



A betting interpretation for probabilities and Dempster–Shafer degrees of belief

Glenn Shafer *

Rutgers Business School, Newark and New Brunswick, 1 Washington Park, Newark, NJ 07102, USA
Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK

ARTICLE INFO

Article history:

Received 27 April 2009

Accepted 22 May 2009

Available online 22 June 2009

ABSTRACT

There are at least two ways to interpret numerical degrees of belief in terms of betting:

1. You can offer to bet at the odds defined by the degrees of belief.
2. You can make the judgement that a strategy for taking advantage of such betting offers will not multiply the capital it risks by a large factor.

Both interpretations can be applied to ordinary additive probabilities and used to justify updating by conditioning. Only the second can be applied to Dempster–Shafer degrees of belief and used to justify Dempster's rule of combination.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

The meaning of numerical probability has long been a matter of contention. Siméon Denis Poisson (1781–1840) distinguished between objective and subjective probabilities [12]. One recent philosophical introduction to probability lists five competing interpretations: classical, frequency, propensity, logical, and subjective [8].

The classical and subjective interpretations both involve betting. In the classical interpretation, the probability of an event is the correct price for a payoff that will equal one monetary unit if the event happens and zero otherwise. In the subjective interpretation, it is the price an individual is willing to pay for this payoff.

This article explains another betting interpretation of probability. Here I call it the *Ville interpretation*, in recognition of Jean André Ville (1910–1989), who first formulated it in his book on collectives [22]. Probabilities are prices under the Ville interpretation, just as they are under the classical and subjective interpretations. But instead of asserting that these prices are correct in some unspecified sense (as in the classical interpretation) or that some individual will pay them (as in the subjective interpretation), we assert that no strategy for taking advantage of them will multiply the capital it risks by a large factor. The Ville interpretation derives from an older interpretation of probability, neglected in the English-language literature, which I call the *Cournot interpretation* after Antoine-Augustin Cournot (1801–1877). According to the Cournot interpretation, the meaning of a probabilistic theory lies in the predictions that it makes with high probability.

As I explain in this article, the Ville interpretation can be applied both to ordinary additive probabilities and to the non-additive degrees of belief of the Dempster–Shafer calculus of belief functions. It works for Dempster–Shafer degrees of belief in ways that the subjective interpretation does not.

* Address: Rutgers Business School, Newark and New Brunswick, 180 University Avenue, Newark, NJ 07102, USA.
E-mail address: gshafer@rutgers.edu
URL: <http://www.glennshafer.com>

2. The Ville interpretation

This section reviews how the Ville interpretation emerges from older ideas and how it extends probability theory beyond its classical domain to games where the probabilities given and prices offered fall short of defining a probability distribution for all events of interest. In Section 2.1, I review briefly the history of the Cournot interpretation of ordinary probabilities. In Section 2.2, I explain how the Ville interpretation is related to the Cournot interpretation. In Section 2.3, I illustrate the power of the Ville interpretation using the example of probability forecasting, and in Section 2.4, I explain its role more generally in game-theoretic probability.

2.1. Cournot

The standard procedure for testing a probabilistic theory involves picking out an event to which the theory gives very small probability: we reject the theory if the event happens. In fact, this seems to be the only way to test a probabilistic theory. Because Cournot was the first to state that mathematical probability makes contact with phenomena only by ruling out events given very small probability [3, p. 58], the prediction that

an event of very small probability will not happen (1)

has been called *Cournot's principle*. In the first half of the twentieth century, many European scholars, including Émile Borel, Paul Lévy, Maurice Fréchet, and Andrei Kolmogorov, contended that Cournot's principle is fundamental to the meaning and use of mathematical probability [20]. As Borel said, we evoke “the only law of chance” when we single out an event of very small probability and predict it will not happen. (Or when, equivalently, we single out an event of very high probability and predict that it will happen.) Let us call the thesis that such predictions constitute the meaning of probability the *Cournot interpretation of probability*.

Cournot, Fréchet, and Kolmogorov are often called frequentists. This is misleading. These authors did believe that the probability of an event will be approximated by the frequency with which it happens in independent trials, but they considered this “law of large numbers” a consequence of Cournot's principle together with Bernoulli's theorem, which gives very high probability to the approximation holding. The true frequentists, such as John Venn, saw no sense in Bernoulli's theorem; probability is frequency, they believed, and so it is silly to try to prove that frequency will approximate probability [21].

Of course, events of very small probability do happen. An experiment may have a very large number of possible outcomes, each of which has very small probability, and one of which must happen. So Cournot's principle makes sense only if we are talking about particular events of very small probability that are salient for some reason: perhaps because they are so simple, perhaps because they have high probability under a plausible alternative hypothesis, or perhaps simply because they were specified in advance. There may be a substantial number of events that are salient in this way, but this is not a problem if we set our threshold for small probability low enough, because the disjunction of a number of events with very small probability will still have reasonably small probability.

In order to put the Cournot interpretation into practice, we must also decide how small a probability we can neglect. This evidently depends on the context. Borel distinguished between what was negligible at the human level, at the terrestrial level, and at the cosmic level [2].

