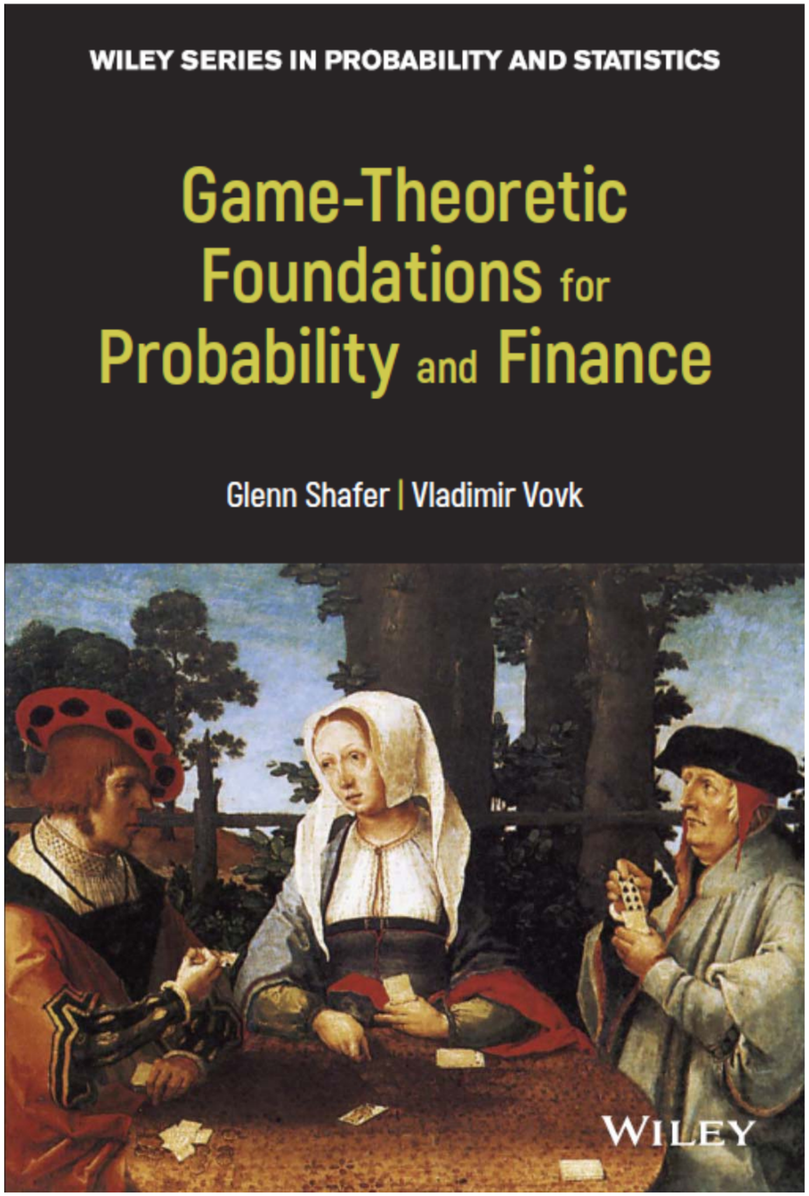
Game-Theoretic Statistics

Glenn Shafer

CWI, Amsterdam, December 17, 2018

- Game theory as an alternative foundation for probability
- Using the game-theoretic foundation in statistics


WILEY SERIES IN PROBABILITY AND STATISTICS
Game-Theoretic Foundations for Probability and Finance
Glenn Shafer | Vladimir Vovk
WILEY

May 2019

ON THE BACK COVER

Ever since Kolmogorov's *Grundbegriffe*, the standard mathematical treatment of probability theory has been measure-theoretic. In this ground-breaking work, Shafer and Vovk give a game-theoretic foundation instead. While being just as rigorous, the game-theoretic approach allows for vast and useful generalizations of classical measure-theoretic results, while also giving rise to new, radical ideas for prediction, statistics and mathematical finance without stochastic assumptions. The authors set out their theory in great detail, resulting in what is definitely one of the most important books on the foundations of probability to have appeared in the last few decades.

