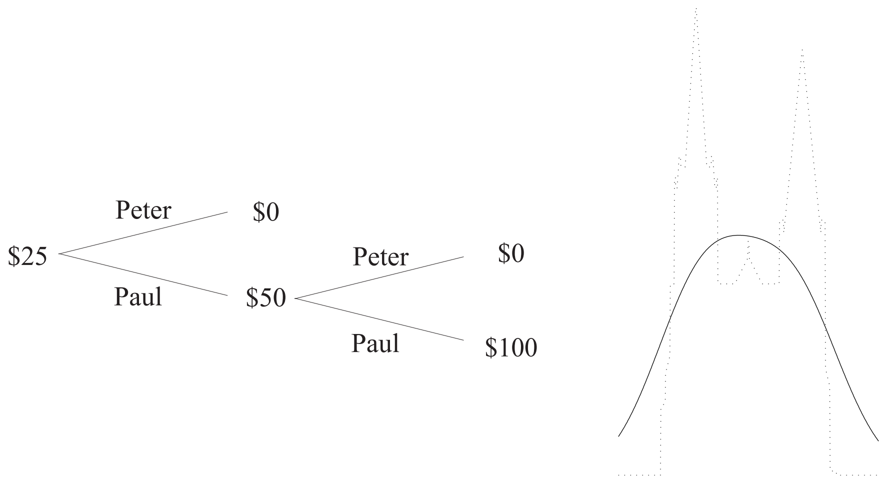
How to base probability theory on perfect-information games

Glenn Shafer, Vladimir Vovk, and Roman Chychyla



The Game-Theoretic Probability and Finance Project

Working Paper #32

First posted December 13, 2009. Last revised December 14, 2009.

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

Abstract

The standard way of making probability mathematical begins with measure theory. This article reviews an alternative that begins with game theory. We discuss how probabilities can be calculated game-theoretically, how probability theorems can be proven and interpreted game-theoretically, and how this approach differs from the measure-theoretic approach.

Contents

1 Introduction

We can make probability into a mathematical theory in two ways. One begins with measure theory, the other with the theory of perfect-information games. The measure-theoretic approach has long been standard. This article reviews the game-theoretic approach, which is less developed.

In §2, we recall that both measure theory and game theory were used to calculate probabilities long before probability was made into mathematics in the modern sense. In letters they exchanged in 1654, Pierre Fermat calculated probabilities by counting equally possible cases, while Blaise Pascal calculated the same probabilities by backward recursion in a game tree.

In §3, we review the elements of the game-theoretic framework as we formulated it in our 2001 book [21] and subsequent articles. This is the material we are most keen to communicate to computer scientists.

In §4, we compare the modern game-theoretic and measure-theoretic frameworks. As the reader will see, they can be thought of as dual descriptions of the same mathematical objects so long as one considers only the simplest and most classical examples. Some readers may prefer to skip over this section, because the comparison of two frameworks for the same body of mathematics is necessarily an intricate and second-order matter. It is also true that the intricacies of the measure-theoretic framework are largely designed to handle continuous time models, which are of little direct interest to computer scientists. The discussion of open systems in §4.5 should be of interest, however, to all users of probability models.

In §5, we summarize what this article has accomplished and mention some new ideas that have been developed from game-theoretic probability.

We do not give proofs. Most of the mathematical claims we make are proven in [21] or in papers at `http://probabilityandfinance.com`.

2 Two ways of calculating probabilities

Mathematical probability is often traced back to two French scholars, Pierre Fermat (1601–1665) and Blaise Pascal (1623–1662). In letters exchanged in 1654, they argued about how to do some simple probability calculations. They agreed on the answers, but not on how to derive them. Fermat's methodology can be regarded as an early form of measure-theoretic probability, Pascal's as an early form of game-theoretic probability.

Here we look at some examples of the type Pascal and Fermat discussed. In §2.1 we consider a simple case of the problem of points. In §2.2 we calculate the probability of getting two heads in succession before getting two tails in succession when flipping a biased coin.

2.1 The problem of points

Consider a game in which two players play many rounds, with a prize going to the first to win a certain number of rounds, or points. If they decide to break off the game while lacking different numbers of points to win the prize, how should they divide it?

Suppose, for example, that Peter and Paul are playing for 64 pistoles, Peter needs to win one more round, and Paul needs to win two. If Peter wins the next round, the game is over; Peter gets the 64 pistoles. If Paul wins the next round, then they play another round, and the winner of this second round gets the 64 pistoles. Figure 1 shows Paul's payoffs for the three possible outcomes: (1) Peter wins the first round, ending the game, (2) Paul wins the first round and Peter wins the second, and (3) Paul wins two rounds.

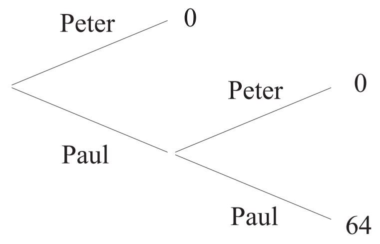


Figure 1: Paul wins either 0 or 64 pistoles.

If they stop now, Pascal asked Fermat, how should they divide the 64 pistoles? Fermat answered by imagining that Peter and Paul play two rounds regardless of how the first comes out. There are four possible cases:

1. Peter wins the first round, Peter the second. Peter gets the 64 pistoles.
2. Peter wins the first round, Paul wins second. Peter gets the 64 pistoles.
3. Paul wins the first round, Peter the second. Peter gets the 64 pistoles.
4. Paul wins the first round, Paul the second. Paul gets the 64 pistoles.

Paul gets the 64 pistoles in only one of the four cases, Fermat said, so he should get only 1/4 of the 64 pistoles, or 16 pistoles.

Pascal agreed with the answer, 16 pistoles, but not with the reasoning. There are not four cases, he insisted. There are only three, because if Peter wins the first round, Peter and Paul will not play a second round. A better way of getting the answer, Pascal argued, was to reason backwards in the tree, as shown in Figure 2. After Paul has just won the first round, he has the same chance as Peter at winning the 64 pistoles, and so his position is worth 32 pistoles. At the beginning, then, he has an equal shot at 0 or 32, and this is worth 16.

Pascal and Fermat did not use the word "probability". But they gave us methods for calculating probabilities. In this example, both methods give 1/4 as the probability for the event that Paul will win the 64 pistoles.

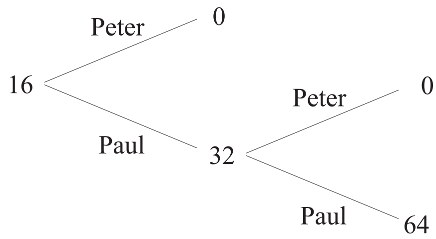


Figure 2: Pascal's backward recursion.

Fermat's method is to count the cases where an event A happens and the cases where it fails; the ratio of the number where it happens to the total is the event's probability. This has been called the classical definition of probability. In the 20th century, it was generalized to a measure-theoretic definition, in which an event is identified with a set and its probability with the measure of the set.

Pascal's method, in contrast, treats a probability as a price. Let A be the event that Paul wins both rounds. We see from Figure 2 that if Paul has 16 pistoles at the beginning, he can bet it in a way that he will have 64 pistoles if A happens, 0 if A fails. (He bets the 16 pistoles on winning the first round, losing it if he loses the round, but doubling it to 32 if he does win, in which case he bets the 32 on winning the second round.) Rescaling so that the prize is 1 rather than 64, we see that 1/4 is what he needs at the beginning in order to get a payoff equal to 1 if A happens and 0 if A fails. This suggests a general game-theoretic definition of probability for a game in which we are offered opportunities to gamble: the probability of an event is the cost of a payoff equal to 1 if the event happens and 0 if it fails.

2.2 Two heads before two tails

Let us apply Fermat's and Pascal's competing methods to a slightly more difficult problem. Suppose we repeatedly flip a coin, with the probability of heads being 1/3 each time (regardless of how previous flips come out). What is the probability we will get two successive heads before we get two successive tails?

Fermat's combinatorial method is to list the ways the event (two heads before two tails) can happen, calculate the probabilities for each, and add them up. The number of ways we can get two heads before two tails is countably infinite; here are the first few of them, with their probabilities:

$$\begin{array}{ll} \text{HH} & \left(\frac{1}{3}\right)^2 \\[1em] \text{THH} & \left(\frac{1}{3}\right)^2 \quad \frac{2}{3} \end{array}$$

HTHH	$\left(\frac{1}{3}\right)^3 \frac{2}{3}$
THTHH	$\left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^2$
HTHTHH	$\left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right)^2$
etc.	

Summing the infinite series, we find that the total probability for two heads before two tails is 5/21.

To get the same answer game-theoretically, we start with the game-theoretic interpretation of the probability 1/3 for a head on a single flip: it is the price for a ticket that pays 1 if the outcome is a head and 0 if it is a tail. More generally, as shown in Figure 3, $(1/3)x + (2/3)y$ is the price for x if a head, y if a tail.

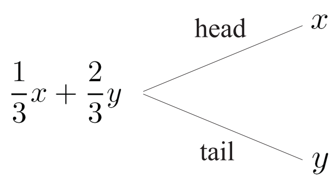


Figure 3: The game-theoretic meaning of probability 1/3 for a head.

Let A be the event that there will be two heads in succession before two tails in succession, and consider a ticket that pays 1 if A happens and 0 otherwise. The probability p for A is the price of this ticket at the outset. Suppose now that we have already started flipping the coin but have not yet obtained two heads or two tails in succession. We distinguish between two situations, shown in Figure 4:

- In Situation H, the last flip was a head. We write a for the value of the ticket on A in this situation.
- In Situation T, the last flip was a tail. We write b for the value of the ticket on A in this situation.

In Situation H, a head on the next flip would be the second head in succession, and the ticket pays 1, whereas a tail would put us in Situation T, where the ticket is worth b. Applying the rule of Figure 3 to this situation, we get

$$a = \frac{1}{3} + \frac{2}{3}b.$$

In Situation T, on the other hand, a head puts us in Situation H, and with a tail the ticket pays 0. This gives

$$b = \frac{1}{3}a.$$

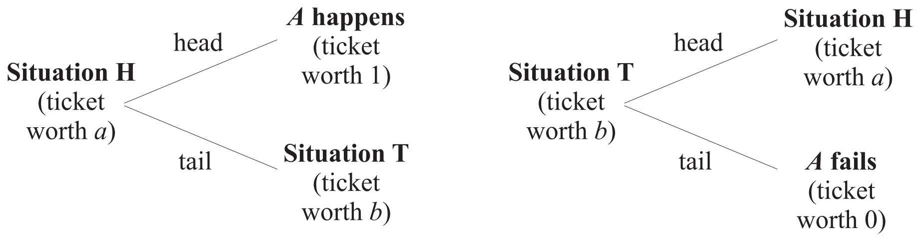


Figure 4: The value of a ticket that pays 1 if A happens and 0 if A fails varies according to the situation.

