Proceedings of the 2007 Winter Simulation Conference
S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, eds.

GAME-THEORETIC PROBABILITY AND DEFENSIVE FORECASTING

Glenn Shafer

Rutgers Business School
180 University Avenue
Newark, N.J. 07102 U.S.A.

ABSTRACT

In 2001, Vladimir Vovk and I demonstrated how game theory can replace measure theory as a foundation for classical probability theory, discrete and continuous (*Probability and Finance: Its Only a Game!*, Wiley 2001). In the game-theoretic framework, classical probability theorems are proven by betting strategies that make a player rich without risking bankruptcy if the theorem's prediction fails. These strategies can be specified explicitly, and so the theory has a constructive flavor that lends itself to applications in economics and statistics.

Defensive forecasting is one of the most interesting of these applications. It identifies a comprehensive betting strategy, which becomes rich if the probabilities fail in a relevant way (say by being uncalibrated or having poor resolution), and it chooses probabilities to defeat this comprehensive betting strategy. The fact that this is possible gives us new insight into the very meaning of probability.

1 INTRODUCTION

Cournot's principle says that an event of small or zero probability singled out in advance will not happen. From the turn of the twentieth century through the 1950s, many mathematicians, including Aleksandr Chuprov, Émile Borel, Maurice Fréchet, Paul Lévy, and Andrei Kolmogorov, saw this principle as fundamental to the application and meaning of probability. In their view, a probability model gains empirical content only when it rules out an event by assigning it small or zero probability.

In his doctoral dissertation, published in 1939, Jean Ville showed that Cournot's principle can be given a game-theoretic interpretation. Shafer and Vovk (2001), (`<www.probabilityandfinance.com>`) extend Ville's game-theoretic approach to cases where successive forecasts fall short of a full probability distribution for the quantities forecast.

2 THE ORIGINS OF COURNOT'S PRINCIPLE

An event with very small probability is *morally impossible*; it will not happen. Equivalently, an event with very high probability is *morally certain*; it will happen. This principle was first formulated within mathematical probability by Jacob Bernoulli. In his *Ars Conjectandi*, published posthumously in 1713, Bernoulli proved that in a sufficiently long sequence of independent trials of an event, there is a very high probability that the frequency with which the event happens will be close to its probability. Bernoulli explained that we can treat the very high probability as moral certainty and so use the frequency of the event as an estimate of its probability.

Augustin Cournot, a mathematician now remembered as an economist and a philosopher of science (Martin 1996, Martin 1998), gave the discussion a nineteenth-century cast in his 1843 treatise on probability (Cournot 1843). Because he was familiar with geometric probability, Cournot could talk about probabilities that are vanishingly small. He brought physics to the foreground. It may be mathematically possible, he argued, for a heavy cone to stand in equilibrium on its vertex, but it is physically impossible. The event's probability is vanishingly small. Similarly, it is physically impossible for the frequency of an event in a long sequence of trials to differ substantially from the event's probability (Cournot 1843, pp. 57 and 106).

At the turn of the twentieth century, it was a commonplace among statisticians that one must decide what level of probability will count as practical certainty in order to apply probability theory. We find this stated explicitly in 1901, for example, in the articles by Georg Bohlmann and Ladislaus von Bortkiewicz in the section on probability in the *Encyklopädie der mathematischen Wissenschaften* (von Bortkiewicz 1901, p. 825) (Bohlmann 1901, p. 861). Aleksandr Chuprov, professor of statistics in Petersburg, was the champion of Cournot's principle in Russia. He called it Cournot's lemma (Chuprov 1910, p. 167) and declared it a basic principle of the logic of the probable (Sheynin 1996, pp. 95–96).

Saying that an event of very small or vanishingly small probability will not happen is one thing. Saying that probability theory gains empirical meaning only by ruling out the happening of such events is another. Cournot may have been the first to make this second assertion:

> . . . *The physically impossible event is therefore the one that has infinitely small probability*, and only this remark gives substance—objective and phenomenal value—to the theory of mathematical probability (Cournot 1843, p. 78).

Paul Lévy, a French mathematician who began writing on probability in the 1920s, stands out for the clarity of his articulation of the thesis that Cournot's principle is the only way of connecting a probabilistic theory with the world outside mathematics (Lévy 1925). Lévy's views were widely shared in France. In the 1940s, Émile Borel called Cournot's principle first "the only law of chance" (la loi unique du hasard) (Borel 1943, Borel 1950). Neither Lévy nor Borel used the name "Cournot's principle," which was coined by Maurice Fréchet in 1949. Fréchet's inspiration was Oskar Anderson, who had talked about the Cournotsche Lemma (Cournot's lemma) and the Cournotsche Brücke (Cournot's bridge) (Anderson 1935, Anderson 1949). Anderson was following his teacher Chuprov in the use of "lemma." Fréchet felt that "lemma," like "theorem," should be reserved for purely mathematical results and so suggested "principe de Cournot." Fréchet's coinage was used in the 1950s in French, German, and English (de Finetti 1951, von Hirsch 1954, Richter 1954, Richter 1956).