In using the Cournot interpretation, we must also bear in mind its role in testing and giving meaning to a probabilistic theory as a whole. Strictly speaking, it gives direct meaning only to probabilities that are very small (the event will not happen) or very large (the event will happen). It gives no meaning to a probability of 40%, say. But when a probabilistic theory says that many successive events are independent and all have probability 40%, it gives probabilities close to one for many aspects of this sequence of events. Probabilistic theories in which probabilities evolve (stochastic processes) also give probabilities close to one to many statements concerning what happens over time, so they can also be tested and acquire meaning by Cournot's principle.

Although it was widely accepted in continental Europe in the middle of the twentieth century, the Cournot interpretation never gained a significant foothold in the English-language literature, and awareness of it receded as English became the language of science and mathematics after World War II. We find only isolated affirmations of it after about 1970. In the article on probability in the Soviet Mathematical Encyclopedia, for example, we find the assertion that only probabilities close to zero or one have empirical meaning [13]. For more on the history of the Cournot interpretation see [9–11,16,20].

2.2. From Cournot to Ville

When a probability distribution is used to set betting odds, there is a well known relationship between the happening of events of small probability and the success of betting strategies. The event that a given betting strategy multiplies the capital it risks by $1/\alpha$ or more has probability α or less. Conversely for every event of probability α or less there is a bet that multiplies the capital it risks by $1/\alpha$ or more if the event happens. So it is natural to consider, as an alternative to Cournot's principle, the principle that

a strategy will not multiply the capital it risks by a large factor. (2)

Let us call this *Ville's principle*. Let us call the thesis that predictions of the form (2) constitute the meaning of probability the *Ville interpretation of probability*.

Ville's principle is equivalent to Cournot's principle whenever a probability distribution is given for the events being considered and the two principles are made specific, with the specific event and small probability mentioned in Cournot's principle matching the specific strategy and large factor mentioned in Ville's principle. But when the two principles are considered more abstractly, without α and the particular event or strategy being specified, they differ in two important respects:

1. Ville's principle gives us more guidance than Cournot's principle. It tells us to specify a strategy for betting, not merely a single event of small probability. We found it necessary to elaborate Cournot's principle by saying that the event of very small probability should be specified in advance. The corresponding coda for Ville's principle is also needed, but it is less easily overlooked, because a betting strategy cannot be implemented unless it is specified in advance.
2. Ville's principle has a broader scope than Cournot's principle. Cournot's principle applies only when there is a probability distribution for the events under discussion. Ville's principle applies whenever prices for gambles are given, even if these prices fall short of defining probabilities for events.

To see some of the implications of Ville's principle giving us more guidance, consider how testing is usually implemented. A test of a probabilistic theory usually begins with a test statistic, say $T(y)$, where y is an outcome that is to be observed. If the theory specifies a probability distribution P for y , then we reject the theory at the significance level α when we observe a value y such that

$$T(y) \geq c,$$

where c is a number such that $P\{T(y) \geq c\} \leq \alpha$. Ville's principle tells us to implement this idea in a particular way: our test statistic is the capital $\mathcal{K}(y)$ achieved by a specified betting strategy that starts with some initial capital \mathcal{K}_0 and does not risk losing more than \mathcal{K}_0 . We reject the theory at the significance level α when we observe a value y such that

$$\mathcal{K}(y) \geq \mathcal{K}_0/\alpha.$$

Markov's inequality tells us that $P\{\mathcal{K}(y) \geq \mathcal{K}_0/\alpha\} \leq \alpha$.¹

When we adopt a betting strategy with which to test a probability distribution P , we are implicitly specifying an alternative hypothesis Q that we can plausibly adopt if we reject P . To see that this is so, let us suppose, for simplicity, that $\mathcal{K}_0 = 1$ (the strategy risks one unit of capital), and that there are only finitely or countably many possible values for y . In this case, we can define Q by

$$Q(y) := \mathcal{K}(y)P(y). \quad (3)$$

It is easy to see that Q is a probability distribution: (1) $Q(y) \geq 0$ because $P(y)$ is a probability and $\mathcal{K}(y)$ is the final capital for a betting strategy that does not risk its capital becoming negative, and (2) $\sum_y Q(y) = 1$ because it is the expected payoff under P of a gamble that costs one unit. Eq. (3) tells us that the final capital $\mathcal{K}(y)$ is the likelihood ratio $Q(y)/P(y)$, a measure of how much the observed outcome y favors Q over P .

2.3. Probability forecasting

As a first example of how Ville's principle and the Ville interpretation apply even when prices offered fall short of defining a probability distribution P for all events of interest, consider a game in which a forecaster announces probabilities successively, observing the outcome of each preceding event before giving the next probability.

2.3.1. Probability forecasting game

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

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

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

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

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

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

(4)

¹ In general, Markov's inequality says that a nonnegative random variable X satisfies

$$P(X \geq E(X)/\alpha) \leq \alpha.$$

Because the betting strategy uses the odds set by P , the expected value of the final capital $\mathcal{K}(y)$ is the initial capital \mathcal{K}_0 . Because the strategy risks only the initial capital, the final capital $\mathcal{K}(y)$ cannot be negative.

This is a perfect-information game; the three players move in sequence, and they all see each move as it is made. The game continues for N rounds.

