic core" from such a proof, which is then seen to be a proof of some assertion about a perfect-information game, and thus directly relevant to the prequential result.

In this appendix we shall "translate" the usual proof of Doob's convergence theorem for non-negative martingales; for simplicity, we shall only consider forecasting task $\mathscr{G}^{\mathrm{mean}}$. Our aim here is to demonstrate, on a very simple example, how one can use the mathematical machine of the conventional theory of martingales in the game-theoretic framework of this paper. We shall also explain the main ideas used to prove the strong law of large numbers.

The *sample space* of forecasting task $\mathscr{G}^{\mathrm{mean}}$ is the set $[-1, 1]^{\infty}$ of all possible sequences $(m_1, x_1, m_2, x_2, \ldots)$ of Forecaster's and Nature's moves. A *random variable* is a real-valued function on the sample space, and an *adapted stochastic process* is a sequence $(\xi_0, \xi_1, \ldots)$ of random variables such that $\xi_i(m_1, x_1, m_2, x_2, \ldots)$ is determined by $(m_1, x_1, \ldots, m_i, x_i)$ alone, for all i. We say that an adapted stochastic process $(\xi_0, \xi_1, \ldots)$ is a *farthingale* if it is the capital process of some strategy for Statistician in $\mathscr{G}^{\mathrm{mean}}$; in other words, if there exists Statistician's strategy $\mathscr{S}$ in $\mathscr{G}^{\mathrm{mean}}(\xi_0)$ that ensures the following: for every $i \geqslant 1$ and every possible sequence $(m_1, x_1, m_2, x_2, \ldots)$ of Forecaster's and Nature's moves, Statistician's capital $\mathscr{K}_i$ at stage i is exactly $\xi_i(m_1, x_1, m_2, x_2, \ldots)$. A *substantive event* is any subset of the sample space. A substantive event E is *null* (and its complement E^{c} is *full*) if there exists a non-negative farthingale $(\xi_0, \xi_1, \ldots)$ with $\xi_0 = 1$ such that

$$\lim_{i \to \infty} \xi_i(m_1, x_1, m_2, x_2, \ldots) = \infty, \; \forall (m_1, x_1, m_2, x_2, \ldots) \in E$$

(cf. Vovk (1993a, Section 3)). This definition is equivalent to that given in Section 7.1. Now we can state a prequential counterpart of Doob's convergence theorem for non-negative martingales.

Lemma 2. *Every non-negative farthingale is convergent on a full event.*

Proof. Let $(\xi_0, \xi_1, \ldots)$ be a non-negative farthingale; we are required to construct a non-negative farthingale $(\eta_0, \eta_1, \ldots)$ with $\eta_0 = 1$ such that $\eta_i \to \infty$ $(i \to \infty)$ whenever ξ_i does not converge. Without loss of generality we assume that $\xi_0 = 1$. Let $(\xi_0, \xi_1, \ldots)$ be the capital process of Statistician's strategy $\mathscr{S}$. For every non-empty interval (a, b) with positive rational end points, let $\mathscr{S}^{(a,b)}$ be the following strategy. Define $\tau_0 = 0$ and, for $j = 1, 2, \ldots$,

$$\sigma_j = \min\{i > \tau_{j-1} \colon \xi_i > b\}, \qquad \tau_j = \min\{i > \sigma_j \colon \xi_i < a\};$$

the move recommended by $\mathscr{S}^{(a,b)}$ is

$$H_i^{(a,b)} = \begin{cases} H_i, & \text{if } \tau_{j-1} < i \leqslant \sigma_j \text{ for some } j, \\ 0, & \text{otherwise,} \end{cases}$$

where H_i is the move recommended by $\mathscr{S}$. The capital process $\xi_i^{(a,b)}$ of $\mathscr{S}^{(a,b)}$ in $\mathscr{G}^{\mathrm{mean}}(1)$ is a non-negative farthingale that tends to infinity when

$$\liminf_{i \to \infty} \xi_i < a, \qquad \limsup_{i \to \infty} \xi_i > b.$$

Therefore, the capital process of Statistician's strategy

$$\sum_{k=1}^{\infty} 2^{-k} \mathscr{S}^{(a_k, b_k)} \tag{47}$$

(where $((a_1, b_1), (a_2, b_2), \ldots)$ is some enumeration of all non-empty intervals with positive rational end-points) tends to infinity when ξ_i is divergent. (Series (47) converges because the move $H_i^{(a,b)}$ recommended by $\mathscr{S}^{(a,b)}$ never exceeds in absolute value the move H_i recommended by $\mathscr{S}$.) $\qquad\qquad\square$