3 VILLE'S THEOREM

Vovk and I (Shafer and Vovk 2001) use Cournot's principle in a game-theoretic form: a strategy for placing bets without risking bankruptcy will not multiply the bettor's capital by a large or infinite factor. In the case where the bettor can buy or sell any random variable for its expected value, this is equivalent to the classical form of the principle; Jean Ville demonstrated the equivalence in 1939 (Ville 1939).

Consider a sequence $Y_1, Y_2, \dots$ of binary random variables with a joint probability distribution P. Suppose, for simplicity, that P assigns every finite sequence $y_1, \dots, y_n$ of 0s and 1s positive probability, so that its conditional probabilities for Y_n given values of the preceding variables are always unambiguously defined. Following Jean Ville (Ville 1939), consider a gambler who begins with \$1 and is allowed to bet as he pleases on each round, provided that he does not risk bankruptcy. We can formalize this with the following protocol, where betting on Y_n is represented as buying some number s_n (possibly zero or negative) of tickets that cost \$$\mathrm{P}\{Y_n = 1 | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}\}$ and pay \$$Y_n$.

BINARY PROBABILITY PROTOCOL
Players: Reality, Skeptic
Protocol:

$\mathscr{K}_0 := 1$.
FOR $n = 1, 2, \dots$:
Skeptic announces $s_n \in \mathbb{R}$.
Reality announces $y_n \in \{0, 1\}$.
$\mathscr{K}_n := \mathscr{K}_{n-1}$
$+ s_n(y_n - \mathrm{P}\{Y_n = 1 | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}\})$.

Restriction on Skeptic: Skeptic must choose the s_n so that his capital is always nonnegative ($\mathscr{K}_n \geq 0$ for all n) no matter how Reality moves.

This is a perfect-information sequential protocol; moves are made in the order listed, and each player sees the other player's moves as they are made. The sequence $\mathscr{K}_0, \mathscr{K}_1, \dots$ is Skeptic's capital process.

Ville showed that Skeptic's getting rich in this protocol is equivalent to an event of small probability happening, in the following sense:

1. When Skeptic follows a measurable strategy (a rule that gives s_n as a function of $y_1, \dots, y_{n-1}$),
$$\mathrm{P}\{\sup_n \mathscr{K}_n \geq \frac{1}{\varepsilon}\} \leq \varepsilon \tag{1}$$
for every $\varepsilon > 0$. (This is because the capital process $\mathscr{K}_0, \mathscr{K}_1, \dots$ is a non-negative martingale; Equation (1) is sometimes called *Doob's inequality*.)
2. If A is a measurable subset of $\{0,1\}^\infty$ with $\mathrm{P}(A) \leq \varepsilon$, then Skeptic has a measurable strategy that guarantees
$$\liminf_{n \to \infty} \mathscr{K}_n \geq \frac{1}{\varepsilon}$$
whenever $(y_1, y_2, \dots) \in A$.

We can summarize these results by saying that Skeptic's being able to multiply his capital by a factor of $1/\varepsilon$ or more is equivalent to the happening of an event with probability ε or less.

Ville's work was motivated by von Mises's notion of a collective (von Mises 1919, von Mises 1928, von Mises 1931). Von Mises had argued that a sequence $y_1, y_2, \dots$ of 0s and 1s should be considered random if no subsequence with a different frequency of 1s can be picked out by a gambler to whom the *y*s are presented sequentially; this condition, von Mises felt, would keep the gambler from getting rich by deciding when to bet. Ville showed that von Mises's condition is insufficient, inasmuch as it does not rule out the gambler's getting rich by varying the direction and amount to bet.

4 THE GAME-THEORETIC FRAMEWORK

Although the preceding explanation of Ville's ideas was limited to the binary case, Ville made it clear that these ideas apply whenever conditional probabilities from a joint probability distribution for a sequence of random variables are used to make successive probability predictions. The framework of (Shafer and Vovk 2001) generalizes the ideas further. The generalization has three aspects:

- Instead of beginning with a probability measure and using its conditional probabilities or expected values as prices on each round, we allow another player, Forecaster, to set the prices as play proceeds. This makes the framework "prequential" (Dawid 1984); there is no need to specify what the price on the nth round would be had Reality moved differently on earlier rounds.
- When convenient, we make explicit additional information, say x_n, that Reality provides to Forecaster and Skeptic before they make their nth moves.
- We allow the story to be multi-dimensional, with Reality making several moves and Forecaster pricing them all.

A convenient level of generality for the present discussion is provided by the following protocol, where $\mathbb{R}^k$ is k-dimensional Euclidean space, $\mathbf{Y}$ is a subset of $\mathbb{R}^k$, and $\mathbf{X}$ is an arbitrary set.

