Game-theoretic foundations for imprecise probabilities

Vladimir Vovk

Department of Computer Science
Royal Holloway, University of London

ISIPTA 2019
Ghent, 4 July, 2019

Main points of this talk

- There are two natural ways to define probability: in terms of measure and in terms of gambling.
- The two definitions are dual to each other and both are important.
- At this time the standard picture of probability is heavily tilted towards measure.
- Restoring the balance would be healthy, in both standard and imprecise probability theories.

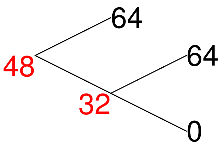
MTP vs GTP

- Measure-theoretic probability (MTP): probability is a basic notion, defined implicitly by the axioms. Axioms: probability is a measure with total mass 1 (and so precise by definition).
- You can introduce imprecision by introducing sets of measures (or, e.g., modifying the axioms).
- Game-theoretic probability (GTP): probability is a derivative notion (defined in terms of frequencies, martingales, etc.); it is intrinsically imprecise.
- Glenn Shafer talked about GTP on July 17 at ISIPTA 2007. My talk will complement his (but is self-contained).

Plan

The problem of points and Pascal's solution

- The two approaches to probability to back to Pascal and Fermat and earlier!
- Two gamblers play three throws; each puts 32 pistoles at stake. The first has two (points) and the other one, and they have to stop the game. How should they divide the 64 pistoles?
- Pascal to Fermat on 29 July, 1654:



This is a martingale.

Fermat's solution

Imagine (some of Pascal's and Fermat's contemporaries protested!) that the two gamblers make two more throws. The first's expected win is

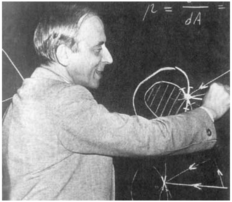
$$64 \times \frac{3}{4} + 0 \times \frac{1}{4} = 48.$$

The same answer but the methods are different: Pascal's can be regarded as a precursor of game-theoretic probability, and Fermat's as a precursor of measure-theoretic probability (in this case, the measure assigned to each pair of outcomes is $1/4$).

Game-theoretic vs measure-theoretic foundations of probability

Hilbert's sixth problem: to treat axiomatically, after the model of geometry, those parts of physics in which mathematics already played an outstanding role, especially probability and mechanics.

- Richard von Mises (1919, 1928, 1931): probability is a derivative notion, based on the notion of a gambling system.
- Andrei Kolmogorov (1931, 1933): probability is axiomatized directly as a special case of measure.



von Mises (1883–1957)

Kolmogorov (1903–1987)

Limitations of von Mises's concept of gambling

The claim that von Mises's gambling systems provide a satisfactory foundation for probability was refuted by Jean Ville (1939): they are not sufficient to derive the law of the iterated logarithm.

Von Mises's gambling systems choose a subsequence of trials on which to bet. Ville: more sophisticated gambling systems which also vary the amount of the bet and the outcome on which to bet. He called the capital processes of such strategies martingales.

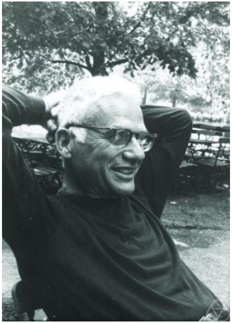
Jean Ville (1910–1988)

Further developments

Joseph Doob reviewed Ville's 1939 book about martingales for Mathematical Reviews. He then translated the notion of martingale into measure-theoretic probability and developed it greatly.

Now the measure-theoretic theory of martingales is the centrepiece of many areas of probability theory. A prominent role in developing the theory of stochastic processes based on martingales was played by Kiyosi Itô (his integration and calculus).

Martingales flourish in the mathematics (but not the foundations or philosophy) of probability.