Having proved an analogue of Doob's martingale convergence theorem, it is easy to prove the strong law of large numbers for $\mathscr{G}^{\mathrm{mean}}$ (Theorem 1(a)). We shall briefly describe the main steps of one of many possible proofs (for details see, for example, Vovk (1996, Section 5) or Shiryayev (1984, Section VII.5)); another proof will be given in Appendix 3.

An adapted stochastic process $(\xi_0, \xi_1, \ldots)$ is *increasing* if $\xi_0 = 0$ and $\xi_0 \leqslant \xi_1 \leqslant \ldots$. If an adapted stochastic process (ξ_i) can be represented as a farthingale minus an increasing adapted stochastic process, we say that (ξ_i) is a *superfarthingale*. (In other words, a superfarthingale is a farthingale in the game $\mathscr{G}^{\mathrm{mean}'}$ whose only difference from $\mathscr{G}^{\mathrm{mean}}$ is that Statistician is allowed to throw away money.)

It is obvious that in the above definition of a null event we can replace "farthingale" by "superfarthingale" and that Lemma 2 can be strengthened to the following.

Lemma 3. *Every non-negative superfarthingale is convergent on a full event.*

(A proof of Lemma 3 can be obtained by a simple modification of the proof of Lemma 2; also, Lemma 3 can be easily deduced from the statement of Lemma 2.)

A *predictable stochastic process* is a sequence $(\xi_0, \xi_1, \xi_2, \ldots)$ of random variables such that $\xi_i(m_1, x_1, m_2, x_2, \ldots)$ is determined by $(m_1, x_1, \ldots, m_{i-1}, x_{i-1}, m_i)$ alone (so the predictable stochastic processes constitute a subclass of the adapted stochastic processes). An adapted stochastic process (ξ_i) is a *semifarthingale* if it can be represented as the sum of a superfarthingale and an increasing predictable stochastic process (and any such increasing predictable stochastic process is called a *supercompensator* of (ξ_i)). It is not difficult to strengthen Lemma 3 as follows.

Lemma 4. *If (ξ_i) is a non-negative semifarthingale and (α_i) is its supercompensator, then the event*

$$\alpha_\infty < \infty \Rightarrow \xi_i \text{ is convergent}$$

is full.

(As usual, we let "$A \Rightarrow B$" stand for the event "(not A) or B".) Lemma 4 immediately implies the following lemma.

Lemma 5. *If* (ξ_i) *is a farthingale and* (ξ_i^2) *is a semifarthingale with supercompensator* (α_i), *then*

$$\alpha_\infty < \infty \Rightarrow \xi_i \text{ is convergent}$$

is a full event.

Proof. It suffices to apply Lemma 4 to the non-negative semifarthingales (ξ_i^2) and $\{(\xi_i + 1)^2\}$ with the same supercompensator (α_i) and note that $\xi_i = \frac{1}{2}\{(\xi_i + 1)^2 - \xi_i^2 - 1\}$. □

Now it is easy to obtain Theorem 1 (a); applying Lemma 5 to the farthingale

$$\xi_n := \sum_{i=1}^n \frac{X_i - M_i}{i}$$

and noticing that (ξ_n^2) is a semifarthingale with supercompensator $\alpha_n = \sum_{i=1}^n (2/i)^2$ (so that $\alpha_\infty < \infty$ everywhere), we obtain that it is a full event that ξ_n converges; it remains to apply Kronecker's lemma (see, for example, Stout (1974, Lemma 3.2.3)). □

It is natural to try to apply the idea of this proof to the "one-sided" Theorem 1 (c). However, this idea does not work in the "one-sided" case. For the game $\mathcal{G}^{\text{mean}^+}$, where Statistician is only allowed to buy a non-negative amount of E tickets, (ξ_n^2) is no longer a semifarthingale with supercompensator $\sum_{i=1}^n (2/i)^2$. (Also, it is not clear whether there exists any "one-sided" variant of Kronecker's lemma.) A very strong form of Theorem 1 (c) will be proved in Appendix 3.