LINEAR FORECASTING PROTOCOL
Players: Reality, Forecaster, Skeptic
Protocol:
$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots, N$:
Reality announces $x_n \in \mathbf{X}$.
Forecaster announces $f_n \in \mathbb{R}^k$.
Skeptic announces $s_n \in \mathbb{R}^k$.
Reality announces $y_n \in \mathbf{Y}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + s_n \cdot (y_n - f_n)$.
Restriction on Skeptic: Skeptic must choose the s_n so that his capital is always nonnegative ($\mathcal{K}_n \geq 0$ for all n) no matter how the other players move.

Here $s_n \cdot (y_n - f_n)$ is the dot product of the k-dimensional vectors s_n and $y_n - f_n$. Notice also that play stops on the Nth round rather than continuing indefinitely. This is a convenient assumption in this section, where we emphasize the finitary picture; we will return to the infinitary picture later.

The linear forecasting protocol covers many prediction problems considered in statistics (where x and y are often called *independent* and *dependent* variables, respectively) and machine learning (where x is called the *object* and y the *label*) (Vovk, Gammerman, and Shafer 2005, Hastie, Tibshirani, and Friedman 2001, Vapnik 1996). Market games can be included by taking f_n to be a vector of opening prices and y_n the corresponding vector of closing prices for the nth trading period.

A strategy for Skeptic in the linear forecasting protocol is a rule that gives each of his moves s_n as a function of the preceding moves by Reality and Forecaster, $(x_1, f_1, y_1), \ldots, (x_{n-1}, f_{n-1}, y_{n-1}), x_n, f_n$. A strategy for Forecaster is a rule that gives each of his moves f_n as a function of the preceding moves by Reality and Skeptic, $(x_1, s_1, y_1), \ldots, (x_{n-1}, s_{n-1}, y_{n-1}), x_n$. One way of prescribing a strategy for Forecaster is to choose a probability distribution for $(x_1, y_1), (x_2, y_2), \ldots$ and set f_n equal to the conditional expected value of y_n given $(x_1, y_1), \ldots, (x_{n-1}, y_{n-1}), x_n$. We will look at other interesting strategies for Forecaster in §8.

How can one express confidence in Forecaster? The natural way is to assert Cournot's principle: say that a legal strategy for Skeptic (one that avoids $\mathcal{K}_n < 0$ no matter how the other players move) will not multiply Skeptic's initial capital by a large factor.

Once we adopt Cournot's principle in this form, it is natural to scale the implications of our confidence in Forecaster the same way we do in classical probability. This means treating an event that happens only when a specified legal strategy multiplies the capital by $1/\varepsilon$ as no more likely than an event with probability ε.

As in classical probability, we can combine Cournot's principle with a form of Bernoulli's theorem to obtain a statement about relative frequency in a long sequence of events. In a sufficiently long sequence of events with upper probability 0.1 or less, for example, it is morally certain that no more than about 10% of the events will happen (Shafer and Vovk 2002, §5.3). This is a martingale-type result; rather than insist that the events be independent in some sense, we assume that the upper probability for each event is calculated at the point in the game where the previous event is settled.

5 EXTENDING THE CLASSICAL LIMIT THEOREMS

One of the main contributions of Shafer & Vovk (Shafer and Vovk 2001) was to show that game theory can replace measure theory as a foundation for classical probability.

We showed in particular that classical limit theorems, especially the strong law of large numbers and the law of the iterated logarithm, can be proven constructively within a purely game-theoretic framework. From Ville's work, we know that for any event with probability zero, there is a strategy for Skeptic that avoids bankruptcy for sure and makes him infinitely rich if the event fails. But constructing the strategy is another matter. In the case of the events of probability zero associated with the classical theorems, we did construct the requisite strategies; they are computable and continuous.

We provided similar constructions for classical results that do not require an infinite number of rounds of play to be meaningful: the weak law of large numbers, finitary versions of the law of the iterated logarithm, and the central limit theorem.

6 IMPLICATIONS FOR MARKET PRICES

Organized exchanges, in which a buyer or seller can always find a ready price for a particular commodity or security, are forecasting games. So we can ask whether Cournot's principle holds in such exchanges, and we can consider the implications of its holding. It is often said that in an efficient market, an investor cannot make a lot of money without taking undue risk. Cournot's principle makes this precise by saying that he will not make a lot of money without risking bankruptcy; he starts with a certain initial capital, and on each round of trading he risks at most a portion of his current capital. This principle alone can explain certain stylized facts about prices that are often explained using stochasticity.

6.1 The $\sqrt{dt}$ effect

Consider first the stylized fact that changes in market prices over an interval of time of length dt scale as $\sqrt{dt}$. In a securities market where shares are traded 252 days a year, for example, the typical change in price of a share from one year to the next is $\sqrt{252}$, or about 16, times as large as the typical change from one day to the next. There is a standard way of explaining this. We begin by assuming that price changes are stochastic, and we argue that successive changes must be uncorrelated; otherwise someone who knew the correlation (or learned it by observation) could devise a trading strategy with positive expected value. Uncorrelatedness of 252 successive daily price changes implies that their sum, the annual price change, has variance 252 times as large and hence standard deviation, or typical value, $\sqrt{252}$ times as large. This is a simple argument, but stochastic ideas intervene in two places, first when price changes are assumed to be stochastic, and then when market efficiency is interpreted as the absence of a trading strategy with positive expected value. As I now explain, we can replace this stochastic argument with a purely game-theoretic argument, in which Cournot's principle expresses the assumption of market efficiency.