– Peter Grünwald, CWI and the University of Leiden

Two foundations for probability:

- Measure theory
- Game theory

The game-theoretic foundation goes deeper:

- Probabilities are derived from a perfect-information game.
- To prove a theorem, you construct a strategy in the game.

Twentieth century statistics was based on measure-theoretic probability.

The statistician has only partial knowledge of the probabilities.

R. A. Fisher in 1922: The statistician only knows that the true probability measure is one of an indexed class $(P_\theta)_{\theta \in \Theta}$.

Game-theoretic statistics goes deeper. The statistician may have only partial knowledge of the perfect-information game. She may be looking at the game from outside, seeing only part of what happens.

The statistician may be outside the perfect-information game, seeing only part of what happens.

- The goal of this talk is to explain this understanding of game-theoretic statistics.
- First I review game-theoretic probability.

Our book begins with the notion of a testing protocol.

Example of a testing protocol

Consider a game with three players: Forecaster, Skeptic, and Reality. On each round of the game,

- Forecaster decides and announces the price m for a payoff y,
- Skeptic decides and announces how many units, say M, of y he will buy,
- Reality decides and announces the value of y, and
- Skeptic receives the net gain $M(y - m)$, which may be positive, negative, or zero.

Perfect information:

- Players move in turn.
- Each sees the other's moves.

Each player can also receive other information, possibly private.

Consider a game with three players: Forecaster, Skeptic, and Reality. On each round of the game,

- Forecaster decides and announces the price m for a payoff y,
- Skeptic decides and announces how many units, say M, of y he will buy,
- Reality decides and announces the value of y, and
- Skeptic receives the net gain $M(y - m)$, which may be positive, negative, or zero.

Protocol 1.1.

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

We can *specialize* to probability forecasting:

Protocol 1.1.
Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.
FOR $n = 1, 2, \ldots$:
Forecaster announces $m_n \in [-1, 1]$.
Skeptic announces $M_n \in \mathbb{R}$.
Reality announces $y_n \in [-1, 1]$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(y_n - m_n)$.

Protocol 1.8.
Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.
FOR $n = 1, 2, \ldots$:
Forecaster announces $p_n \in [0, 1]$.
Skeptic announces $M_n \in \mathbb{R}$.
Reality announces $y_n \in \{0, 1\}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(y_n - p_n)$.

We call any restriction on Skeptic's opponents a *specialization*.

Protocol 1.8 is a specialization of Protocol 1.1.

If Skeptic can accomplish something in one protocol, this remains true in any specialization.

Protocol 1.1.

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

Three specializations:

Protocol 1.3.

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

Protocol 1.8.

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

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

Protocol 1.1.

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

A *testing protocol* allows Skeptic to test Forecaster's reliability.

If Skeptic multiplies the capital he risks by a large factor, then Forecaster does not look good.

"Skeptic multiplies the capital he risks by 1000 in the first N trials" means

1. $\mathcal{K}_n > 0$.
2. $\mathcal{K}_{n-1} + M_n(1 - m_n) \geq 0$ and $\mathcal{K}_{n-1} + M_n(-1 - m_n) \geq 0$ for $n = 1, \ldots, N$ and
3. $\mathcal{K}_N / \mathcal{K}_0 \geq 1000$.

The inequalities in Point 2 are equivalent to

$$-\frac{\mathcal{K}_{n-1}}{1 - m_n} \leq M_n \leq \frac{\mathcal{K}_{n-1}}{1 + m_n}.$$

Skeptic tests Forecaster

Protocol 1.8.

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

Set $\overline{y}_N := \frac{\sum_{n=1}^{N} y_n}{N}$ and $\overline{p}_N := \frac{\sum_{n=1}^{N} p_n}{N}$.

Skeptic has strategies that multiply the capital he risks by a large factor unless $|\overline{y}_N - \overline{p}_N|$ is small.

One such strategy is $M_n := \frac{\overline{y}_{n-1} - \overline{p}_{n-1}}{2} \mathcal{K}_{n-1}$.

Skeptic tests Forecaster

Skeptic has strategies that multiply the capital he risks by a large factor unless $|\overline{y}_N - \overline{p}_N|$ is small.