With very minor changes, the method of proof given above also proves the following assertion (analogous to Kolmogorov's strong law of large numbers): in forecasting task $\mathcal{G}^{\text{mean\&var}'}$, whose difference from $\mathcal{G}^{\text{mean\&var}}$ is that X_i and M_i are now allowed to be chosen from the whole of $\mathbb{R}$, it is a full event that

$$\sum_{i=1}^\infty \frac{D_i}{i^2} < \infty \Rightarrow \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n (X_i - M_i) = 0.$$

Appendix 3: Law of the iterated logarithm

In this appendix we shall prove the "upper half" of the prequential law of the iterated logarithm: in forecasting task $\mathcal{G}^{\text{mean\&var}}$, it is a full event that

$$\limsup_{n\to\infty} \sum_{i=1}^n (X_i - M_i) \Big/ \left\{ 2\left(\sum_{i=1}^n D_i\right) \ln\ln\left(\sum_{i=1}^n D_i\right) \right\}^{1/2} \leq 1 \quad \text{or} \quad \sum_{i=1}^\infty D_i < \infty. \quad (48)$$

Our proof will follow Ville (1939, Section V.3). (In the case of two-valued X_i, this result has

been proved by Minozzo (1996, Section 5.6); for an "algorithmic" variant of Ville's proof, see Vovk (1988).) This proof is very natural, although this fact may be obscured by technical details. In this appendix we first give a formal proof with little explanation, then we explain the simple ideas behind the proof. It may be helpful to read the formal proof and the informal explanation in parallel.

Formal proof. Fix temporarily some (small) $\delta > 0$. For every $\kappa \in (0, \frac{1}{2})$, Statistician has a strategy $\mathscr{S}^{(\kappa)}$ in game $\mathscr{G}^{\text{mean\&var}}(1)$ such that, for all i, Statistician's capital $\mathscr{K}_i^{(\kappa)}$ satisfies $\mathscr{K}_0^{(\kappa)} = 1$ and

$$\mathscr{K}_{i+1}^{(\kappa)} = \mathscr{K}_i^{(\kappa)} \frac{1 + \kappa(X_i - M_i) + (1 + \delta)\kappa^2 (X_i - M_i)^2/2}{1 + (1 + \delta)\kappa^2 D_i/2} \tag{49}$$

when he is using that strategy. Indeed, (49) is equivalent to

$$\mathscr{K}_{i+1}^{(\kappa)} = \mathscr{K}_i^{(\kappa)} + \frac{\mathscr{K}_i^{(\kappa)}}{1 + (1 + \delta)\kappa^2 D_i/2} \left(\kappa(X_i - M_i) + (1 + \delta)\frac{\kappa^2}{2}\{(X_i - M_i)^2 - D_i\} \right);$$

so $\mathscr{S}^{(\kappa)}$ recommends buying

$$\frac{\mathscr{K}_i^{(\kappa)} \kappa}{1 + (1 + \delta)\kappa^2 D_i/2} \tag{50}$$

E tickets and

$$\frac{\mathscr{K}_i^{(\kappa)}(1 + \delta)\kappa^2/2}{1 + (1 + \delta)\kappa^2 D_i/2} \tag{51}$$

var tickets at every stage i. Note that (49) is always positive. Put

$$\kappa(k) = \{2(\ln k)(1 + \delta)^{-k}\}^{1/2}, \tag{52}$$

$$\mathscr{S} = \sum_{k=K}^{\infty} k^{-1-\delta} \mathscr{S}^{(\kappa(k))}, \tag{53}$$

where K is chosen so that $\kappa(k) \in (0, \frac{1}{2})$ for $k \geqslant K$ (we shall impose additional requirements on K below). Let $\mathscr{K}_i > 0$ be the capital process corresponding to $\mathscr{S}$, with initial capital $\mathscr{K}_0 = \sum_{k=K}^{\infty} k^{-1-\delta}$ (so we have $\mathscr{K}_i = \sum_{k=K}^{\infty} k^{-1-\delta} \mathscr{K}_i^{(\kappa(k))}$ for all i). We know that the event $\sup_i \mathscr{K}_i < \infty$ is full, which implies that

$$\sup_i \mathscr{K}_i^{(\kappa(k))} = O(k^{1+\delta}) \tag{54}$$

is also full.