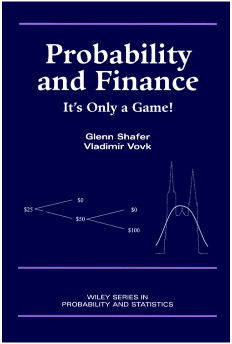


Doob (1910–2004)

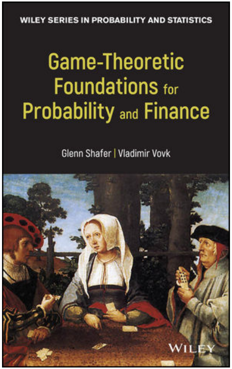
Itô (1915–2008)

Moving away from Kolmogorov's axioms

- A. Philip Dawid: prequential (predictive sequential) statistics. Calls for evaluating a probability forecaster only using his actual forecasts. (He might not even have a strategy.)
- Sits uneasily with Kolmogorov's axioms of probability.
- Glenn Shafer's and my books: systematization.



Our 2001 book



Our 2019 book
(3 players of the right sexes)

Foremost contributors to GTP

To a large degree, our 2019 book was based on the research by:

- Tokyo group led by Takeuchi and Takemura;
- Ghent group led by Gert de Cooman (including Jasper De Bock and Natan T'Joens);
- in continuous time (in particular, Itô calculus): Perkowski, Prömel, Łochowski.

Duality in imprecise probability: GTP side

- Let me give a simple example illustrating the duality between game-theoretic and measure-theoretic probability.
- Suppose we have a number of available gambles whose outcome depends on an unknown state of nature j; A_{ij} is the payoff of gamble i in state j; A is a finite matrix. Gamble i costs c_i.
- There are two ways to define the upper price of a new gamble with payoff b_j in state j.
- Pascal-type: as the solution to

$$\sum_i c_i x_i \to \min \quad \text{subject to} \quad \forall j : \sum_i x_i A_{ij} \geq b_j$$

(we are "superhedging" the new gamble).

Duality: MTP side

- The dual LP problem is

$$\sum_j b_j y_j \to \max \quad \text{subject to}$$

$$\forall i : \sum_j A_{ij} y_j = c_i \text{ and } \forall j : y_j \geq 0.$$

- This is a Fermat-type solution: we are looking at all measures correctly pricing the available gambles.
- y is a probability measure if "doing nothing" is one of the available gambles.

Making it non-trivial

- The simple duality I have just talked about is common-place in imprecise probabilities (Williams 1975, Walley 1991, Troffaes and de Cooman 2014) and mathematical finance.
- In this talk we will be interested in the sequential setting (perhaps with infinite horizon and continuous time).
- In our 2019 book we explore a far-reaching sequential (with infinite horizon) generalization of the simple dual picture (but there are severe limitations!).

Plan

Advantages of game-theoretic probability

In the preface to our 2019 book, we enumerate seven new ideas coming out of game-theoretic probability. In this section I cover the core ones:

- Strategies for testing. Theorems of probability are made constructive.
- Limited betting opportunities (imprecise probabilities): the assumptions required for those results can be weakened.
- Strategies for Reality, also constructive.
- Strategies for Forecaster, also constructive.

Perfect-information games as an alternative foundation for probability

Example: Kolmogorov's strong law of large numbers (SLLN) as a typical limit theorem.

Forecasting protocol:

$\mathcal{K}_0 = 1$
FOR $n = 1, 2, \dots$:
  Forecaster announces $m_n \in \mathbb{R}$ and $v_n \geq 0$
  Sceptic announces $M_n \in \mathbb{R}$ and $V_n \geq 0$
  Reality announces $y_n \in \mathbb{R}$
  $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(y_n - m_n) + V_n((y_n - m_n)^2 - v_n)$

Example: predicting temperature (imprecise theory).

Player's goals

A strategy for Sceptic forces an event E if it guarantees both

- $\mathcal{K}_n \geq 0$ for all n
- either $\mathcal{K}_n \to \infty$ or E happens.

A strategy for Reality forces E if it defeats Sceptic in the sense of

- $\mathcal{K}_n < 0$ for some n, or
- $\mathcal{K}_n$ is bounded and E happens.