For simplicity, consider the following protocol, which describes a market in shares of a corporation. Investor plays the role of Skeptic; he tries to make money, and Cournot's principle says he cannot get very rich following the rules, which do not permit him to risk bankruptcy. Market plays the roles of Forecaster (by giving opening prices) and Reality (by giving closing prices). For simplicity, we suppose that today's opening price is yesterday's closing price, so that Market gives only one price each day, at the end of the day. When Investor holds s_n shares during day n, he makes $s_n(y_n - y_{n-1})$, where y_n is the price at the end of day n.

THE MARKET PROTOCOL
Players: Investor, Market
Protocol:
$\mathscr{K}_0 := 1$.
Market announces $y_0 \in \mathbb{R}$.
FOR $n = 1, 2, \ldots, N$:
Investor announces $s_n \in \mathbb{R}$.
Market announces $y_n \in \mathbb{R}$.
$\mathscr{K}_n := \mathscr{K}_{n-1} + s_n(y_n - y_{n-1})$.
Restriction on Investor: Investor must choose the s_n so that his capital is always nonnegative ($\mathscr{K}_n \geq 0$ for all n) no matter how Market moves.

For simplicity, we ignore the fact that the price y_n of a share cannot be negative.

Since there is no stochastic assumption here, we cannot appeal to the idea of the variance of a probability distribution for price changes to explain what $\sqrt{dt}$ scaling means. But we can use

$$\sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - y_{n-1})^2} \tag{2}$$

as the typical daily change, and we can compare it to the magnitude of the change we see over the whole game, say

$$\max_{0<n\leq N} |y_n - y_0| \tag{3}$$

The quantity (3) should have the same order of magnitude as $\sqrt{N}$ times the quantity (2). Equivalently, we should have

$$\sum_{n=1}^{N}(y_n - y_{n-1})^2 \sim \max_{0<n\leq N}(y_n - y_0)^2, \tag{4}$$

where $\sim$ is understood to mean that the two quantities are of the same order of magnitude.

Does Cournot's principle give us any reason to think that (4) should hold? Indeed it does. As it turns out, Investor has a legal strategy (one avoiding bankruptcy) that makes a lot of money if (4) is violated. Market (who here represents all the other investors and speculators) wants to set prices so that Investor will not make a lot money, and we shall see, in §8 that he can more or less do so. So we may expect (4) to hold.

The strategy that makes money if (4) is violated is an average of two strategies, one a momentum strategy (holding more shares after the price goes up), the other a contrarian strategy (holding more shares after the price goes down).

1. The momentum strategy is based on the assumption that Investor can count on $\sum(y_n - y_{n-1})^2 \leq E$ and $\max(y_n - y_0)^2 \geq D$, where D and E are known constants. On this assumption, the strategy is legal and turns \$1 into $\$D/E$ or more for sure.
2. The contrarian strategy is based on the assumption that Investor can count on $\sum(y_n - y_{n-1})^2 \geq E$ and $\max(y_n - y_0)^2 \leq D$, where D and E are known constants. On this assumption, the strategy is legal and turns \$1 into $\$E/D$ or more for sure.

If the assumptions about $\sum(y_n - y_{n-1})^2$ and $\max(y_n - y_0)^2$ fail, then the strategy fails to make money, but Investor can still avoid bankruptcy. For details, see (Vovk and Shafer 2003).

6.2 The game-theoretic CAPM

The Capital Asset Pricing Model (CAPM), popular in finance theory for almost forty years, assumes that a firm whose shares are traded in a securities market has a stable level of risk relative to the market as a whole. The risk for a security s is defined in terms of a probability model for the returns of all the securities in the market; it is the theoretical regression coefficient

$$\beta_s = \frac{\mathrm{Cov}(R_s, R_m)}{\mathrm{Var}(R_m)}, \tag{5}$$

where R_s is a random variable whose realizations are s's returns, and R_m is a random variable whose realizations are a market index's returns.(Here "return" means simple return; $R = (p_{n+1} - p_n)/p_n$, where p_n is the the price of the share (or the level of the market index) at time n. All expected values, variances, and covariances are with respect to probabilities conditional on information known at time n.) The CAPM says that

$$\mathrm{E}(R_s) = r + \beta_s(\mathrm{E}(R_m) - r), \tag{6}$$

where r is rate of interest on government debt, assumed to be constant (Copeland and Weston 1988, p. 197). Because $\mathrm{E}(R_m) - r$ is usually positive, this equation suggests that securities with higher β have higher average returns. The equation has found only weak empirical confirmation, but it continues to be popular because it suggests plausible ways of analyzing decision problems faced by financial managers.