Consider a sequence of moves such that both (54) and $\sum_{i=1}^{\infty} d_i = \infty$ hold. For such a sequence, and n large enough, we take

$$k = k(n) = \left\lfloor \log_{1+\delta} \sum_{i=1}^{n} d_i \right\rfloor \tag{55}$$

and obtain

$$\ln \mathcal{H}_n^{(\kappa(n))} \leq (1 + \delta) \ln k(n) + O(1) \leq (1 + \delta) \ln \ln \sum_{i=1}^{n} d_i + O(1),$$

where $\kappa(n) = \kappa(k(n))$. By (49), we further obtain

$$\ln \prod_{i=1}^{n} \frac{1 + \kappa(n)(x_i - m_i) + (1 + \delta)\kappa^2(n)(x_i - m_i)^2/2}{1 + (1 + \delta)\kappa^2(n)d_i/2} \leq (1 + \delta) \ln \ln \sum_{i=1}^{n} d_i + O(1),$$

i.e.

$$\sum_{i=1}^{n} \ln\left(1 + \kappa(n)(x_i - m_i) + (1 + \delta)\frac{\kappa^2(n)}{2}(x_i - m_i)^2\right) \leq \sum_{i=1}^{n} \ln\left(1 + (1 + \delta)\frac{\kappa^2(n)}{2}d_i\right)$$

$$+ (1 + \delta) \ln \ln \sum_{i=1}^{n} d_i + O(1). \tag{56}$$

For t small enough in absolute value we have

$$\ln\left(1 + t + (1 + \delta)\frac{t^2}{2}\right) \geq t;$$

so (56) implies that (remember that $\kappa(n) \to 0$ as $n \to \infty$)

$$\kappa(n) \sum_{i=1}^{n}(x_i - m_i) \leq \sum_{i=1}^{n} \ln\left(1 + (1 + \delta)\frac{\kappa^2(n)}{2}d_i\right) + (1 + \delta) \ln \ln \sum_{i=1}^{n} d_i + O(1).$$

Since always $\ln(1 + t) \leq t$, we further obtain

$$\sum_{i=1}^{n}(x_i - m_i) \leq (1 + \delta)\frac{\kappa(n)}{2}\sum_{i=1}^{n} d_i + \frac{1 + \delta}{\kappa(n)} \ln \ln \sum_{i=1}^{n} d_i + O\left(\frac{1}{\kappa(n)}\right). \tag{57}$$

For n large enough,

$$\frac{1}{1 + \delta}\left\{\left(2 \ln \ln \sum_{i=1}^{n} d_i\right) \middle/ \sum_{i=1}^{n} d_i\right\}^{1/2} \leq \kappa(n) \leq (1 + \delta)\left\{\left(2 \ln \ln \sum_{i=1}^{n} d_i\right) \middle/ \sum_{i=1}^{n} d_i\right\}^{1/2}, \tag{58}$$

which gives

$$\sum_{i=1}^{n}(x_i - m_i) \leq (1 + \delta)^2\left(2 \sum_{i=1}^{n} d_i \ln \ln \sum_{i=1}^{n} d_i\right)^{1/2} + O\left[\left\{\sum_{i=1}^{n} d_i \middle/ \left(\ln \ln \sum_{i=1}^{n} d_i\right)\right\}^{1/2}\right]. \tag{59}$$

Therefore, for any $\delta > 0$ we have exhibited (see (49) and (53)) Statistician's strategy $\mathscr{S} = \mathscr{S}(\delta)$ which, starting from initial capital $\mathscr{K}_0 = \mathscr{K}_0(\delta)$, will make an infinite gain unless the event

$$\sum_{i=1}^{n}(X_i - M_i) \leqslant (1 + \delta)^2 \left(2\sum_{i=1}^{n} D_i \ln\ln \sum_{i=1}^{n} D_i\right)^{1/2} + O\left[\left\{\sum_{i=1}^{n} D_i \Big/ \left(\ln\ln \sum_{i=1}^{n} D_i\right)\right\}^{1/2}\right]$$

$$\text{or} \quad \sum_{i=1}^{\infty} D_i < \infty$$

(cf. (59)) happens. Taking

$$\mathscr{S} = \sum_{j=1}^{\infty} 2^{-j} \mathscr{S}\left(\frac{1}{j}\right), \tag{60}$$

we obtain a strategy which, starting from

$$\mathscr{K}_0 = \sum_{j=1}^{\infty} 2^{-j} \mathscr{K}_0\left(\frac{1}{j}\right), \tag{61}$$

will make an infinite gain unless event (48) happens. It follows that (48) is full.