Game-theoretic version of Kolmogorov's SLLN

Theorem

Sceptic can force

$$\sum_{n=1}^{\infty} \frac{v_n}{n^2} < \infty \Longrightarrow \frac{1}{n} \sum_{i=1}^{n} (y_i - m_i) \to 0.$$

Reality can force

$$\sum_{n=1}^{\infty} \frac{v_n}{n^2} = \infty \Longrightarrow \frac{1}{n} \sum_{i=1}^{n} (y_i - m_i) \not\to 0.$$

The strategies constructed in the proofs are explicit (and computable; in particular measurable).

SLLN for bounded variables

- If we bound Reality's moves, $|y_n| \leq C$, we can drop v_n and still claim that Sceptic can force

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - m_i) \to 0.$$

- A standard interpretation for a trusted Forecaster: we do not expect Sceptic to become infinitely rich, and so expect $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(x_i - m_i) = 0$ (calibration).
- Curious example: predictions in a prediction market either allow us to become infinitely rich or are calibrated. Which?

Connection with measure-theoretic probability

Kolmogorov's SLLN is an easy corollary of the game-theoretic version (in combination with the measurability of Sceptic's strategy): if $\xi_1, \xi_2, \ldots$ are independent random variables with expected values $\mathbb{E}(\xi_1), \mathbb{E}(\xi_2), \ldots$ and variances $\text{var}(\xi_1), \text{var}(\xi_2), \ldots$,

$$\sum_{n=1}^{\infty} \frac{\text{var}(\xi_n)}{n^2} < \infty \Longrightarrow \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (\xi_i - E(\xi_i)) = 0 \quad \text{a.s.}$$

Proof: If Reality follows the randomized strategy $y_n := \xi_n$, Forecaster always chooses $m_n := \mathbb{E}(\xi_n)$ and $v_n := \text{var}(\xi_n)$, and Sceptic's follows his winning strategy, $\mathcal{K}_n$ will be a nonnegative martingale. According to a standard result, $\mathcal{K}_n \to \infty$ with probability 0.

Strategies for Reality

- They do not have any analogues in the existing measure-theoretic probability.
- And they are often extremely simple; e.g., the one forcing Kolmogorov's SLLN is:
  - If $\mathcal{K}_{n-1} + V_n(n^2 - v_n) > 1$, set $y_n := 0$.
  - If $\mathcal{K}_{n-1} + V_n(n^2 - v_n) \leq 1$ and $M_n \leq 0$, set $y_n := n$.
  - If $\mathcal{K}_{n-1} + V_n(n^2 - v_n) \leq 1$ and $M_n > 0$, set $y_n := -n$.
- Takemura and Miyabe obtained lots of beautiful results about forcing by Reality (some described in our book).

Strategies for Forecaster

- This is known as defensive forecasting; covered by Glenn Shafer at ISIPTA 2007.
- Under weak assumptions, testing strategies can be turned into successful prediction strategies.
- Idea: Forecaster can defend Reality against a known strategy for Sceptic, making sure $\mathcal{K}_n$ does not grow.
- If Sceptic is testing calibration, Forecaster can enforce calibration.

Game-theoretic probability 1

Sceptic's part of the game-theoretic SLLN can be restated as:

$$\sum_{n=1}^{\infty} \frac{v_n}{n^2} < \infty \Longrightarrow \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (x_i - m_i) = 0 \tag{1}$$

holds with lower probability one, for a new notion of probability.

Or: the complement of (1) holds with upper probability zero.

Game-theoretic probability 2

For an event E,

$$\overline{\mathbb{P}}(E) := \inf\{\epsilon : \exists \text{ allowed strategy for Sceptic such that } \sup_n \mathcal{K}_n \geq 1/\epsilon \text{ on the event } E\}$$

(upper game-theoretic probability; equivalent definition with lim sup or lim inf in place of sup) and

$$\underline{\mathbb{P}}(E) := 1 - \overline{\mathbb{P}}(E^c)$$

(lower game-theoretic probability).

For many interesting sets E, $\overline{\mathbb{P}}(E) = \underline{\mathbb{P}}(E)$ (in continuous time) and $\overline{\mathbb{P}}(E) \approx \underline{\mathbb{P}}(E)$ (in discrete time).