Protocol 1.8.

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

This is the law of large numbers. We can make it precise asymptotically or finitely.

- Asymptotic: The difference tends to zero unless Skeptic becomes infinitely rich (*aka* except on a set of probability zero) .
- Finite: For large enough *N*, the difference is less than a certain small amount unless Skeptic multiplies his capital by a certain large factor (*aka* except on a set of small probability).

These laws of large numbers are proven constructively, by constructing strategies for Skeptic.

Cournot's principle

Cournot's principle:

Forecaster is *reliable*

=

Skeptic is unable to multiply the capital he risks substantially.

A forecasting strategy is *valid*

=

it withstands Skeptic's strategies.

If Forecaster is *reliable* if Skeptic cannot multiply the capital he risks substantially.

This is an aspect of *Cournot's principle*.

Skeptic has strategies that multiply the capital he risks by a large factor unless $|\overline{y}_N - \overline{p}_N|$ is small.

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

By Cournot's principle, reliablity of Forecaster means that $|\overline{y}_N - \overline{p}_N|$ will be small.

If Forecaster uses a *valid* probability measure as a strategy, then Skeptic will not multiply the capital he risks substantially.

This is another aspect of *Cournot's principle*.

Skeptic has strategies that multiply the capital he risks by a large factor unless $|\overline{y}_N - \overline{p}_N|$ is small.

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

By Cournot's principle, validity of probability measure as strategy for Forecaster means that $|\overline{y}_N - \overline{p}_N|$ will be small.

Cournot's principle:

- Forecaster is *reliable* when Skeptic fails to multiply the capital he risks substantially.
- A forecasting strategy is *valid* if it withstands Skeptic's strategies.

This game-theoretic version of Cournot's principle generalizes the traditional version:

- Betting offers can be tested by a series of bets whether or not the bets boil down to a bet on a single event.
- The test need not follow a pre-specified strategy.

Upper expectations

In general, testing protocols use *upper expectations* (aka *upper previsions*).

Given a nonempty set $\mathcal{Y}$, we call a functional $\overline{\mathbf{E}} : \overline{\mathbb{R}}^{\mathcal{Y}} \to \overline{\mathbb{R}}$ an *upper expectation on* $\mathcal{Y}$ if it satisfies these five axioms:

Axiom E1. If $f_1, f_2 \in \overline{\mathbb{R}}^{\mathcal{Y}}$, then $\overline{\mathbf{E}}(f_1 + f_2) \leq \overline{\mathbf{E}}(f_1) + \overline{\mathbf{E}}(f_2)$.

Axiom E2. If $f \in \overline{\mathbb{R}}^{\mathcal{Y}}$ and $c \in (0, \infty)$, then $\overline{\mathbf{E}}(cf) = c\overline{\mathbf{E}}(f)$.

Axiom E3. If $f_1, f_2 \in \overline{\mathbb{R}}^{\mathcal{Y}}$ and $f_1 \leq f_2$, then $\overline{\mathbf{E}}(f_1) \leq \overline{\mathbf{E}}(f_2)$.

Axiom E4. For each $c \in \mathbb{R}$, $\overline{\mathbf{E}}(c) = c$.

Axiom E5. If $f_1 \leq f_2 \leq \cdots \in [0, \infty]^{\mathcal{Y}}$, then $\overline{\mathbf{E}}\left(\lim_{k\to\infty} f_k\right) = \lim_{k\to\infty} \overline{\mathbf{E}}(f_k)$.

We call Axiom E5 the *continuity axiom*. We call $\overline{\mathbf{E}}(f)$ *f's upper expected value*.

Axiom E5 is optional. Almost never really needed.

Axiom E1. If $f_1, f_2 \in \mathbb{R}^{\mathcal{Y}}$, then $\mathbf{E}(f_1 + f_2) \leq \mathbf{E}(f_1) + \mathbf{E}(f_2)$.

Axiom E2. If $f \in \overline{\mathbb{R}}^{\mathcal{Y}}$ and $c \in (0, \infty)$, then $\overline{\mathbf{E}}(cf) = c\overline{\mathbf{E}}(f)$.