The number p_n can be thought of as the price of a ticket that pays the amount y_n . Skeptic can buy any number s_n of the tickets. Since he pays p_n for each ticket and receives y_n in return, his net payoff is $s_n(y_n - p_n)$. The number s_n can be positive or negative. By choosing s_n positive, Skeptic buys tickets; by choosing s_n negative, he sells tickets.

Within the game, the p_n are simply prices. But we think of them as Forecaster's probabilities: p_n is Forecaster's probability that Reality will choose $y_n = 1$. On the other hand, Forecaster need not have a joint probability distribution P for Reality's moves y_1, \dots, y_N . He simply chooses p_n as he pleases at each step.

Skeptic tests Forecaster's p_n by trying to increase his capital using them as prices. If Skeptic succeeds – i.e., if he makes \mathcal{K}_N large without risking more than his initial capital \mathcal{K}_0 , then we conclude that Forecaster is not a good probability forecaster. Ville's principle says that if Forecaster is a good forecaster, then Skeptic will not achieve a large value for his final capital \mathcal{K}_N without risking more than \mathcal{K}_0 .

What does it mean for Skeptic not to risk more than \mathcal{K}_0 ? It means that his moves do not allow Reality to make his final capital \mathcal{K}_N negative. Since Reality can always keep Skeptic from making money (by choosing $y_n = 0$ if s_n is positive and $y_n = 1$ if s_n is negative), she can make \mathcal{K}_N negative as soon as Skeptic lets \mathcal{K}_n become negative for any n . So in order to deny Reality the option of making \mathcal{K}_N negative, Skeptic must choose each s_n so as to deny Reality the option of making \mathcal{K}_n negative. By (4), this means choosing s_n in the interval

$$-\mathcal{K}_{n-1}/p_n \leq s_n \leq \mathcal{K}_{n-1}/(1 - p_n). \quad (5)$$

For brevity, let us say that Skeptic plays *safely* if he always chooses s_n satisfying (5), and let us call a strategy for Skeptic *safe* if it always prescribes s_n satisfying (5).

We can get back to classical probability by assuming that Forecaster follows a strategy based on a joint probability distribution P for y_1, \dots, y_N and perhaps other events outside the game, the strategy being to set p_n equal to P 's conditional probability for $y_n = 1$ given what has been observed so far. But Ville's principle is powerful even in the absence of a specified strategy for Forecaster. It is all we need in order to derive various relations, such as the law of the large numbers, the law of the iterated logarithm, and the central limit theorem, that classical probability theory says will hold between the probabilities p_1, \dots, p_N and the outcomes y_1, \dots, y_N . It turns out, for example, that Skeptic can play safely in such a way that either the relative frequency of 1s among y_1, \dots, y_N , $\sum_{n=1}^N y_n/N$, approximates the average probability forecast, $\sum_{n=1}^N p_n/N$, or else \mathcal{K}_N becomes very large [19, p. 125]. Because it tells us that \mathcal{K}_N will not become very large, Ville's principle therefore implies that $\sum_{n=1}^N y_n/N$ will approximate $\sum_{n=1}^N p_n/N$. This is a version of the law of large numbers.

2.4. Game-theoretic probability

Probability forecasting is only one example where prices fall short of defining a probability distribution. In many other examples, the shortfall is substantially greater.

One class of such examples arises in finance theory, where the price for a security at the beginning of the day can be thought of as the price for a ticket that pays what the security is worth at the end of the day. Here the roles of Forecaster and Reality are both played by the market that sets the prices, and the role of Skeptic is played by a speculator. Over a period of N days, they play a perfect-information game much like our probability forecasting game:

2.4.1. Market game

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

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

Market announces opening price $p_n \in [0, \infty)$.

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

Market announces closing price $y_n \in [0, \infty)$.

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

Here $s_n > 0$ when Speculator goes long in the security, and $s_n < 0$ when he goes short.

A cornerstone of finance theory is the *efficient market hypothesis*, which states that a speculator cannot expect to make money using publicly available information. Efforts to formulate this hypothesis more precisely usually start with the questionable assumption that market prices are governed, in some sense, by a probability distribution. Ville's principle offers an alternative way of making the hypothesis precise: we can say that Speculator will not make \mathcal{K}_N large while playing safely. This version of the efficient market hypothesis can be tested directly, without making any probabilistic assumptions [27]. It also implies a number of stylized facts about financial markets, including the \sqrt{dt} effect [23] and the relation between the volatility and average of simple returns called the CAPM [25].

Shafer and Vovk [19] give other examples of games where prices fall short of defining a probability distribution. It turns out that many of the usual results of probability theory can be extended to such games, provided that we adopt Ville's principle. In general, we call the study of such games *game-theoretic probability*.

The results in [19] are concerned with strategies for Skeptic or Speculator in a probability game; they say that this player can multiply his capital by a large factor if some result in probability theory or finance theory does not hold. It is also fruitful, however, to consider how Forecaster or Market can play against such strategies for Skeptic or Speculator. It turns out that they can do this effectively, and this gives a new method of making predictions, called *defensive forecasting* [26,24].

3. The judgement of irrelevance in updating by conditioning

How should Forecaster's probabilities change when he learns new information?