As it turns out, a purely game-theoretic argument based on Cournot's principle leads to an analogous equation involving only observed returns, with no reference to a probability distribution. The game-theoretic equation is

$$\bar{r}_s \sim r' + b_s(\bar{r}_m - r'), \tag{7}$$

where

$$\bar{r}_s := \frac{1}{N}\sum_{n=1}^{N} s_n, \qquad \bar{r}_m := \frac{1}{N}\sum_{n=1}^{N} m_n,$$

and

$$b_s := \frac{\sum_{n=1}^{N} s_n m_n}{\sum_{n=1}^{N} m_n^2}, \qquad r' := \bar{r}_m - \frac{1}{N}\sum_{n=1}^{N} m_n^2,$$

s_n and m_n being the actual returns of s and the market index, respectively, over period n. This is analogous to (6), inasmuch as r' measures the performance of the market as a whole, and the other quantities are empirical analogues of the theoretical quantities in (6).

The interpretation of (7) is similar to the interpretation of the game-theoretic version of $\sqrt{dt}$ scaling, equation (4); a speculator can make money to the extent it is violated. Given the approximations in the derivation of (7), as well as the existence of transaction costs and other market imperfections, we can expect the relation to hold only loosely, but we can ask whether it is any looser in practice than the empirical relations implied by CAPM. If not, then the very approximate confirmation of CAPM that has been discerned in data might be attributed to (7), leaving nothing that can be interpreted as empirical justification for the stochastic assumptions in CAPM. For details, see (Vovk and Shafer 2002).

7 THE IDEA OF A QUASI-UNIVERSAL TEST

If two events have very small probability, their union also has reasonably small probability. The analogous idea in game-theoretic probability is that of averaging strategies: if one strategy for Skeptic makes him very rich without risking bankruptcy if one event happens, and another makes him very rich without risking bankruptcy if a second event happens, then the average of the two strategies will make him reasonably rich without risking bankruptcy if either of the events happens. This leads us to the notion of a quasi-universal strategy: we list the most important extreme events that we want to rule out, and by averaging the strategies that rule each out, we obtain a strategy that rules them all out.

Leaving aside how this idea has been developed in the past within measure-theoretic probability, let us consider how it can be developed measure-theoretically in this protocol of binary forecasting:

BINARY PROBABILITY PROTOCOL WITH FORECASTER AND OBJECTS

Players: Reality, Forecaster, Skeptic

Protocol:

$\mathcal{K}_0 := 1$.

FOR $n = 1, 2, \ldots$:

Reality announces $x_n \in \mathbf{X}$.

Forecaster announces $p_n \in [0, 1]$.

Skeptic announces $s_n \in \mathbb{R}$.

Reality announces $y_n \in \{0, 1\}$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + s_n(y_n - p_n)$.

Restriction on Skeptic: Skeptic must choose the s_n so that his capital is always nonnegative ($\mathcal{K}_n \geq 0$ for all n) no matter how the other players move.

In this protocol, where Forecaster gives a probability p_n on each round, taking into account the previous outcomes $y_1, \ldots, y_{n-1}$ and auxiliary information $x_1, \ldots, x_n$, we are mainly interested in two aspects of the agreement between the probabilities p_n and the outcomes y_n:

Calibration. Whenever there are a large number of rounds on which p_n is close to some fixed probability p^*, we want the frequency with which $y_n = 1$ on those rounds to be approximately equal to p^*.

Resolution. We want this approximate equality between frequency and p^* to remain true when we consider only rounds where p_n is close to p^* and also x_n is close to some fixed value x^* in the object space $\mathbf{X}$.

As it turns out (Vovk, Takemura, and Shafer 2005), we can often average strategies that reject Forecaster's performance over a grid of values of (x^*, p^*) that are sufficiently dense to capture all deviations of practical interest. This average strategy, which is testing for calibration and resolution, will not necessarily test for more subtle deviations by $y_1, y_2, \ldots$ from the forecasts $p_1, p_2, \ldots$, such as those associated with the law of the iterated logarithm or Ville's refutation of von Mises's theory, but these more subtle deviations may hold little interest. So the average strategy can be regarded, for practical purposes, as a universal test. To avoid confusion, I call it a *quasi-universal strategy*.

8 DEFENSIVE FORECASTING

In cases where we have a quasi-universal strategy, a new opportunity opens up for Forecaster. Forecaster will do well enough if he can avoid rejection by that strategy. Formally, he needs a winning strategy in a version of the game where Skeptic is required to follow the quasi-universal strategy but Reality is free to move as she pleases. Does Forecaster have such a winning strategy? The surprising answer is yes.

This is easiest to see in the case where the quasi-universal strategy gives a move for the nth round that is continuous in the forecast p_n. As it happens, this is not an unreasonable requirement. We can construct quasi-universal strategies for calibration and resolution that are continuous in this respect, and there is even a philosophical argument for ruling out any discontinuous strategy for Skeptic: discontinuous functions are not really computable (Brouwer 1918, Martin-Löf 1970).