Solving these two equations in the two unknowns a and b, we obtain $a = 3/7$ and $b = 1/7$.

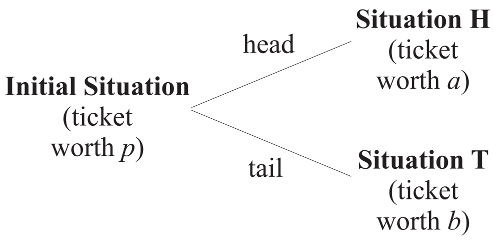


Figure 5: The initial value p is equal to $5/21$.

Figure 5 describes the initial situation, before we start flipping the coin. With probability $1/3$, the first flip will put us in a situation where the ticket is worth $3/7$; with probability $2/3$, it will put us in a situation where it is worth $1/7$. So the initial value is

$$p = \frac{1}{3} \cdot \frac{3}{7} + \frac{2}{3} \cdot \frac{1}{7} = \frac{5}{21},$$

in agreement with the combinatorial calculation.

2.3 Why did Pascal and Fermat get the same answers?

We will address a more general version of this question in §4.3, but on this first pass let us stay as close to our two examples as possible. Let us treat both examples as games where we flip a coin, either fair or biased, with a rule for stopping that determines a countable set Ω of sequences of heads and tails as possible outcomes. In our first example, $\Omega = \{\text{H},\text{TH},\text{TT}\}$, where H represents Peter's winning, and T represents Paul's winning. In our second example, $\Omega = \{\text{HH},\text{TT},\text{HTT},\text{THH},\text{HTHH},\text{THTT},\dots\}$.[1]

[1] To keep things simple, we assume that neither of the infinite sequences HTHTHT... or THTHTH... will occur.

Suppose p is the probability for heads on a single flip. The measure-theoretic approach assigns a probability to each element ω of Ω by multiplying together as many ps as there are Hs in ω and as many $(1-p)$s as there are Ts. For example, the probability of HTHH is $p^3(1-p)$. The probability for a subset A of Ω is then obtained by adding the probabilities for the ω in A.

The game-theoretic approach defines probability differently. Here the probability of A is the initial capital needed in order to obtain a certain payoff at the end of the game: 1 if the outcome ω is in A, 0 if not. To elaborate a bit, consider the *capital process* determined by a certain initial capital together with a strategy for gambling. Formally, such a capital process is a real-valued function L defined on the set $\mathbb{S}$ consisting of the sequences in Ω and all their initial segments, including the empty sequence $\Box$. For each $x \in \mathbb{S}$, $L(x)$ is the capital the gambler would have right after x happens if he starts with $L(\Box)$ and follows the strategy. In the game where we wait for two heads or two tails in succession, for example, $L(\text{HTHT})$ is the capital the gambler would have after HTHT, where the game is not yet over, and $L(\text{HH})$ is the capital he would have after HH, where the game is over. We can rewrite our definition of the probability of A as

$$P(A) := L(\Box), \text{ where } L \text{ is the unique capital process with } L(\omega) = I_A(\omega) \text{ for all } \omega \in \Omega. \quad (1)$$

Here I_A is the indicator function for A, the function on Ω equal to 1 on A and 0 on $\Omega \setminus A$.

We can use Equation (1) to explain why Pascal's method gives the same answers as Fermat's.

1. If you bet all your capital on getting a head on the next flip, then you multiply it by $1/p$ if you get a head and lose it if you get a tail. Similarly, if you bet all your capital on getting a tail on the next flip, then you multiply it by $1/(1-p)$ if you get a tail and lose it if you get a head. So Equation (1) gives the same probability to a single path in Ω as the measure-theoretic approach. For example, if $A = \{\text{HTHH}\}$, we can get the capital I_A at the end of the game by starting with capital $p^3(1-p)$, betting it all on H on the first flip, so that we have $p^2(1-p)$ if we do get H; then betting all this on T on the second flip, so that we have p^2 if we do get T, and so on, as in Figure 6.

2. We can also see from Equation (1) that the probability for a subset A of Ω is the sum of the probabilities for the individual sequences in A. This is because we can add capital processes. Consider, for example, a doubleton set $A = \{\omega_1, \omega_2\}$, and consider the capital processes L_1 and L_2 that appear in Equation (1) for $A = \{\omega_1\}$ and $A = \{\omega_2\}$, respectively. Starting with capital $P(\{\omega_1\})$ and playing one strategy produces L_1 with final capital $I_{\{\omega_1\}}(\omega)$ for all $\omega \in \Omega$, and starting with capital $P(\{\omega_2\})$ and playing another strategy produces L_2 with final capital $I_{\{\omega_2\}}(\omega)$ for all $\omega \in \Omega$. So starting with capital $P(\{\omega_1\}) + P(\{\omega_2\})$ and playing the sum

of the two strategies[2] produces the capital process $L_1 + L_2$, which has final capital $I_{\{\omega_1\}}(\omega) + I_{\{\omega_2\}}(\omega) = I_{\{\omega_1,\omega_2\}}(\omega)$ for all $\omega \in \Omega$.

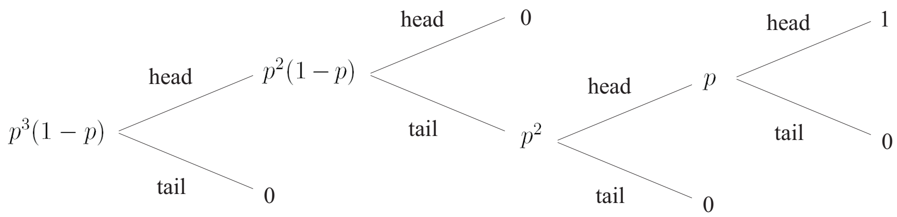


Figure 6: We get the payoff 1 if the sequence of outcomes is HTHH.

We can generalize Equation (1) by replacing I_A with a real-valued function ξ on Ω. This gives a formula for the initial price $\mathbb{E}(\xi)$ of the uncertain payoff $\xi(\omega)$:

$$\mathbb{E}(\xi) := L(\square), \text{ where } L \text{ is a capital process with } L(\omega) = \xi(\omega) \text{ for all } \omega \in \Omega. \quad (2)$$

If ξ is bounded, such a capital process exists and is unique. Christian Huygens explained the idea of Equation (2) very clearly in 1657 [13], shortly after he heard about the correspondence between Pascal and Fermat.

The fundamental idea of game-theoretic probability is to generalize Equation (2) as needed to more complicated situations, where there may be more or fewer gambles from which to construct capital processes. If we cannot count on finding a capital process whose final value will always exactly equal the uncertain payoff $\xi(\omega)$, let alone a unique one, we write instead

$$\overline{\mathbb{E}}(\xi) := \inf\{L(\square) | L \text{ is a capital process } \& \ L(\omega) \geq \xi(\omega) \text{ for all } \omega \in \Omega\}, \quad (3)$$

and we call $\overline{\mathbb{E}}(\xi)$ the *upper price* of ξ.[3] In games with infinite horizons, where play does not necessarily stop, we consider instead capital processes that equal or exceed ξ asymptotically, and in games where issues of computability or other considerations limit our ability to use all our current capital on each round, we allow some capital to be discarded on each round. But Pascal's and Huygens's basic idea remains.

3 Elements of game-theoretic probability

Although game-theoretic reasoning of the kind used by Pascal and Huygens never disappeared from probability theory, Fermat's idea of counting equally

[2]This means that we always make both the bets specified by the first strategy and the bets specified by the second strategy.

[3]As we explain in §3.1, there is a dual and possibly smaller price for ξ, called the *lower price*. The difference between the two is somewhat analogous to the bid-ask spread in a financial market.

likely cases became the standard starting point for the theory in the 19th century and then evolved, in the 20th century, into the measure-theoretic foundation for probability now associated with the names of Andrei Kolmogorov and Joseph Doob [14, 9, 22]. The game-theoretic approach re-emerged only in the 1930s, when Jean Ville used it to improve Richard von Mises's definition of probability as limiting frequency [17, 18, 27, 1]. Our formulation in 2001 [21] was inspired by Ville's work and by A. P. Dawid's work on prequential probability [7, 8] in the 1980s.

Whereas the measure-theoretic framework for probability is a single axiomatic system that has every instance as a special case, the game-theoretic approach begins by specifying a game in which one player has repeated opportunities to bet, and there is no single way of doing this that is convenient for all possible applications. So we begin our exposition with a game that is simple and concrete yet general enough to illustrate the power of the approach. In §3.1, we describe this game, a game of bounded prediction, and define its game-theoretic sample space, its variables and their upper and lower prices, and its events and their upper and lower probabilities. In §3.2, we explain the meaning of upper and lower probabilities. In §3.3, we extend the notions of upper and lower price and probability to situations after the beginning of the game and illustrate these ideas by stating the game-theoretic form of Lévy's zero-one law. Finally, in §3.4, we discuss how our definitions and results extend to other probability games.

3.1 A simple game of prediction

Here is a simple example, borrowed from Chapter 3 of [21], of a precisely specified game in which probability theorems can be proven.

The game has three players: Forecaster, Skeptic, and Reality. They play infinitely many rounds. Forecaster begins each round by announcing a number μ, and Reality ends the round by announcing a number y. After Forecaster announces μ and before Reality announces y, Skeptic is allowed to buy any number of tickets (even a fractional or negative number), each of which costs μ and pays back y. For simplicity, we require both y and μ to be in the interval $[0, 1]$. Each player hears the others' announcements as they are made (this is the assumption of perfect information). Finally, Skeptic is allowed to choose the capital $\mathcal{K}_0$ with which he begins.

We summarize these rules as follows.

PROTOCOL 1. BOUNDED PREDICTION

Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.
FOR $n = 1, 2, \ldots$:
  Forecaster announces $\mu_n \in [0, 1]$.
  Skeptic announces $M_n \in \mathbb{R}$.
  Reality announces $y_n \in [0, 1]$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(y_n - \mu_n)$.