An important school of thought, called *Bayesian* in recent decades, contends that when we learn A , we should update our probability for B from $P(B)$ to

$$\frac{P(A \& B)}{P(A)}. \quad (6)$$

The change is called *conditioning*. Bayesians acknowledge that it is appropriate only if we judge A to be the only relevant information we have learned [5, Section 11.2.2, 1, p. 45].²

In this section, I review arguments for the updating rule (6), with attention to how they account for the judgement of relevance and irrelevance. I consider the argument originally given by Abraham De Moivre, the variation given by Bruno de Finetti, and another variation that is based on Ville's principle. Only the argument from Ville's principle uses the judgement of relevance.

3.1. De Moivre's argument

Abraham De Moivre was the first to state the *rule of compound probability*. In the second edition of his *Doctrine of Chances*, published in 1738 [6], he stated the rule as follows:

...the Probability of the happening of two Events dependent, is the product of the Probability of the happening of one of them, by the Probability which the other will have of happening, when the first shall have been consider'd as having happen'd...

This rule can be written

$$P(A \& B) = P(A)P(B|A), \quad (7)$$

where $P(A \& B)$ is the probability of the happening of A and B , $P(A)$ is the probability of the happening of A , and $P(B|A)$ is the probability which B will have of happening, when A shall have been considered as having happened.

The 20th century abandoned De Moivre's way of talking about probabilities. Now we call $P(B|A)$ the *conditional probability* of B given A , and we say that it is defined by the equation

$$P(B|A) := \frac{P(A \& B)}{P(A)}, \quad (8)$$

provided that $P(A) \neq 0$. This makes (7) a trivial consequence of a definition. But for De Moivre (7) was more substantive. It was a consequence of how probability is related to price.

De Moivre gave an argument for the rule of compound probability on pp. 5–6 of his second edition. He used a language that is somewhat unfamiliar today; he talked about the values of gamblers' expectations. But it is true to his thinking to say that the probability of an event is the price (or the fair price, if you prefer) for a ticket that pays 1 if the event happens and 0 if it does not happen. (An expectation is the possession of a ticket with an uncertain payoff, and its value is the price you should pay for the ticket.) Using the language of tickets, payoffs, and price, we can express his argument as follows:

1. The price of a ticket that pays 1 if A happens is $P(A)$.
2. Assume one can buy or sell any number of such tickets, even fractional amounts. So $P(A)x$ is the price of a ticket that pays x if A happens, where x is any real number. (Buying a negative amount means selling.)
3. After A happens (or everyone learns that A has happened and nothing else), $P(B|A)$ is the price of a ticket that pays 1 if B happens.
4. So starting with $P(A)P(B|A)$, you can get 1 if $A \& B$ happens. You use the $P(A)P(B|A)$ to buy a ticket that pays $P(B|A)$ if A happens, and then, if A does happen, you use the $P(B|A)$ to buy a ticket that pays 1 if B also happens.
5. So $P(A)P(B|A)$ is the value of a ticket that pays 1 if $A \& B$ happens.

² The authors just cited, de Finetti and Bernardo and Smith, go on to say that irrelevance usually fails; when we learn A we usually learn other information that will also modify our judgement concerning B . Nevertheless, updating by (6) is widely taught and implemented.

De Moivre's argument is unconvincing to modern readers because we do not accept his starting point – his unexamined assumption that an expectation has a well defined numerical value. Our positivist heritage demands that such numbers be cashed out in some way that can be observed.

3.2. De Finetti's version of the argument

Bruno de Finetti (1906–1985) had a way of responding to the positivist challenge. For him, probability is specific to an individual. An individual's probability for an event A is the price the individual sets for a ticket that returns 1 if A happens – the price at which he is willing to trade in such tickets, buying or selling as the occasion arises.

As for the conditional probability $P(B|A)$, de Finetti proposed a betting interpretation that avoids references to a situation after A has happened or is known to have happened. For him, $P(B|A)$ is the price of a conditional ticket – the price of a ticket that pays 1 if B happens, with the understanding that the transaction is cancelled (the price is refunded and no payoff is made if B happens) if A does not happen.

With these interpretations, de Finetti was able to formulate a version of De Moivre's argument that leaves aside the notion of changing probabilities. We situate ourselves at the beginning of the game, as it were, and argue as follows:

1. $P(A)$ is the price at which I am willing to buy or sell tickets that pay 1 if A happens.
2. I am willing to buy or sell any number of such tickets, even fractional amounts. So $P(A)x$ is the price I will pay for a ticket that pays x if A happens, where x is any real number.
3. $P(B|A)$ is the price I am willing to pay for a ticket that pays 1 if B happens, with the understanding that this price is refunded if A does not happen.
4. It follows that I am willing to pay $P(A)P(B|A)$ to get back 1 if A and B both happen. You can prove this by selling me two tickets:
 - For $P(A)P(B|A)$, a ticket that pays $P(B|A)$ if A happens.
 - For $P(B|A)$, a ticket that pays 1 if B and A both happen, with the price being refunded if A does not happen.
 If A and B both happen, I end up with 1, less the $P(A)P(B|A)$ I paid for the first ticket; the payoff from the first ticket is cancelled by the cost of the second. If A does not happen, I lose only the $P(A)P(B|A)$, the second purchase having been cancelled. If A happens but B does not, I again lose only the $P(A)P(B|A)$, the cost of the second purchase being cancelled by the payoff on the first.
5. So $P(A)P(B|A)$ is the price I am willing to pay for 1 if $A \& B$ happens – i.e., my probability for $A \& B$.