As it turns out, it is easy to show that for any forecast-continuous strategy for Skeptic there exists a strategy for Forecaster that does not allow Skeptic's capital to grow, regardless of what Reality does. Let me repeat the simple proof given in (Vovk, Nouretdinov, Takemura, and Shafer 2005, Vovk, Takemura, and Shafer 2005). It begins by simplifying so that Forecaster's job seems to be even a little harder. Instead of requiring that the entire forecast-continuous strategy for Skeptic be announced at the beginning of the game, we ask only that Skeptic announce his strategy for each round before Forecaster's move on that round. And we drop the restriction that Skeptic avoid risk of bankruptcy. This produces the following protocol:

BINARY FORECASTING AGAINST CONTINUOUS TESTS

Players: Reality, Forecaster, Skeptic

Protocol:

$\mathcal{K}_0 := 1$.

FOR $n = 1, 2, \ldots$:

Reality announces $x_n \in \mathbf{X}$.

Skeptic announces continuous $S_n : [0, 1] \to \mathbb{R}$.

Forecaster announces $p_n \in [0, 1]$.

Reality announces $y_n \in \{0, 1\}$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + S_n(p_n)(y_n - p_n)$.

Here S_n is Skeptic's strategy for the nth round; it gives his move as a function of Forecaster's not-yet-announced move p_n.

Theorem 1 *Forecaster has a strategy that ensures* $\mathcal{K}_0 \geq \mathcal{K}_1 \geq \mathcal{K}_2 \geq \cdots$.

Proof Because S_n is continuous, Forecaster can use the following strategy:

- if the function $S_n(p)$ takes the value 0, choose p_n so that $S_n(p_n) = 0$;
- if S_n is always positive, take $p_n := 1$;
- if S_n is always negative, take $p_n := 0$.

This guarantees that $S_n(p_n)(y_n - p_n) \leq 0$, so that $\mathcal{K}_n \leq \mathcal{K}_{n-1}$. ∎

Some readers may question the philosophical rationale for requiring that S_n be continuous. As it turns out, dropping this requirement does not cost us much; Forecaster can still win if we allow him to randomize (Vovk and Shafer 2005). This means that instead of telling Reality his probability p_n, Forecaster may give Reality only a probability distribution P_n for p_n, with the value p_n to be drawn from P_n out of sight of Reality or perhaps after Reality has selected y_n.

A strategy for Forecaster is what one usually calls a probability model; given the previous outcomes $y_1, \ldots, y_{n-1}$ and auxiliary information $x_1, \ldots, x_n$, it gives a probability p_n for $y_n = 1$. Such probabilities can be used in any repetitive decision problem (Vovk 2005). So Theorem 1's guarantee that they are valid, in the sense that they pass any reasonable test of calibration and resolution, has immense practical significance.

8.1 PHILOSOPHICAL IMPLICATIONS

Until the middle of the twentieth century, specialists in mathematical probability generally assumed that any probability can be known, either a priori or by observation. Those who understood probability as a measure of belief did not question the presumption that one can know one's beliefs. Those who understood probability as relative frequency assumed that one can observe frequencies. Those who interpreted probability using Cournot's principle did so on the assumption that they would know the probabilities they wanted to test; you would not check whether an event of small probability happened unless you had conjectured it had small probability.

The observations necessary for estimating a numerical probability may be hard to come by. But at worst, Cournot suggested, they could be made by a superior intelligence who represents the limits of what humans can observe (Martin 1996, pp. 146–150). Here Cournot was drawing an analogy with the classical understanding of determinism. Classical determinism required more than the future being determined in some theological sense; it required that the future be predictable by means of laws that can be used by a human, or at least by a superior intelligence whose powers of calculation and observation are human-like.

The presumption that probabilities be knowable leads to the apprehension that some events may not have probabilities. Perhaps there are three categories of events:

1. Those we can predict with certainty.
2. Those we can predict only probabilistically.
3. Those that we can predict neither with certainty nor probabilistically.

Most probabilists did think that there are events in the third category. Kolmogorov said so explicitly, and he did not speak of them as events whose probabilities cannot be known; he spoke of them as events that do not have probabilities (Kolmogorov 1983, p. 1). John Maynard Keynes and R. A. Fisher, each in his own way, also insisted that not every event has a numerical probability (Keynes 1921, Keynes 1937, Fisher 1956).

Doob's success in formalizing the concept of a probability measure for an arbitrary stochastic process destabilized this consensus. As I have already emphasized, there are many cases where we cannot repeat an entire stochastic process—cases where there is only one realization, one time series. In these cases, the probability measure assigns probabilities to many events that are not repeated. Having no direct frequency interpretation, these probabilities cannot be verified in any direct way. Because Doob did not appeal to Cournots principle or provide any other guidance about their meaning, his followers looked in other directions for understanding. Many looked towards mechanisms, such as well-balanced dice, that produce or at least simulate randomness. As they saw it, phenomena must be produced in some way. Deterministic phenomena are produced by deterministic mechanisms, indeterministic phenomena by chance mechanisms. The probabilities, even if unverifiable and perhaps unknowable, are meaningful because they have this generative task.