Axiom E3. If $f_1, f_2 \in \overline{\mathbb{R}}^{\mathcal{Y}}$ and $f_1 \leq f_2$, then $\overline{\mathbf{E}}(f_1) \leq \overline{\mathbf{E}}(f_2)$.

Axiom E4. For each $c \in \mathbb{R}$, $\overline{\mathbf{E}}(c) = c$.

Axiom E5. If $f_1 \leq f_2 \leq \cdots \in [0, \infty]^{\mathcal{Y}}$, then $\overline{\mathbf{E}}\left(\lim_{k\to\infty} f_k\right) = \lim_{k\to\infty} \overline{\mathbf{E}}(f_k)$.

Protocol 6.12.

PARAMETER: Nonempty set $\mathcal{Y}$

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

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

Forecaster announces an upper expectation $\overline{\mathbf{E}}_n$ on $\mathcal{Y}$.

Skeptic announces $f_n \in \overline{\mathbb{R}}^{\mathcal{Y}}$ such that $\overline{\mathbf{E}}_n(f_n) = \mathcal{K}_{n-1}$.

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

$\mathcal{K}_n := f_n(y_n)$.

Protocol 6.12.

PARAMETER: Nonempty set $\mathcal{Y}$

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

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

Forecaster announces an upper expectation $\overline{\mathbf{E}}_n$ on $\mathcal{Y}$.

Skeptic announces $f_n \in \overline{\mathbb{R}}^{\mathcal{Y}}$ such that $\overline{\mathbf{E}}_n(f_n) = \mathcal{K}_{n-1}$.

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

$\mathcal{K}_n := f_n(y_n)$.

It is often convenient to allow Skeptic to give up some of his capital. We call this a *slackening*. It does not affect what Skeptic can accomplish.

PARAMETER: Nonempty set $\mathcal{Y}$

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

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

Forecaster announces an upper expectation $\overline{\mathbf{E}}_n$ on $\mathcal{Y}$.

Skeptic announces $f_n \in \overline{\mathbb{R}}^{\mathcal{Y}}$ such that $\overline{\mathbf{E}}_n(f_n) \leq \mathcal{K}_{n-1}$.

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

$\mathcal{K}_n := f_n(y_n)$.

slackening

Protocol 6.12.

PARAMETER: Nonempty set $\mathcal{Y}$

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

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

Forecaster announces an upper expectation $\overline{\mathbf{E}}_n$ on $\mathcal{Y}$.

Skeptic announces $f_n \in \overline{\mathbb{R}}^{\mathcal{Y}}$ such that $\overline{\mathbf{E}}_n(f_n) \leq \mathcal{K}_{n-1}$.

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

$\mathcal{K}_n := f_n(y_n)$.

Protocol 1.1.

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

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

Forecaster announces $m_n \in [-1, 1]$.

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

Reality announces $y_n \in [-1, 1]$.

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

The slackening of Protocol 1.1 is a specialization of Protocol 6.12.

$$\overline{\mathbf{E}}^m(f) := \inf\{\alpha \in \overline{\mathbb{R}} \mid \exists M \in \overline{\mathbb{R}}\, \forall y \in [-1,1] : f(y) \leq M(y - m) + \alpha\}$$

Axiom E1. If $f_1, f_2 \in \mathbb{R}^{\mathcal{Y}}$, then $\mathbf{E}(f_1 + f_2) \leq \mathbf{E}(f_1) + \mathbf{E}(f_2)$.

Axiom E2. If $f \in \overline{\mathbb{R}}^{\mathcal{Y}}$ and $c \in (0, \infty)$, then $\overline{\mathbf{E}}(cf) = c\overline{\mathbf{E}}(f)$.

Axiom E3. If $f_1, f_2 \in \overline{\mathbb{R}}^{\mathcal{Y}}$ and $f_1 \leq f_2$, then $\overline{\mathbf{E}}(f_1) \leq \overline{\mathbf{E}}(f_2)$.

Axiom E4. For each $c \in \mathbb{R}$, $\overline{\mathbf{E}}(c) = c$.