There are no probabilities in this game, only limited opportunities to bet. But we can define prices and probabilities in Pascal's sense.

The following definitions and notation will help.

- A *path* is a sequence $\mu_1 y_2 \mu_2 y_2 \ldots$, where the μs and ys are all in $[0, 1]$.
- We write Ω for the set of all paths, and we call Ω the *sample space*.
- An *event* is a subset of Ω, and a *variable* is a real-valued function on Ω.
- We call the empty sequence $\Box$ the *initial situation*.
- We call a sequence of the form $\mu_1 y_1 \ldots \mu_{n-1} y_{n-1} \mu_n$ a *betting situation*.
- We call a sequence of the form $\mu_1 y_1 \ldots \mu_n y_n$ a *clearing situation*. We write $\mathcal{S}$ for the set of all clearing situations. We allow $n = 0$, so that $\Box \in \mathcal{S}$.
- A strategy Sstrat for Skeptic specifies his capital in the initial situation ($\mathcal{K}_0$ in $\Box$) and his move $\mathsf{Sstrat}(\mu_1 y_1 \ldots \mu_{n-1} y_{n-1} \mu_n)$ for every betting situation $\mu_1 y_1 \ldots \mu_{n-1} y_{n-1} \mu_n$.
- Given a strategy Sstrat for Skeptic, we define a function L_{Sstrat} on $\mathcal{S}$ by $L_{\mathsf{Sstrat}}(\Box) := \mathsf{Sstrat}(\Box)$ and

$$\begin{aligned} L_{\mathsf{Sstrat}}(\mu_1 y_1 \ldots \mu_n y_n) := L_{\mathsf{Sstrat}}&(\mu_1 y_1 \ldots \mu_{n-1} y_{n-1}) \\ &+ \mathsf{Sstrat}(\mu_1 y_1 \ldots \mu_{n-1} y_{n-1} \mu_n)(y_n - \mu_n). \end{aligned}$$

  We call L_{Sstrat} the *capital process* determined by Sstrat.[4] If Skeptic follows Sstrat, then $L_{\mathsf{Sstrat}}(\mu_1 y_1 \ldots \mu_n y_n)$ is his capital $\mathcal{K}_n$ after clearing in the situation $\mu_1 y_1 \ldots \mu_n y_n$.
- We write $\mathcal{L}$ for the set of all capital processes.
- Given $\omega \in \Omega$, say $\omega = \mu_1 y_2 \mu_2 y_2 \ldots$, we write ω^n for the clearing situation $\mu_1 y_1 \ldots \mu_n y_n$.

In the spirit of Equation (3) in §2.3, we say that the *upper price* of a bounded variable ξ is

$$\overline{\mathbb{E}}(\xi) := \inf\{L(\Box) \mid L \in \mathcal{L} \text{ and } \liminf_{n\to\infty} L(\omega^n) \geq \xi(\omega) \text{ for all } \omega \in \Omega\}. \qquad (4)$$

We get the same number $\overline{\mathbb{E}}(\xi)$ if we replace the lim inf in (4) by lim sup or lim. In other words,

$$\begin{aligned} \overline{\mathbb{E}}(\xi) &= \inf\{L(\Box) \mid L \in \mathcal{L} \text{ and } \limsup_{n\to\infty} L(\omega^n) \geq \xi(\omega) \text{ for all } \omega \in \Omega\} \\ &= \inf\{L(\Box) \mid L \in \mathcal{L} \text{ and } \lim_{n\to\infty} L(\omega^n) \geq \xi(\omega) \text{ for all } \omega \in \Omega\}. \end{aligned} \qquad (5)$$

(The inequality $\lim_{n\to\infty} L(\omega^n) \geq \xi(\omega)$ means that the limit exists and satisfies the inequality.) For a proof, which imitates the standard proof of Doob's convergence theorem, see [23]. The essential point is that if a particular strategy for

[4] In [27] and [21], such a capital process is called a *martingale*.

Skeptic produces capital that is sufficient in the sense of lim sup but oscillates on some paths rather than reaching a limit, Skeptic can exploit the successive upward oscillations, thus obtaining a new strategy whose capital tends to infinity on these paths.

If someone from outside the game pays Skeptic $\overline{\mathbb{E}}(\xi)$ at the beginning of the game, Skeptic can turn it into $\xi(\omega)$ or more at the end of the game. (Here we neglect, for simplicity, the fact that the infimum in (5) may not be attained.) So he can commit to giving back $\xi(\omega)$ at the end of the game without risking net loss. He cannot do this if he charges any less. So $\overline{\mathbb{E}}(\xi)$ is, in this sense, Skeptic's lowest safe selling price for ξ.

We set $\underline{\mathbb{E}}(\xi) := -\overline{\mathbb{E}}(-\xi)$ and call $\underline{\mathbb{E}}(\xi)$ the *lower price* of ξ. Because selling $-\xi$ is the same as buying ξ, $\underline{\mathbb{E}}(\xi)$ is the highest price at which Skeptic can buy ξ without risking loss.

The names "upper" and "lower" are justified by the fact that

$$\underline{\mathbb{E}}(\xi) \leq \overline{\mathbb{E}}(\xi). \tag{6}$$

To prove (6), consider a strategy Sstrat_1 that begins with $\underline{\mathbb{E}}(\xi)$ and returns at least ξ and a strategy Sstrat_2 that begins with $\overline{\mathbb{E}}(\xi)$ and returns at least $-\xi$. (We again neglect the fact that the infimum in (5) may not be attained.) Then $\mathsf{Sstrat}_1 + \mathsf{Sstrat}_2$ begins with $\overline{\mathbb{E}}(\xi) + \overline{\mathbb{E}}(-\xi)$ and returns at least 0. This implies that $\overline{\mathbb{E}}(\xi) + \overline{\mathbb{E}}(-\xi) \geq 0$, because there is evidently no strategy for Skeptic in Protocol 1 that turns a negative initial capital into a nonnegative final capital for sure. But $\overline{\mathbb{E}}(\xi) + \overline{\mathbb{E}}(-\xi) \geq 0$ is equivalent to $\underline{\mathbb{E}}(\xi) \leq \overline{\mathbb{E}}(\xi)$.

As we noted in §2.3, probability is a special case of price. We write $\overline{\mathbb{P}}(A)$ for $\overline{\mathbb{E}}(I_A)$, where I_A is the indicator function for A, and we call it A's *upper probability*. Similarly, we write $\underline{\mathbb{P}}(A)$ for $\underline{\mathbb{E}}(I_A)$, and we call it A's *lower probability*. We can easily show that

$$0 \leq \underline{\mathbb{P}}(A) \leq \overline{\mathbb{P}}(A) \leq 1 \tag{7}$$

for any event A. The inequality $\underline{\mathbb{P}}(A) \leq \overline{\mathbb{P}}(A)$ is a special case of (6). The inequalities $0 \leq \underline{\mathbb{P}}(A)$ and $\overline{\mathbb{P}}(A) \leq 1$ are special cases of the general rule that $\overline{\mathbb{E}}(\xi_1) \leq \overline{\mathbb{E}}(\xi_1)$ whenever $\xi_1 \leq \xi_2$, a rule that follows directly from (4). Notice also that

$$\underline{\mathbb{P}}(A) = 1 - \overline{\mathbb{P}}(A^c) \tag{8}$$

for any event A, where $A^c := \Omega \setminus A$. This equality is equivalent to $\overline{\mathbb{E}}(I_{A^c}) = 1 + \overline{\mathbb{E}}(-I_A)$, which follows from the fact that $I_{A^c} = 1 - I_A$ and from another rule that follows directly from (4): when we add a constant to a variable ξ, we add the same constant to its upper price.

If $\underline{\mathbb{E}}(\xi) = \overline{\mathbb{E}}(\xi)$, then we say that ξ is *priced*; we write $\mathbb{E}(\xi)$ for the common value of $\overline{\mathbb{E}}(\xi)$ and $\underline{\mathbb{E}}(\xi)$ and call it ξ's *price*. Similarly, if $\underline{\mathbb{P}}(A) = \overline{\mathbb{P}}(A)$, we write $\mathbb{P}(A)$ for their common value and call it A's *probability*.

3.2 The interpretation of upper and lower probabilities

According to the 19th century philosopher Augustin Cournot, as well as many later scholars [19], a probabilistic theory makes contact with the world only by predicting that events assigned very high probability will happen. Equivalently, those assigned very low probability will not happen.

In the case where we have only upper and lower probabilities rather than probabilities, we make these predictions:

1. If $\underline{\mathbb{P}}(A)$ is equal or close to one, A will happen.

2. If $\overline{\mathbb{P}}(A)$ is equal or close to zero, A will not happen.

It follows from (8) that Conditions 1 and 2 are equivalent. We see from (7) that these conditions are consistent with Cournot's principle. When $\underline{\mathbb{P}}(A)$ is one or approximately one, $\overline{\mathbb{P}}(A)$ is as well, and since we call their common value the probability of A, we may say that A has probability equal or close to one. Similarly, when $\overline{\mathbb{P}}(A)$ is zero or approximately zero, we may say that A has probability equal or close to zero.

In order to see more clearly the meaning of game-theoretic probability equal or close to zero, let us write $\mathfrak{L}^+$ for the subset of $\mathfrak{L}$ consisting of capital processes that are nonnegative—i.e., satisfy $L(\omega^n) \geq 0$ for all $\omega \in \Omega$ and $n \geq 0$. We can then write

$$\overline{\mathbb{P}}(A) := \inf\{L(\Box) \mid L \in \mathfrak{L}^+ \text{ and } \lim_{n\to\infty} L(\omega^n) \geq 1 \text{ for all } \omega \in A\}. \tag{9}$$

When $\overline{\mathbb{P}}(A)$ is very close to zero, (9) says that Skeptic has a strategy that will multiply the capital it risks by a very large factor $(1/L(\Box))$ if A happens. (The condition that $L(\omega^n)$ is never negative means that only the small initial capital $L(\Box)$ is being put at risk.) If Forecaster does a good job of pricing the outcomes chosen by Reality, Skeptic should not be able to multiply the capital he risks by a large factor. So A should not happen.

If an event has lower probability exactly equal to one, we say that the event happens *almost surely*. Here are two events that happen almost surely in Protocol 1:

- The subset A_1 of Ω consisting of all sequences $\mu_1 y_1 \mu_2 y_2 \ldots$ such that

  $$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (y_i - \mu_i) = 0. \tag{10}$$

  The assertion that A_1 happens almost surely is proven in Chapter 3 of [21]. It is a version of the *strong law of large numbers*: in the limit, the average of the outcomes will equal the average of the predictions.