In order to make the above proof watertight we have to check that the various infinite mixtures appearing in it are convergent. First we prove that (61) is finite. This is easy:

$$\sum_{j=1}^{\infty} 2^{-j} \mathscr{K}_0\left(\frac{1}{j}\right) = \sum_{j=1}^{\infty} 2^{-j} \sum_{k=K(1/j)}^{\infty} k^{-1-1/j} \leqslant \sum_{j=1}^{\infty} 2^{-j} = 1,$$

provided that $K(1/j)$ are so large that

$$\sum_{k=K(1/j)}^{\infty} k^{-1-1/j} \leqslant 1, \qquad \forall j.$$

It remains to prove the convergence of series (53) and (60). Our task is facilitated by the fact that every strategy considered recommends buying *non-negative* amounts of E tickets and var tickets. It suffices to prove that the resulting strategy $\mathscr{S}$ recommends buying *finite* number of E tickets and var tickets at every stage.

First we note that $\delta \in (0, 1]$, $\kappa \in (0, \frac{1}{2})$ and (49) imply that

$$\sup_{k,\delta} \mathscr{K}_i^{(\kappa(k,\delta))} < \infty, \qquad \forall i.$$

(Now we explicitly indicate δ in our notation for $\kappa = \{2(\ln k)(1 + \delta)^{-k}\}^{1/2}$.) Since $\mathscr{S}^{(\kappa(k,\delta))}$ recommends buying (50) E tickets and (51) var tickets, it is enough to prove that

$$\sum_{j=1}^{\infty} 2^{-j} \sum_{k=K(1/j)}^{\infty} k^{-1-1/j} \kappa\left(k, \frac{1}{j}\right) < \infty \tag{62}$$

and

$$\sum_{j=1}^{\infty} 2^{-j} \sum_{k=K(1/j)}^{\infty} k^{-1-1/j} \kappa^2\left(k, \frac{1}{j}\right) < \infty.$$

The second inequality follows from the first; so all we are required to prove is (62). Since the inner series is convergent, we can ensure that

$$\sum_{k=K(1/j)}^{\infty} k^{-1-1/j} \kappa\left(k, \frac{1}{j}\right) \leqslant 1, \qquad \forall j,$$

taking K large enough. The proof is complete. $\square$

Informal explanation. Requirement (49) is about the same as

$$\mathscr{K}_{i+1}^{(\kappa)} = \mathscr{K}_i^{(\kappa)} \exp\left(\kappa(X_i - M_i) - \frac{\kappa^2 D_i}{2}\right)$$

(κ and δ are small). Therefore, the adapted stochastic process

$$\xi_n = \prod_{i=1}^{n} \exp\left(\kappa(X_i - M_i) - \frac{\kappa^2 D_i}{2}\right) = \exp\left(\kappa \sum_{i=1}^{n}(X_i - M_i) - \frac{\kappa^2}{2} \sum_{i=1}^{n} D_i\right) \qquad (63)$$

is nearly a farthingale. Suppose for some large n we have

$$\sum_{i=1}^{n}(x_i - m_i) \approx \left\{2\left(\sum_{i=1}^{n} d_i\right) \ln\ln\left(\sum_{i=1}^{n} d_i\right)\right\}^{1/2}$$

(this is the maximal value allowed by the law of the iterated logarithm). If we knew $\sum_{i=1}^{n} d_i$ in advance, we could use one of the "near farthingales" (63) to earn a lot of money. It is easy to see that the optimal value of κ would be

$$\kappa = \sum_{i=1}^{n}(x_i - m_i) \bigg/ \sum_{i=1}^{n} d_i$$

$$\approx \left\{\left(2\ln\ln \sum_{i=1}^{n} d_i\right) \bigg/ \sum_{i=1}^{n} d_i\right\}^{1/2}$$

(cf. (58)). It would be even sufficient to know $\sum_{i=1}^{n} d_i$ approximately (with a small relative error); therefore, (52) with k defined by (55) is a natural expression for κ. The problem is that even k is unknown in advance; the solution is to take the mixture of the "farthingales" corresponding to different k with weights $w(k)$ shrinking (as $k \to \infty$) as slowly as possible; we took $w(k) = k^{-1-\delta}$ (see (53)).