As a coda, we may add de Finetti's argument for the price being unique. De Moivre had taken it for granted that the value of a thing is unique. De Finetti, using his assumption that we are willing to buy and sell any amount, argued that we must make the probability unique in order to prevent an opponent from extracting an indefinite amount of money from us.

De Finetti's version of the argument comes closer to modern mathematical rigor than De Moivre's, because it leaves aside the notion of something being "considered as having happened", for which De Moivre gave no set-theoretic exegesis. But some such notion must still be used in order to extend the argument to a justification for using conditional probabilities as one's new probabilities after something new is learned. We must explain why the price $P(B|A)$ for the conditional ticket on B given A should not change when A and nothing else is learned. There is a large literature on how convincingly this argument can be made; some think it requires that a protocol for new information be fixed and known in advance. See [15] and references therein.

3.3. Making the argument from Ville's principle

Ville's principle, like Cournot's, can usually be applied directly only to a run of events, in which a strategy has time to multiply the capital it risks substantially (or, in the case of Cournot's principle, we can identify an event of very small probability). So in order to apply Ville's principle to the problem of changing probabilities that are neither very small nor very large, we must imagine them being embedded in a longer sequence of similar probabilities for similar events. This is how probability judgments are often made: we judge that an event is like an event in some repetitive process for which we know probabilities [18].

In de Finetti's picture, we make a probability judgement $P(A) = p$ by saying that p is the price at which we are willing to buy or sell tickets that pay 1 if A happens. (I omit needed caveats: that we buy and sell only to people who have the same knowledge as ourselves, that this is only the price we might be inclined to set if we were inclined to gamble, etc.) In Ville's picture, we make a probability judgement $P(A) = p$ by saying that if we do offer such bets on A , and on a sequence of similar events in similar but independent circumstances, then an opponent *will not succeed in multiplying the capital they risk in betting against us by a large factor*. Let us abbreviate this to the statement that an opponent *will not beat the probability*.

In this terminology, our task is to show that the following claim holds:

Suppose we are in a situation where we judge that an opponent will not beat $P(A)$ and $P(A \& B)$. Suppose we then learn A and nothing more. Then we can include $P(A \& B)/P(A)$ as a new probability for B among the probabilities that we judge an opponent will not beat. (9)

In one respect, we are following De Moivre more faithfully than de Finetti did. De Finetti's mathematical argument is concerned only with prices in a single situation. Here we propose, like De Moivre, to give an argument that relates prices over two situations: an initial situation and a subsequent situation where our additional knowledge is A and nothing more. This is normal for the game-theoretic framework reviewed in Sections 2.3 and 2.4; there we apply Ville's principle to games with many rounds.

Here is the argument for (9) from Ville's principle:

1. An opponent will not beat the probabilities $P(A)$ and $P(A \& B)$. This means that a strategy for the opponent that buys and sells tickets on A and $A \& B$ at these prices, along with similar tickets on other events, will not multiply the capital risked by a large factor.
2. We need to show that this impossibility of multiplying the capital risked still holds for strategies that are also allowed to use $P(A \& B)/P(A)$ as a new probability for B after A and nothing more is known.
3. It suffices to show that if \mathcal{S} is a strategy against all three probabilities ($P(A)$ and $P(A \& B)$ in the initial situation and $P(A \& B)/P(A)$ later), then there exists a strategy \mathcal{S}' against the two probabilities ($P(A)$ and $P(A \& B)$ in the initial situation) alone that risks no more capital and has the same payoffs as \mathcal{S} .
4. Let M , which may be positive or negative or zero, be the amount of B tickets \mathcal{S} buys after learning A . To construct \mathcal{S}' from \mathcal{S} , we delete this purchase of B tickets and add

$$M \text{ tickets on } A \& B \quad \text{and} \quad -M \frac{P(A \& B)}{P(A)} \text{ tickets on } A \quad (10)$$

to \mathcal{S}' 's purchases of tickets on A and $A \& B$ in the initial situation.

- The tickets in (10) have zero net cost:

$$MP(A \& B) - M \frac{P(A \& B)}{P(A)} P(A) = 0.$$

So \mathcal{S}' uses the same capital in the initial situation as \mathcal{S} .

- The payoffs of the tickets in (10) are the same as the net payoffs of the M tickets deleted from \mathcal{S} :

0 if A does not happen;

$$-M \frac{P(A \& B)}{P(A)} \text{ if } A \text{ happens but not } B;$$

$$M \left(1 - \frac{P(A \& B)}{P(A)} \right) \text{ if } A \text{ and } B \text{ both happen.}$$

so \mathcal{S}' uses no more capital than \mathcal{S} after the initial situation and has the same payoffs in the end.

5. By hypothesis, \mathcal{S}' will not multiply the capital it risks by a large factor. So \mathcal{S} , which risks the same capital and has the same payoffs, does not either.

See [17] for an extension of this argument to Peter Walley's updating principle for upper and lower previsions.

3.4. The judgement of irrelevance

The argument from Ville's principle for using conditional probability as one's new probability uses the role and implications of knowledge in a way that de Finetti's argument does not.