- The subset A_2 of Ω consisting of all sequences $\mu_1 y_1 \mu_2 y_2 \ldots$ such that if $\lim_{n\to\infty} |J_{n,a,b}| = \infty$, where a and b are rational numbers and $J_{n,a,b}$ is the set of indices i such that $0 \leq i \leq n$ and $a \leq \mu_i \leq b$, then

  $$a \leq \liminf_{n\to\infty} \frac{\sum_{i\in J_{n,a,b}} y_i}{|J_{n,a,b}|} \quad \text{and} \quad \limsup_{n\to\infty} \frac{\sum_{i\in J_{n,a,b}} y_i}{|J_{n,a,b}|} \leq b.$$

The assertion that A_2 happens almost surely is an assertion of *calibration*: in the limit, the average of the outcomes for which the predictions are in a given interval will also be in that interval. See [36].

In [21], we also give examples of events in Protocol 1 that have lower probability close to one but not exactly equal to one. One such event, for example, is the event, for a large fixed value of N, that $\frac{1}{N}\sum_{i=1}^{N}(y_i - \mu_i)$ is close to zero. The assertion that this event will happen is a version of *Bernoulli's theorem*, sometimes called the *weak law of large numbers*.

The almost sure predictions we make (A will happen when $\underline{\mathbb{P}}(A) = 1$, and A will not happen when $\overline{\mathbb{P}}(A) = 0$) will be unaffected if we modify the game by restricting the information or choices available to Skeptic's opponents. If Skeptic has a winning strategy in a given game, then he will still have a winning strategy when his opponents are weaker. Here are three interesting ways to weaken Skeptic's opponents in Protocol 1:

- **Probability forecasting.** Require Reality to make each y_n equal to 0 or 1. Then μ_n can be interpreted as Forecaster's probability for $y_n = 1$, and the strong law of large number, (10), says that the frequency of 1s gets ever closer to the average probability.

- **Fixing the probabilities.** Require Forecaster to follow some strategy known in advance to the other players. He might be required, for example, to make all the μ_n equal to $1/2$. In this case, assuming that Reality is also required to set each y_n equal to 0 or 1, we have the familiar case where (10) says that the frequency of 1s will converge to $1/2$.

- **Requiring Reality's neutrality.** Prevent Reality from playing strategically. This can be done by hiding the other players' moves from Reality, or perhaps by requiring that Reality play randomly (whatever we take this to mean).

Weakening Skeptic's opponents in these ways makes Protocol 1 better resemble familiar conceptions of the game of heads and tails, but it does not invalidate any theorems we can prove in the protocol about upper probabilities being small ($\overline{\mathbb{P}}(A) = 0$, for example) or about lower probabilities being large ($\underline{\mathbb{P}}(A) = 1$, for example). These theorems assert that Skeptic has a strategy that achieves certain goals regardless of his opponents' moves. Additional assumptions about how his opponents move (stochastic models for their behavior, for example) might enable us to prove that Skeptic can accomplish even more, perhaps raising some lower prices or lowering some upper prices, but they will not invalidate any conclusions about what happens almost surely or with high probability.

It is also noteworthy that the almost sure predictions will not be affected if some or all of the players receive additional information in the course of the game. If Skeptic can achieve a certain goal regardless of how the other players move, then it makes no difference if they have additional information on which to base their moves. We will comment on this point further in §4.5.

The framework also applies to cases where Forecaster's moves μ_n and Reality's moves y_n are the result of the interaction of many agents and influences. One such case is that of a market for a company's stock, μ_n being the opening price of the stock on day n, and y_n its closing price. In this case, Skeptic plays the role of a day trader who decides how many shares to hold after seeing the opening price. Our theorems about what Skeptic can accomplish will hold regardless of the complexity of the process that determines μ_n and y_n. In this case, the prediction that A will not happen if $\overline{\mathbb{P}}(A)$ is very small can be called an *efficient market hypothesis*.

3.3 Price and probability in a situation

We have defined upper and lower prices and probabilities for the initial situation, but the definitions can easily be adapted to later situations. Given a situation s let us write $\Omega(s)$ for the set of paths for which s is a prefix. Then a variable ξ's upper price in the situation s is

$$\overline{\mathbb{E}}(\xi \mid s) := \inf\{L(s) \mid L \in \mathfrak{L} \text{ and } \lim_{n\to\infty} L(\omega^n) \geq \xi(\omega) \text{ for all } \omega \in \Omega(s)\}.$$

This definition can be applied both when s is a betting situation ($s = \mu_1 y_1 \dots \mu_n$ for some n) and when s is a clearing situation ($s = \mu_1 y_1 \dots \mu_n y_n$ for some n).

We may define $\underline{\mathbb{E}}(\xi \mid s)$, $\overline{\mathbb{P}}(A \mid s)$, and $\underline{\mathbb{P}}(A \mid s)$ in terms of $\overline{\mathbb{E}}(\xi \mid s)$, just as we have defined $\underline{\mathbb{E}}(\xi)$, $\overline{\mathbb{P}}(A)$, and $\underline{\mathbb{P}}(A)$ in terms of $\overline{\mathbb{E}}(\xi)$. We will not spell out the details. Notice that $\overline{\mathbb{E}}(\xi)$, $\underline{\mathbb{E}}(\xi)$, $\overline{\mathbb{P}}(A)$, and $\underline{\mathbb{P}}(A)$ are equal to $\overline{\mathbb{E}}(\xi \mid \Box)$, $\underline{\mathbb{E}}(\xi \mid \Box)$, $\overline{\mathbb{P}}(A \mid \Box)$, and $\underline{\mathbb{P}}(A \mid \Box)$, respectively.

In [23], we show that if the upper and lower prices for a variable ξ are equal, then this remains true almost surely in later situations: if $\overline{\mathbb{E}}(\xi) = \underline{\mathbb{E}}(\xi)$, then $\overline{\mathbb{E}}(\xi \mid \omega^n) = \underline{\mathbb{E}}(\xi \mid \omega^n)$ for all n almost surely.

The game-theoretic concepts of probability and price in a situation are parallel to the concepts of conditional probability and expected value in classical probability theory.[5] In order to illustrate the parallelism, we will state the game-theoretic form of Paul Lévy's zero-one law [16], which says that if an event A is determined by a sequence $X_1, X_2, \dots$ of variables, its conditional probability given the first n of these variables tends, as n tends to infinity, to one if A happens and to zero if A fails.[6] More generally, if a bounded variable ξ is

[5] In this paragraph, we assume that the reader has some familiarity with the concepts of conditional probability and expected value, even if they are not familiar with the measure-theoretic formalization of the concept that we will review briefly in §4.1.

[6] For those not familiar with Lévy's zero-one law, here is a simple example of its application to the problem of the gambler's ruin. Suppose a gambler plays many rounds of a game, losing or winning 1 pistole on each round. Suppose he wins each time with probability 2/3, regardless of the outcomes of preceding rounds, and suppose he stops playing only if and when he goes bankrupt (loses all his money). A well known calculation shows that when he has k pistoles, he will eventually lose it all with probability $(1/2)^k$. Suppose he starts with 1 pistole, and let $Y(n)$ be the number of pistoles he has after round n. Then his probability of going bankrupt is equal to 1/2 initially and to $(1/2)^{Y(n)}$ after the nth round. Levy's zero-one law, applied to the event A that he goes bankrupt, says that with probability one, either he goes bankrupt, or else $(1/2)^{Y(n)}$ tends to zero and hence $Y(n)$ tends to infinity. The probability that $Y(n)$ oscillates forever, neither hitting 0 nor tending to infinity, is zero.

determined by $X_1, X_2, \dots$, the conditional expected value of ξ given the first n of the X_i tends to ξ almost surely. In [23], we illustrate the game-theoretic concepts of price and probability in a situation by proving the game-theoretic version of this law. It says that

$$\liminf_{n\to\infty} \overline{\mathbb{E}}(\xi \,|\, \omega^n) \geq \xi(\omega) \tag{11}$$

almost surely. If ξ's initial upper and lower prices are equal, so that its upper and lower prices are also equal in later situations almost surely, we can talk simply of its price in situation s, $\mathbb{E}(\xi \,|\, s)$, and (11) implies that

$$\lim_{n\to\infty} \mathbb{E}(\xi \,|\, \omega^n) = \xi(\omega) \tag{12}$$

almost surely. This is Lévy's zero-one law in its game-theoretic form.

3.4 Other probability games

The game-theoretic results we have discussed apply well beyond the simple game of prediction described by Protocol 1. They hold for a wide class of perfect-information games in which Forecaster offers Skeptic gambles, Skeptic decides which gambles to make, and Reality decides the outcomes.

Let us assume, for simplicity, that Reality chooses her move from the same space, say $\mathcal{Y}$, on each round of the game. Then a gamble for Skeptic can be specified by giving a real-valued function f on $\mathcal{Y}$: if Skeptic chooses the gamble f and Reality chooses the outcome y, then Skeptic's gain on the round of play is $f(y)$. Forecaster's offer on each round will be a set of real-valued functions on $\mathcal{Y}$ from which Skeptic can choose.

Let us call a set $\mathcal{C}$ of real-valued functions on a set $\mathcal{Y}$ a *pricing cone* on $\mathcal{Y}$ if it satisfies the following conditions:

1. If $f_1 \in \mathcal{C}$, f_2 is a real-valued function on $\mathcal{Y}$, and $f_2 \leq f_1$, then $f_2 \in \mathcal{C}$.
2. If $f \in \mathcal{C}$ and $c \in [0, \infty)$, then $cf \in \mathcal{C}$.
3. If $f_1, f_2 \in \mathcal{C}$, then $f_1 + f_2 \in \mathcal{C}$.
4. If $f_1, f_2, \ldots \in \mathcal{C}$, $f_1(y) \leq f_2(y) \leq \cdots$ for all $y \in \mathcal{Y}$, and $\lim_{n\to\infty} f_n(y) = f(y)$ for all $y \in \mathcal{Y}$, where f is a real-valued function on $\mathcal{Y}$, then $f \in \mathcal{C}$.
5. If $f \in \mathcal{C}$, then there exists $y \in \mathcal{Y}$ such that $f(y) \leq 0$.

Let us write $\mathbf{C}_{\mathcal{Y}}$ for the set of all pricing cones on $\mathcal{Y}$.