Other game-theoretic results

For example:

- law of the iterated logarithm [lower probability one]
- zero-one law [lower probability one]
- central limit theorem [general game-theoretic probability]
- ...

Imply their measure-theoretic counterparts.

Plan

Three more ideas in game-theoretic probability

- Open protocols for science (not everything is modelled).
- Insuring against loss of evidence by Sceptic.
- Game-theoretic probability in continuous time.

Introduction

Game-theoretic probability provides both a new mathematical theory of probability (like measure-theoretic probability) and a new interpretation of probability (the martingale interpretation). But it interprets sources of probabilities (such as probabilistic theories), not individual probabilities. (In this respect the martingale interpretation is close to the frequentist: probabilities are interpreted *en masse*.)

We are often interested in two dual questions:

- What does a probabilistic theory tell us about reality? (What are its predictions?)
- When does a probabilistic theory contradict reality? (How do we test it?)

Main features of the von Neumann Quantum Mechanics

This is the universally accepted core of QM.

- the random character of QM shows at the moments of measurements (the current wave function randomly collapses producing a result of the measurement)
- between measurements the isolated physical system develops deterministically (according to the Schrödinger equation).

Testing the theory

Parameter of the protocol: t_0 (initial time)

$\mathcal{K}_0 = 1$
FOR $n = 1, 2, \dots$:
Observer announces an observable A_n and a time $t_n > t_{n-1}$
QM announces a probability measure P_n on $\mathbb{R}$
Sceptic announces $f_n : \mathbb{R} \to [0, \infty]$ such that $\int f_n \, \mathrm{d}P_n \le \mathcal{K}_{n-1}$
At time t_n, Reality announces the result $y_n \in \mathbb{R}$
$\mathcal{K}_n := f_n(y_n)$

The empirical content of the theory according to the martingale interpretation

Sceptic's capital $\mathcal{K}_n$ is the amount of evidence found against the theory by time n.

If Sceptic follows a fixed strategy, then as a function of Reality's and Forecaster's moves, $\mathcal{K}_n$ is a game-theoretic martingale. (If Forecaster's strategy is also fixed, this becomes a function of Reality's moves only.) If $\mathcal{K}_n(\text{theory}, \text{data})$ is very large:

- If we trust the data (e.g., we have just observed it ourselves) but are uncertain about the theory, we should reject the theory: the testing mode.
- If we trust the theory (it has been well tested) but not the data (e.g., it is still in the future), we will reject the data: the prediction mode.

Example of prediction

The forecasting protocol covers many other forecasting protocols. For example, in the case of the QM protocol: if Observer repeatedly measures observable bounded in absolute value by some constant C, Sceptic can force the event

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left(y_n-\int yP_n(\mathrm{d}y)\right)=0$$

(the equality is regarded as satisfied if there are only finitely many measurements).

Open theories

One feature of the von Neumann QM is typical of a scientific theory: it is open, in that it assigns probabilities only to some observations; and there is no way to isolate the “probabilized” observations (the outcomes of measurements) from the “unprobabilized” observations (what and when is measured).

Kolmogorov’s axioms of probability assume that each event is assigned a probability, and so cannot be cleanly applied to open theories.

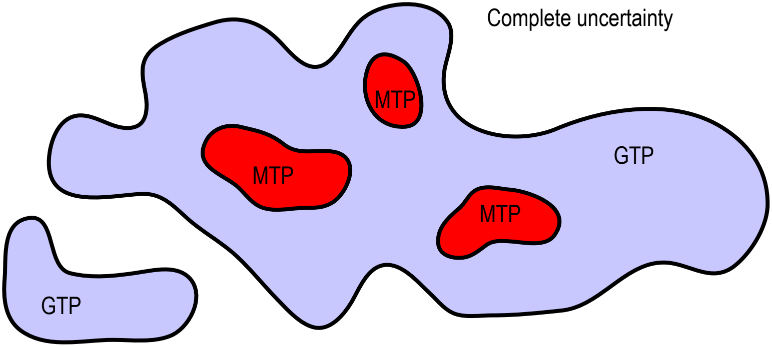
Experimental vs observational studies

- Open theories are not a problem for experimental studies: we use a stochastic strategy for choosing the unprobabilized observations (as in, e.g., randomized clinical trials). The "combined theory" (the original theory + the probabilities coming from the experimental design) is closed, and every event is now assigned a probability.
- In observational or mixed studies, there is no natural way to "close" an open theory.