- De Finetti argued for the conditional probability $P(B|A)$ being the price in the initial situation for a conditional purchase – a purchase of a B ticket on the condition that A happens. He then merely asserted, with no argument, that it should remain the price for this purchase after we learn that A happens and nothing more.
- The Ville argument, in contrast, is truly an argument for $P(B|A)$ being the price for a B ticket in the new situation where we have learned that A happened and nothing more.

The Ville argument is able to bring knowledge into the story because it looks what can be accomplished by different strategies. What a strategy can accomplish depends on what information is available.

It is important to understand how the caveat “nothing more” enters into the Ville argument. The argument depends on constructing a strategy \mathcal{S}' for the initial situation alone that is equivalent to a strategy \mathcal{S} that makes additional bets in the

later situation where A is known. If something more than A is known, and \mathcal{S} uses this additional information as well (its choice of the number M depends on it), then the construction is not possible.

We can of course relax the requirement that nothing more be known than A 's happening. The essential requirement is that nothing more be known that can help an opponent multiply his capital. In this case, we may say that the happening of A is our only *relevant information*. We may have learned many other things by the time or at the time when we learned A , but none of them can provide further help to a strategy for betting against the probabilities.

4. Judgements of irrelevance in the Dempster–Shafer calculus

The Dempster–Shafer theory of belief functions extends conditional probability to a calculus for combining probability judgements based on different bodies of evidence. Judgements of irrelevance enter into this calculus explicitly and pervasively. These judgements can be explained in terms of Ville's principle in the same way as the judgement of irrelevance in the case of updating by conditioning on A : they are judgements that once certain information is taken into account, other information is of no help to a strategy for betting against certain probabilities.

In this section, I list the Ville judgements of irrelevance required by various operations in the Dempster–Shafer calculus (Section 4.1), and I discuss how attention to these judgements in applications can strengthen the calculus's usefulness (Section 4.2).

4.1. Basic operations

The Dempster–Shafer calculus derives from a series of articles by Dempster, recently republished along with other classic articles on the calculus in [28]. The calculus was described in detail in [14] and reviewed in [7]. Without reviewing the examples and details readers can find in these references, I give here an overview of four related operations: the transfer of belief, conditioning, independent combination, and Dempster's rule of combination. In each case, I explain the judgement of irrelevance involved.

I omit two other important operations, natural extension and marginalization, because they do not require judgements of irrelevance.

4.1.1. Transfer of belief

Suppose \mathbf{X} is a variable, whose possible values form the set \mathcal{X} , and suppose P is a probability distribution on \mathcal{X} , expressing our probability judgements about the value of \mathbf{X} . Suppose ω is another variable, with the set of possible values Ω , for which we do not have a probability distribution. Suppose further that Γ is a multivalued mapping from \mathcal{X} to Ω (a mapping from \mathcal{X} to non-empty subsets of Ω). Then we can define a function Bel on subsets of Ω by setting

$$\text{Bel}(A) := P\{x | \Gamma(x) \subseteq A\}. \quad (11)$$

A function defined in this way is called a *belief function*. We call $\text{Bel}(A)$ its *degree of belief* in A .

We can give Bel 's degrees of belief a Ville interpretation under the following conditions:

1. The probability distribution P has a Ville interpretation: no betting strategy will beat the probabilities it gives for \mathbf{X} .
2. The multivalued mapping Γ has this meaning:

$$\text{If } \mathbf{X} = x, \text{ then } \omega \in \Gamma(x). \quad (12)$$

3. Learning the relationship (12) between \mathbf{X} and ω does not affect the impossibility of beating the probabilities for \mathbf{X} . (This is the irrelevance judgement.)

The Ville interpretation that follows from these conditions is one-sided: a strategy that buys for $\text{Bel}(A)$ tickets that pay 1 if $\omega \in A$ (and makes similar bets on the strength of similar evidence) will not multiply the capital its risks by a large factor.

4.1.2. Conditioning

Suppose we modify the preceding setup by allowing the subset $\Gamma(x)$ of Ω to be empty for some x . In this case, condition (12) tells us that the event $\{x | \Gamma(x) \neq \emptyset\}$ happened, and if we judge that we have learned nothing else that can help a strategy beat P 's probabilities, then we are entitled to condition P on this event. This results in replacing (11) by

$$\text{Bel}(A) = \frac{P\{x | \Gamma(x) \subseteq A \ \& \ \Gamma(x) \neq \emptyset\}}{P\{x | \Gamma(x) \neq \emptyset\}}.$$

The judgements of irrelevance that justify this equation can be summarized by saying that aside from the impossibility of the x for which $\Gamma(x) = \emptyset$, learning (12) does not provide any other information that can help a strategy beat the probabilities for \mathbf{X} .

4.1.3. Independent combination

Suppose P_1 and P_2 are probability distributions on \mathcal{X}_1 and \mathcal{X}_2 , respectively, expressing our probability judgements about the values of the variables \mathbf{X}_1 and \mathbf{X}_2 , respectively. What judgement is involved when we say further that the product probability measure $P_1 \times P_2$ on $\mathcal{X}_1 \times \mathcal{X}_2$ expresses our probability judgement about \mathbf{X}_1 and \mathbf{X}_2 jointly?