If we require Skeptic to offer a pricing cone on each round of the game, then our protocol has the following form:

Protocol 2. General prediction
Parameter: Reality's move space $\mathcal{Y}$
  Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.
  FOR $n = 1, 2, \ldots$:

Forecaster announces $\mathcal{C}_n \in \mathbf{C}_{\mathcal{Y}}$.
Skeptic announces $f_n \in \mathcal{C}_n$.
Reality announces $y_n \in \mathcal{Y}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + f_n(y_n)$.

The probability games studied in [21] and in the subsequent working papers at `http://probabilityandfinance.com` are all essentially of this form, although sometimes Forecaster or Reality are further restricted in some way. As we explained in §3.2, our theorems state that Skeptic has a strategy that accomplishes some goal, and such theorems are not invalidated if we give his opponents less flexibility. We may also alter the rules for Skeptic, giving him more flexibility or restricting him in a way that does not prevent him from following the strategies that accomplish his goals.

In the case of Protocol 1, the outcome space $\mathcal{Y}$ is the interval $[0,1]$. Forecaster's move is a number $\mu \in [0,1]$, and Skeptic is allowed to choose any payoff function f that is a multiple of $y - \mu$. It will not invalidate our theorems to allow him also to choose any payoff function that always pays this much or less, so that his choice is from the set

$$\mathcal{C} = \{f : [0,1] \to \mathbb{R} \mid \text{there exists } \mu \in [0,1] \text{ and } M \in \mathbb{R} \\ \text{such that } f(y) \leq M(y-\mu) \text{ for all } y \in [0,1]\}.$$

This is a pricing cone; Conditions 1–5 are easy to check. So we have an instance of Protocol 2.

As we have just seen, Condition 1 in our definition of a pricing cone (the requirement that $f_2 \in \mathcal{C}$ when $f_1 \in \mathcal{C}$ and $f_2 \leq f_1$) is of minor importance; it sometimes simplifies our reasoning. Conditions 2 and 3 are more essential; they express the linearity of probabilistic pricing. Condition 4 plays the same role as countable additivity (sometimes called continuity) in measure-theoretic probability; it is needed for limiting arguments such as the ones used to prove the strong law of large numbers. Condition 5 is the condition of *coherence*; it rules out sure bets for Skeptic.

At first glance, it might appear that Protocol 2 might be further generalized by allowing Reality's move space to vary from round to round. This would not be a substantive generalization, however. If Reality is required to choose from a set $\mathcal{Y}_n$ on the nth round, then we can recover the form of Protocol 2 by setting $\mathcal{Y}$ equal to the union of the $\mathcal{Y}_n$; the fact that Reality is restricted on each round to some particular subset of the larger set $\mathcal{Y}$ does not, as we noted, invalidate theorems about what Skeptic can accomplish.

4 Contrasts with measure-theoretic probability

For the last two hundred years at least, the mainstream of probability theory has been measure-theoretic rather than game-theoretic. We need to distinguish, however, between classical probability theory, developed during the nineteenth and early twentieth centuries, and the more abstract measure-theoretic framework, using σ-algebras and filtrations, that was developed in the twentieth century, in large part by Kolmogorov [14] and Doob [9]. Classical probability theory, which starts with equally likely cases and combinatorial reasoning as Fermat did and extends this to continuous probability distributions using the differential and integral calculus, is measure-theoretic in a broad sense. The more abstract Kolmogorov-Doob framework qualifies as measure-theoretic in a more narrow mathematical sense: it uses the modern mathematical theory of measure.

Although there is a strong consensus in favor of the Kolmogorov-Doob framework among mathematicians who work in probability theory per se, many users of probability in computer science, engineering, statistics, and the sciences still work with classical probability tools and have little familiarity with the Kolmogorov-Doob framework. So we provide, in §4.1, a concise review of the Kolmogorov-Doob framework. Readers who want to learn more have many excellent treatises, such as [2, 24], from which to choose. For additional historical perspective on the contributions of Kolmogorov and Doob, see [22, 12].

In §4.2 and §4.3, we discuss some relationships between the game-theoretic and measure-theoretic pictures. As we will see, these relationships are best described not in terms of the abstract Kolmogorov-Doob framework but in terms of the concept of a forecasting system. This concept, introduced by A. P. Dawid in 1984, occupies a position intermediate between measure theory and game theory. A forecasting system can be thought of as a special kind of strategy for Forecaster, which always gives definite probabilities for Reality's next move. The Kolmogorov-Doob framework, in contrast, allows some indefiniteness, inasmuch as its probabilities in new situations can be changed arbitrarily on any set of paths of probability zero. The game-theoretic framework permits a different kind of indefiniteness; it allows Forecaster to make betting offers that determine only upper and lower probabilities for Reality's next move. In §4.2, we discuss how the game-theoretic picture reduces to a measure-theoretic picture when we impose a forecasting system on Forecaster. In §4.3, we discuss the duality between infima from game-theoretic capital processes and suprema from forecasting systems.

In §4.4, we discuss how continuous time can be handled in the game-theoretic framework. In §4.5, we point out how the open character of the game-theoretic framework allows a straightforward use of scientific theories that make predictions only about some aspects of an observable process.

4.1 The Kolmogorov-Doob framework

The basic object in Kolmogorov's picture [14, 22] is a *probability space*, which consists of three elements:

1. A set Ω, which we call the *sample space*.

2. A σ-*algebra* $\mathcal{F}$ on Ω – i.e., a set of subsets of Ω that contains Ω itself, contains the complement $\Omega \setminus A$ whenever it contains A, and contains the intersection and union of any countable set of its elements.

3. A *probability measure* P on $\mathcal{F}$ – i.e., a mapping from $\mathcal{F}$ to $[0,\infty)$ that satisfies
   (a) $P(\Omega) = 1$,
   (b) $P(A \cup B) = P(A) + P(B)$ whenever $A, B \in \mathcal{F}$ and $A \cap B = \emptyset$, and
   (c) $P(\cap_{i=1}^{\infty} A_i) = \lim_{i\to\infty} P(A_i)$ whenever $A_1, A_2, \dots \in \mathcal{F}$ and $A_1 \supseteq A_2 \supseteq \cdots$.

Condition (c) is equivalent, in the presence of the other conditions, to countable additivity: if $A_1, A_2, \dots$ are pairwise disjoint elements of $\mathcal{F}$, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Only subsets of Ω that are in $\mathcal{F}$ are called *events*. An event A for which $P(A) = 1$ is said to happen *almost surely* or for *almost all* ω.

A real-valued function ξ on the sample space Ω that is *measurable* (i.e., $\{\omega \in \Omega \mid \xi(\omega) \le a\} \in \mathcal{F}$ for every real number a) is called a *random variable*. If the Lebesgue integral of ξ with respect to P exists, it is called ξ's *expected value* and is denoted by $\mathbb{E}_P(\xi)$.

We saw examples of probability spaces in §2. In the problem of two heads before two tails, $\Omega = \{$HH,TT,HTT,THH,HTHH,THTT,…$\}$, and we can take $\mathcal{F}$ to be the set of all subsets of Ω. We defined the probability for an element ω of Ω by multiplying together as many ps as there are Hs in ω and as many $(1-p)$s as there are Ts, where p is the probability of getting a head on a single flip. We then defined the probability for a subset of Ω by adding the probabilities for the elements of the subset.

In general, as in this example, the axiomatic properties of the probability space $(\Omega, \mathcal{F}, P)$ make no reference to the game or time structure in the problem. Information about how the game unfolds in time is hidden in the identity of the elements of Ω and in the numbers assigned them as probabilities.

Doob [9] suggested bringing the time structure back to the axiomatic level by adding what is now called a *filtration* to the basic structure $(\Omega, \mathcal{F}, P)$. A filtration is a nested family of σ-algebras, one for each point in time. The σ-algebra $\mathcal{F}_t$ for time t consists of the events whose happening or failure is known at time t. We assume that $\mathcal{F}_t \subseteq \mathcal{F}$ for all t, and that $\mathcal{F}_t \subseteq \mathcal{F}_u$ when $t \le u$; what is known at time t is still known at a later time u. The time index t can be discrete (say $t = 0, 1, 2, \dots$) or continuous (say $t \in [0,\infty)$ or $t \in \mathbb{R}$).

Kolmogorov and Doob used the Radon-Nikodym theorem to represent the idea that probabilities and expected values change with time. This theorem implies that when ξ is a random variable in $(\Omega, \mathcal{F}, P)$, $\mathbb{E}_P(\xi)$ exists and is finite, and $\mathcal{G}$ is a σ-algebra contained in $\mathcal{F}$, there exists a random variable ζ that is measurable with respect to $\mathcal{G}$ and satisfies

$$\mathbb{E}_P(\xi I_A) = \mathbb{E}_P(\zeta I_A) \tag{13}$$

for all $A \in \mathcal{G}$. This random variable is unique up to a set of probability zero: if ζ_1 and ζ_2 are both measurable with respect to $\mathcal{G}$ and $\mathbb{E}_P(\xi I_A) = \mathbb{E}_P(\zeta_1 I_A) = \mathbb{E}_P(\zeta_2 I_A)$ for all $A \in \mathcal{G}$, then the event $\zeta_1 \ne \zeta_2$ has probability zero. We write

$\mathbb{E}_P(\xi \mid \mathcal{G})$ for any version of ζ, and we call it the *conditional expectation* of ξ given $\mathcal{G}$.

In the case where each element ω of Ω is a sequence, and we learn successively longer initial segments $\omega^1, \omega^2, \dots$ of ω, we may use the discrete filtration $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$, where $\mathcal{F}_n$ consists of all the events in $\mathcal{F}$ that we know to have happened or to have failed as soon as we know ω^n. In other words,

$$\mathcal{F}_n := \{A \in \mathcal{F} \mid \text{if } \omega_1 \in A \text{ and } \omega_2 \notin A, \text{ then } \omega_1^n \neq \omega_2^n\}.$$

It is also convenient to assume that $\mathcal{F}$ is the smallest σ-algebra containing all the $\mathcal{F}_n$. In this case, the measure-theoretic version of Lévy's zero-one law says that for any random variable ξ that has a finite expected value $\mathbb{E}_P(\xi)$,

$$\lim_{n\to\infty} \mathbb{E}_P(\xi \mid \mathcal{F}_n)(\omega) = \xi(\omega)$$

for almost all ω.[7] This is similar to the game-theoretic version of the law, Equation (12) in §3.3:

$$\lim_{n\to\infty} \mathbb{E}(\xi \mid \omega^n) = \xi(\omega)$$

almost surely. In the game-theoretic version, it is explicit in the notation that $\mathbb{E}(\xi \mid \omega^n)$ depends on ω only through what is known at the end of round n, namely ω^n. In the measure-theoretic version, we know that the value $\mathbb{E}_P(\xi \mid \mathcal{F}_n)(\omega)$ of the random variable $\mathbb{E}_P(\xi \mid \mathcal{F}_n)$ depends on ω only through ω^n because this random variable is measurable with respect to $\mathcal{F}_n$.