In the above argument the value

$$\kappa \approx \left\{\left(2\ln\ln \sum_{i=1}^{n} d_i\right) \bigg/ \sum_{i=1}^{n} d_i\right\}^{1/2}$$

was obtained from the statement of the law of the iterated logarithm. A less *ad hoc* way to obtain it is to minimize the sum of the first two addends in the right-hand side of (57) with respect to κ. $\square$

Extensions.

(i) We first show how the above proof implies Theorem 1 (c). Without loss of generality we can assume that at every stage i of game $\mathscr{G}^{99\%}$ Statistician can buy non-negative amounts of E tickets and var tickets; an E ticket pays £$\mathbf{I}(X_i \notin [L_i, U_i])$ and costs £0.01 and a var ticket pays £$\{\mathbf{I}(X_i \notin [L_i, U_i]) - 0.01\}^2$ and costs £0.0099. (The E tickets are exactly the "primary" tickets of game $\mathscr{G}^{99\%}$, and the pay-off

$$\{\mathbf{I}(X_i \notin [L_i, U_i]) - 0.01\}^2 - 0.0099 = 0.98\{\mathbf{I}(X_i \notin [L_i, U_i]) - 0.01\}$$

of a var ticket can be simulated as the pay-off of a "portfolio" consisting of 0.98 E tickets.) Noting that the above strategy for Statistician requires buying non-negative amounts of E tickets and var tickets (see (50) and (51)), we can see (cf. (48)) that

$$\limsup_{n\to\infty} \left(\sum_{i=1}^{n} \{\mathbf{I}(X_i \notin [L_i, U_i]) - 0.01\}/\{2(0.0099 n) \ln \ln n\}^{1/2} \right) \leq 1,$$

which is equivalent to

$$\liminf_{n\to\infty} \left(\sum_{i=1}^{n} \{\mathbf{I}(X_i \in [L_i, U_i]) - 0.99\}/(n \ln \ln n)^{1/2} \right) \geq -(0.0198)^{1/2}.$$

This inequality is much stronger than (34) (and it is easy to see that the constant $(0.0198)^{1/2}$ cannot be improved).

(ii) Finally, we shall deduce Theorem 1 (a) from our proof of the law of the iterated logarithm. Without loss of generality we can assume that at every stage i of the game $\mathscr{G}^{\mathrm{mean}}$ Statistician is allowed to buy any non-negative amount of var tickets (paying £$(X_i - M_i)^2$) for £4 each (indeed, a var ticket will bring $(X_i - M_i)^2 \leq 4$ to Statistician; so it never hurts Forecaster to sell them for £4). Recalling that (51) is never negative, we deduce (cf. (48)) that the event

$$\limsup_{n\to\infty} \left(\sum_{i=1}^{n} (X_i - M_i) \Big/ (8n \ln \ln n)^{1/2} \right) \leq 1$$

is full, which implies that the event

$$\limsup_{n\to\infty} \left(\left| \sum_{i=1}^{n} (X_i - M_i) \right| \Big/ (8n \ln \ln n)^{1/2} \right) \leq 1$$

is full. This is much stronger than the assertion that (31) is full. (Note, however, that this approach to the strong law of large numbers strongly relies on the assumption that the X_i have uniformly bounded support; this assumption can be weakened, but still this approach cannot be used to prove the extension to arbitrary real-valued quantities described at the end of Appendix 2.)

Acknowledgements

References

Billingsley, P. (1986) *Probability and Measure*, 2nd edn. New York: Wiley.

Cox, D.R. and Lewis, P.A.W. (1966) *Statistical Analysis of Series of Events*. London: Methuen.

Dawid, A.P. (1982) The well-calibrated Bayesian (with discussion). *J. Amer. Statist. Ass.*, **77**, 605–613.