This question is not answered simply by saying that \mathbf{X}_1 and \mathbf{X}_2 are probabilistically independent, because probabilistic independence, in modern probability theory, is a property of a joint probability distribution for two variables, not a judgement outside the mathematics that justifies adopting the product distribution as a joint probability distribution for them.

De Finetti's betting interpretation of probability does give an answer to the question: we should adopt the product distribution if learning the value of one of the variables and nothing else will not change the prices we are willing to offer on the other variable.

The Ville interpretation gives an analogous answer: we should adopt the product distribution if we make the judgement that knowing the value of one of the variables and nothing more would not help a strategy beat the probabilities for the other variable.

Dempster–Shafer theory extends the idea of independent combination to belief functions, by considering two multi-valued mappings, say a mapping Γ_1 from \mathcal{X}_1 to non-empty subsets of Ω_1 and a mapping Γ_2 from \mathcal{X}_2 to non-empty subsets of Ω_2 . Suppose Γ_1 and Γ_2 have these meanings, where ω_1 and ω_2 are variables that take values in Ω_1 and Ω_2 , respectively:

$$\text{If } \mathbf{X}_1 = x, \text{ then } \omega_1 \in \Gamma_1(x). \quad (13)$$

$$\text{If } \mathbf{X}_2 = x, \text{ then } \omega_2 \in \Gamma_2(x). \quad (14)$$

Then we can form a belief function Bel for the pair (ω_1, ω_2) :

$$\text{Bel}(A) = (P_1 \times P_2)\{(x_1, x_2) | \Gamma_1(x_1) \times \Gamma_2(x_2) \subseteq A\}$$

for $A \subseteq \Omega_1 \times \Omega_2$. To justify this, we must make the Ville judgement justifying the formation of the product distribution $P_1 \times P_2$ and also the judgement that learning (13) and (14) does not help beat the probabilities given by P_1 or P_2 . This goes beyond the individual judgements that learning (13) does not help beat P_1 and that learning (14) does not help beat P_2 .

4.1.4. Dempster's rule of combination

Dempster's rule concerns the combination of two bodies of evidence bearing on the same variable ω . Given the ideas we have just reviewed, it is most easily stated by considering two multivalued mappings from the different probability spaces to the same space Ω , say Γ_1 from (\mathcal{X}_1, P_1) and Γ_2 from (\mathcal{X}_2, P_2) . They have the usual meaning:

$$\text{If } \mathbf{X}_1 = x, \text{ then } \omega \in \Gamma_1(x). \quad (15)$$

$$\text{If } \mathbf{X}_2 = x, \text{ then } \omega \in \Gamma_2(x). \quad (16)$$

Even if both $\Gamma_1(x_1)$ and $\Gamma_2(x_2)$ are always non-empty, their intersection may be empty. When we learn (15) and (16), we learn that the event $\{(x_1, x_2) | \emptyset \neq \Gamma_1(x_1) \cap \Gamma_2(x_2)\}$ has happened.

Conditioning on the intersection being non-empty, we obtain the belief function Bel on Ω given by

$$\text{Bel}(A) := \frac{(P_1 \times P_2)\{(x_1, x_2) | \emptyset \neq \Gamma_1(x_1) \cap \Gamma_2(x_2) \subseteq A\}}{(P_1 \times P_2)\{(x_1, x_2) | \emptyset \neq \Gamma_1(x_1) \cap \Gamma_2(x_2)\}}.$$

In this case, the required Ville judgements are those involved in forming the product measure, along with the judgement that learning (15) and (16) does not help beat the probabilities given by the product measure aside from providing the information that $\{(x_1, x_2) | \emptyset \neq \Gamma_1(x_1) \cap \Gamma_2(x_2)\}$ has happened.

4.2. Discussion

In [14], I stated that Dempster's rule of combination is appropriate when the bodies of evidence underlying individual belief functions are independent. The Ville judgements I have just detailed elaborate this notion of independence, in a way that should be useful in applications.

In our various writings on belief functions and in debates with critics, Dempster and I frequently took the view that the notions of independence and conditioning involved in Dempster's rule are the same as in ordinary probability theory. The analysis of this article vindicates this view in some degree, insofar as it has shown that the judgements of irrelevance required for Dempster's rule have the same general form as judgements of irrelevance that justify the formation of product measures in ordinary probability theory and updating by conditioning in Bayesian reasoning. The analysis has also revealed, however, the complexity that can be involved in judgements of this general form.

The critics often demanded, of course, explanations of independence and conditioning that were consistent with de Finetti's explanation of the meaning of these concepts in the Bayesian calculus. Here I have argued that de Finetti's explanations are not as convincing as sometimes thought even for Bayesian updating: they justify the pricing of conditional tickets but not the changes in price from one state of knowledge to another. In any case, they surely do not extend to the Dempster–Shafer case, where no embedding of the rules in a static picture seems to be possible. For the process of combining evidence, we need a more dynamic picture, which is provided by the Ville interpretation.