A few other examples of open probabilistic theories

- Quantum computing.
- Statistical mechanics of the gas in a vessel that changes its shape non-stochastically.
- Efficiency of a particular financial market.
- Cox's survival model.

A caricature of the world of probabilities



A danger

- If $\mathcal{K}_0 = 1$, we regard $\mathcal{K}_n$ as the amount of evidence against Forecaster (e.g., a scientific theory).
- But what if Sceptic continues betting and loses all of this evidence?
- Let us see how he can retain at least some of this evidence.
- Idea: Sceptic should modify his strategy laying aside part of his capital at each step.

Saving protocol

Let us bundle all the old players (Forecaster, Sceptic, Reality) into one player.

$\mathcal{K}_0 = 1$ and $\mathcal{L}_0 = 1$
FOR $n = 1, 2, \dots$:
  Rival Sceptic announces $p_n \in [0, 1]$
  Old players announce $\mathcal{K}_n \in \mathbb{R}$
  $\mathcal{L}_n := (1 - p_n)\mathcal{L}_{n-1}\mathcal{K}_n/\mathcal{K}_{n-1} + p_n\mathcal{L}_{n-1}$

(with $a/0 := 0$ for any a)

- Rival Sceptic keeps fraction p_n of his current capital in cash.
- The rest is invested in Sceptic's move (assumed, implicitly, scalable).

What Rival Sceptic can achieve

- Set $\mathcal{K}_n^* := \max_{0 \le i \le n} \mathcal{K}_i$.
- An increasing function $H : [1, \infty) \to [0, \infty)$ is a lookback calibrator if Rival Sceptic has a strategy that guarantees

$$\mathcal{L}_n \ge H(\mathcal{K}_n^*), \quad n = 0, 1, \ldots .$$

Proposition

An increasing function $H : [1, \infty) \to [0, \infty)$ is a lookback calibrator if and only if

$$\int_1^\infty \frac{H(v)}{v^2} \mathrm{d}v \le 1.$$

Two examples

- Of course, $H(v) := v$ is not a lookback calibrator; but we would like to get as close to it as possible.
- An interesting parametric family of maximal lookback calibrators:

$$G_\kappa(v) := \kappa v^{1-\kappa} \text{ for all } v \in [1, \infty),$$

where $\kappa \in (0, 1)$. Asymptotically as $v \to \infty$, good values are $\kappa \approx 0$.

- Even better asymptotically:

$$H_\kappa(v) := \begin{cases} \kappa(1+\kappa)^\kappa v \ln^{-1-\kappa} v & \text{if } v \geq e^{1+\kappa} \\ 0 & \text{otherwise,} \end{cases}$$

where $\kappa \in (0, \infty)$.

Bayes factors and p-values

- Lookback calibrators H can also be used for transforming p-values p into Bayes factors against the null hypothesis ($H(1/p)$ is a Bayes factor).
- This is relevant to the current debate about p-values.
- Glenn Shafer covered some of these ideas in his poster in the morning.

Basic game

Players: Reality (market) and Sceptic (speculator).

Time: $[0, \infty)$.

Two steps of the game:

- Sceptic chooses his trading strategy.
- Reality chooses a continuous function $\omega : [0, \infty) \to \mathbb{R}$ (the price trajectory).

The market is often a metaphor for us, but in this part (Part IV) of our 2019 book we prefer to take it seriously.

Allowed strategies I

A simple trading strategy G consists of:

- an increasing sequence of stopping times $\tau_1 \le \tau_2 \le \dots$ such that $\tau_n = \tau_n(\omega) \to \infty$ as $n \to \infty$ (for all ω)
- for each $n = 1, 2, \dots$, a bounded $\mathcal{F}_{\tau_n}$-measurable h_n.