Dawid, A.P. (1983) Inference, statistical: I. In S. Kotz, N.L. Johnson and C.B. Read (eds), *Encyclopedia of Statistical Sciences*, Vol. 4, pp. 89–105. New York: Wiley–Interscience.

Dawid, A.P. (1984) Statistical theory: the prequential approach (with discussion). *J. Roy. Statist. Soc. A*, **147**, 278–292.

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

Dawid, A.P. (1986) Probability forecasting. In S. Kotz, N.L. Johnson and C.B. Read (eds), *Encyclopedia of Statistical Sciences*, Vol. 7, pp. 210–218. New York: Wiley–Interscience.

Dawid, A.P. (1991) Fisherian inference in likelihood and prequential frames of reference (with discussion). *J. Roy. Statist. Soc. B*, **53**, 79–109.

Dawid, A.P. (1992a) Prequential data analysis. In M. Ghosh and P.K. Pathak (eds), *Current Issues in Statistical Inference: Essays in Honor of D. Basu* pp. 113–126. IMS Lecture Notes—Monograph Ser. 17. Hayward, CA: Institute of Mathematical Statistics.

Dawid, A.P. (1992b) Prequential analysis, stochastic complexity and Bayesian inference (with discussion). In J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith (eds), *Bayesian Statistics 4*, pp. 109–125. Oxford: Oxford University Press.

de Finetti, B. (1974) *Theory of Probability*, Vol. 1. London: Wiley.

Doob, J.L. (1953) *Stochastic Processes*. New York: Wiley.

Hill, T.P. (1982) Conditional generalizations of strong laws which conclude the partial sums converge almost surely. *Ann. Probab.*, **10**, 828–830.

Loveland, D. (1969) On minimal program complexity measures. *Proceedings of the First ACM Symposium on the Theory of Computing*, pp. 61–66. New York: ACM Press.

Martin, D. (1975) Borel determinacy. *Ann. Math.*, **102**, 363–371.

Martin, D. (1990) An extension of Borel determinacy. *Ann. Pure Appl. Logic*, **49**, 279–293.

Minozzo, M. (1996) On some aspects of the prequential and algorithmic approaches to probability and statistical theory. PhD Thesis, University of London.

Rosenblatt, M. (1952) Remarks on a multivariate transformation. *Ann. Math. Statist.*, **23**, 470–472.

Schnorr, C.P. (1970) Klassifikation der Zufallsgesetze nach Komplexität und Ordnung. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **16**, 1–21.

Schnorr, C.P. (1971a) A unified approach to the definition of random sequences. *Math. Systems Theory*, **5**, 246–258.

Schnorr, C.P. (1971b) *Zufälligkeit und Wahrscheinlichkeit*. Berlin: Springer-Verlag.

Seillier-Moiseiwitsch, F. and Dawid, A.P. (1993) On testing the validity of sequential probability forecasts. *J. Am. Statist. Assoc.*, **88**, 355–359.

Shafer, G. (1976) *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press.

Shafer, G. (1996) *The Art of Causal Conjecture*. Cambridge, MA: MIT Press.

Shiryayev, A.N. (1984) *Probability*. New York: Springer-Verlag.

Stout, W.F. (1970) A martingale analogue of Kolmogorov's law of the iterated logarithm. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **15**, 279–290.

Stout, W.F. (1974) *Almost Sure Convergence*. New York: Academic Press.

Ville, J. (1939) *Etude Critique de la Notion de Collectif*. Paris: Gauthier–Villars.

Vovk, V.G. (1987) The law of the iterated logarithm for random Kolmogorov, or chaotic, sequences. *Theory Probab. Applic.*, **32**, 413–425.

Vovk, V.G. (1988) Kolmogorov–Stout law of the iterated logarithm. *Math. Notes*, **44**, 502–507.

Vovk, V.G. (1993a) A logic of probability, with application to the foundations of statistics (with discussion). *J. Roy. Statist. Soc. B*, **55**, 317–351.

Vovk, V.G. (1993b) Forecasting point and continuous processes: prequential analysis. *Test*, **2**, 189–217.

Vovk, V.G. (1996) A purely martingale version of Kolmogorov's strong law of large numbers. *Theory Probab. Applic.*, **41**, 605–608.