There are additional differences between the measure-theoretic and game-theoretic concepts. In the game-theoretic picture, a variable ξ may have only upper and lower prices, $\overline{\mathbb{E}}(\xi \mid s)$ and $\underline{\mathbb{E}}(\xi \mid s)$, but these are well defined even if the probability of arriving in the situation s was initially zero. Moreover, in the special case where upper and lower prices are equal, they behave as expected values are supposed to behave: $\mathbb{E}(\xi_1 + \xi_2 \mid s) = \mathbb{E}(\xi_1 \mid s) + \mathbb{E}(\xi_2 \mid s)$, etc. In contrast, the measure-theoretic quantity $\mathbb{E}_P(\xi \mid \mathcal{F}_n)(\omega)$ is undefined (i.e., can be chosen arbitrarily) if ω^n has initial probability zero, and the abstract definition (13) does not guarantee that the quantities $\mathbb{E}_P(\xi \mid \mathcal{F}_n)(\omega)$ will behave like expected values when ω is fixed and ξ is varied, or even that they can be chosen so that they do so.

The extent to which conditional expectations can fail to behave like expected values was a matter of some consternation when it was discovered in the 1940s and 1950s [22]. But in the end, the awkward aspects of the concept of conditional expectation have been tolerated, because the measure-theoretic framework is very general, applying to continuous as well as discrete time, and the usefulness of its theorems for sensible probability models is not harmed by the existence of less attractive models that also satisfy its axioms.

[7] If we do not assume that $\mathcal{F}$ is the smallest σ-algebra containing the $\mathcal{F}_n$, then we can say only that $\lim_{n\to\infty} \mathbb{E}_P(\xi \mid \mathcal{F}_n)(\omega) = \mathbb{E}_P(\xi \mid \mathcal{F}_\infty)(\omega)$ for almost all ω, where $\mathcal{F}_\infty$ is the smallest σ-algebra containing the $\mathcal{F}_n$. Lévy's own statement of his law, first published in 1937 [16], was simpler. He wrote that $\lim_{n\to\infty} E_n(\xi) = \xi$ almost surely, where $E_n(\xi)$ is ξ's expected value after $\omega_1 \dots \omega_n$ are known. Lévy had his own theory of conditional probability and expected value, slightly different from the one Kolmogorov published in 1933 [14].

4.2 Forecasting systems

In many applications of probability to logic and computer science, we consider an infinite sequence of 0s and 1s. If we write $\mu(y_1 \dots y_n)$ for the probability that the sequence will start with $y_1 \dots y_n$, then we should have:

- $0 \leq \mu(y_1 \dots y_n) \leq 1$, and
- $\mu(y_1 \dots y_n) = \mu(y_1 \dots y_n 0) + \mu(y_1 \dots y_n 1)$

for all finite sequences $y_1 \dots y_n$ of zeroes and ones. Let us call a function μ satisfying these two rules a *binary probability distribution.*

Standard expositions of the Kolmogorov-Doob framework show how to construct a probability space $(\Omega, \mathcal{F}, P)$ from a binary probability distribution μ:

- Ω is the set of all infinite sequences of zeroes and ones: $\Omega = \{0,1\}^\infty$.
- $\mathcal{F}$ is the smallest σ-algebra of subsets of Ω that includes, for every finite sequence $y_1 \dots y_n$ of zeroes and ones, the set consisting of all $\omega \in \Omega$ that begin with $y_1 \dots y_n$. (In this case, we say that $y_1 \dots y_n$ is a *prefix* of ω.)
- P is the unique probability measure on $\mathcal{F}$ that assigns, for every finite sequence $y_1 \dots y_n$ of zeroes and ones, the probability $\mu(y_1 \dots y_n)$ to the set consisting of all $\omega \in \Omega$ that have $y_1 \dots y_n$ as a prefix.

Given a bounded random variable ξ in $(\Omega, \mathcal{F}, P)$, let us write $\mathbb{E}_\mu(\xi)$ instead of $\mathbb{E}_P(\xi)$ for its expected value.

Let us call a binary probability distribution μ *positive* if $\mu(y_1 \dots y_n)$ is always strictly positive. In this case, conditional probabilities for y_n given the preceding values $y_1 \dots y_{n-1}$ are well defined. Let us write $\mu_{y_1 \dots y_{n-1}}(y_n)$ for these conditional probabilities:

$$\mu_{y_1 \dots y_{n-1}}(y_n) := \frac{\mu(y_1 \dots y_{n-1} y_n)}{\mu(y_1 \dots y_{n-1})} \tag{14}$$

for any sequence $y_1 \dots y_n$ of zeroes and ones.

Now consider the variation on Protocol 1 where Reality must choose each of her moves y_n from $\{0,1\}$ (rather than from the larger set $[0,1]$). In this case, Forecaster's move μ_n can be thought of as Forecaster's probability, after he has seen $y_1 \dots y_n$, that Reality will set y_n to equal 1. This thought reveals how Forecaster can use a positive binary probability distribution μ as a strategy in the game: he sets his move μ_n equal to $\mu_{y_1 \dots y_{n-1}}(1)$. If we assume that Forecaster plays this strategy, then we can replace him by the strategy in the protocol, reducing it to the following:

PROTOCOL 3. USING A POSITIVE BINARY PROBABILITY DISTRIBUTION AS A STRATEGY FOR BOUNDED PROBABILITY PREDICTION
Parameter: Positive binary probability distribution μ
Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.
FOR $n = 1, 2, \dots$:

Skeptic announces $M_n \in \mathbb{R}$.
Reality announces $y_n \in \{0,1\}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(y_n - \mu_{y_1 \ldots y_{n-1}}(1))$.

The sample space for this protocol is the space we just discussed: $\Omega = \{0,1\}^\infty$. The upper price in this protocol of a bounded variable, if it is measurable, is the same as its expected value in $(\Omega, \mathcal{F}, P)$ ([21], Proposition 8.5).

In the case of a binary probability distribution μ that is not positive, the denominator in Equation (14) will sometimes be zero, and so μ will not determine a strategy for Forecaster in our game. To avoid this difficulty, it is natural to replace the concept of a binary probability distribution with the concept of a *forecasting system*, which gives directly the required conditional probabilities $\mu_{y_1 \ldots y_{n-1}}(y_n)$. A binary probability distribution μ can be constructed from such a system:

$$\mu(y_1 \ldots y_n) := \mu_\Box(y_1)\mu_{y_1}(y_2) \cdots \mu_{y_1 \ldots y_{n-1}}(y_n).$$

If $\mu_{y_1 \ldots y_{n-1}}(y_n) = 0$ for some $y_1 \ldots y_{n-1} y_n$, then the forecasting system carries more information than the binary probability distribution.

The concept of a forecasting system generalizes beyond probability prediction (the variation on Protocol 1 where the y_n are all either zero or one) to Protocol 2. Fix a σ-algebra $\mathcal{G}$ on Reality's move space $\mathcal{Y}$, and write $\mathbf{P}_\mathcal{Y}$ for the set of all probability measures on $(\mathcal{Y}, \mathcal{G})$. Write $\mathcal{Y}^*$ for the set of all finite sequences of elements of $\mathcal{Y}$. In symbols: $\mathcal{Y}^* := \cup_{n=0}^\infty \mathcal{Y}^n$. Then a *forecasting system* is a mapping μ from $\mathcal{Y}^*$ to $\mathbf{P}_\mathcal{Y}$ that is measurable in an appropriate sense. Such a system μ determines a measure-theoretic object on the one hand and game-theoretic object on the other:

- It determines a probability measure P on the sample space $\mathcal{Y}^\infty$, and in each later situation a probability measure whose expected values form conditional expectations with respect to P and that situation.
- It determines a strategy for Forecaster in the protocol: in the situation $y_1 \ldots y_n$, Forecaster announces the pricing cone consisting of every real-valued function g on $\mathcal{Y}$ such that $f \le g$ for some random variable g on $(\mathcal{Y}, \mathcal{G})$ such that
$$\mathbb{E}_{\mu(y_1 \ldots y_n)}(g) \le 0.$$

The two objects agree on global pricing: the game-theoretic upper price of a bounded random variable on $\mathcal{Y}^\infty$ will be equal to its expected value with respect to P.

With respect to our game-theoretic protocols, however, the pricing cones determined by a forecasting system are rather special. In Protocol 1, for example, Forecaster is asked to give only a single number μ_n as a prediction of $y_n \in [0,1]$, not a probability distribution for y_n. The pricing cone thus offered to Skeptic (tickets that cost μ_n and pay y_n) is much smaller than the pricing cone defined by a probability distribution for y_n that has μ_n as its expected value. In Protocol 2, Forecaster has the option on each move of offering a pricing cone defined by a probability distribution for Reality's move, but he also has the option of offering a smaller pricing cone.

4.3 Duality

Using the concept of a forecasting system, we can see how game-theoretic and measure-theoretic probability are dual to each other. The quantity $\overline{\mathbb{E}}(\xi)$ represented in Equation (4) as an infimum over a class of capital processes is also a supremum over a class of forecasting systems.