Axiom E5. If $f_1 \leq f_2 \leq \cdots \in [0, \infty]^{\mathcal{Y}}$, then $\overline{\mathbf{E}}\left(\lim_{k\to\infty} f_k\right) = \lim_{k\to\infty} \overline{\mathbf{E}}(f_k)$.

Protocol 6.12.

PARAMETER: Nonempty set $\mathcal{Y}$

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

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

Forecaster announces an upper expectation $\overline{\mathbf{E}}_n$ on $\mathcal{Y}$.

Skeptic announces $f_n \in \overline{\mathbb{R}}^{\mathcal{Y}}$ such that $\overline{\mathbf{E}}_n(f_n) \leq \mathcal{K}_{n-1}$.

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

$\mathcal{K}_n := f_n(y_n)$.

Forecaster is de Finetti's **You**.

Upper and lower expectations are called *upper* and *lower previsions* in the theory of imprecise probabilities. Because the theory takes You's viewpoint, it emphasizes lower rather than upper previsions. As noted on p. 102, buying X for α is the same as selling $-X$ for $-\alpha$. But because most people buy more often than they sell, ordinary language is more developed for buying than for selling, and we tend to develop theories in terms of buying prices. As we have learned, Skeptic's buying prices are given by the upper functional. You being Skeptic's counterparty, his buying prices are given by the lower functional.

Chapter 8: A testing protocol determines a global upper expectation.

Protocol 6.12.

PARAMETER: Nonempty set $\mathcal{Y}$

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

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

Forecaster announces an upper expectation $\overline{\mathbf{E}}_n$ on $\mathcal{Y}$.

Skeptic announces $f_n \in \overline{\mathbb{R}}^{\mathcal{Y}}$ such that $\overline{\mathbf{E}}_n(f_n) \leq \mathcal{K}_{n-1}$.

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

$\mathcal{K}_n := f_n(y_n)$.

The global upper expectation is an upper expectation on the space of all sequences of moves by Skeptic's opponents.

How it is defined: The upper expected value of a function $X(\overline{\mathbf{E}}_1, y_1, \overline{\mathbf{E}}_2, y_2, \ldots)$ is the least initial capital $\mathcal{K}_0$ needed for a strategy for which $\lim \mathcal{K}_n$ is always at least $X(\overline{\mathbf{E}}_1, y_1, \overline{\mathbf{E}}_2, y_2, \ldots)$.

A testing protocol determines a global upper expectation.

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

The global upper expectation on $\{0,1\}^\infty$ is essentially Lebesgue measure on $[0,1]$. (See Borel 1909.)

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

The global upper expectation on $([0,1] \times \{0,1\})^\infty$ is not a probability measure.

New argument for the generalization from probabilities to “imprecise probabilities”:
The generalization arises naturally even in the description of probability forecasting.

Chapter 10 of *Game-Theoretic Foundations for Probability and Finance*

Using Testing Protocols in Science and Technology **175**

10.1 Signals in Open Protocols 176
10.2 Cournot's Principle 179
10.3 Daltonism 180
10.4 Least Squares 185
10.5 Parametric Statistics with Signals 188
10.6 Quantum Mechanics 191

Signals in testing protocols

Protocol 1.8.

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

Skeptic's strategy

$$M_n := \frac{\overline{y}_{n-1} - \overline{p}_{n-1}}{2} \mathcal{K}_{n-1}$$

multiplies the capital it risks by a large factor unless $|\overline{y}_N - \overline{p}_N|$ is small.

Protocol 10.1.

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

signal

The presence of the signals x_n does not invalidate this theorem.

But when the p_n are fixed as functions of the x_n, the theorem becomes a statement relating the x_n and the y_n.

Least squares

(assuming only bounded errors, in the spirit of Lai and Wei, 1982)

Tze Leung Lai and Ching Zong Wei. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics*, 10(1):154–166, 1982.

Protocol 10.7.

PARAMETER: $w \in \mathbb{R}^K$

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

Reality announces $x_n \in \mathbb{R}^K$.

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

Reality announces $\epsilon_n \in [-1, 1]$ and sets $y_n := \langle w, x_n \rangle + \epsilon_n$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n \epsilon_n$.

dot product

The statistician stands outside the protocol, not seeing the parameter w or the ϵ_n. She sees only the x_n and the y_n.