The growing importance of this way of seeing the world is evidenced by a pivotal article published by Jerzy Neyman in 1960 (Neyman 1960). According to Neyman, science was moving into a period of dynamic indeterminism,

> . . . characterized by the search for evolutionary chance mechanisms capable of explaining the various frequencies observed in the development of phenomena studied. The chance mechanism of carcinogenesis and the chance mechanism behind the varying properties of the comets in the Solar System exemplify the subjects of dynamic indeterministic studies. One might hazard the assertion that every serious contemporary study is a study of the chance mechanism behind some phenomena. The statistical and probabilistic tool in such studies is the theory of stochastic processes. . .

As this quotation confirms, Neyman was a frequentist. But his rhetoric suggests that the initial meaning of probabilities lies in their relation to how phenomena are generated rather than in their relation to frequencies. He wants to explain frequencies, but he does not ask that every probability have a frequency interpretation. Perhaps it is enough that successive probability predictions be well calibrated and have good resolution in the sense explained in §7.

What is most striking about Neyman's vision is that stochastic processes appear as the only alternative to deterministic models. The third category of phenomena, those we can predict neither with certainty nor probabilistically, has disappeared. This way of thinking has become ever more dominant since 1960. In many branches of science, we now hear casual references to “true,” “physical,” or “objective” probabilities, without any hesitation about their existence. An indeterministic process is assumed to be a stochastic process, regardless of whether we do or even can know the probabilities. The naïveté derided by von Kries 120 years ago is once again orthodoxy.

Our game-theoretic results provide a framework for regaining the philosophical sophistication of von Kries, Keynes, Fisher, and Kolmogorov, without abandoning the successes achieved by the theory of stochastic processes. Whenever we test a stochastic process empirically, we are applying Cournot’s principle to known (hypothetical) probabilities. When we have less than a stochastic process, a model giving only limited prices or probabilities, we can still test it via Cournot’s principle, without regarding it as part of some unknowable yet somehow still meaningful full stochastic process.

The possibility of defensive forecasting reveals that in a certain limited sense, our third category is indeed empty. Any quantity or event that can be placed in a series (in a time series, not necessarily a series of independent repetitions) can be predicted probabilistically, at least with respect to that series. This suggests that talk about chance mechanisms is also empty. Defensive forecasting works for any time series, regardless of how it is generated. The idea of a chance mechanism adds nothing.

REFERENCES

Anderson, O. N. 1935. *Einführung in die mathematische Statistik*. Vienna: Springer.

Anderson, O. N. 1949. Die Begründung des Gesetzes der grossen Zahlen und die Umkehrung des Theorems von Bernoulli. *Dialectica* 3 (9/10): 65–77.

Bernoulli, J. 1713. *Ars conjectandi*. Basel: Thurnisius. Edith Sylla’s English translation, *The Art of Conjecturing, together with Letter to a Friend on Sets in Court Tennis*, was published by Johns Hopkins University Press in 2005. Oscar Sheynin’s English translation of Part IV, dated 2005, is at <www.sheynin.de>.

Bohlmann, G. 1901. Lebensversicherungs-Mathematik. In *Encyklopädie der mathematischen Wissenschaften, Bd. I, Teil 2*, 852–917. Leipzig: Teubner.

Borel, E. 1943. *Les probabilités et la vie*. Paris: Presses Universitaires de France. Later editions in 1946, 1950, 1958, and 1967. The 1958 edition was translated into English by Maurice Baudin and published as *Probabilities and Life* by Dover, New York, in 1962.

Borel, E. 1950. *Probabilité et certitude*. Paris: Presses Universitaires de France. An English translation, *Probability and Certainty*, was published in 1963 by Walker, New York.

Brouwer, L. E. J. 1918. Begründung der Mengenlehre unabhängig vom logischen Satz vom ausgeschlossenen Dritte. Erster Teil. Allgemeine Mengelehre. *Koninklijke Nederlandse Akademie van Wetenschschappen Verhandelingen* 5:1–43.

Chuprov, A. A. 1910. Очерки по теории статистики *(Essays on the theory of statistics)*. Second ed. Saint Petersburg: Sabashnikov. The first edition appeared in 1909. The second edition was reprinted by the State Publishing House, Moscow, in 1959.

Copeland, T. E., and J. F. Weston. 1988. *Financial theory and corporate policy*. Third ed. Reading, MA: Addison-Wesley.

Cournot, A.-A. 1843. *Exposition de la théorie des chances et des probabilités*. Paris: Hachette. Reprinted in 1984 as Volume I (B. Bru, editor) of (Cournot 1984).

Cournot, A.-A. 1973–1984. *Œuvres complètes*. Paris: Vrin. 10 volumes, with an eleventh to appear.

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

de Finetti, B. 1951. Recent suggestions for the reconciliation of theories of probability. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, ed. J. Neyman, 217–225. Berkeley and Los Angeles: University of California Press.

Fisher, R. A. 1956. *Statistical methods and scientific inference*. Edinburgh: Oliver and Boyd.

Hastie, T., R. Tibshirani, and J. Friedman. 2001. *The elements of statistical learning theory: Data mining, inference, and prediction*. New York: Springer.