To such G and initial capital $c \in \mathbb{R}$ corresponds the simple capital process

$$\mathcal{K}_t^{G,c}(\omega) := c + \sum_{n=1}^{\infty} h_n(\omega)\big(\omega(\tau_{n+1} \wedge t) - \omega(\tau_n \wedge t)\big)$$

(interpretation: the interest rate is zero).

$h_n(\omega)$: Sceptic's bet (or stake) at time τ_n
$\mathcal{K}_t^{G,c}(\omega)$: Sceptic's capital at time t

Allowed strategies II

A nonnegative capital process is a process S with values in $[0, \infty]$ that can be represented as

$$S_t(\omega) := \sum_{n=1}^{\infty} \mathcal{K}_t^{G_n, c_n}(\omega),$$

where

- the simple capital processes $\mathcal{K}_t^{G_n, c_n}(\omega)$ are required to be nonnegative, for all t and ω
- the nonnegative series $\sum_{n=1}^{\infty} c_n$ is required to converge

Null events

- An event E is said to be null if there exists a nonnegative capital process S with $S_0 = 1$ such that $S_t(\omega) \to \infty$ as $t \to \infty$ for all $\omega \in E$.
- Intuition: you can become infinitely rich risking only €1 if the event is singled out in advance and happens.
- But this does not really mean that we do not expect E to happen. Infinity is far away. . . .

Instantly blockable and instantly enforceable events

- Let us say that a property E of time t and ω is instantly blockable if there exists a nonnegative capital process S such that $S_0 = 1$ and

$$(t, \omega) \in E \Longrightarrow S_t(\omega) = \infty.$$

  Now such E are really not expected to happen.
- Complements of such E hold with instant enforcement (w.i.e.).

Share prices in efficient markets

The following proposition says that share prices must look like Brownian motion.

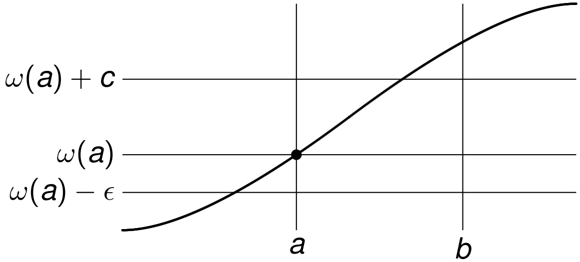
Proposition

The trader can instantly block ω being monotonic in an open interval where it is not constant.

Proof.

Idea: To see that we can profit from a price trajectory ω that is too regular, see the next slide (which assumes that ω strictly increases over an interval). □

How to profit from an increasing price trajectory



Strategy (a, b, c, ϵ): at time a buy $h := 1/c$ shares selling them at b or when the price reaches $\omega(a) + c$ or when the price reaches $\omega(a) - \epsilon$ (whatever happens earlier).

Enumerate all rational triples: (a_k, b_k, c_k), $k = 1, 2, \ldots$, and set $\epsilon_k := 2^{-k} c_k$. Launch $(a_k, b_k, c_k, \epsilon_k)$ with initial capital 2^{-k} for all k.

What else holds w.i.e.

We have w.i.e. (in a streamlined sense):

- the existence of the Itô integral
- Itô's formula.

This allows us to state GTP analogues of

- Girsanov's theorem
- CAPM
- stochastic portfolio theory,

all of them probability-free.

Emergence of precise probabilities

- Let me state (somewhat informally) a game-theoretic analogue of the Dubins–Schwarz result in MTP.
- Suppose a measurable set E of trajectories is invariant w.r. to adding constants and continuous monotonic time transformations.
- Then the game-theoretic probability [defined essentially as before] is precise: $\overline{\mathbb{E}}(E) = W(E)$, where W is the standard Wiener measure.
- This implies the previous proposition and lots of other interesting properties.

Summary: some advantages of GTP over MTP

- All 3 players have interesting strategies.
- All strategies are constructive, and Sceptic's may be imprecise.
- Probability-free theory of (non-)stochastic processes.

Bibliography

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

A growing list of working papers (now $54 + 20$):

`http://www.probabilityandfinance.com`

Thank you for your attention!