The goal is to find w.

Protocol 10.7.

PARAMETER: $w \in \mathbb{R}^K$

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

Reality announces $x_n \in \mathbb{R}^K$.

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

Reality announces $\epsilon_n \in [-1, 1]$ and sets $y_n := \langle w, x_n \rangle + \epsilon_n$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n \epsilon_n$.

- Treat w and x_n as column vectors.
- Let X_n be the $n \times K$ matrix whose ith row is x'_i.
- Let Y_n be the n-dimensional column vector whose ith element is y_i.
- The *least squares estimate* of w is

$$w_n := (X'_n X_n)^{-1} X'_n Y_n$$

  if $X'_n X_n$ is invertible.
- Write $\lambda_n^{\max}$ and $\lambda_n^{\min}$ for $X'_n X_n$'s largest and smallest eigenvalues, respectively.

Theorem. Skeptic has a strategy that multiplies his capital infinitely unless

$$(\lambda_n^{\min} \to \infty \ \& \ \ln \lambda_n^{\max} / \lambda_n^{\min} \to 0) \implies \|w_n - w\| = O\left(\sqrt{\ln \lambda_n^{\max} / \lambda_n^{\min}}\right) = o(1).$$

where $\|\cdot\|$ is the Euclidean norm.

The statistician, believing that Skeptic cannot multiply the capital he risks infinitely, concludes that the least squares estimate is consistent.

Parametric Statistics

Convenient general protocol for testing a probability forecaster

PARAMETERS: Nonempty set $\mathcal{X}$ and measurable space $\mathcal{Y}$

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

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

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

Forecaster announces $P_n \in \mathcal{P}(\mathcal{Y})$.

Skeptic announces $f_n \in [0, \infty]^{\mathcal{Y}}$ such that $P_n(f_n) = 1$.

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

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

$\mathcal{P}(\mathcal{Y})$ is the set of probability measures on $\mathcal{Y}$.

Here Skeptic is not allowed to risk bankruptcy or waste money.

PARAMETERS: Nonempty set $\mathcal{X}$ and measurable space $\mathcal{Y}$

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

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

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

Forecaster announces $P_n \in \mathcal{P}(\mathcal{Y})$.

Skeptic announces $f_n \in [0, \infty]^{\mathcal{Y}}$ such that $P_n(f_n) = 1$.

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

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

The statistician might…

1. …be Forecaster, making probability predictions.
2. …be Skeptic, testing a theory that plays the role of Forecaster.
3. …stand outside the protocol, prescribing strategies to Forecaster and Skeptic.

PARAMETERS: Nonempty set $\mathcal{X}$ and measurable space $\mathcal{Y}$

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

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

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

Forecaster announces $P_n \in \mathcal{P}(\mathcal{Y})$.

Skeptic announces $f_n \in [0, \infty]^{\mathcal{Y}}$ such that $P_n(f_n) = 1$.

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

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

1. Suppose P_n always has a density p_n.
2. Set $q_n = f_n p_n$.
3. Then $\int f_n dP_n = 1$ implies that q_n is also a density.
4. So $f_n(y_n) = \dfrac{q_n(y_n)}{p_n(y_n)}$ and $\mathcal{K}_n = \dfrac{\prod_{i=1}^n q_i(y_i)}{\prod_{i=1}^n p_i(y_i)}$.

Skeptic's capital is a likelihood ratio!

Protocol 10.11.

PARAMETERS: Nonempty sets Θ and $\mathcal{X}$, measurable space $\mathcal{Y}$

Reality announces $\theta \in \Theta$.

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

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

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

Forecaster announces $P_n \in \mathcal{P}(\mathcal{Y})$.

Skeptic announces $f_n \in [0, \infty]^{\mathcal{Y}}$ such that $P_n(f_n) = 1$.

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

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

The statistician does not quite stand inside the protocol; she observes the x_n and y_n as they are announced, but she does not observe Reality's announcement of θ.

Suppose Forecaster follows a strategy, known to the statistician, that specifies each of his moves as a function of Reality's previous moves. To fix ideas, assume that the strategy specifies a density $p_n(\theta, x_1, y_1, \ldots, x_n)$ for P_n with respect to a fixed underlying probability measure.