As a first step to understanding this duality, consider how pricing cones on $\mathcal{Y}$ are related to probability measures on $\mathcal{Y}$. For simplicity, assume $\mathcal{Y}$ is finite, let $\mathcal{G}$ be the σ-algebra consisting of all subsets of $\mathcal{Y}$, and again write $\mathbf{P}_{\mathcal{Y}}$ for the set of all probability measures on $(\mathcal{Y}, \mathcal{G})$. Given a pricing cone $\mathcal{C}$ on $\mathcal{Y}$, set

$$\mathcal{P}_{\mathcal{C}} := \{P \in \mathbf{P}_{\mathcal{Y}} \mid \mathbb{E}_P(f) \leq 0 \text{ for all } f \in \mathcal{C}\}. \tag{15}$$

Given a real valued function ξ on $\mathcal{Y}$, we can show that

$$\mathcal{C} = \{f : \mathcal{Y} \to \mathbb{R} \mid \mathbb{E}_P(f) \leq 0 \text{ for all } P \in \mathcal{P}_{\mathcal{C}}\} \tag{16}$$

and that

$$\begin{aligned} &\sup\{\mathbb{E}_P(\xi) | P \in \mathcal{P}_{\mathcal{C}}\} \\ &\qquad = \inf\{\alpha \in \mathbb{R} \mid \exists f \in \mathcal{C} \text{ such that } \alpha + f(y) \geq \xi(y) \text{ for all } y \in \mathcal{Y}\} \\ &\qquad = \inf\{\alpha \in \mathbb{R} \mid \xi - \alpha \in \mathcal{C}\}. \end{aligned} \tag{17}$$

Equations (15) and (16) express one aspect of a duality between pricing cones and sets of probability measures. Equation (17) says that an upper price defined by taking an infimum over a pricing cone can also be obtained by taking a supremum over the dual set of probability measures.[8]

The concept of a filtration, because of the way it handles probabilities conditional on events of probability zero, does not lend itself to simple extension of (17) to a probability game with more than one round. Simple formulations in discrete time are possible, however, using the concept of a forecasting system.

For simplicity, assume again that $\mathcal{Y}$ is finite, and let us also assume that the game ends after N rounds. Write $\mathcal{Y}^*$ for the set of all finite sequences of elements of $\mathcal{Y}$ of length less than N. In symbols: $\mathcal{Y}^* := \cup_{n=0}^{N-1}\mathcal{Y}^n$. A *forecasting system with horizon N* is a mapping from $\mathcal{Y}^*$ to $\mathbf{P}_{\mathcal{Y}}$. Here, as in the binary case we just studied more closely, a forecasting system μ determines a probability measure P_μ on $\mathcal{Y}^N$ that has the probabilities given by μ as its conditional probabilities when these are well defined. Let us write $\mathbf{F}_{\mathcal{Y},N}$ for the set of all forecasting systems with horizon N.

We modify Protocol 2 by stopping play after round N and fixing a strategy for Forecaster, say Fstrat, that ignores the moves by Skeptic and chooses $\mathcal{C}_n$ based only on Reality's previous moves $y_1 \dots y_{n-1}$; this means that Fstrat is a mapping from $\mathcal{Y}^*$ to $\mathbf{C}_{\mathcal{Y}}$. Since Forecaster's strategy is fixed, we may remove him from the protocol, writing it in this form:

[8] Because of the finiteness of $\mathcal{Y}$ and Condition 4 in our definition of a pricing cone, the infimum and the supremum in (17) are attained.

Protocol 4. Finite horizon & fixed forecasts

Parameters: N, Reality's move space $\mathcal{Y}$, Forecaster's strategy Fstrat

Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.

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

Skeptic announces $f_n \in \mathsf{Fstrat}(y_1 \ldots y_{n-1})$.

Reality announces $y_n \in \mathcal{Y}$.

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

In this finite-horizon protocol, $\Omega = \mathcal{Y}^N$, and our definition of the upper price of a variable ξ, (4), simplifies to

$$\overline{\mathbb{E}}(\xi) := \inf\{L(\Box) \mid L \in \mathcal{L} \text{ and } L(\omega) \geq \xi(\omega) \text{ for all } \omega \in \Omega\}.$$

We can show that

$$\overline{\mathbb{E}}(\xi) = \sup\{\mathbb{E}_\mu(\xi) \mid \mu \in \mathbf{F}_{\mathcal{Y},N} \text{ and } \mu_{y_1\ldots y_n} \in \mathcal{P}_{\mathsf{Fstrat}(y_1\ldots y_n)} \text{ for all } (y_1 \ldots y_n) \in \mathcal{Y}^*\}.$$

This is the duality we announced at the outset: the infimum over initial stakes for different capital processes available to Skeptic that attain ξ equals the supremum over expected values of ξ for different forecasting systems that respect the offers made to Skeptic. See [6] for proofs and further comments on this duality.

4.4 Continuous time

It would be out of place to emphasize continuous-time processes in an introduction to game-theoretic probability for computer scientists. But these processes are very important in the measure-theoretic framework, and we would be selling the game-theoretic framework short if we did not take the time to point out that it can make a contribution in this domain.

How can we adapt the idea of a probability game to the case where Reality chooses a continuous-time path y_t instead of merely a sequence of moves $y_1 y_2 \ldots$? One answer, which uses non-standard analysis, was developed in [21]. In more recent work, which seems more promising, one supposes that Skeptic divides his capital among many strategies, all of which make bets at discrete points in time, but some of which operate at a much higher frequency than others. This approach has been dubbed *high-frequency limit-order trading* by Takeuchi [25].

Some of the continuous-time results require surprisingly little structure: we merely assume that Reality outputs a continuous path y_t that Skeptic observes as time passes, and that and at each time t Skeptic is allowed to buy an arbitrary number of tickets (negative, zero, or positive) that will pay him $S_{t'} - S_t$ at a future time t' of his choice. (Imagine that S_t is the price at time t in a security traded in an idealized financial market.) This assumption, combined with our definition of *almost sure* (an event happens almost surely if there is a strategy for Skeptic that multiplies the capital it risks by an infinite factor when the event fails) allows us to derive numerous qualitative properties that have been proven for Brownian motion and other martingales in the measure-theoretic framework. For example, we can show that S_t almost surely has no point of increase [33].[9] We can also show that S_t will almost surely have the jaggedness of Brownian motion in any interval of time in which it is not constant [34, 25, 32].[10] It appears that volatility is created by trading itself: if the price is not constant, there must be volatility. In general, a result analogous to that obtained by Dubins and Schwarz in 1965 for continuous martingales in measure-theoretic probability holds in this game-theoretic picture for S_t: any event that is invariant under transformations of the time scale has a game-theoretic probability, which is equal to its probability under Brownian motion [10, 31].

We can add additional structure to this game-theoretic picture by adding another player, Forecaster, who offers Skeptic additional betting opportunities. In this way, we can construct game-theoretic analogs to well known stochastic processes, including counting processes and Brownian motion [28]. The game-theoretic treatment of stochastic differential equations, sketched using non-standard analysis in [21] has yet to be undertaken in the high-frequency limit-order trading model.

The contribution here goes beyond showing that game-theoretic probability can obtain results already obtained by measure-theoretic probability. The game-theoretic approach clarifies the assumptions needed: the notion that Reality behaves stochastically is reduced to the assumption that Skeptic cannot multiply the capital he risks by a large or infinite factor. And because Skeptic tests Reality by betting at discrete points of time, the game-theoretic approach makes the continuous-time picture directly testable.

4.5 Open systems

An important aspect of the game-theoretic framework for probability is the open character of the protocols with which it works. Our protocols require only that the three players move in the order given and that Skeptic see the other players' moves. The players may receive other information, some of it private. Our theorems, such as the law of large numbers and Lévy's zero-one law, are not affected by such additional information.

In some applications, it is useful to make additional information explicit. We sometimes elaborate Protocol 2, for example, by having Reality give the other players information x_n before they move on the nth round. If we write $\mathcal{X}$ for the space from which this information is drawn, the protocol looks like this:

Protocol 5. Prediction with auxiliary information

[9] We say that t is a point of increase for S_t if there exists $\delta > 0$ such that $S_{t_1} < S_t < S_{t_2}$ for all $t_1 \in (t-\delta, t)$ and $t_2 \in (t-\delta, t)$. In 1961 Dvoretzky, Erdős, and Kakutani [11] proved that Brownian motion almost surely has no point of increase, and in 1965 Dubins and Schwarz [10] noticed that this is true for any continuous martingale. The game-theoretic argument in [33] imitates the measure-theoretic argument given in 1990 by Burdzy [4].

[10] This is made precise in different ways in the different references cited. In [32], a measure-theoretic construction by Bruneau [3] is adapted to show that the p-variation index of S_t is equal to 2 almost surely if S_t is not constant.

Parameters: Reality's information space $\mathcal{X}$, Reality's move space $\mathcal{Y}$

Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.
FOR $n = 1, 2, \dots$:
Reality announces $x_n \in \mathcal{X}$.
Forecaster announces $\mathcal{C}_n \in \mathbf{C}_{\mathcal{Y}}$.
Skeptic announces $f_n \in \mathcal{C}_n$.
Reality announces $y_n \in \mathcal{Y}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + f_n(y_n)$.

Putting the protocol in this form allows us to discuss strategies for Forecaster and Skeptic that use the x_n, but it does not invalidate the theorems for Protocol 2 that we have discussed. These theorems say that Skeptic can achieve certain goals using only the information about past y_n, regardless of how his opponents move and regardless of their additional information.

In many scientific and engineering applications of probability and statistical theory, only certain aspects $y_1 y_2 \dots$ of a process are given probabilities, while other aspects $x_1 x_2 \dots$, although they may affect the probabilities for the y, are not themselves given probabilities. Examples include:

- Quantum mechanics, where measurements y_n have probabilities only after we decide on the circumstances x_n under which we make measurements. See section 8.4 of [21].
- Genetics, where probabilities for the allele y_n of the next child are specified only after the next parents to have a child, x_n, are specified.
- Decision analysis, where in general outcomes y_n have probabilities only after decisions x_n have been made.
- Regression analysis, where each new outcome y_n is modeled only conditionally on a vector x_n of predictor variables.

In these examples, we can say we are using measure theory. Our model, we can say, is a class of probability measures – all the probability measures for $x_1 y_1 x_2 y_2 \dots$ in which the conditional probabilities for y_n given $x_1 y_1 \dots x_{n-1} y_{n-1} x_n$ satisfy certain conditions, the conditional probabilities for x_n given $x_1 y_1 \dots x_{n-1} y_{n-1}$ not being restricted at all. This formulation is, however, pragmatically and philosophically awkward. Pragmatically awkward because many results of mathematical statistics are applied in this way to situations where they do not necessarily hold. Philosophically awkward because we may not really want to say that the x_n follow some completely unknown or unspecified probability model. What is the content of such a statement?

The game-theoretic approach deals with these examples more straightforwardly. We specify bets on each y_n based on what is known just before it is announced. Using Cournot's principle we can give these bets an objective interpretation: no opponent will multiply the capital they risk by a large factor. Or we can settle for a subjective interpretation, either by weakening Cournot's principle (*we believe* that no opponent will multiply the capital they risk by a large factor) or by asserting, in the spirit of de Finetti, that we are willing to make the bets. There is no need to imagine unspecified bets on the x_n.