Keynes, J. M. 1921. *A Treatise on Probability*. London: Macmillan.

Keynes, J. M. 1937. The general theory of employment. *Quarterly Journal of Economics* 51:209–233.

Kolmogorov, A. N. 1983. Combinatorial foundations of information theory and the calculus of probabilities. *Russian Mathematical Surveys* 38 (4): 29–40.

Lévy, P. 1925. *Calcul des probabilités*. Paris: Gauthier-Villars.

Martin, T. 1996. *Probabilités et critique philosophique selon Cournot*. Paris: Vrin.

Martin, T. 1998. *Bibliographie cournotienne*. Besançon: Annales littéraires de l’Université Franche-Comté.

Martin-Löf, P. 1970. *Notes on constructive mathematics*. Stockholm: Almqvist & Wiksell.

Neyman, J. 1960. Indeterminism in science and new demands on statisticians. *Journal of the American Statistical Association* 55:625–639.

Richter, H. 1954. Zur Begründung der Wahrscheinlichkeitsrechnung. *Dialectica* 8:48–77.

Richter, H. 1956. *Wahrscheinlichkeitstheorie*. Berlin: Springer.

Shafer, G., and V. Vovk. 2001. *Probability and finance: It’s only a game*. New York: Wiley.

Shafer, G., and V. Vovk. 2002. On a review by V. N. Tutubalin. <www.probabilityandfinance.com/reviews>.

Sheynin, O. 1996. *Aleksandr A. Chuprov: Life, work, correspondence. The making of mathematical statistics*. Göttingen: Vandenhoeck & Ruprecht.

Vapnik, V. 1996. *The nature of statistical learning theory*. New York: Springer.

Ville, J.-A. 1939. *Étude critique de la notion de collectif*. Paris: Gauthier-Villars. This differs from Ville's dissertation, which was defended in March 1939, only in that a 17-page introductory chapter replaces the dissertation's one-page introduction.

von Bortkiewicz, L. 1901. Anwendungen der Wahrscheinlichkeitsrechnung auf Statistik. In *Encyklopädie der mathematischen Wissenschaften, Bd. I, Teil 2*, 821–851. Leipzig: Teubner.

von Hirsch, G. 1954. Sur un aspect paradoxal de la théorie des probabilités. *Dialetica* 8:125–144.

von Mises, R. 1919. Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift* 5:52–99.

von Mises, R. 1928. *Wahrscheinlichkeitsrechnung, Statistik und Wahrheit*. Vienna: Springer. Second edition 1936, third 1951. A posthumous fourth edition, edited by his widow Hilda Geiringer, appeared in 1972. English editions, under the title *Probability, Statistics and Truth*, appeared in 1939 and 1957.

von Mises, R. 1931. *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*. Leipzig and Vienna: Deuticke.

Vovk, V. 2005. Competitive on-line learning with a convex loss function. Working Paper No. 14, <www.probabilityandfinance.com>.

Vovk, V., A. Gammerman, and G. Shafer. 2005. *Algorithmic learning in a random world*. New York: Springer.

Vovk, V., I. Nouretdinov, A. Takemura, and G. Shafer. 2005. Defensive forecasting for linear protocols. In *AISTATS 2005: Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics, Barbados, January 6-8, 2005*, 365–372. <www.gatsby.ucl.ac.uk/aistats>. Extended version: Working Paper No. 10, <www.probabilityandfinance.com>.

Vovk, V., and G. Shafer. 2002. The game-theoretic capital asset pricing model. Working Paper No. 1, <www.probabilityandfinance.com>.

Vovk, V., and G. Shafer. 2003. A game-theoretic explanation of the $\sqrt{dt}$ effect. Working Paper No. 5, <www.probabilityandfinance.com>.

Vovk, V., and G. Shafer. 2005. Good randomized sequential probability forecasting is always possible. *Journal of the Royal Statistical Society, Series B* 67:747–764. Extended version: Working Paper No. 7, <www.probabilityandfinance.com>.

Vovk, V., A. Takemura, and G. Shafer. 2005. Defensive forecasting. In *Algorithmic Learning Theory: Proceedings of the 16th International Conference, ALT 2005, Singapore, October 8-11, 2005*, 459: Springer-Verlag. Lecture Notes in Computer Science, Volume 3734. Extended version: Working Paper No. 8, <www.probabilityandfinance.com>.

AUTHOR BIOGRAPHY

GLENN SHAFER is Board of Governors Professor at the Rutgers Business School – Newark and New Brunswick, where he serves as faculty director of the doctoral program. He is also Professor in the Computer Learning Centre at Royal Holloway University of London. He previously held positions in the Business School and the Mathematics Department at the University of Kansas and in the Statistics Department at Princeton University. He lives in Newark and serves on the board of the Newark Boys Chorus School <www.newark-boys-chorus-school.net>. He is best known for his contributions to the Dempster-Shafer theory, but his primary interest is in the history and philosophy of probability and statistics. His web page is <www.glennshafer.com>.