It is easy to construct examples in which the Ville irrelevance judgements required for Dempster's rule are unreasonable or clearly wrong. It is also easy enough to construct examples in which these judgements are reasonable; I gave some such examples in the 1980s (see for example [18]). Existing applications of the Dempster–Shafer calculus would be enriched, however, by a systematic examination of the reasonableness of the irrelevance judgements they require. A clearer understanding of these judgements might also help us construct Dempster–Shafer models for complex scientific problems where the irrelevance judgements need to justify ordinary probabilities and Bayesian reasoning seem unreasonably strong.

References

- [1] Jose Bernardo, Adrian F.M. Smith, *Bayesian Theory*, Wiley, New York, 1994.
- [2] Emile Borel, *Valeur pratique et philosophie des probabilités*, Gauthier-Villars, Paris, 1939.
- [3] Antoine-Augustin Cournot, *Exposition de la théorie des chances et des probabilités*, Hachette, Paris, 1843 (Reprinted in 1984 as vol. I of [4]).
- [4] Antoine-Augustin Cournot, *CŒuvres complètes*, Vrin, Paris, 1973.
- [5] Bruno de Finetti, *Teoria Delle Probabilità*, English translation: *Theory of Probability*, two volumes, 1974 and 1975, Wiley, London, 1970.
- [6] Abraham De Moivre, *The Doctrine of Chances: or, A Method of Calculating the Probabilities of Events in Play*, Pearson, London, 1718 (second ed. 1738 and third 1756).
- [7] A.P. Dempster, The Dempster–Shafer calculus for statisticians, *International Journal of Approximate Reasoning* 48 (2008) 365–377.
- [8] Maria Carla Galavotti, *Philosophical Introduction to Probability*, CSLI Publications, Stanford, CA, 2005.
- [9] Thierry Martin, *Probabilités et critique philosophique selon Cournot*, Vrin, Paris, 1996.
- [10] Thierry Martin, *Probabilité et certitude*, in: Thierry Martin (Ed.), *Probabilités subjectives et rationalité de l'action*, CNRS Editions, Paris, 2003, pp. 119–134.
- [11] Thierry Martin, Cournot, philosophe des probabilités, in: Thierry Martin (Ed.), *Actualité de Cournot*, Vrin, Paris, 2005, pp. 51–68.
- [12] Simeon-Denis Poisson, *Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédés des règles générales du calcul des probabilités*, Bachelier, Paris, 1837.
- [13] Yu.V. Prokhorov, B.A. Sevast'yanov, *Probability theory*, in: M. Hazewinkel (Ed.), *Encyclopaedia of Mathematics: An updated and Annotated Translation of the Soviet Mathematical Encyclopaedia*, vol. 7, Reidel, Boston, 1987/1994, pp. 307–313.
- [14] Glenn Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1976.
- [15] Glenn Shafer, Conditional probability, *International Statistical Review* 53 (1985) 261–277.
- [16] Glenn Shafer, From Cournot's principle to market efficiency, in: Jean-Phillipe Touffut (Ed.), *Augustin Cournot: Modelling Economics*, Edward Elgar, 2007, pp. 55–95. A pre-publication version is at <www.probabilityandfinance.com> as Working Paper 15.
- [17] Glenn Shafer, Peter R. Gillett, Richard B. Scherl, A new understanding of subjective probability and its generalization to lower and upper prevision, *International Journal of Approximate Reasoning* 33 (2003) 1–49.
- [18] Glenn Shafer, Amos Tversky, Languages and designs for probability judgment, *Cognitive Science*, 1985.
- [19] Glenn Shafer, Vladimir Vovk, *Probability and Finance: It's Only a Game*, Wiley, New York, 2001.
- [20] Glenn Shafer, Vladimir Vovk, The sources of Kolmogorov's Grundbegriffe, *Statistical Science* 21(1) (2006) 70–98. A longer version is at <www.probabilityandfinance.com> as Working Paper 4.
- [21] John Venn, *The Logic of Chance*, third ed., MacMillan, London and New York, 1888.
- [22] Jean-André Ville, *Étude critique de la notion de collectif*, Gauthier-Villars, Paris, 1939.
- [23] Vladimir Vovk, Glenn Shafer, A game-theoretic explanation of the \sqrt{dt} effect, Working Paper 5 <www.probabilityandfinance.com>, 2003.
- [24] Vladimir Vovk, Glenn Shafer, Good randomized sequential probability forecasting is always possible, *Journal of the Royal Statistical Society, Series B* 67 (2005) 747–764. A longer version is at <www.probabilityandfinance.com> as Working Paper 7.
- [25] Vladimir Vovk, Glenn Shafer, The game-theoretic capital asset pricing model, *International Journal of Approximate Reasoning* 49(1) (2008) 175–197. See also Working Paper 1 at <www.probabilityandfinance.com>.
- [26] Vladimir Vovk, Akimichi Takemura, Glenn Shafer, Defensive forecasting, in: Tenth International Workshop on Artificial Intelligence and Statistics, 2005. <www.gatsby.ucl.ac.uk/aistats/>. See also Working Paper 8 at <www.probabilityandfinance.com>.
- [27] Wei Wu, Glenn Shafer, Testing lead-lag effects under game-theoretic efficient market hypotheses, Working Paper 23 <www.probabilityandfinance.com>, November 2007.
- [28] Ronald Yager, Liping Liu (Eds.), *Classic Works of the Dempster–Shafer Theory of Belief Functions*, Springer, New York, 2008.