5 Conclusion

In this article, we have traced game-theoretic probability back to Blaise Pascal, and we have explained, with simple examples, how it generalizes classical probability. In particular, we have stated game-theoretic versions of the strong law of large numbers, Lévy's zero-one law, and the law of calibration. We have also spelled out various relationships with the measure-theoretic framework for probability.

When a field of mathematics is formalized in different ways, the different frameworks usually treat topics at the core of the field similarly but extend in different directions on the edges. This is the case with the game-theoretic and measure-theoretic frameworks for probability. They both account for the central results of classical probability theory, and the game-theoretic framework inherits very naturally the modern branches of measure-theoretic probability that rely on the concept of a martingale. But outside these central topics, the two frameworks offer more unique perspectives. Some topics, such as ergodic theory, are inherently measure-theoretic and seem to offer little room for fresh insights from the game-theoretic viewpoint. In other areas, the game-theoretic framework offers important new perspectives. We have already pointed to new perspectives on Brownian motion and other continuous-time processes. Other topics where the game-theoretic viewpoint is promising include statistical testing, prediction, finance, and the theory of evidence.

In the thesis he defended in 1939 [27], Jean Ville explained how we can test a probabilistic hypothesis game-theoretically. The classical procedure is to reject the hypothesis if a specified event to which it assigns very small probability happens. Ville pointed out that we can equivalently specify a strategy for gambling at prices given by the hypothesis and reject the hypothesis if this strategy multiplies the capital it risks by a large factor. In other words, we reject the hypothesis if a nonnegative capital process – a nonnegative *martingale*, in the now familiar terminology that Ville introduced – becomes many times as large as its initial value. Ville also pointed out that we can average martingales (this corresponds to averaging the gambling strategies) to obtain a more or less universal martingale, one that becomes very large if observations diverge from the probabilities in any important way. In the 1960s, Per Martin-Löf and Klaus-Peter Schnorr rediscovered and developed the idea of a universal test or universal martingale. The game-theoretic framework allows us to make these ideas practical. As we showed in [21], we can construct martingales that test violations of classical laws. The notion of a universal test is only an ideal notion; Martin-Löf's universal test and Schnorr's universal martingale are not computable. But by combining gambling strategies that test classical laws implied by a statistical hypothesis, we can construct martingales that are more or less universal in a practical sense.

In 1976 [15], Leonid Levin realized that for any test, including any universal test, there is a forecasting system guaranteed to pass the test.[11] So there is an ideal forecasting system, one that passes a universal test and hence passes every test. Like the universal test that defines it, Levin's ideal forecasting system is not computable. But in game-theoretic probability, we can implement practical versions of Levin's idea. For a wide class of prediction protocols, every computable game-theoretic law of probability defines a computable forecasting system that produces forecasts that conform to the law. By choosing suitable laws of probability, we can ensure that our forecasts agree with reality in all the ways we specify. We call this method of defining forecasting strategies *defensive forecasting*. It works well in many settings. It extends to decision problems, because the decisions that are optimal under forecasts that satisfy appropriate laws of probability will have satisfactory empirical performance, and it compares well with established methods for prediction with expert advice [36, 29, 30, 5].

We noted some of game-theoretic probability's implications for the theory of finance in §4.4. Other work has shown that versions of some of the standard results in finance can be obtained from the game-theoretic framework alone, without the introduction of stochastic assumptions. In [35], an empirical version of CAPM, which relates the average returns from securities to their correlations with a market portfolio, is derived game-theoretically. In [26], asymmetries in the movement of stock prices up and down are tested game-theoretically. In [37], observed correlations in stock returns are subjected to purely game-theoretic tests, and it is concluded that apparent inefficiencies are due to transaction costs.

A central question in the theory of evidence is the meaning and appropriateness of the judgements involved in updating and the combination of evidence. What judgements are involved, for example, when we use Bayes's theorem, Walley's rule for updating upper and lower probabilities, or Dempster's rule for combining belief functions? A game-theoretic answer to these questions is formulated in [20].

Acknowledgments

We are grateful to Yuri Gurevich and Akimichi Takemura for encouragement.

References

[1] Laurent Bienvenu, Glenn Shafer, and Alexander Shen. On the history of martingales in the study of randomness. *Electronic Journal for History of Probability and Statistics*, 5(1), June 2009.

[11] Levin's terminology was different, of course. His picture was not game-theoretic; instead of a forecasting system, he considered something like a probability measure, which he called a semimeasure. He showed that there is a semimeasure with respect to which every sequence of outcomes looks random.

[2] Patrick Billingsley. *Probability and Measure*. Wiley, New York, third edition, 1995. Previous editions appeared in 1979 and 1986.

[3] Michel Bruneau. Sur la p-variation des surmartingales. *Séminaire de probabilités de Strasbourg*, 13:227–232, 1979. Available free of charge at `http://www.numdam.org`.

[4] Krzysztof Burdzy. On nonincrease of Brownian motion. *Annals of Probability*, 18:978–980, 1990.

[5] Alexey Chernov and Vladimir Vovk. Prediction with expert evaluators' advice. In Ricard Gavaldà, Gábor Lugosi, Thomas Zeugmann, and Sandra Zilles, editors, *Proceedings of the Twentieth International Conference on Algorithmic Learning Theory*, volume 5809 of *Lecture Notes in Artificial Intelligence*, pages 8–22, Berlin, 2009. Springer. Full version: Technical report `arXiv:0902.4127` [cs.LG], `arXiv.org` e-Print archive, February 2009.

[6] Roman Chychyla. On the duality between game-theoretic and measure-theoretic probability. `http://probabilityandfinance.com`, Frequently Asked Question #5, 2009.

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

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

[9] Joseph L. Doob. *Stochastic Processes*. Wiley, New York, 1953.

[10] Lester E. Dubins and Gideon Schwarz. On continuous martingales. *Proceedings of the National Academy of Sciences*, 53:913–916, 1965.

[11] Aryeh Dvoretzky, Paul Erdős, and Shizuo Kakutani. Nonincrease everywhere of the Brownian motion process. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume II (Contributions to Probability Theory), pages 103–116, Berkeley, CA, 1961. University of California Press.

[12] Ronald Getoor. J. L. Doob: Foundations of stochastic processes and probabilistic potential theory. *Annals of Probability*, 37:1647–1663, 2009.

[13] Christian Huygens. *De ratiociniis in ludo aleae*. 1657.

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

[15] Leonid A. Levin. Uniform tests of randomness. *Soviet Mathematics Doklady*, 17:337–340, 1976.

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

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

[18] Richard von Mises. *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik.* F. Deuticke, Leipzig and Vienna, 1931.

[19] Glenn Shafer. From Cournot's principle to market efficiency. `http://probabilityandfinance.com`, Working Paper 15, March 2006.

[20] Glenn Shafer. A betting interpretation of Dempster-Shafer degrees of belief. *International Journal of Approximate Reasoning*, 2009.

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

[22] Glenn Shafer and Vladimir Vovk. The sources of Kolmogorov's *Grundbegriffe*. *Statistical Science*, 21:70–98, 2006. A longer version is at `http://probabilityandfinance.com` as Working Paper 4.

[23] Glenn Shafer, Vladimir Vovk, and Akimichi Takemura. Lévy's zero-one law in game-theoretic probability. `http://probabilityandfinance.com`, Working Paper 29, May 2009. Also arXiv:0905.0254v1.

[24] Albert N. Shiryaev. *Probability.* Springer, New York, second edition, 1996. First English edition 1984.

[25] Kei Takeuchi, Masayuki Kumon, and Akimichi Takemura. A new formulation of asset trading games in continuous time with essential forcing of variation exponent. Technical Report `arXiv:0708.0275` [math.PR], `arXiv.org` e-Print archive, August 2007. To appear in *Bernoulli.*

[26] Kei Takeuchi, Akimichi Takemura, and Masayuki Kumon. New procedures for testing whether stock price processes are martingales. Technical Report `arXiv:0907.3273v1` [math.PR], `arXiv.org` e-Print archive, August 2009.

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

[28] Vladimir Vovk. Forecasting point and continuous processes: prequential analysis. *Test*, 2:189–217, 1993.

[29] Vladimir Vovk. Predictions as statements and decisions. Technical Report `arXiv:cs/0606093` [cs.LG], `arXiv.org` e-Print archive, June 2006. Abstract in *Proceedings of the Nineteenth Annual Conference on Learning Theory*, volume 4005 of *Lecture Notes in Artificial Intelligence*, page 4, Berlin 2006. Springer.

[30] Vladimir Vovk. Predictions as statements and decisions. In Gábor Lugosi and Hans Ulrich Simon, editors, *Proceedings of the Nineteenth Annual Conference on Learning Theory*, volume 4005 of *Lecture Notes in Artificial Intelligence*, page 4, Berlin, 2006. Springer. Full version: Technical Report `arXiv:cs/0606093` [cs.LG], `arXiv.org` e-Print archive, June 2006.

[31] Vladimir Vovk. Continuous-time trading and emergence of probability. Technical Report `arXiv:0801.0000` [math.PR], `arXiv.org` e-Print archive, January 2008.

[32] Vladimir Vovk. Continuous-time trading and the emergence of volatility. *Electronic Communications in Probability*, 13:319–324, 2008.

[33] Vladimir Vovk. Continuous-time trading and the emergence of randomness. *Stochastics*, 81:455–466, 2009.

[34] Vladimir Vovk and Glenn Shafer. A game-theoretic explanation of the $\sqrt{dt}$ effect. `http://probabilityandfinance.com`, Working Paper 5, January 2003.

[35] Vladimir Vovk and Glenn Shafer. The game-theoretic capital asset pricing model. *International Journal of Approximate Reasoning*, 49:175–197, 2008.

[36] Vladimir Vovk, Akimichi Takemura, and Glenn Shafer. Defensive forecasting. In Robert G. Cowell and Zoubin Ghahramani, editors, *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics, January 6–8, 2005, Savannah Hotel, Barbados*, pages 365–372. Society for Artificial Intelligence and Statistics, 2005. Available electronically at `http://www.gatsby.ucl.ac.uk/aistats/`.

[37] Wei Wu and Glenn Shafer. Testing lead-lag effects under game-theoretic efficient market hypotheses. `http://probabilityandfinance.com`, Working Paper 23, November 2007.