If Skeptic also follows a strategy that is a function of Reality's previous moves, say $f_n(\theta, x_1, y_1, \ldots, x_n)$, and $q_n := f_n p_n$, then

$$\mathcal{K}_n = \frac{\prod_{i=1}^n q_i(\theta, x_1, y_1, \ldots, x_i)(y_i)}{\prod_{i=1}^n p_i(\theta, x_1, y_1, \ldots, x_i)(y_i)}.$$

David Cox called this a "partial likelihood".

ABSTRACT

Fermat and Pascal's two different methods for solving the problem of division lead to two different mathematical foundations for probability theory: a measure-theoretic foundation that generalizes the method of counting cases used by Fermat, and a game-theoretic foundation that generalizes the method of backward recursion used by Pascal. The game-theoretic foundation has flourished in recent decades, as documented by my forthcoming book with Vovk, *Game-Theoretic Probability and Finance*. In this book's formulation, probability typically involves three players, a player who offers betting rates (Forecaster), a player who tests the reliability of the forecaster by trying to multiply the capital he risks betting at these rates (Skeptic), and a player who decides the outcomes (Reality).

Game-theoretic statistics is less developed but appears to offer powerful and flexible resources for applications. One way of using the game between Forecaster, Skeptic, and Reality in applications is to suppose there are multiple Forecasters, each making forecasts according to a given probability model. This makes the picture look like standard statistical modeling in the tradition of Karl Pearson and R. A. Fisher, but it is only one possibility. In this talk, based on Chapter 10 of *Game-Theoretic Probability and Finance*, I will explore some other possibilities, drawing on examples from least squares, survival analysis, and quantum computing.

BACK COVER

Game-theoretic probability and finance come of age.

Glenn Shafer and Vladimir Vovk's *Probability and Finance*, published in 2001, showed that perfect-information games can be used to define mathematical probability. Based on fifteen years of further research, *Game-Theoretic Foundations for Probability and Finance* presents a mature view of the foundational role game theory can play. Its account of probability theory opens the way to new methods of prediction and testing and makes many statistical methods more transparent and widely usable. Its contributions to finance theory include purely game-theoretic accounts of Ito's stochastic calculus, the capital asset pricing model, the equity premium, and portfolio theory.

Game-Theoretic Foundations for Probability and Finance is a book of research. It is also a teaching resource. Each chapter is supplemented with carefully designed exercises and notes relating the new theory to its historical context.

BACK COVER

Praise from early readers

Ever since Kolmogorov's *Grundbegriffe*, the standard mathematical treatment of probability theory has been measure-theoretic. In this ground-breaking work, Shafer and Vovk give a game-theoretic foundation instead. While being just as rigorous, the game-theoretic approach allows for vast and useful generalizations of classical measure-theoretic results, while also giving rise to new, radical ideas for prediction, statistics and mathematical finance without stochastic assumptions. The authors set out their theory in great detail, resulting in what is definitely one of the most important books on the foundations of probability to have appeared in the last few decades.

– Peter Grünwald, University of Leiden

Shafer and Vovk have thoroughly re-written their 2001 book on the game-theoretic foundations for probability and for finance. They have included an account of the tremendous growth that has occurred since, in the game-theoretic and pathwise approaches to stochastic analysis and in their applications to continuous-time finance. This new book will undoubtedly spur a better understanding of the foundations of these very important fields, and we should all be grateful to its authors.

– Ioannis Karatzas, Columbia University

R. A. Fisher made the notion of a *parametric statistical model* central to mathematical statistics. Here the statistician knows only that the true probability distribution for a certain phenomenon is in a known class

$$(P_\theta)_{\theta \in \Theta} .$$

Fisher coined the name *parameter* for θ.

From the game-theoretic perspective, Fisher's picture is oversimplified. A statistician's knowledge and role with respect to a testing protocol can take several forms.

"On the mathematical foundations of theoretical statistics", by R. A. Fisher, *Phil. Trans. R. Soc. Lond. A* 222:309–368, 1922.